Proceedings of the 11th International Conference on Decision Support System Technology (ICDSST 2025)

Decision Support System Technology in the AI Era

Editors: Sandro Radovanović, Boris Delibašić. Ana Paula Cabral Seixas Costa, Jason Papathanasiou, María Teresa Escobar

Belgrade, Serbia, May 26th – May 29th, 2025

Preface

We are pleased to present the Local Proceedings of the 11th International Conference on Decision Support System Technology (ICDSST 2025), held in Belgrade, Serbia, in May 2025. Organized by the University of Belgrade, Faculty of Organizational Sciences, in collaboration with the EURO Working Group on Decision Support Systems (EWG-DSS), this year's conference was dedicated to the theme: "Decision Support System Technology in the AI Era."

The local proceedings provide a comprehensive overview of the research and practical contributions showcased during the conference, with a particular emphasis on the wide-ranging applications of **artificial intelligence, machine learning, big data analytics**, and **multi-criteria decision-making (MCDM)** in the development of modern decision support systems. These contributions demonstrate how DSS is evolving in response to the accelerating pace of digital transformation and increasing societal demand for intelligent, ethical, and transparent decision-making tools.

This volume includes 27 peer-reviewed contributions across full papers and poster presentations, organized into key thematic sections:

- AI, Machine Learning, and Big Data in Decision Support
- MCDM, Optimization, and Advanced Analytics
- Environment, Agriculture, and Natural Resource Management
- Blockchain, Supply Chain, and Logistic
- Public Sector, Education, and Administration
- DSS Design, User Experience, and Business/Process Frameworks
- Posters

In addition to research papers, the conference was enriched by insightful talks from distinguished invited speakers, including Fátima Dargam, Pascale Zaraté, Zoran Obradović, Marko Bohanec, and Lily Xu, whose contributions helped illuminate critical advances and open challenges in the DSS landscape.

We extend our sincere gratitude to all authors, reviewers, keynote speakers, and members of the organizing and program committees for their valuable contributions to the success of ICDSST 2025. Special thanks also go to our institutional sponsors and international partners who supported the realization of this event.

We hope that the works collected in this volume will serve as a catalyst for further discussion, collaboration, and innovation in the decision support systems community.

Belgrade, Serbia

May 2025

Sandro Radovanović, Boris Delibašić. Ana Paula Cabral Seixas Costa, Jason Papathanasiou, María Teresa Escobar

Organizers

The main purpose of the EWG-DSS – EURO Working Group on Decision Support Systems is to establish a platform for encouraging state-of-the-art high quality research and collaboration work within the DSS community. Other aims of the EWG-DSS are to:

- Encourage the exchange of information among practitioners, end-users, and researchers in the area of Decision Systems;
- Enforce the networking among the DSS communities available and facilitate activities that are essential for the start-up of international cooperation research and projects;
- Facilitate professional academic and industrial opportunities for its members;
- Favour the development of innovative models, methods and tools in the field Decision Support and related areas;
- And actively promote the interest on Decision Systems in the scientific community by organizing dedicated workshops, seminars, mini-conferences and conference streams in major conferences, as well as editing special and contributed issues in relevant scientific journals.

The EWG-DSS has already established a recognized tradition is organizing high level conferences and workshops concerning themes of Decision Support Systems, their applications and development approaches. More details about all the organized events of the EWG-DSS can be found in the organization website.

Organizing Committee

Honorary Chairs

Pascale Zarate, University Toulouse Capitole-IRIT, France

Fatima Dargam, SimTech-Reach Consulting, Austria

Conference Chairs

Boris Delibašić, University of Belgrade, Serbia

Sandro Radovanović, University of Belgrade, Serbia

Organizing Committee

Dražen Drašković, University of Belgrade, Serbia

Miloš Jovanović, University of Belgrade, Serbia

Marko Mišić, University of Belgrade, Serbia

Andrija Petrović, University of Belgrade, Serbia

Milan Vukićević, University of Belgrade, Serbia

Program Committee

Shaofeng Liu, University of Plymouth, UK

Ana Paula Cabral Seixas Costa, Federal University of Pernambuco, Brazil

Jason Papathanasiou, University of Macedonia, Greece

Pavlos Delias, Kavala Institute of Technology, Greece

María Teresa Escobar, University of Zaragoza, Spain

Sandro Radovanović, University of Belgrade, Serbia

Sergio Pedro Duarte, University of Porto, Portugal

Program Committee Members

Adiel Teixeira de Almeida, Federal University of Pernambuco, Brazil

Alberto Turón, University of Zaragoza, Spain

Alexander Smirnov, Russian Academy of Sciences, Russia

Alexis Tsoukias, University Paris Dauphine, France

Ana Paula Cabral Seixas Costa, Federal University of Pernambuco, Brazil

Andrija Petrović, University of Belgrade, Serbia

Andy Wong, University of Strathclyde, UK

Antonio Lobo, Faculdade de Engenharia da Universidade do Porto, Porto

Boris Delibašić, University of Belgrade, Serbia

Ben C.K. Ngan, Worcester Polytechnic Institute, USA

Daouda Kamissoko, IMT Mines Albi, University of Toulouse, France

Dragana Bečejski-Vujaklija, Serbian Society for Informatics, Serbia

Fátima Dargam, SimTech Simulation Technology/ILTC, Austria

Femi Olan, University of Essex, UK

Fernando Tricas Garcia, University of Zaragoza, Spain

Festus Oderanti, Liberty University, USA

Francisco Marmolejo-Cossio, Boston College, USA

George Tsaples, University of Macedonia, Greece

Guoqing Zhao, Swansea University, UK

Isabelle Linden, University of Namur, Belgium

Jason Papathanasiou, University of Macedonia, Greece

João Lourenço, Universidade de Lisboa, Portugal

Jorge Freire de Sousa, Engineering University of Porto, Portugal

José Maria Moreno-Jiménez, Zaragoza University, Spain

Karim Soliman, Arab Academy for Science, Technology, and Maritime Transport, Egypt

Kathrin Kirchner, Technical University of Denmark, Denmark

Konstantinos Vergidis, University of Macedonia, Greece

Marc Kilgour, Wilfrid Laurier University, Canada

María Teresa Escobar, University of Zaragoza, Spain

Marko Bohanec, Jozef Stefan Institute, Slovenia

Md Asaduzzaman, Staffordshire University, UK

Milan Vukićević, University of Belgrade, Serbia

Miloš Jovanović, University of Belgrade, Serbia

Panagiota Digkoglou, University of Macedonia, Greece

Pascale Zaraté, IRIT/Toulouse University, France

Pavlos Delias, Kavala Institute of Technology, Greece

Rudolf Vetschera, University of Vienna, Austria

Rita Ribeiro, UNINOVA – CA3, Portugal

Sandro Radovanović, University of Belgrade, Serbia

Sean Eom, Southeast Missouri State University, USA

Sérgio Pedro Duarte, University of Porto, Portugal

Shaofeng Liu, University of Plymouth, UK

Uchitha Jayawickrama, Loughborough University, UK

Vlasta Sikimić, Eindhoven University of Technology, Netherlands

Poster Session Committee Members

Aditya Gupta, Amazon, USA

Balaji Ingole, Gainwell Technologies LLC, USA

Jayant Tyagi, Salesforce, USA

Manan Buddhadev, IEEE, USA

Mohit Menghnani, Twilio, USA

Naveen Achyuta, Roblox, USA

Sandeep Ravindra Tengali, Atlassian, USA

Organizers

We would like to express our gratitude to our sponsors, whose generous support has been instrumental in the successful organization of this conference. Their commitment to fostering research, innovation, and collaboration has significantly contributed to the advancement of our field. We deeply appreciate their invaluable contribution, which has enabled us to create a platform for meaningful dialogue and impactful solutions. We look forward to continuing this partnership in promoting excellence and progress within our community.

Conference Venue

The conference will be held across three distinguished locations, each selected for its historical, cultural, and intellectual significance.

Day 1 – Cathedral of the Blessed Virgin Mary, Belgrade

The opening day will take place at the Cathedral of the Blessed Virgin Mary, situated in the heart of Belgrade. This distinguished ecclesiastical setting offers a reflective and dignified atmosphere, providing an inspiring commencement to the conference proceedings.

Days 2 & 3 – Palace of Science

The second and third days will be hosted at the Palace of Science, a prominent institution known for fostering academic excellence and interdisciplinary dialogue. Its state-of-the-art facilities and scholarly ambiance make it an ideal venue for the main sessions of the conference.

Day 4 – Viminacium Archaeological Site

The final day will be held at Viminacium, one of the most significant archaeological sites in the region and a former Roman provincial capital. This historically rich location offers a unique setting for concluding the conference, encouraging reflection on the enduring legacy of knowledge and civilization.

Publications

The 11th International Conference on Decision Support System Technology (ICDSST 2025) will have a special issue on Decision Support System Technology in the Artificial Intelligence Era in the International Transactions in Operational Research (ITOR), which is the official journal of IFORS. ITOR bridges the gap between theory and application in operational research, fostering collaboration between academics and practitioners. With a 2023 Journal Impact Factor of 3.1, it ranks highly in Operations Research & Management Science and Management categories.

Call for Papers

Digital technologies in the AI (Artificial Intelligence) era reshape how organizations and individuals make decisions in an increasingly dynamic and complex environment. In this context, Decision Support plays a central role by providing methods, models, and tools integrating emerging technologies to improve decision-making processes. With the possibilities created by AI tools and the advancement of digital technologies, new challenges and opportunities arise within the domain of decision support systems.

This special issue welcomes articles on theoretical advancements, practical applications, case studies, and methodological developments in decision support in the AI era.

The deadline for submissions is October 31, 2025.

Although this Call for Papers is open to the entire community of academics and practitioners, we particularly encourage the authors who presented their work at the EWG-DSS 11th International Conference on Decision Support System Technology (ICDSST2025) to submit their manuscripts.

Papers will be peer-reviewed according to the editorial policy of the International Transactions in Operational Research (ITOR), published by the International Federation of Operational Research Societies. Contributions should be prepared according to the instructions to authors in the journal homepage. Authors should upload their contributions using the submission site https://wiley.atyponrex.com/submission/dashboard?siteName=ITOR, indicating in their cover letter that the paper is intended for this special issue.

Other inquiries should be sent directly to the Guest Editors: Ana Paula Cabral Seixas Costa (apcabral@cdsid.org.br), Boris Delibašić (boris.delibasic@fon.bg.ac.rs), and José Maria Moreno Jiménez (moreno@unizar.es)

Table of Contents

Invited Speakers

Fátima Dargam

Fátima Dargam

REACH Innovation

CEO & Research Managing Director / Co-founder

Honorary Chair and Invited Speaker

Biography:
Dr. Fátima Dargam is an expert in Decision Support Systems and Artificial Intelligence, with a Ph.D. from Imperial College London and extensive experience as an R&D IT Manager at SimTech in Austria. Her work spans intelligent decision-making, multi-agent systems, and big data analytics, with significant contributions through research, publications, and her role as a founding member and former coordinator of the EURO Working Group on DSS.

Title:
Tracing the Evolution of AI Integration in Decision Support: A Review of EWG-DSS and ICDSST Contributions

Abstract:
Over the past three decades, Decision Support Systems (DSS) have increasingly incorporated Artificial Intelligence (AI) to enhance Decision-Making (DM) processes. The EURO Working Group on Decision Support Systems (EWG-DSS) has been at the forefront of this evolution. Already in the EWG-DSS foundation, at the EURO Summer Institute on DSS in the Madeira Island in1989, I presented a paper extracted from my Master Thesis, about an Expert System developed for the Brazilian Navy Research Institute (IPqM), which instantiated the way that AI could be used in *DSS for military applications*[1]. Following that, the EWG-DSS has promoted a rich body of research, through its annual organized Workshops and the International Conferences on DSS Technology (ICDSST), capturing how AI techniques have become integral to modern DSS. The current Abstract-Review synthesizes key insights from EWG-DSS and ICDSSTs published contributions, including Workshops and Conferences Proceedings; Springer LNBIP volumes; and Journal Special Issues, to trace the historical evolution of AI in DSS within our scenario. This abstract also highlights major technical and methodological innovations, identifying emerging thematic trends, like: Cognitive DSS, Sustainability, Data Visualization, Collaborative Decision-Making, among others; and emphasizing the role and importance of international and

[1] F.C.C. Dargam et al. "Decision Support Systems for Military Applications", EJOR European Journal of Operational Research 55 (1991) 403-408.

multidisciplinary research collaboration in those developments.

When we review the evolution of AI in Decision Support Systems, it is to be noticed that early DSS research in the 1990s began fusing AI with traditional decision models, for example by integrating expert systems and knowledge bases into DSS architectures. During the 2000s, the rise of the Internet and Groupware shifted focus towards Collaborative and Networked DSS, enabling distributed teams to make decisions jointly. In this period, the EWG-DSS promoted publications and initiatives emphasized formal decision methodologies and collaborative technologies. As an example, we can cite the 2009 Special Issue in EJOR, which addressed "Formal Tools and Methodologies for DSS"[2], laying groundwork for the so-called "Intelligent DSS". In the 2010s, rapid advances in Data Science brought Data Analytics and Machine Learning to the forefront. The community responded with dedicated forums on Big Data and Predictive Analytics in Decision Support. In this respect, from 2015 we can cite the Springer LNBIP-216 publication "*Decision Support Systems V – Big Data Analytics for Decision Making*"[3], which marked the 1st ICDSST in 2015 also in Belgrade, organized by the same core-team of the ICDSST 2025 now. By that time, themes around Knowledge Management were also prominent, illustrated in the publication of the Springer LNBIP-221 "*DSS IV – Information and Knowledge Management in Decision Processes*"[4], reflecting the integration of AI for handling knowledge and information. By the late 2010s, DSS research had expanded to address global and complex contexts, from "*DSS for Sustainability and Societal Challenges" (2016)*, *Springer LNBIP-250*[5], to *"Main Developments and Future Trends" (2019)*[6], culminating in recent focus on Cognitive Decision Support Systems that leverage AI to mimic or augment human cognition, with significant publications derived from EWG-DSS promoted conferences: ICDSST-2020 *LNBIP-384*[7], and *IJDSST(13) Special Issue*[8]. In the 2020s, the tendency within our promoted research work followed (*in 2021, Springer LNBIP-414*)[9] the research lines of: evidence-based reasoning for decision making; probabilistic inference and explainable machine learning (ML); digital resilience fake news, and management of misinformation during crisis (post-covid phase); fair machine learning models, and ML approaches for sentiment analysis; the use of AI to industrial planning, with implemented case-studies; web-based platforms supporting decision systems in different engineering and management applications; and the use of AI for validation and interpretation of physical documents. In 2022 (*Springer LNBIP-447*)[10], the most predominant research within the publications derived from EWG-DSS promoted events followed the lines of: DSS addressing modern industrial challenges; real-time anomaly detection systems; blockchain and AI technologies in DSS; sustainable engagement and co-planning multicriteria DSS; and multicriteria models

[2] P. Zaraté (editor). Formal Tools and Methodologies for DSS, Special issue EJOR European Journal of Operational Research, Vol. 195 N.3, June 2009, Elsevier.

[3] B. Delibašić, et al. (editors). LNBIP 216 – Decision Support Systems V – Big Data Analytics for Decision Making, (2015). Published by Springer in Lecture Notes in Business Information Processing (LNBIP), volume 216. ISBN: 978-3-319-18532-3.

[4] I. Linden et al. (editors). LNBIP 221 – Decision Support Systems IV – Information and Knowledge Management in Decision Processes, (2015). Published by Springer in Lecture Notes in Business Information Processing (LNBIP), volume 221. ISBN: 978-3-319-21536-5.

[5] S. Liu et al. (editors). LNBIP 250 "Decision Support Systems VI – Addressing Sustainability and Societal Challenges" (2016). Published by Springer in Lecture Notes in Business Information Processing (LNBIP), volume 250. ISBN: 978-3-319-32876-8, 2016.

[6] P. Freitas et al. (editors). LNBIP 348 "Decision Support Systems IX: Main Developments and Future Trends" (2019). Proceedings of the 5th International Conference on Decision Support System Technology, EmC-ICDSST 2019, Funchal, Madeira, Portugal, May 27–29, 2019. Published by Springer in Lecture Notes in Business Information Processing (LNBIP), volume 348. ISBN: 978-3-030-18818-4, 2019.

[7] J. Moreno-Jiménez et al. (editors). LNBIP 384 "Decision Support Systems X: Cognitive Decision Support Systems and Technologies" (2020). Proceedings of the 6th International Conference on Decision Support System Technology, ICDSST 2020, Zaragoza, Spain, May 27–29, 2020. Published by Springer in Lecture Notes in Business Information Processing (LNBIP), volume 384. ISBN: 978-3-030-46223-9, 2020.

[8] B. Delibašić, et al. (editors). IJDSST 13(1) Special Issue on "AI for Intelligent Decision Support Systems" in the IJDSST International Journal of Decision Support System Technology (2019).

[9] U. Jayawickrama, et al. (editors). LNBIP 414 "Decision Support Systems XI: Decision Support Systems, Analytics and Technologies in Response to Global Crisis Management" (2021). Springer Book, Lecture Notes in Business Information Processing (LNBIP, volume 414). eBook ISBN978-3-030-73976-8. Published: 17 May 2021.

[10] A. Costa, et al. editors). LNBIP 447 "Decision Support Systems XII: Decision Support Addressing Modern Industry, Business, and Societal Needs" (2022). Springer Book, Lecture Notes in Business Information Processing (LNBIP, volume 447). eBook ISBN978-3-031-06530-9Published: 12 May 2022.

and systems for various energy sector applications. In 2023 (*Springer LNBIP-474*)[11], the tendency among the ICDSST-2023 publications followed the lines of: DSS applied to uncertainties; the contribution of digital twins to DSS and decision making; knowledge management processes framework applied to decisions pertinent to manufacturing supply chains; data-driven stakeholder-centred DSS; DSS applications for sustainability in health, energy and transportation sectors; and time aware optimization models for DSS, among others. As societal dependence on digital platforms and the real-time information offerings grows, there is an accentuated demand for a customized and intelligent decision support able of meeting the needs of modern industry, business, and society. In 2024, two major challenges were targeted to motivate the communities of the EWG-DSS and the GDN groups to present research results, namely: using technology to improve the decision process; as well as ensuring that decisions really support the best interest of the actors involved. Both challenges needed to consider the evolution of machine learning and AI in decision making, as well as their benefits, and the need to ensure that humans remain the main beneficiaries of new generated services, systems, and policies. In this period, the most predominant research within the publications derived from the joint event of EWG-DSS (*Springer LNBIP-506*)[12] and GDN (*Springer LNBIP-509*)[13], followed, among others, the research lines of: Argumentation-based approaches for reasoning and decision making under uncertainty, and in explainability of AI models; Advanced analytics practice with Gen.AI; Group Decision Negotiation in social and informational sciences; group decision-making process for cyberattack mitigation; ML for predictive detection of negotiation progress; Gen.AI in digital negotiations; zero-trust architecture for negotiation-based DSS; impact of Gen.AI tools on group discussions and decision-making; Human-AI collaboration - understanding expectation disconfirmation and trust; market evaluation based on descriptive and comparative information; Grey Neural Network applied for cost estimation of new market-products; Network Analysis of International Conflicts; ML models used for network attacks prediction and for supporting decision making; Quantum extensions of classical games and decision optimization; and DSS applied to politics and socio-environmental issues, all based on the topic of "Human-Centric Decision and Negotiation Support for Societal Transitions".

Considering the technical and methodological innovations in the AI-based DSS reviewed, we notice that parallel to this historical evolution, major technical and methodological innovations have emerged within AI-based DSS. Knowledge-driven DSS were an early innovation, combining rule-based AI systems with decision models to capture expert reasoning. Over time, these have been complemented by data-driven approaches that harness large datasets and machine learning algorithms. For instance, the integration of Big Data analytics into DSS has enabled more predictive and adaptive decision support. Optimization and simulation techniques from the AI and operations research domains have been incorporated to evaluate decision scenarios under uncertainty. Another key innovation was the development of intelligent agents and recommender systems within DSS, facilitating automated information filtering and negotiation support in multi-actor decisions (as seen in special issues on negotiation and networked decision-making). Methodologically, the field has seen advances in multi-criteria decision analysis (MCDA) and knowledge management frameworks integrated with AI, improving how preferences and expertise are modelled in DSS. Together, these innovations have transformed DSS into intelligent decision platforms

[11] S. Liu, et al. (editors). LNBIP-474 "Decision Support Systems XIII. Decision Support Systems in An Uncertain World: The Contribution of Digital Twins" (LNBIP, volume 474). eBook ISBN978-3-031-32534-2Published: 17 May 2023.

[12] S. Duarte, et al. (editors). LNBIP 506 "Decision Support Systems XIV. Human-Centric Group Decision, Negotiation and Decision Support Systems for Societal Transitions – EWG-DSS Publications" (2024). Springer Book, Lecture Notes in Business Information Processing (LNBIP, volume 506). eBook ISBN978-3-031-59376-5Published: 01 May 2024.

[13] M. Ferreira, et al. (editors). LNBIP 509 "Decision Support Systems XIV. Human-Centric Group Decision, Negotiation and Decision Support Systems for Societal Transitions – GDN Publications" (2024). Springer Book, Lecture Notes in Business Information Processing (LNBIP, volume 509). eBook ISBN978-3-031-59373-4Published: 10 May 2024.

that combine data-centric AI algorithms with domain knowledge and robust decision models, including also the use of Large Language Models (LLM) within Gen.AI tools to support DSS.

About emerging thematic-trends in DSS research within the promoted publications of EWG-DSS and ICDSST, it is evident that most recent contributions reveal several emerging trends that shape the future of AI-integrated DSS research, like: Cognitive DSS, as there is growing emphasis on this topic to emulate human cognitive processes and learning capabilities. These systems apply AI techniques like natural language processing, machine learning, and reasoning engines to provide more context-aware and adaptive support. The ICDSST 2020 theme, "Cognitive Decision Support Systems and Technologies", exemplifies this trend, with the presentation of DSS that can interact with users in more human-like ways and continuously improve by learning from data and feedback. This also became explicit with the growth of the integration of Generative AI in DSS approaches (more visible in the ICDSST-2024).

In terms of sustainability in DSS development over the years, it is observed that sustainable decision-making has become a core theme, reflecting global challenges such as environmental management and social responsibility. AI-powered DSS are now designed to incorporate sustainability criteria and long-term impact analysis. The community's focus is evident in dedicated efforts like "DSS VI – Addressing Sustainability and Societal Challenges" (2016) and special issues on sustainable DSS applications (e.g.: in agriculture and industry). These works highlight how multidisciplinary approaches (combining AI, environmental science, and policy) are enabling DSS to support the UN Sustainable Development Goals and other sustainability initiatives.

Data visualization and data analytics also hold important position in the DSS approaches reviewed, as the systems become increasingly data-driven, transforming complex data into intuitive visuals and insights is critical. Data visualization has therefore emerged as an important research front. Recent conferences and publications have stressed interactive data exploration and visual analytics in decision support. For example, "DSS VII – Data, Information and Knowledge Visualization in DSS" (2017) focused on techniques to improve how data and model outputs are communicated to decision-makers. This trend recognizes that AI algorithms must be paired with human-centric design – dashboards, visual decision maps, and explainable AI – to truly augment human decision-making.

Building on the longstanding interest of the EWG-DSS, the community continues to advance in Collaborative Decision-Making systems. Modern collaborative DSS facilitate real-time, distributed teamwork and consensus-building, often integrating social media, cloud platforms, and AI-driven coordination aids. The importance of this theme is reflected in special issues such as "Collaborative Decision Processes and Analysis" (EURO Journal on Decision Processes, 2015) and earlier works on networking decision support. Current research extends these ideas with AI features – for instance, intelligent facilitation agents or recommender systems to mediate group discussions – aiming to improve the effectiveness of team decisions across geographic and disciplinary boundaries.

Overall, the progression of the usage of AI-based approached reviewed here illustrates a clear trajectory from early experiments in AI-driven decision support to today's sophisticated AI-rich DSS paradigms, most recently with the integration of Generative AI to support decision making in various applications.

For over 3 years, the AI research area has been experiencing growing development and usage of Generative AI (Gen.AI) tools. This "boom" came to be optimized, expanded, regulated,

envisaging a long-term and successful history in the AI applications roadmap. It is well known that the integration of Gen.AI into Decision Support Systems (DSS) enhances the system's ability to assist human decision-making through more dynamic, adaptive, and creative outputs. Such integrated systems can include key contributions in various implementation features, like: Scenario Generation, creating multiple hypothetical scenarios to test decisions under uncertainty; Data Augmentation, filling gaps in incomplete datasets by generating synthetic data to improve model reliability: Natural Language Interfaces, allowing users to interact with the DSS using conversational queries, making insights more accessible; Automated Reporting, drafting summaries, recommendations, and visualizations to support decisions; as well as Simulation and Modelling, generating simulations or synthetic models to explore consequences of decisions in complex environments. The major benefits of DSS integrating generative AI can be cited as: Faster and more adaptable decision-making; Greater personalization of insights; Ability to handle unstructured or incomplete data; and Enhanced creativity in exploring alternatives. There are, however, several challenges and risks to be faced and mitigated, like for instance: The risk of generating biased or misleading outputs; the need for trustworthiness, explainability and transparency; the integration into existing workflows and validation of generated content, among others. The research in this area is highly important, despite the challenges, as Generative AI can significantly empower DSS by expanding from static data retrieval to proactive, creative support, offering richer tools for decision-makers while raising important considerations around trust, ethics, and interpretability.

In conclusion, the evolution of DSS, as traced through the perspective of the EWG-DSS and the ICDSST contributions, showcases a dynamic interplay between Decision Science and AI Technology. From early knowledge-based systems to today's cognitive and data-driven decision platforms, AI integration has continually expanded the capabilities and scope of DSS. Notably, this progress has been driven by robust international collaboration and multidisciplinary efforts. The EWG-DSS network itself, grown from a small 1989 gathering to over 350 members worldwide, exemplifies how diverse expertise (spanning operational research, computer science, and domain specialties) has merged to push DSS research forward. Such collaboration has yielded significant contributions / publications; and has also fostered innovation in methodologies and applications of DSS. Hence, the past thirty years of the EWG-DSS work reveal a field that has continually reinvented itself by embracing AI advancements and new decision challenges. This ongoing synergy between AI and DSS promises to further enhance decision support technologies for complex problem-solving in the years ahead.

Pascale Zaraté

Pascale Zaraté

Université Toulouse Capitole- IRIT

Full Professor

Honorary Chair and Invited Speaker

Biography:
Pascale Zaraté is a professor at Toulouse Capitole University, France. She conducts her research at the IRIT Laboratory and chair the Artificial Intelligence department. She holds a PhD. at University Paris Dauphine, Paris (1991) and a Habilitation to Conduct Research at Institut national Polytechnic Toulouse. Pascale Zaraté's current research interests include: decision support systems, cooperative decision-making, Group Decision Making. She has been coordinator of the European working group on DSS for 20 years. She is currently President of the INFORMS GDN section. She edited more than 30 special issues in several international journals, several books and proceedings of international conferences. She has published 3 books and more than 40 articles in several international journals, seven chapters in collective works, a more than 60 articles in international conferences. She is a member of the scientific editorial board of seven international journals. She organised several international conferences focusing on Decision Support Systems.

Title:
Decision Support: History and trends

Abstract:
Decision Support Systems (DSS) were born in the 70's. Since this decade, we can find in the literature several tracks of evolution for these systems. We will draw the evolution of DSS: describing the architecture, the usability, the functionalities, the programming techniques as well as the way to use these systems. We will show how Artificial Intelligence and Machine Learning influenced the development of DSSs. Different projects will be presented related to DSSs.

Zoran Obradović

Zoran Obradović

Director, Center for Data Analytics and Biomedical Informatics

Temple University

Invited Speaker

Biography:
Zoran Obradović is a Distinguished Professor and a Center director at Temple University, an Academician at the Academia Europaea (the Academy of Europe), and a Foreign Academician at the Serbian Academy of Sciences and Arts. He mentored about 50 postdoctoral fellows and Ph.D. students, many of whom have independent research careers at academic institutions (e.g. Northeastern Univ., Ohio State Univ,) and industrial research labs (e.g. Amazon, eBay, Facebook, Hitachi Big Data, IBM T.J.Watson, Microsoft, Yahoo Labs, Uber, Verizon Big Data, Spotify). Zoran is the editor-in-chief of the Big Data journal and the steering committee chair for the SIAM Data Mining conference. He is also an editorial board member of 13 journals and was the general chair, program chair, or track chair for 11 international conferences. His research interests include data science and complex networks in decision support systems addressing challenges related to big, heterogeneous, spatial, and temporal data analytics motivated by applications in healthcare management, power systems, earth, and social sciences. His studies were funded by AFRL, DARPA, DOE, NIH, NSF, ONR, PA Department of Health, US Army ERDC, US Army Research Labs, and industry. He published about 450 articles and is cited more than 34,000 times (H-index 70).

Title:
Lifting the Fog by Harnessing Interactions Across Diverse Systems

Abstract:
This presentation will discuss our machine learning approaches for identifying, categorizing, and forecasting events of interest from limited data and imprecise labels. This is accomplished by integrating multimodal structured and unstructured information and jointly modeling multiple types of interactions in temporal multilayer networks. Our findings indicate that a learning framework that integrates information from various sources of differing quality and resolution can significantly enhance informed decision-making.

Marko Bohanec

Marko Bohanec

Scientific Councillor

Department for Knowledge Tehnologies, Jožef Stefan Institute

Invited Speaker

Biography:
Marko Bohanec received a Ph.D. in Computer Science from the University of Ljubljana, Slovenia, in 1991. He is a leading Slovenian researcher in decision support models and systems. He currently serves as a Scientific Councillor at the Department of Knowledge Technologies, Jožef Stefan Institute, and a Full Professor of Computer Science at the University of Nova Gorica, Slovenia. His research interests include decision making, decision analysis, decision support, decision modeling, artificial intelligence, expert systems, machine learning, and data mining. He is a co-author of the method DEX (Decision EXpert) and its associated software, widely used for qualitative hierarchical multi-attribute decision modeling. Over the past decade, he has contributed as a decision support expert and developer of decision support systems in European projects spanning food and feed production, distribution, and quality control; healthcare disease management; nuclear safety; and sustainable mobility.

Title:
Exploring Synergies Between Human and Artificial Intelligence: The Case of the DEX Decision Modeling Method

Abstract:
Knowledge-driven Decision Support Systems (DSS) rely primarily on expert knowledge, rules, and heuristics to analyze situations and provide recommendations. Traditionally, they have been developed through Expert Modeling (EM), the process of capturing, formalizing, and structuring human expertise. However, the rise of Machine Learning (ML) has often overshadowed EM, creating the illusion that complex decision-making can be reduced to simply “pressing a button” to generate models from data. While ML is considered modern and efficient, critical requirements such as correctness, completeness, consistency, comprehensibility, and convenience often necessitate the integration of EM. This lecture explores the synergy between ML and EM, emphasizing the importance of combining AI with human expertise to develop robust DSS. We will examine how Large Language Models (LLMs) can assist in structuring decision criteria and modeling preferences, while acknowledging their limitations in capturing human decision preferences. These aspects will be illustrated using the qualitative decision modeling method DEX (Decision EXpert).

Lily Xu

Lily Xu

Postdoc at Oxford

(soon Assistant Professor at Columbia)

Invited Speaker

Biography:
Lily Xu develops methods across machine learning, optimization, and causal inference for planetary health challenges, with a focus on biodiversity conservation. She aims to enable practitioners to make effective decisions in the face of limited data, taking actions that are robust to uncertainty, effective at scale, and future-looking. In her work, Lily partners closely with NGOs to bridge research and practice, serving as AI Lead for the SMART Partnership. Since 2020, she has co-organized the EAAMO research initiative, committed to advancing Equity and Access in Algorithms, Mechanisms, and Optimization. Lily will join Columbia IEOR as an Assistant Professor in July 2025.

Title:
High-stakes decisions from low-quality data: AI decision-making for planetary health

Abstract:
Planetary health recognizes the inextricable link between human health and the health of our planet. Our planet's growing crises include biodiversity loss, with animal population sizes declining by an average of 70% since 1970, and maternal mortality, with 1 in 49 girls in low-income countries dying from complications in pregnancy or birth. Overcoming these crises will require effectively allocating and managing our limited resources. My research develops data-driven AI decision-making methods to do so, overcoming the messy data ubiquitous in these settings. Here, I'll present technical advances in multi-armed bandits, robust reinforcement learning, and causal inference, addressing research questions that emerged from on-the-ground challenges across conservation and maternal health. I'll also discuss bridging the gap from research and practice, with anti-poaching field tests in Cambodia, field visits in Belize and Uganda, and large-scale deployment with SMART conservation software.

AI, Machine Learning, and Big Data in Decision Support

Evaluating Generative Models for Synthetic Tabular Data: A Comparative Analysis of Fidelity, Diversity, and Generalization

Zoran Mahovac, Andrija Petrović, Sandro Radovanović, Boris Delibašić

Faculty of Organizational Sciences, Jove Ili´ca 154, Belgrade, 11000, Serbia
zm20200109@student.fon.bg.ac.rs, andrija.petrovic@fon.bg.ac.rs,
sandro.radovanovic@fon.bg.ac.rs, boris.delibasic@fon.bg.ac.rs

Abstract: Tabular data, such as relational tables, Web tables and CSV files, is among the most primitive and essential forms of data in machine learning, characterized by excellent structural properties, readability, and interpretability. However, acquiring substantial amounts of high-quality tabular data for ML model training remains a persistent challenge. This study evaluates the performance of six generative models — TVAE, RTVAE, CTGAN, ADSGAN, BNN, and Marginal Distributions on synthetic data generation. The evaluation is based on three key metrics: Fidelity, Diversity, and Generalization. Fidelity measures the quality of synthetic data, Diversity assesses how well the samples cover the variability of the real dataset, and Generalization quantifies the risk of overfitting. The research applies these metrics to four datasets: Abalone, Acute Inflammation, Census Income, and Pittsburgh Bridges. Results show that CTGAN consistently outperforms other models measured by IPα and IRβ metrics, while RTVAE excels in the Census Income dataset in terms of Generalization. Marginal Distributions stands out in preserving data authenticity. This study offers a refined method of evaluating generative models, emphasizing precision-recall analysis grounded in minimum volume sets, thus providing a deeper understanding of model performance across multiple dimensions.

Keywords: Generative models, synthetic data, Fidelity, Diversity, Generalization, tabular data generation

1. Introduction

Generative AI (GenAI) is a class of machine learning (ML) algorithms that can learn from content such as text, images, and audio in order to generate new content. In contrast to discriminative ML algorithms, which learn decision boundaries, GenAI models produce artifacts as output, which can have a wide range of variety and complexity. [1] A central problem in statistical inference is to calculate a posterior distribution of interest. Given a likelihood function, $p(y \mid \theta)$ or a forward model $y = f(\theta)$, and a prior distribution $\pi(\theta)$, the goal of an inverse probability calculation is to compute the posterior distribution $p(\theta|y)$.

This is notoriously hard for high-dimensional models. MCMC methods solve this by generating samples from the posterior using density evaluation. Generative AI techniques, on the other hand, directly learn the mapping from the a uniform to the distribution of the interest. The main advantage of generative AI is that it is model free and doesn't require the use of iterative density methods. [2]

Tabular data, such as relational tables, Web tables and CSV files, is among the most primitive and essential forms of data in machine learning (ML), characterized by excellent structural properties, readability, and interpretability. [3] However, acquiring substantial amounts of high-quality tabular data for ML model training remains a persistent challenge. [4] Additionally, in the industrial sector where tabular data is most commonly used, the availability of data is often limited due to privacy concerns. [5]

In this paper, we aimed to investigate to what extent generative models applied to tabular data can synthesize data that has high fidelity, meaning it belongs to the distribution of real data, is sufficiently divergent to describe the entire variability of the original data, and demonstrates the model's ability to generalize, i.e., not just produce copies of the training data. Generative models used for synthetic data generation include Bayesian Neural Network (BNN), Adversarial Deep Synthetic Generative Adversarial Network (ADSGAN), Conditional Generative Adversarial Network (CTGAN), Tabular

Variational Autoencoder (TVAE), Recurrent Temporal Variational Autoencoder (RTVAE), as well as a simpler model that learns the marginal distributions of all attributes individually. We used the implementation of these models from the synthcity [6] library.

To evaluate the performance of the generative models, we used metrics that rep-resent its performance as points in a 3-dimensional space, where each dimension corresponds to a different aspect of the model's quality. These qualities are: Fidelity, Diversity and Generalization. Fidelity measures the quality of a model's synthetic samples, and Diversity is the extent to which these samples cover the full variability of real samples, whereas Generalization quantifies the extent to which a model overfits (copies) training data. Since statistical comparisons of complex data types in the raw input space are difficult, the evaluation pipeline starts by embedding original training data and synthetic data into a meaningful feature space through an evaluation embedding function, and evaluating Fidelity, Diversity and Generalization on the embedded features [7] We verified the effectiveness of these models on four datasets: Abalone, Acute Inflammations, Pittsburgh Bridges, and Census Income.

2. Related work

Generative Models Based on Variational Autoencoder (VAE) VAEs are generative models that adopt variational inference and graphical models. VAE has two components: an encoder and a decoder. The encoder transforms high dimensional data to a low-dimensional latent space with an approximate tractable posterior distribution. The decoder samples from this distribution and transforms the sample back to the original dimension. [8] In variational methods, the functions used as prior and posterior distributions are restricted to those that lead to tractable solutions. For any choice of a tractable distribution q(Z), the distribution of the latent variable, the following decomposition holds:

$$\log \mathrm{p}_\theta\,(\mathrm{X}) = \mathcal{L}(\mathrm{q}(\mathrm{Z}), \theta) + \mathrm{D}_{\mathrm{KL}}\big(\mathrm{q}(\mathrm{Z}) \parallel \mathrm{p}_\theta(\mathrm{Z}|\mathrm{X})\big) \tag{1}$$

where DKL represents the Kullback-Leibler (KL) divergence. Instead of maximizing the loglikelihood $p_\theta(X)$, with respect to the model parameters θ, the variational inference approach maximizes its variational evidence lower bound ELBO [9]:

$$\mathcal{L}(\mathrm{q}, \theta) = \mathrm{E}_{\mathrm{q(Z)}}[\log \mathrm{p}_\theta\,(\mathrm{X}|\mathrm{Z})] - \mathrm{D}_{\mathrm{KL}}\big(\mathrm{q}(\mathrm{Z}) \parallel \mathrm{p}_\theta(\mathrm{Z})\big). \tag{2}$$

The Synthcity implementation of TVAE is based on a combination of a tabular encoder and a VAE to form a generative model for tabular data.

Notable approach in tabular data generation includes a novel approach based on Variational Autoencoders (VAEs) enhanced with a Bayesian Gaussian Mixture (BGM) model, which trains the VAE conventionally and then applies the BGM model to the learned latent space. This allows for a more accurate representation of the underlying data distribution during data generation. Moreover, this approach offers enhanced flexibility by accommodating various differentiable distributions for individual features, enabling the handling of continuous and discrete data types. [10]

An additional noteworthy proposition is a robust variational autoencoder with β divergence for tabular data (RTVAE) with mixed categorical and continuous features. In order to ensure robust variational inference, this approach modifies reconstruction term LREC in order to make it more robust to outliers for categorical data. In this approach, the empirical distribution $\hat{p}(X)$ is defined as a sum of Dirac delta functions representing discrete data points. Then, KL-divergence is used, which measures the difference between this empirical distribution and the generative distribution $p_\theta(X|z)$. In essence, KL-divergence quantifies the difference between the true data distribution and the model's distribution. KL-divergence is sensitive to outliers because for data with very low probability (e.g., outliers), the negative log-likelihood can become very large, disrupting the model's learning process. Instead of using KL-divergence, β-divergence is used, which is defined as [11]:

$$D_\beta\big(\hat{p}(X) \parallel p_\theta(X|z)\big) = \frac{1}{\beta}\int \hat{p}(X)^{\beta+1} dX - \frac{\beta}{\beta+1}\int \hat{p}(X) p_\theta(X|z)^\beta dX + \int p_\theta(X|z)^{\beta+1} dX \tag{3}$$

β-divergence is a generalization of KL-divergence, which converges to KL-divergence as $\beta \to 0$. By adjusting the β parameter, this approach becomes more robust to outliers, providing greater flexibility in modeling the data distribution. β-divergence is related to β-cross-entropy, which measures the similarity between the real and modeled distributions. Therefore, the reconstruction loss is expressed as [11]:

$$\mathcal{L}_{\mathcal{REC}-\beta} = \mathrm{E}_{\mathrm{z}\sim \mathrm{q}_{\phi}(\mathrm{z}|\mathrm{X})}\left[\mathcal{H}_{\beta}(\mathrm{x}_{\mathrm{i}}|\mathrm{z})\right] \quad (4)$$

Generative Adversarial Networks (GAN). In this framework two models are trained simultaneously: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. [12] To learn the generator's distribution p_g over data x, we define a prior on input noise variables $p_z(z)$, then represent a mapping to data space as $G(z;\theta_g)$, where G is a differen-tiable function represented by a multilayer perceptron with parameters θg. We also define a second multilayer perceptron $D(x;\theta_d)$ that outputs a single scalar. $D(x)$ represents the probability that x came from the data rather than pg. In other words, D and G play the following two-player minimax game with value function V(G, D): [12]

$$\min_G \max_D V(D,G) = E_{x\sim p_{\mathrm{dt}}(x)}[\log D(x)] + E_{z\sim p_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right] \quad (5)$$

Tabular data usually contains a mix of discrete and continuous columns. Continuous columns may have multiple modes whereas discrete columns are sometimes imbalanced making the modeling difficult. Existing statistical and deep neural network models fail to properly model this type of data. A conditional GAN (CTGAN) as a synthetic tabular data generator is proposed to address issues mentioned above. Xu et al. [13] invented mode-specific normalization to over-come the non-Gaussian and multimodal distribution. This method is applied for each continuous column in three steps; (1) for each continuous column Ci variational Gaussian mixture model (VGM) is being used to estimate the num-ber of modes mi and fit a Gaussian mixture. The learned Gaussian mixture is $P_{C_i}(c_i, j) = \sum_{k=1}^{3} \mu_k\, N(c_{i,j}; \eta_k, \phi_k)$, where μ_k and ϕ_k are the weight and standard deviation of a mode, respectively. (2) Then, for each value $c_{i,j}$ in C_i, compute the probability of $c_{i,j}$ coming from each mode. (3) Sample one mode from given the probability density and use the sampled mode to normalize the value. [13] The traditional GAN generator takes a random vector sampled from a standard multivariate normal distribution (MVN) as input. During training, the generator learns to map this distribution to the data distribution. However, this approach does not account for the potential imbalance in categorical variables. If the data is randomly sampled, underrepresented categories may not be sufficiently represented, leading to poor generator training. Alternatively, if the data is resampled to balance categories, the generator may learn an altered distribution that does not reflect the true data distribution. The solution to this issue is the use of a Conditional Generator in GANs. It generates data conditioned on specific categorical values, ensuring balanced representation during training. This approach learns the true conditional distribution of the data, maintaining the integrity of the original data distribution during both training and testing.

In the medical and machine learning communities, AI has the potential to trans-form personalized treatment and decision-making. However, legal and ethical concerns regarding unconsented patient data and privacy limit data sharing, restricting access to electronic health records (EHR). To address this, a novel framework, ADS-GAN, uses conditional generative adversarial networks (GANs) to generate synthetic data that closely approximates the original EHR dataset while minimizing patient identifiability. The framework introduces a quantifiable definition of "identifiability" based on re-identification probabilities. ADS-GAN outperforms existing methods, providing a legally and ethically sound solution for open data sharing and AI development. [14]

Bayesian Neural Networks (BNN). Bayesian Neural Networks (BNNs) rep-resent a powerful generative approach that integrates uncertainty estimation into the modeling process. Unlike standard

neural networks, which learn fixed weight values, BNNs assign probabilistic distributions to weights, enabling a principled way to quantify uncertainty. This characteristic makes them particularly useful for synthetic data generation, where capturing both the underlying data struc-ture and variability is essential. A natural extension of BNNs in this context involves their formulation within Directed Acyclic Graphs (DAGs), where prob-abilistic dependencies between variables are explicitly modeled. By structuring BNNs as DAGs, the generative process becomes interpretable, as each node rep-resents a latent or observed variable, and directed edges capture conditional dependencies. This setup is particularly advantageous in tabular data synthesis, where relationships between features are often complex and hierarchical. [16]

3. Experimental setup

The main goal of this research is to evaluate and compare the performance of different generative models for synthetic data in the context of their ability to represent the structure of the original data. Specifically, the models TVAE, RT-VAE, CTGAN, ADSGAN, BNN, and Marginal Distributions are analyzed using the metrics Fidelity, Divergence, and Generalization. Fidelity measures the quality of a model's synthetic samples, and Diversity is the extent to which these samples cover the full variability of real samples, whereas Generalization quantifies the extent to which a model overfits (copies) training data. [7] This is an alternative approach to evaluating generative models, where instead of assessing the generative distribution by looking at all synthetic samples collectively to compute likelihood or statistical divergence, we classify each sample individually as being of high or low quality. [7] The main contribution proposed by Alaa et al [7] is a refined precision-recall analysis of the Fidelity and Diversity performance of generative models that is grounded in the notion of minimum volume sets. Given a probability measure P and a reference measure μ, one is often interested in the minimum μ - measure set with P - measure at least α. Minimum volume sets of this type summarize the regions of greatest probability mass of P, and are useful for detecting anomalies and constructing confidence regions. [15] As mentioned, the approaches were evaluated on the following datasets: the Abalone, Acute Inflammation, Census Income, and Pittsburgh Bridges datasets.

We first preprocessed the data using two distinct pipelines: one for continuous variables, and another for categorical variables. The data were then split into training and test sets with an 80:20 ratio. Hyperparameter optimization was performed by sampling hyperparameter values over five trials, with the best parameters selected based on evaluation metrics. We trained the evaluation neural network on the training data, while synthetic data were passed through the trained network, and evaluation metrics were computed on both the original and synthetic embeddings.

4. Discussion

In this section, we will discuss the data preprocessing steps applied to handle continuous and categorical features using different strategies for imputation and transformation. We will also cover the neural network architecture imple-mentation used for embedding representations and how it was trained. Finally, we will describe results of the following approaches: TVAE, RTVAE, CTGAN, ADSGAN, BNN, and MD using evaluation metrics Fidelity, Diversity, and Generalization on a dataset that includes the Abalone, Acute Inflammation, Census Income, and Pittsburgh Bridges datasets.

Data preprocessing. For data preprocessing, two pipelines were created: (1) for continuous data, which involves filling missing values with *SimpleImputer* using the constant strategy and *StandardScaler* for data standardization, and (2) for categorical data, which uses *SimpleImputer* with the most frequent strategy, along with *OneHotEncoder* for transformation. For each dataset, the data were split with a training-to-test ratio of 80:20. Hyperparameter optimization was performed by sampling hyperparameter values for each model from their respective hyperparameter spaces in five trials, with the best hyperparameters selected based on the evaluation metric values when comparing the training dataset and synthesized examples. The network was trained on the training data, while the synthetic data were passed through the trained network, and evaluation metrics were computed on the original and synthetic embeddings.

Neural Architecture Evaluator. To evaluate the performance of all mod-els, both synthetic and original training data were transformed into embedding representations using a neural network. The

architecture of the network was determined by modules composed of Dropout, Linear, and Nonlinear layers (activation functions). Our architecture consisted of 3 such modules with a dropout probability of 0.4. The hidden layer dimension of the linear transformations was set to 200, and the ReLU activation function was used for the nonlinear trans-formation. The embedding representations were centered around a point defined by a vector of value 1 in all dimensions, ensuring they belong to the interior of a hypersphere with a radius of 1. The learning rate parameter used when fitting the neural network was 0.0001. The neural network was trained for 100 epochs using the AdamW optimizer with a weight decay parameter value of 0.01.

Fidelity implementation. The implementation of evaluation metrics is based on [7]. IPα is implemented by integrating the absolute difference between α -precision and the alpha value describing the support, over all α - support values from 0 to 1 and a scalar in the range [0, 1/2] is obtained. α - precision is implemented by calculating the Euclidean distance of each real point from the center of the hypersphere using the norm function from the torch library, and the radius containing α * 100% of the training embedding distances from the center of the hypersphere is returned by the quantile function. Then, the Euclidean distances of the synthetic points' embeddings from the center of the hypersphere are calculated, and the total number of such distances less than the radius describing the α - quantile of the training points' embeddings is summed (the total number of reconstructions belonging to the training support). Finally, α precision is defined as the proportion of these points in the total number of embeddings of synthetic points.

Diversity implementation. For the implementation of the Diversity metric, the center of the synthetic data embeddings was calculated as the mean of all the vectors across all dimensions. A radius was determined that encompasses the beta quantile of all distances between the synthetic data embeddings and their center. All embeddings of synthetic data points belonging to the beta support, defined by the hypersphere with the center and radius described above, were filtered. Using these, a *NearestNeighbors* model was fitted, and for each training data embedding, the closest embedding from the synthetic data was found. Additionally, another *NearestNeighbors* model was fitted on the embeddings of real data, with the number of neighbors set to k = 5. Finally, if the distance from the closest point in the beta support to the real data embedding is smaller than the distance of that embedding to its k nearest neighbors, it can be concluded that the variability of the real data point is represented in the synthetic examples. The proportion of real data points whose variance is explained by the synthetic examples, relative to the total number of real data points for a given value of the beta support, represents the beta recall.

Generalization implementation. The Generalization metric was also calculated by comparing the distances between real and synthetic data point embed-dings using the *NearestNeighbors* algorithm. For each real data point, the closest real neighbor was identified, and the distance to its nearest real neighbor was recorded. Similarly, for each synthetic data point, the distance to its closest real data point was calculated. A mask was then applied to check if the synthetic data point's distance to the nearest real neighbor was greater than the real data point's distance to its nearest neighbor. The authenticity ratio was computed as the proportion of synthetic data points satisfying this condition, indicating how well the synthetic data approximates real data.

The implementation of generative models is based on the implementation of existing models from the Synthcity [6] library.

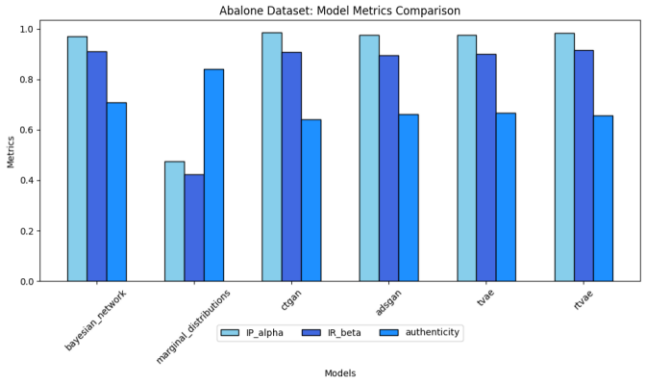


Figure 1. Model metrics comparison on Abalone Dataset

Results. For the Abalone dataset, the CTGAN model emerged as the best, achieving the highest values in IPα (0.9856) and IRβ (0.9076), with an Authenticity score of 0.6411. This model demonstrated exceptional effectiveness in pre-serving information and data representation. On the other hand, the Marginal Distributions model showed significantly lower values for IPα and IRβ, although it excelled in the Authenticity metric (0.8402), indicating its ability to preserve the authenticity of the data, even though it wasn't the most efficient in other metrics. The visual representation of the metric values for the Abalone dataset is shown in 1.

In the Census Income dataset, the RTVAE model showed outstanding performance with values of 0.8475 for IPα, 0.7578 for IRβ, and 0.75 for Authenticity, making it the most successful model in this dataset. Marginal Distributions once again showed weakness in the IPα (0.2899) and IRβ (0.5831) metrics, but recorded a very good result for Authenticity (0.8583), suggesting its ability to generate authentic data. The visual representation of the metric values for the Census Income dataset is shown in 2.

For the Acute Inflammations dataset, the CTGAN again achieved the best results with an IPα value of 0.9382, IRβ value of 0.7906, and Authenticity value of 0.5729. These results highlight its high effectiveness in preserving information and generating relevant data. In comparison, the Bayesian Network had a higher value for IRβ (0.8449), but did not achieve the best results in other metrics. The visual representation of the metric values for the Acute Inflammations dataset is shown in 3.

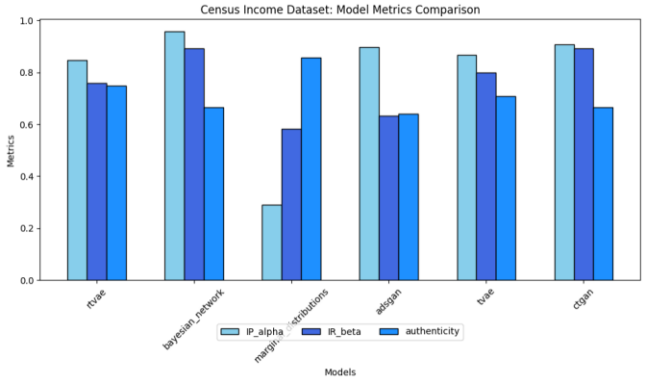


Figure 2. Model metrics comparison on Census Income Dataset

On the final dataset, Pittsburgh Bridges, CTGAN once again demonstrated exceptional results, with the highest values for IRβ (0.8733) and Authenticity (0.7143), while RTVAE delivered solid values across all three metrics, but did not reach the same heights as CTGAN. ADSGAN had the lowest values in all metrics, indicating that this model was not the most efficient on this dataset. The visual representation of the metric values for the Pittsburgh Bridges dataset is shown in 4.

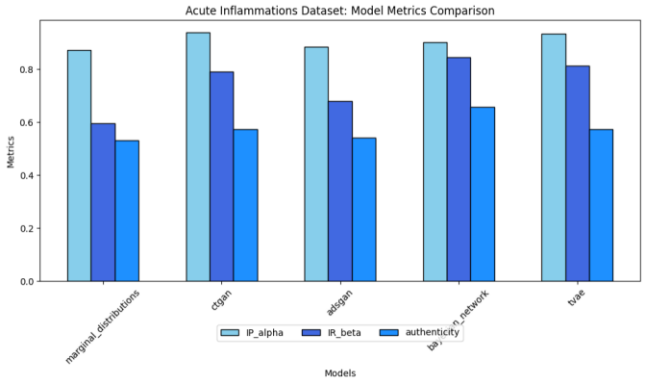


Figure 3. Model metrics comparison on Acute Inflammations Dataset

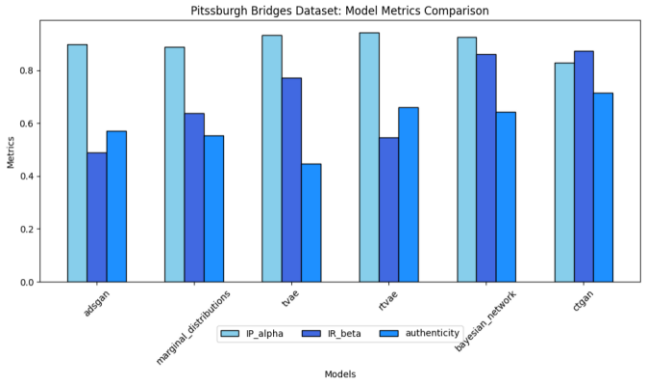


Figure 4. Model metrics comparison on Pittsburgh Bridges Dataset

In conclusion, CTGAN was generally the most effective model in this study, achieving the highest values in IPα and IRβ across most datasets, while RTVAE was the best on the Census Income dataset, particularly in terms of Generalization. Marginal Distributions was competitive in preserving the authenticity of the data, while Bayesian Network and ADSGAN had various strengths in certain metrics but did not reach the highest values in all cases. The detailed values of the metrics can be found in the tables below.

Table 1. Results for Abalone Dataset

Model	**IPα**	**IRβ**	**Authenticity**
Bayesian network	0.9706	0.9111	0.7085
Marginal distributions	0.4746	0.4225	**0.8402**
CTGAN	**0.9856**	0.9076	0.6411
ADSGAN	0.9772	0.8962	0.6618
TVAE	0.9764	0.9009	0.6675
RTVAE	0.9841	**0.9165**	0.6561

Table 2. Results for Census Income Dataset

Model	**IPα**	**IRβ**	**Authenticity**
RTVAE	0.8475	0.7578	0.7500
Bayesian network	**0.9572**	0.8928	0.6667
Marginal distributions	0.2900	0.5831	**0.8583**
ADSGAN	0.8987	0.6322	0.6417
TVAE	0.8665	0.8000	0.7083
CTGAN	0.9070	**0.8928**	0.6667

Table 3. Results for Acute Inflammations Dataset

Model	IPα	IRβ	Authenticity
Marginal distributions	**0.8722**	0.5950	0.5313
CTGAN	0.9382	0.7906	0.5729
ADSGAN	0.8845	0.6798	0.5417
Bayesian network	0.9026	**0.8450**	**0.6563**
TVAE	0.9346	0.8137	0.5729
RTVAE	0.8213	0.7511	0.5822

Table 4. Results for Pittsburgh Bridges Dataset

Model	IPα	IRβ	Authenticity
ADSGAN	0.8982	0.4890	0.5714
Marginal distributions	0.8884	0.6366	0.5536
TVAE	0.9322	0.7712	0.4464
RTVAE	**0.9419**	0.5463	0.6607
Bayesian network	0.9260	0.8615	0.6429
CTGAN	0.8278	**0.8733**	**0.7143**

5. Conclusion

In this research, CTGAN emerged as the most effective model overall, with superior results in the Fidelity and Diversity metrics across most datasets. RTVAE demonstrated excellent performance on the Census Income dataset, particularly in Generalization, making it a strong contender for real-world applications where minimizing overfitting is crucial. While the Marginal Distributions model was not the best in most metrics, it excelled at preserving the authenticity of the data, which may be valuable in specific scenarios where data authenticity is a priority. Other models, such as BNN and ADSGAN, showed competitive strengths but did not dominate in all evaluated metrics. Overall, this study highlights the im-portance of a comprehensive evaluation framework when comparing generative models, emphasizing the need to balance Fidelity, Diversity, and Generalization to effectively generate synthetic data that is both accurate and varied.

References

1. Sun, J., Liao, Q. V., Muller, M., Agarwal, M., Houde, S., Talamadupula, K., Weisz, J. D.: Investigating Explainability of Generative AI for Code through Scenario-based Design. arXiv preprint arXiv:2202.04903, 2022.

2. Polson, N., Sokolov, V.: Generative Modeling: A Review. arXiv preprint arXiv:2501.05458, 2024.

3. Borisov, V., Seßler, K., Leemann, T., Pawelczyk, M., Kasneci, G.: Language Models are Realistic Tabular Data Generators. arXiv preprint arXiv:2210.06280, 2023.

4. Liu, J., Chai, C., Luo, Y., Lou, Y., Feng, J., Tang, N.: Feature Augmentation with Reinforcement Learning. In: 2022 IEEE 38th International Conference on Data En-gineering (ICDE), pp. 3360–3372. IEEE Press, 2022. https://doi.org/10.1109/ ICDE53745.2022.00317

5. Li, Y., Li, Z., Zhang, K., Dan, R., Jiang, S., Zhang, Y.: ChatDoctor: A Medical Chat Model Fine-Tuned on a Large Language Model Meta-AI (LLaMA) Using Medical Domain Knowledge. Cureus 15 (n. d.), e40895. https://doi.org/10.7759/cureus. 40895

6. Qian, Z., Cebere, B.-C., van der Schaar, M.: Synthcity: facilitating innovative use cases of synthetic data in different data modalities. arXiv preprint arXiv:2301.07573, 2023.

7. Alaa, A. M., van Breugel, B., Saveliev, E., van der Schaar, M.: How Faithful is your Synthetic Data? Sample-level Metrics for Evaluating and Auditing Generative Mod-els. arXiv preprint arXiv:2102.08921, 2022. https://arxiv.org/abs/2102.08921

8. Akrami, H., Aydore, S., Leahy, R. M., Joshi, A. A.: Robust Variational Autoencoder for Tabular Data with Beta Divergence. arXiv preprint arXiv:2006.08204, 2020.

9. Akrami, H., Joshi, A. A., Li, J., Aydore, S., Leahy, R. M.: Robust Variational Autoencoder. arXiv preprint arXiv:1905.09961, 2019.

10. Apella´niz, P. A., Parras, J., Zazo, S.: An improved tabular data generator with VAE-GMM integration. arXiv preprint arXiv:2404.08434, 2024. https://arxiv.org/ abs/2404.08434

11. Akrami, H., Aydore, S., Leahy, R. M., Joshi, A. A.: Robust Variational Autoen-coder for Tabular Data with Beta Divergence. arXiv preprint arXiv:2006.08204, 2020. https://arxiv.org/abs/2006.08204

12. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative Adversarial Networks. arXiv preprint arXiv:1406.2661, 2014.

13. Xu, L., Skoularidou, M., Cuesta-Infante, A., Veeramachaneni, K.: Modeling Tab-ular Data using Conditional GAN. arXiv preprint arXiv:1907.00503, 2019.

14. Yoon, J., Drumright, L. N., van der Schaar, M.: Anonymization Through Data Syn-thesis Using Generative Adversarial Networks (ADS-GAN). IEEE J Biomed Health Inform, 24(8):2378-2388, 2020. https://doi.org/10.1109/JBHI.2020.2980262.

15. Scott, C. D., Nowak, R. D.: Learning Minimum Volume Sets. Journal of Machine Learning Research, 7, 665–704, 2006.

16. Goan, E., Fookes, C.: Bayesian Neural Networks: An Introduction and Survey. In: Case Studies in Applied Bayesian Data Science, Springer International Publishing, 45–87, 2020. https://doi.org/10.1007/978-3-030-42553-1 3.

The Role of Big Data Analytics in Enhancing Oil and Gas Drilling Operations Performance: a thematic analysis

Mohammad Jarkhi, Shaofeng Liu, Jonathan Moizer, Jiang Pan

Plymouth Business School, University of Plymouth, Plymouth, United Kingdom
mohammad.jarkhi@plymouth.ac.uk, shaofeng.liu@plymouth.ac.uk, jonathan.moizer@plymouth.ac.uk
jiang.pan@plymouth.ac.uk

Abstract: The global energy market relies heavily on the oil and gas (O&G) sector, a key component of the global economy. Drilling is a critical and costly process within the O&G value chain. It is essential for discovering oil and gas, making every decision during drilling operations vital to avoiding delays. Data can be collected and analysed during drilling from various sources, including logged activities, in-well sensors (real-time data), incident reports, and Daily Drilling Reports (DDR). Nevertheless, international O&G companies continue to face significant challenges in decision-making due to inadequate support from advanced analytics and the vast volume of available data.

This study explores the role of Big Data Analytics (BDA) in supporting decisions related to drilling operations performance. It addresses three research questions: What key performance indicators (KPIs) are used to evaluate drilling operations performance? What are the applications of BDA in optimising drilling operations? What challenges arise in analysing drilling data? To answer these questions, thirty practitioners from various Middle Eastern O&G drilling enterprises participated in semi-structured interviews as part of the empirical data collection. Thematic analysis was used to examine the interview data, leading to the identification of six KPI categories: operational efficiency, cost management, well performance, equipment maintenance, safety and compliance, and environmental impact. The study also identified three stages of BDA: descriptive, diagnostic, and predictive. Finally, both technical and non-technical challenges were found to hinder effective data analysis and decision-making. The study contributes a comprehensive framework that links KPIs to BDA stages and supports improved managerial practices in drilling operations.

Keywords: Oil and Gas industry; Drilling operations; Big data analytics; Key Performance Indicators; Decision Support

1. Introduction and background

Large datasets are increasingly generated in the O&G sectors due to technological advancements, raising concerns among companies about data management (Mohammadpoor & Torabi, 2020). Improvements in drilling operations and modern technologies have increased data collection through advanced measurement tools. The speed of this data flow requires companies to use advanced big data analytics. (Nguyen et al., 2020). Companies' reliance on data collection and analysis using advanced big data analytics enhances the accuracy of monitoring and tracking KPIs while reducing human errors (Murphy et al., 2024). Drilling operations in the O&G sector benefit significantly from the density of data and the technological advancements available for processing it. Companies often utilize third-party software or develop their own to meet their data needs.

The upstream segment of the O&G industry typically applies four main types of big data analytics: descriptive, diagnostic, predictive, and prescriptive. Descriptive analytics summarise past operations; diagnostic analytics investigates the causes of issues; predictive analytics forecasts future scenarios based on historical data; and prescriptive analytics supports decision-making by recommending possible courses of action.(Pandey et al., 2021).Though previous studies have addressed the types of data analysis in the O&G sector and the types of performance, few have examined their direct

relationship with performance outcomes in drilling operations, particularly regarding the practical challenges faced by industry practitioners (Nguyen et al., 2020).

Big data, as discussed in prior literature, is not merely a technological advancement but also a cultural and analytical shift. Moorthy et al. (2015) describe it as a fusion of algorithms, computing infrastructure, databases, and analytical tools. In this context, research depicts the characteristics of big data as follows: the amount of data (volume), the number of data sources and their types (diversity), the frequency with which data is recorded (velocity), and the reliability of the data (veracity). (Chen et al., 2017; Hartmann et al., 2016). Furthermore, the challenges faced by companies while using BDA in the O&G industry are classified into two categories: technical and non-technical. (Nguyen et al., 2020)

Technical challenges pertain to the processes of data collection and analysis, whereas non-technical challenges refer to managerial circumstances that influence data management. Various forms of BDA are used to facilitate decision-making, including descriptive analytics (what happened), diagnostic analytics (why it happened), predictive analytics (what will happen), and prescriptive analytics (how can we make it happen) (Dargam et al., 2021; Pandey et al., 2021). In addition, the efficiency of data analysis is reflected in performance, which is measured through the KPIs in the O&G sector. Drilling KPIs are essential to assess and improve organisational performance from exploration to sales and this requires distinct metrics. For example, drilling KPIs comprise time, quality, and health, safety, and environmental (HSE) considerations, with technological monitoring playing a pivotal role in minimizing non-productive time and improving overall efficiency

(Zhang et al., 2020). The identification of these KPIs enables organisations to sustain operations and attain a competitive edge, as performance measurement highlights the principle that “you can’t improve what you can’t measure”. (Varma et al., 2008).

Against this backdrop, this study seeks to identify the types of data analysis and the types of performance in drilling operations and to reveal the challenges facing practitioners. Specifically, it seeks to address three key research questions: (1) What KPIs are used to evaluate drilling operations performance? (2) How is BDA applied to enhance performance in drilling operations? (3) What challenges arise in analysing drilling data? To answer these questions, semi-structured interviews were conducted with thirty practitioners from a range of departments within the Gulf Cooperation council (GCC) region, specifically between the National Oil Company (NOC) and the International Oil Company (IOC). Participants represented diverse organisational units, including the Project Unit, HSE, Data Management, Performance Management, Budgeting, Planning, Information Solutions, and Information Management, as well as representatives from companies and service providers.

The remainder of this paper is structured as follows. Section 2 outlines the research methodology. Section 3 presents the data analysis and findings, followed by Section 4, which discusses the results and conclusions.

2. Research methodology

The approach used in this study is qualitative, specifically using semi-structured interviews, which allow for a more comprehensive exploration. They also offer insights that quantitative methods may lack and have a greater ability to explore perspectives and answer questions compared to rigidly structured interviews (Cohen et al., 2002). Due to their lower theoretical and technical requirements than other qualitative methodologies, thematic analysis was chosen over content analysis, narrative analysis, and discourse analysis. It also provides a more accessible analysis format and requires fewer prescriptive implementation steps (Zhao et al., 2023).

The data collection for this study took place over a period of four months, specifically from December 2023 to March 2024, during which thirty participants engaged in the research. The GCC was strategically chosen due to the limited studies available on the topic and because the region's relationship dynamics facilitated expert interviews. The oil sector’s privacy adds complexity, making

relationships crucial for access. Highlighting the topic's significance, Kuwait, through the Kuwait Oil Company, organised the "Big Data Galaxy" conference and exhibition in 2023. This event attracted major international companies, discussing digital transformation in oil industry sectors. It was an opportunity for companies to exchange experiences and advance their data management using modern technologies like artificial intelligence. This aligns with the study's vision, paralleling contemporary trends embraced by relevant authorities in the GCC.

The methodology revolved around three main axes: the role of big data analytics, KPIs for measuring performance, and the challenges encountered by practitioners in the field. Important criteria for conducting interviews included the participants' professional experience, requiring a minimum of ten years in the industry. Consequently, thirty individuals were interviewed, all of whom are senior officials and consultants within the oil sector with a focus on data management. The interviewees predominantly consist of engineers engaged in software development, project management, and drilling, with specialisations in health, safety, environmental standards, data management, performance monitoring, budgeting, and planning. Their professional experience spans from 10 to 25 years across various units, including service companies and specific roles in health, safety, environment, and information management. These experts emphasise the importance of efficiency, quality assurance, contract management, and operational supervision, reflecting a diverse yet specialised workforce.

3. Data analysis and findings

The data analysis followed a thematic approach, with codes derived from previous literature (e.g., KPIs from Onyeme et al., 2015; technical and non-technical challenges from Nguyen et al., 2020) and themes emerging from the interview data. This ensured both a theoretical foundation and empirical relevance. The analysis was guided by Braun and Clarke (2006) six-step method: familiarisation with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. NVivo 14 software was used to assist in organising and managing the coding process. As noted by Jugder (2016), combining computer-assisted analysis with manual methods can enhance accuracy and reliability. The analysis identified four key themes that directly address the research questions: (1) KPIs related to drilling operations, (2) stages of BDA, (3) the nature and categories of data utilised, and (4) challenges hindering BDA implementation.

3.1 Key Performance Indicators (KPIs) categories

Measuring performance is vital to business success as it provides a balanced view of the company, offering a complete picture of its health and performance (Onyemeh et al., 2015). The study identified six KPIs: operational efficiency, cost management, well performance, equipment maintenance, safety and compliance, and environmental impact, as depicted in the accompanying Figure 1.

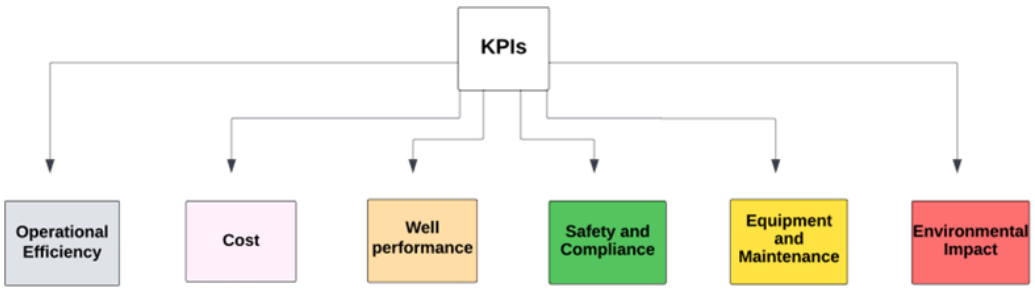


Figure 1: Drilling Operations KPI

3.1.1. Operational Efficiency KPIs

Operational efficiency is vital in O&G drilling, especially in resource-rich countries like the Middle East. It consists of three main KPIs: Drilling Performance, Non-Productive Time (NPT), and Invisible Lost Time (ILT) as shown in Figure 2. Drilling Performance focuses on the rate of penetration (ROP) and rig time, impacting speed and cost savings. NPT identifies unproductive periods, while ILT

measures time management aspects, ensuring projects stay on track and helping minimise delays and optimise workflows.

ROP is particularly significant as it directly correlates with drilling efficiency, reducing costs and downtime. Minimising NPT is equally crucial, as it reduces idle equipment and personnel expenses, fostering improvement. Distinguishing productive from unproductive time, including "invisible lost time," ensures accurate performance assessment. These KPIs, recorded in reports and databases, provide insights for smooth, cost-effective operations.

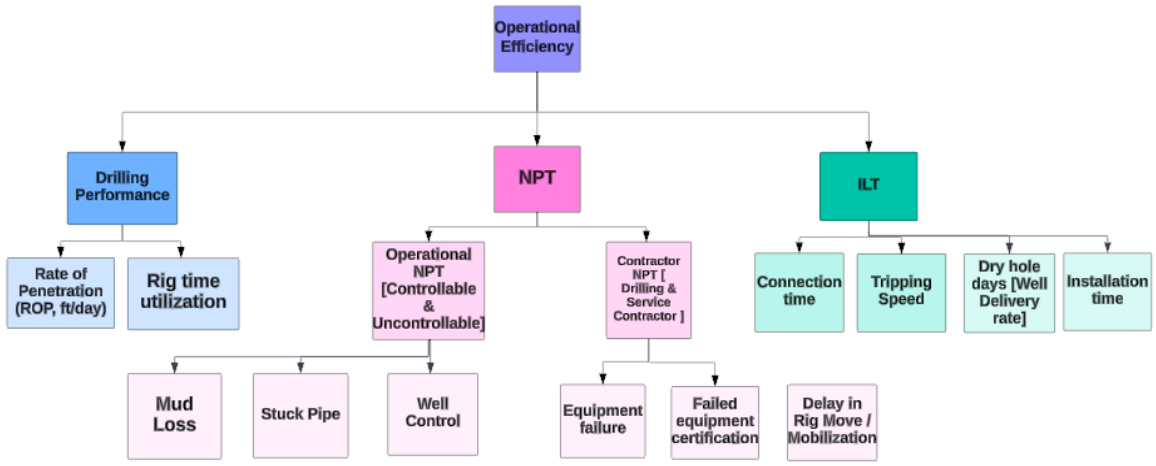


Figure 2: Operational Efficiency KPIs

3.1.2. Cost KPIs

Monitoring the comprehensive cost KPIs focuses on reducing capital expenditures (CAPEX) and operational expenditures (OPEX) to achieve profitability in drilling operations. Costs are divided into operating and capital expenditures, with the finance department monitoring and controlling overall expenditures, as illustrated in Figure 3. Capital expenditures (OPEX) include new project expenses, such as drilling a new well. Expenses include the cost per foot drilled, non-productive time (NPT) costs, actual paid-in costs, cost per foot of drilling, and cost per well. The budget allocated for these activities is adhered to, considering the disbursement of project funds within a specific timeframe. On the other hand, operating expenditures (OPEX) disburse funds based on each objective, such as project maintenance. The finance department monitors the disbursement of funds to the project. Despite low-cost objectives, maintaining quality is essential to prevent any failures that result in additional expenditures.

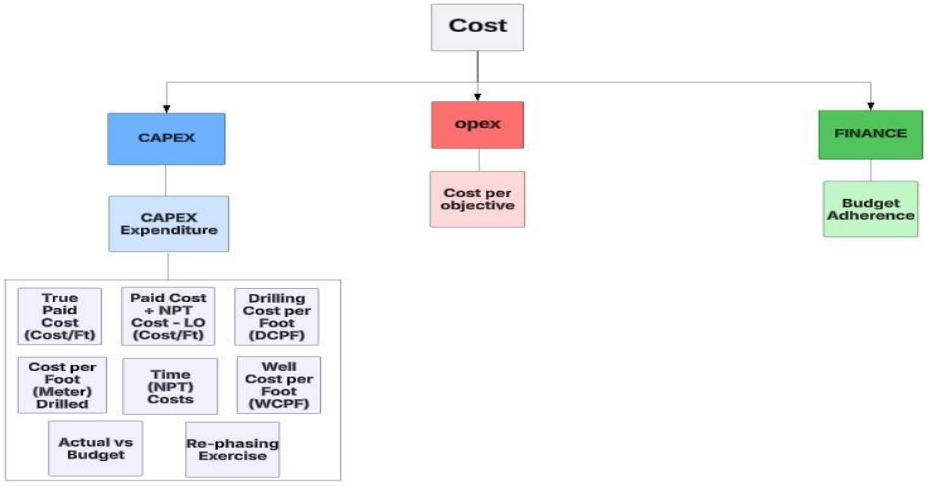


Figure 3: Cost KPIs

3.1.3. Safety and Compliance KPIs

Safety indicators assess the implementation of safety protocols in drilling operations, while compliance indicators monitor adherence to regulations and policies. In the GCC O&G sector, safety and compliance are guided by lagging and leading indicators, as illustrated in Figure 4. Lagging indicators, such as Lost Time Injury Frequency Rate (LTIFR) and Total Recordable Incident Rate (TRIR), Motor Vehicle Accidents (MVA), causes leading to fatalities (Fatality), infringement & traffic violations and a number of spills and leaks (Environmental Incidents), reflect past incidents. Conversely, leading indicators, like near-miss tracking and safety drills, aim to prevent future issues. Other leading indicators such as HSE site visits to ensure compliance and safety, HSE training programmes for workers to ensure understanding of safety standards, stop cards and HSE Initiatives aimed at improving safety practices and compliance, may not need to use BDA technologies as they could be done by using traditional methods. The Health, Safety and Environment Unit ensures that contractors adhere to safety standards, and always gives priority to protecting its employees and the environment, and tracking these indicators provides a safe, low-risk working environment.

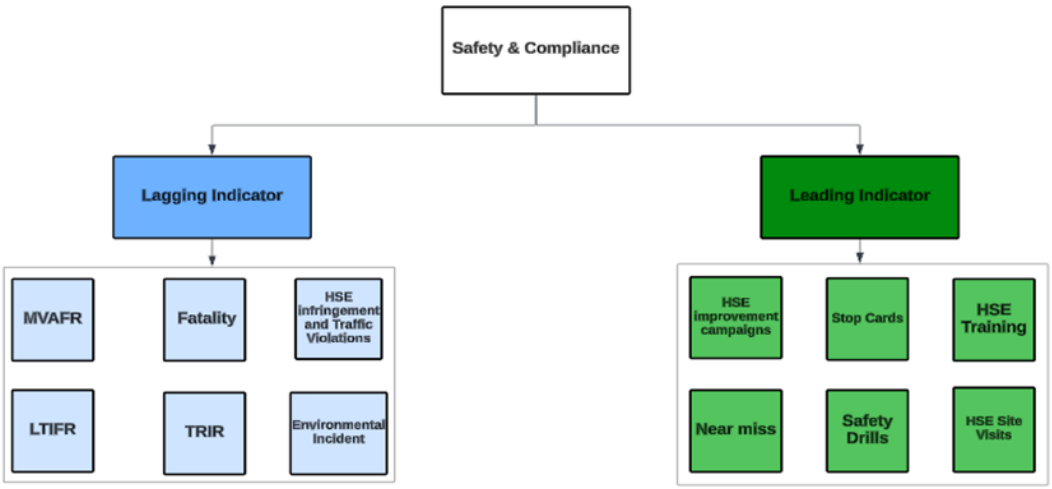


Figure 4: Safety and Compliance KPIs

3.1.4. Well Performance KPIs

Well performance KPIs are essential for assessing drilling operations, focusing on well quality, completion efficiency, and productivity, as presented in Figure 5. These metrics guide practitioners in enhancing well quality and operational success, directly affecting profitability. Companies develop internal programmes tailored to specific regions and conditions, ensuring precise KPI monitoring. Additionally, with the rise of technology, firms are encouraged to create software tools for broader performance tracking. Efficient completion signifies the transition from drilling to production, emphasising safety and environmental considerations. In competitive GCC markets, practitioners prioritise productivity KPIs to evaluate post-completion well performance, considering factors like production volume and geological challenges.

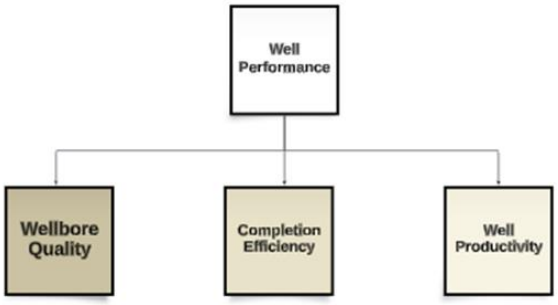


Figure 5: Well performance KPIs

3.1.5. Equipment and Maintenance KPIs

These metrics are essential for evaluating equipment before and during drilling operations, providing a clear vision. Using this data allows companies to take preventive measures that enhance maintenance, as failures can lead to huge losses. Key metrics include mean time between failures (MTBF), equipment downtime, and preventive maintenance completion rate, as shown in Figure 6. It measures the average time between equipment failures to reduce unproductive time by identifying failure patterns. Service companies usually monitor these KPIs because they are concerned with working in the well, and the equipment belongs to them, as this company is keen to perform tasks accurately to avoid financial losses and damage to its reputation.

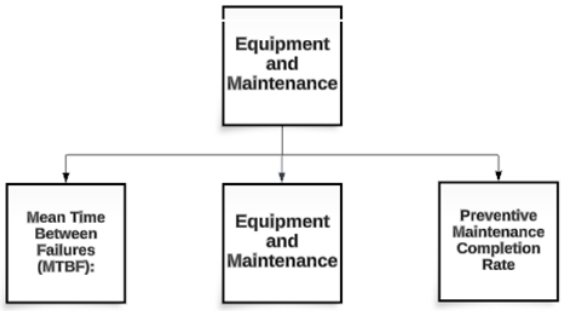


Figure 6: Equipment and Maintenance KPIs

3.1.6. Environmental Impact KPIs

Environmental impact KPIs are vital for promoting eco-friendly practices and evaluating compliance with global standards. Key KPIs include water usage, emissions per well, and waste management

efficiency, as depicted in Figure 7. Monitoring these indicators enhances efficiency, sustainability, and risk mitigation in the O&G sector. The Health, Safety, and Environmental Compliance Department ensures adherence to corporate guidelines during drilling. Environmental and social impact assessments are conducted to evaluate impacts and maintain standards. Air, water, and land quality are monitored biannually, with reports submitted to authorities. Diesel generators significantly contribute to emissions, with daily reports collecting data to estimate emissions. Compliance with environmental regulations reflects the operator's commitment to pollution control and community care.

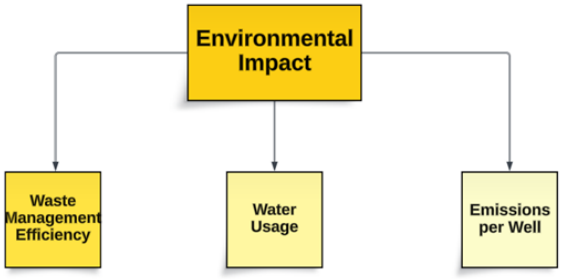


Figure 7: Environmental impact KPIs

3.2 Types of BDA

Practitioners in the drilling industry in the GCC use three stages of Business Data Analysis (BDA): descriptive, diagnostic, and predictive as shown in figure 8. The descriptive stage visualises historical data on dashboards using different platforms, with Power BI being a popular choice in GCC drilling operations for its ease of use. This enables decision-makers to identify issues and predict future events. Operators have developed custom software to enhance data analysis, driving innovation and industry interest. Diagnostic analysis, called "problem finding," overlaps with descriptive analysis as practitioners use monitoring dashboards and daily drilling reports (DDRs) to uncover issues, relying heavily on field experience. Predictive analysis is critical for forecasting future events, with some units still using traditional methods while others embrace AI technologies like machine learning to improve forecasting.

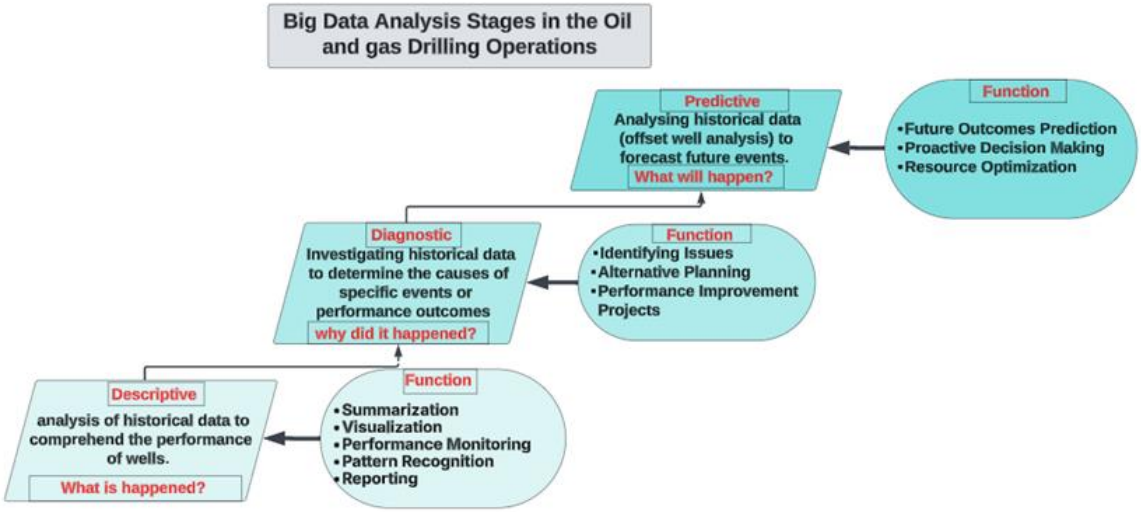


Figure 8: BDA stages

3.3 Types of data in the Drilling Operations

Both high-frequency and low-frequency data are used in drilling operations. While unstructured data, which comes from daily or monthly reports, is more complicated and may need sophisticated tools, like machine learning, for efficient interpretation, structured data, which may be found in logs and spreadsheets, is arranged for simple analysis. To facilitate effective monitoring and decision-making, high-frequency sensor data—such as pressure, torque, and weight on drilling rigs—is arranged and shown as time-series data that is regularly updated in real-time. Financial records, historical data, and daily drilling reports (DDRs) are examples of low-frequency data. Analysis is made more difficult by the fact that DDRs include both structured and unstructured features, such as narrative notes and time-stamped activity logs. While low-frequency data presents difficulties because it is unstructured, high-frequency data enables quick operational decisions.

3.4 BDA Challenges

According to the interview responses, two types of challenges hinder the analysis process, which are technical and non-technical. Technical challenges relate to software and data management, which make data analysis difficult. Many technical challenges were identified in this study, such as the need for software improvements (software), difficulty in determining project costs due to unclear data (data quality and accuracy), data integration challenges from various sources and lack of standardised data (data management and integration), constraints on cloud data storage due to local regulatory requirements (data storage), connectivity or hardware issues leading to data spikes and gaps (connectivity), lack of standardised reporting and standard platforms for accessing data (coordination), and use of machine learning for data quality improvements and predictive analysis (technologies).

Non-technical challenges include administrative issues such as workforce shortages and high turnover rates impacting operations (workforce challenges), lack of communication between departments leading to operational inefficiencies (communication), bureaucratic obstacles that hinder the implementation of software changes (bureaucracy), discrepancies and inadequate information in operational documentation (operational inconsistencies), the impact of working conditions, including hot weather, on data transfer and reporting, as some company men would leave the well without completing their reports according to weather circumstances (environmental and working conditions), and challenges associated with adopting new technologies and processes, particularly AI tools (resistance to change). Despite the challenges facing drilling operations, there is optimism about the positive impact of adopting advanced technologies such as AI to enhance drilling operations.

3.5 Development of a Framework for BDA in the Drilling Operations

The BDA development framework includes three phases of data analysis (descriptive, diagnostic, and predictive) and six categories of KPIs (operational efficiency, cost, well performance, equipment maintenance, safety, compliance, and environmental impact). However, technical and non-technical challenges negatively impact this relationship. Overcoming these challenges by improving data management will enhance the analysis process and positively impact overall performance.

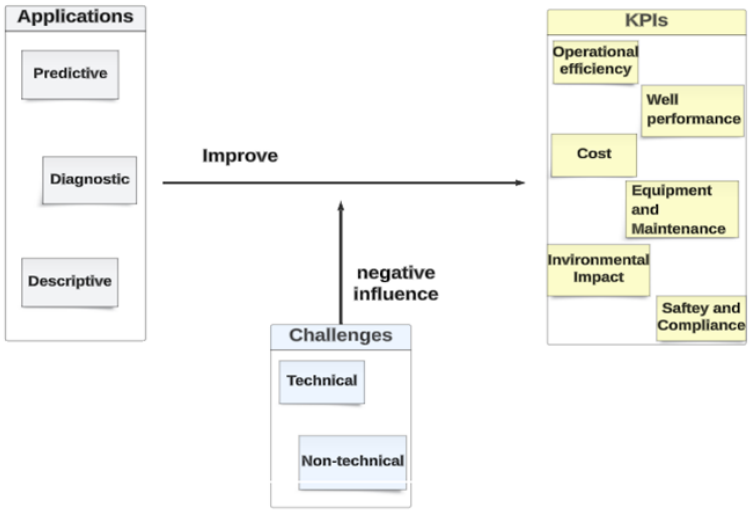


Figure 9: Development framework

4. Discussion and conclusions

This study examines the role of big data analytics in enhancing O&G drilling operations, using a qualitative research methodology based on semi-structured interviews with thirty practitioners from various O&G drilling companies in the GCC. A framework was developed to link six categories of KPIs (operational efficiency, cost, well performance, equipment maintenance, safety and compliance, and environmental impact) to three stages of big data analytics (descriptive, diagnostic, and predictive).

The study shows the potential of big data analytics to improve drilling operations performance, improve decision-making, avoid financial losses, and increase profitability. However, it also identifies technical challenges, such as data integration and software, and non-technical challenges, like managerial circumstances, that harm the data analysis process. Addressing these challenges requires adopting modern artificial intelligence technology, training practitioners, and managing data professionally. Further research is recommended to clarify the relationships between the stages of data analysis, especially (predictive and descriptive), and the types of performance in drilling operations, as it is a sector that produces a flood of data daily, and modern big data analytics technology is emerging, and the relationship between BDA technologies and drilling performance remains somewhat ambiguous.

This study, while providing valuable insights, has limitations. It focuses exclusively on drilling operations and does not include other segments of the oil and gas value chain, such as exploration, refining, or production, each of which generates and utilises data in distinct ways. Furthermore, the research is situated within the context of the GCC, where regulatory conditions, technological infrastructure, and operational cultures may differ significantly from other regions. Additionally, the study reflects the characteristics of qualitative research, with findings derived from semi-structured interviews that depend on participants' personal perspectives and experiences. Despite efforts to include a diverse range of viewpoints, there is potential for bias in participant selection and interpretation. The lack of quantitative performance data indicates that the analysis is based on informed opinion rather than measurable outcomes. Future research would benefit from triangulating these insights with performance metrics to enhance the evidence base and facilitate broader generalisations.

References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *Ieee Access*, *5*, 8869-8879.

Cohen, L., Manion, L., & Morrison, K. (2002). *Research methods in education*. routledge.

Dargam, F., Liu, S., & Ribeiro, R. A. (2021). On the impact of big data analytics in decision-making processes. In *EURO Working Group on DSS: A Tour of the DSS Developments Over the Last 30 Years* (pp. 273-298). Springer.

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data–a taxonomy of data-driven business models used by start-up firms. *International Journal of Operations & Production Management*.

Jugder, N. (2016). The thematic analysis of interview data: An approach used to examine the influence of the market on curricular provision in Mongolian higher education institutions. *Hillary place papers*(3).

Mohammadpoor, M., & Torabi, F. (2020). Big Data analytics in oil and gas industry: An emerging trend. *Petroleum*, *6*(4), 321-328.

Moorthy, J., Lahiri, R., Biswas, N., Sanyal, D., Ranjan, J., Nanath, K., & Ghosh, P. (2015). Big data: Prospects and challenges. *Vikalpa*, *40*(1), 74-96.

Murphy, N., Petrova, E., & Jane, K. (2024). Performance Metrics for Drilling Operations: Key performance indicators (KPIs) in drilling.

Nguyen, T., Gosine, R. G., & Warrian, P. (2020). A systematic review of big data analytics for oil and gas industry 4.0. *IEEE access*, *8*, 61183-61201.

Onyemeh, N. C., Lee, C. W., & Iqbal, M. A. (2015). Key performance indicators for operational quality in the oil and gas industry a case study approach. 2015 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM),

Pandey, R. K., Dahiya, A. K., & Mandal, A. (2021). Identifying applications of machine learning and data analytics based approaches for optimization of upstream petroleum operations. *Energy Technology*, *9*(1), 2000749.

Varma, S., Wadhwa, S., & Deshmukh, S. (2008). Evaluating petroleum supply chain performance: application of analytical hierarchy process to balanced scorecard. *Asia Pacific Journal of Marketing and Logistics*.

Zhang, H., Lu, B., Yang, S., Ke, K., Song, J., Hou, X., Wang, Z., & Jin, X. (2020). A global drilling KPIs analysis system based on modern data science techniques. Abu Dhabi International Petroleum Exhibition & Conference,

Zhao, G., Chen, X., Liu, S., & Xie, X. (2023). The Application of Digital Technologies in the Agri-Food Supply Chain of China: Enablers Identification and Prioritization. *Decision Support System in an Uncertain World: the Contribution of Digital Twins*.

Beyond Gateways: Evaluating LLM capabilities to decouple decision logic from business process flows

Nikolaos Nousias[1], George Tsakalidis[1] and Kostas Vergidis[1*]

[1] Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece
[*] kvergidis@uom.edu.gr

Abstract: Business process modeling is a critical phase in Business Process Management (BPM), enabling organizations to document, analyze, and optimize workflows. A particularly challenging aspect of process modeling is when representing decision logic within process flows. Typically, cascading gateways are used to preserve the intended logic, which in turn increases the complexity and maintenance effort for both process and decision logic. Recent advancements in Large Language Models (LLMs) have introduced new possibilities for automating process modeling, yet their effectiveness in decoupling decision logic from process flows remains largely unexplored. To address this gap, this study evaluates three publicly available LLM-based tools - ProMoAI, BPMN-Chatbot, and Nala2BPMN - assessing their ability to externalize decision logic in accordance with the Separation of Concerns (SoC) paradigm. Using a leave request process as a case study, the authors compare the LLM-generated models against a human-expert model. The findings reveal that while ProMoAI and BPMN-Chatbot consistently failed to separate decision logic, Nala2BPMN demonstrated a greater capacity, successfully externalizing decision logic in one instance. These results underscore both the limitations and potential of current LLM-based tools in handling decision logic, and highlight the need for improved prompting methodologies to better instruct LLMs in decoupling decision logic from process flows.

Keywords: LLM; GPT; BPM; business process modeling; decision modeling

1. Introduction

In today's competitive business environment, operational efficiency and strategic success hinge on effective Business Process Management (BPM) (Tsakalidis & Vergidis, 2021). BPM follows a lifecycle methodology, employing various tools and techniques to continuously optimize processes, reduce costs, and enhance performance. Among its phases, business process modeling plays a critical role (Nousias et al., 2024). This phase involves documenting how businesses execute processes by developing conceptual models that represent activities, events, and control flow logic. Such models facilitate communication, foster shared understanding, and provide essential inputs for subsequent stages, including process analysis, implementation, and execution.

Despite its importance and extensive attention from both academia and industry, business process modeling faces significant challenges that hinder its broader adoption. Traditionally, it requires manually converting textual descriptions - derived from documents, interviews, or workshops - into formal representations, such as Business Process Model and Notation (BPMN) models. This manual approach is time-consuming, with *"AS-IS"* model acquisition accounting for up to 60% of the total time spent on process management projects (Herbst & Karagiannis, 1999). This is mainly because process stakeholders and modelers are typically different individuals, requiring modeling experts to first extract knowledge from stakeholders, create the model, and subsequently validate it.

Furthermore, business process modeling requires expertise in both the semantics of the modeling language and the representation of complex scenarios, making it error-prone and often resulting in low-quality models. One particularly challenging aspect is the representation of decision logic within process flows. Decisions are typically modeled using cascading gateways and multiple control flow elements, which increase complexity, and reduce maintainability. This contributes to Business Process Debt (Nousias et al., 2023), where inefficiently modeled processes accumulate unnecessary complexity,

eventually necessitating costly redesigns. To address this, prior research advocates for the Separation of Concerns (SoC) paradigm (Batoulis et al., 2015; Biard et al., 2015; Hasić et al., 2018), which decouples decision logic from process flows by externalizing it into decision tables, such as Decision Model and Notation (DMN) models, that can be invoked by process models.

To overcome the challenges associated with manual modeling and the need for specialized expertise, researchers have explored automating business process model generation from various sources. Since most organizational knowledge is stored in textual documents, this paper focuses on unstructured text as the primary input for model generation. Traditional approaches - such as rule-based methods, machine learning, and machine translation - have been investigated (Schüler & Alpers, 2024). However, these methods often struggle with accuracy and real-world applicability due to challenges in handling diverse writing styles, domain-specific variations, and the limited availability of high-quality training datasets (Nour Eldin et al., 2025). As a result, recent research has shifted toward leveraging pre-trained Large Language Models (LLMs), which have demonstrated promising results. These AI models, trained on vast amounts of textual data, can comprehend natural language and generate structured representations from unstructured descriptions, making them strong candidates for automating business process modeling (Kourani et al., 2024).

Despite the growing body of research on LLMs for business process modeling, existing studies primarily focus on how users can instruct LLMs to generate process models using different prompting techniques. However, to the best of the authors' knowledge, no prior research has examined how decision logic is represented within these generated models. Overlooking this aspect can lead to increased model complexity and reduced clarity, presenting a research gap that this paper aims to preliminary explore.

Overall, the aim of this paper is to evaluate the capability of publicly available LLM-based tools to model decision logic as a distinct concern in generated business process models - an aspect that has not been explored in prior research. The remainder of the paper is organized as follows: Section 2 discusses the SoC paradigm in business process and decision modeling. Section 3 reviews previous work on LLM applications in business process and decision modeling. Section 4 evaluates existing LLMs, while section 5 concludes with key findings and directions for future research.

2. Separation of Concerns (SoC) in business process and decision modeling

Decision-making is a fundamental component of business processes, directly influencing the flow and outcomes of business operations. To systematically represent decision logic within business process models, it is essential first to distinguish between two primary types of decisions: criteria-based decisions and routing decisions. Criteria-based decisions rely on predefined conditions where specific inputs determine a defined output. These decisions are best represented using decision tables, which explicitly map conditions (inputs) to corresponding actions (outputs). In contrast, routing decisions are primarily concerned with determining the subsequent path in a process flow based on the outcome of a preceding step. Unlike criteria-based decisions, which follow explicit conditions, routing decisions are inherently flow-dependent and are best represented using gateways within process models.

Despite this distinction, business process modeling practices often depict both decision types using gateways. While this approach is appropriate for routing decisions, applying it to criteria-based decisions can lead to the modeling of labyrinths or spaghetti-like diagrams (Batoulis et al., 2015). In such cases, cascading gateways are employed to preserve decision logic, resulting in models that are difficult to interpret, maintain, and modify. In addition, even minor changes to decision criteria can significantly impact the overall structure of the process model, requiring modelers to rearrange elements to accomodate the revised decision logic.

To address this challenge, prior research advocates for the Separation of Concerns (SoC) paradigm, which decouples criteria-based decision logic from process flows. In this approach, decision logic is externalized and encapsulated in decision tables, which process models can invoke at runtime. These tables receive input data, process the decision logic, and return an output to the invoking process, guiding routing decisions and subsequent task execution. Since decision logic is stored in a separate

model, it can be invoked by any process model, provided there is a well-defined integration between the two. This modularity enhances reusability as opposed to hardcoding decisions within a single process model, where they remain confined to a specific decision point (Hasić et al., 2017). Additionally, because decision logic is externalized rather than embedded within process model constructs, model-size-based complexity metrics improve, leading to enhanced understandability and maintainability of process models (Hasić et al., 2018).

Building on this foundation, several studies have explored methods for externalizing decision logic and integrating it with business process models. For example, (Biard et al., 2015) propose leveraging BPMN's business rule tasks to invoke DMN-based decision logic, replacing the need for cascading gateways. (Hasić et al., 2017) presents the Decision as a Service (DaaS) paradigm, which aligns with Service-Oriented Architecture (SOA) principles by treating decisions as externalized services that processes can dynamically invoke. Moreover, (Hasić et al., 2018) introduce five principles for integrated process and decision modeling (5PDM) which can be applied to develop decision-aware and decision-intelligent processes. Overall, the literature concludes that separating but effectively integrating process and decision logic enhances operational flexibility while significantly improving the readability, traceability, and maintainability of both process and decision models (Batoulis et al., 2015; Biard et al., 2015; Hasić et al., 2018).

3. Related work: LLMs for business process and decision modeling

The automatic extraction of business process models from textual descriptions has attracted significant attention in recent years. While various techniques exist for generating process models from text (Schüler & Alpers, 2024), LLMs have the potential to transform these methods, making business process modeling more accessible and widely usable. Initial studies, such as (Fill et al., 2023), have explored the feasibility of using LLMs for conceptual model generation, employing a few-shot prompting strategy with GPT-4. Furthermore, a publicly available tool, ProMoAI[14], was introduced in (Kourani et al., 2024), interfacing with multiple state-of-the-art AI providers (e.g., OpenAI, Google, DeepSeek). ProMoAI incorporates advanced prompt engineering, error handling, and a feedback loop to enable users to iteratively refine the generated process models. While its capabilities appear promising, (Köpke & Safan, 2025) highlights that ProMoAI incurs significant costs due to its extensive use of input and output tokens. To address this limitation, the authors introduce BPMN-Chatbot[15], a tool specifically designed to optimize token usage by leveraging GPT-4. Their evaluation reports a substantial reduction in token consumption - 94% for context tokens and 75% for output tokens - compared to ProMoAI. Additionally, (Nour Eldin et al., 2025) proposes Nala2BPMN[16], a web-based application that interacts with OpenAI's GPT-3.5/4o/4-turbo. This tool employs a modular architecture that decomposes the process modeling task into multiple sub-tasks (e.g., completing process description, extracting process entities, constructing the model), mimicking how a human tackles the process modeling problem. The evaluation demonstrates that this approach not only reduces token usage but also enhances the accuracy and interpretability of the generated models. Collectively, these studies suggest that LLMs serve as effective tools for generating initial process model drafts, which can then be iteratively refined. Nevertheless, expertise in modeling remains crucial for validation and providing meaningful feedback to improve the generated models.

Despite the growing body of research at the intersection of LLMs and business process modeling, their application in decision logic modeling remains largely unexplored. The only relevant study, proposed by (Goossens et al., 2023), investigates an automated approach for generating decision tables from natural language using GPT-3. Across 72 experiments with six problem descriptions, the authors found that GPT-3 performs well in identifying decision contexts and key variables but struggles to generate accurate and complete decision tables. However, this study focuses solely on decision descriptions without incorporating process flow elements.

[14] ProMoAI
[15] BPMN-Chatbot
[16] Nala2BPMN

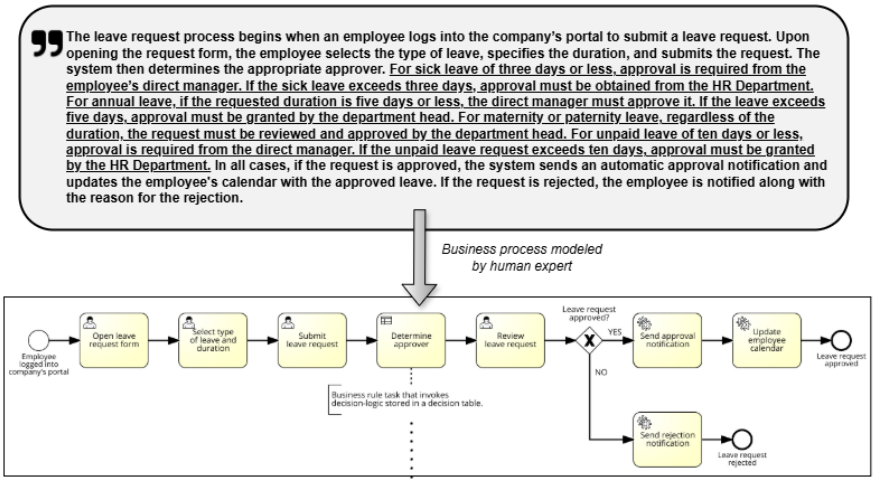
Therefore, it is evident that existing research has examined LLM capabilities for business process modeling and decision-logic modeling separately. However, no prior work has evaluated LLMs' ability to handle simultaneously both criteria-based decision-logic modeling and process flows, representing a significant gap in the literature.

4. Preliminary Experiments

To bridge this gap, the authors conducted preliminary experiments using a business process description that illustrates the leave request process within an organization. Initially, a human expert - independent of the authors - was asked to model this process while following the SoC paradigm. Then, three publicly available LLM-based tools were evaluated for their ability to model the same process. The results were compared to the human-generated model, focusing on how effectively each tool decoupled the criteria-based decision logic from the overall process flow. All experimental results are publicly available online[17].

4.1 Human-expert business process modeling

Figure 1 illustrates the business process description used in this study, which encompasses key elements commonly found in business processes, including manual and automated tasks, criteria-based decision logic, and routing decisions. The authors provided this textual description to a human expert (i.e., 5 years of experience in process modeling), who was instructed to model it according to the SoC paradigm.



Leave Type	Duration	Approver
Sick Leave	≤ 3 days	Direct Manager
Sick Leave	> 3 days	HR Department
Annual Leave	≤ 5 days	Direct Manager
Annual Leave	> 5 days	Department Head
Maternity Leave	Any duration	Department Head
Paternity Leave	Any duration	Department Head
Unpaid Leave	≤ 10 days	Direct Manager
Unpaid Leave	> 10 days	HR Department

Figure 1: Manually modeled process flow and decision table for employee leave request process

The resulting model (Figure 1) reflects this approach by externalizing the identified criteria-based decision logic into a decision table. Through analyzing the process description, the human expert identified that the logic for determining the approver - based on the requested leave type and duration (i.e., underlined in the process description) - is criteria-based and thus should be externalized. To

[17] Study's repository

illustrate the invocation of this decision logic, the expert incorporated a BPMN business rule task *"Determine Approver"*. This task invokes a decision table with the *"Leave Type"* and *"Duration"* as inputs, and receives the *"Approver"* as an output. Subsequently, a user task *"Review Leave Request"* was introduced to indicate that the identified approver is responsible for reviewing the underlying leave request. Finally, the expert incorporated a gateway "*Leave request approved?"* to model the routing decision, determining the process flow based on whether the leave request is approved or rejected. Overall, the authors assess this modeling approach as fully aligned with the SoC paradigm, ensuring a clear separation between process flows and decision logic.

4.2 LLM-based business process modeling

To evaluate the capability of LLM-based tools in decoupling decision logic from process flow, the authors examined three publicly available tools identified in the literature review: ProMoAI, BPMN-Chatbot, and Nala2BPMN. For each tool, GPT-4 was selected as the underlying model. To ensure consistency and account for the inherent randomness of LLM-generated outputs, the process description was provided three times, each in a separate conversation window to prevent the model from using prior interactions as context. The generated models from all three tools were systematically evaluated by the authors, and results are presented in Table 1.

Table 5. Evaluation of LLM-based tools in decoupling decision logic from process flows

LLM-based tool	Run 1	Run 2	Run 3
ProMoAI	No	No	No
BPMN-Chatbot	No	No	No
Nala2BPMN	Partially	Yes	Partially

The evaluation revealed that ProMoAI and BPMN-Chatbot consistently failed to decouple criteria-based decision logic from the process flow across all three runs. Instead, they relied on cascading gateways to represent decision-making, resulting in excessively complex, spaghetti-like diagrams that obscured the principle of separation of concerns (SoC). In contrast, Nala2BPMN demonstrated a notable capacity to externalize decision logic. Specifically, in the second run, the generated model closely resembled the human-designed counterpart, effectively isolating decision logic from the process flow. In the first and third runs, Nala2BPMN partially decoupled decision logic, though with some limitations. In run 1, while the task *"Identify appropriate approver"* was correctly represented, the process flow redundantly included both *"Approve leave request"* and *"Reject leave request"* as separate tasks following the decision gateway, despite both representing the same review action. In run 3, although the task *"Approve leave request"* and the subsequent gateway were accurately depicted, the equivalent of the *"Determine approver"* task from the human-generated model was missing.

The relative proficiency of Nala2BPMN in separating decision logic can likely be attributed to its architectural approach and modeling methodology. Unlike ProMoAI and BPMN-Chatbot, which generate process models in a single step, Nala2BPMN employs a hybrid pipeline that modularizes the modeling process. In this approach, the LLM is invoked multiple times: first to analyze, clarify, and refine the textual description, and then to extract process entities and relationships. The extracted elements are subsequently processed by a structured algorithm that deterministically constructs the process model. By combining the natural language understanding capabilities of LLMs with a deterministic modeling approach, this hybrid methodology improves the robustness and accuracy of model generation.

5. Conclusions

This paper investigated the capability of publicly available LLM-based tools to automate business

process modeling while adhering to the SoC paradigm. Through a series of experiments, the authors evaluated three LLM-based tools - ProMoAI, BPMN-Chatbot, and Nala2BPMN - by comparing their generated models against a human-expert baseline. The results highlight that while LLMs can accelerate the business process modeling activity, manual validation remains essential to ensure model quality. As demonstrated by ProMoAI and BPMN-Chatbot, LLM-generated models can be overly complex, disregarding best practices such as decision logic separation. However, the promising performance of Nala2BPMN suggests that LLMs hold significant potential when guided effectively. Therefore, the authors conclude that the key to leveraging LLMs for high-quality process modeling lies in crafting precise prompts, as these serve as the interface that directs the model to follow best practices.

Despite these insights, this study has several limitations. First, the evaluation was based on a single process description, which may not generalize to more complex scenarios or different domains. Additionally, the assessment was qualitative, relying on expert judgment rather than quantitative complexity metrics. Future research should expand the dataset, incorporating diverse business process descriptions, and employing objective complexity measures to systematically evaluate LLM-generated models. In addition, further refinements in LLM prompting strategies could improve consistency and enhance decision logic representation in automated business process modeling.

References

Batoulis, K., Meyer, A., Bazhenova, E., Decker, G., & Weske, M. (2015). Extracting Decision Logic from Process Models. In J. Zdravkovic, M. Kirikova, & P. Johannesson (Eds.), *Advanced Information Systems Engineering* (pp. 349–366). Springer International Publishing. https://doi.org/10.1007/978-3-319-19069-3_22

Biard, T., Le Mauff, A., Bigand, M., & Bourey, J.-P. (2015). Separation of Decision Modeling from Business Process Modeling Using New "Decision Model and Notation" (DMN) for Automating Operational Decision-Making. In L. M. Camarinha-Matos, F. Bénaben, & W. Picard (Eds.), *Risks and Resilience of Collaborative Networks* (pp. 489–496). Springer International Publishing. https://doi.org/10.1007/978-3-319-24141-8_45

Fill, H.-G., Fettke, P., & Köpke, J. (2023). Conceptual Modeling and Large Language Models: Impressions From First Experiments With ChatGPT. *Enterprise Modelling and Information Systems Architectures (EMISAJ)*, *18*, 3:1-15. https://doi.org/10.18417/emisa.18.3

Goossens, A., Vandevelde, S., Vanthienen, J., & Vennekens, J. (2023). *GPT-3 for Decision Logic Modeling*.

Hasić, F., De Smedt, J., & Vanthienen, J. (2017). A Service-Oriented Architecture Design of Decision-Aware Information Systems: Decision as a Service. In H. Panetto, C. Debruyne, W. Gaaloul, M. Papazoglou, A. Paschke, C. A. Ardagna, & R. Meersman (Eds.), *On the Move to Meaningful Internet Systems. OTM 2017 Conferences* (pp. 353–361). Springer International Publishing. https://doi.org/10.1007/978-3-319-69462-7_23

Hasić, F., De Smedt, J., & Vanthienen, J. (2018). Augmenting processes with decision intelligence: Principles for integrated modelling. *Decision Support Systems*, *107*, 1–12. https://doi.org/10.1016/j.dss.2017.12.008

Herbst, J., & Karagiannis, D. (1999). *An Inductive Approach to the Acquisition and Adaptation of Workflow Models*. https://www.semanticscholar.org/paper/An-Inductive-Approach-to-the-Acquisition-and-of-Herbst-Karagiannis/03d2b48ef0585906618778295297 70de93f30d51

Köpke, J., & Safan, A. (2025). Efficient LLM-Based Conversational Process Modeling. In K. Gdowska, M. T. Gómez-López, & J.-R. Rehse (Eds.), *Business Process Management Workshops* (pp. 259–270). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-78666-2_20

Kourani, H., Berti, A., Schuster, D., & Aalst, W. M. P. van der. (2024). *Process Modeling With Large Language Models* (Vol. 511, pp. 229–244). https://doi.org/10.1007/978-3-031-61007-3_18

Nour Eldin, A., Assy, N., Anesini, O., Dalmas, B., & Gaaloul, W. (2025). A Decomposed Hybrid Approach to Business Process Modeling with LLMs. In M. Comuzzi, D. Grigori, M. Sellami, & Z. Zhou (Eds.), *Cooperative Information Systems* (pp. 243–260). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-81375-7_14

Nousias, N., Nedos, G., Tsakalidis, G., & Vergidis, K. (2023). *Mitigating Business Process Debt with Digital Process Twins*.

Nousias, N., Tsakalidis, G., & Vergidis, K. (2024). Not yet another BPM lifecycle: A synthesis of existing approaches using BPMN. *Information and Software Technology*, *171*, 107471. https://doi.org/10.1016/j.infsof.2024.107471

Schüler, S., & Alpers, S. (2024). State of the Art: Automatic Generation of Business Process Models. In J. De Weerdt & L. Pufahl (Eds.), *Business Process Management Workshops* (pp. 161–173). Springer Nature

Switzerland. https://doi.org/10.1007/978-3-031-50974-2_13

Tsakalidis, G., & Vergidis, K. (2021). A Roadmap to Critical Redesign Choices That Increase the Robustness of Business Process Redesign Initiatives. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(3), 178. https://doi.org/10.3390/joitmc7030178

Machine Learning for Search and Rescue Decision Support

Thomas Laperrière-Robillard, Michael Morin, and Irène Abi-Zeid*

[1]Operations and Decision Systems, Université Laval. Québec, Canada
* irene.abi-zeid@osd.ulaval.ca

Abstract: Maritime search and rescue operations are important humanitarian activities for locating and rescuing survivors in distress at sea. This paper investigates the application of supervised learning techniques to improve search operations planning by estimating the probability of success of a search operation. Four models are evaluated in the study: random forest, K-Nearest Neighbors, Support Vector Machines Regression and Neural Networks. The results show that integrating machine learning can significantly reduce computation time for the allocation of search resources. This can enhance *SAR Optimizer*, the current optimization and evaluation module used in the Canadian Coast Guard decision support system. By improving the quality of search recommendations, our approach has the potential to improve SAR operations and, ultimately, save lives.

Keywords: Machine learning; Simulation; Maritime search and rescue; Canadian Coast Guard; Decision support system.

1. Introduction

Search and rescue (SAR) involves locating and aiding individuals who are in distress or facing immediate danger. Canada is responsible for one of the most challenging and vast SAR zones spanning 18 million square kilometers of land and water. One of the Canadian Coast Guard (CCG) roles is to save and protect lives in the maritime environment. It coordinates an average of 7,000 incidents per year with a rate of 97% of lives saved (Fisheries and Oceans, 2009). Maritime SAR operations in Canada are managed by three joint rescue control centers and two sub-control centers. SAR mission coordinators (SMC) are highly trained individuals who are tasked with the planning, coordination, control and management of operations. One of the biggest challenges for SMCs is deciding where to send search resources. Search planning is time-critical, as survivors must be found quickly due to the rapid decrease of survival rates (Xu et al., 2011).

In order to support SMCs in search planning, the CCG developed the Advanced Search Planning Tool (ASPT) decision support system (DSS), also known as CANSARP (Abi-Zeid, et al., 2019). This DSS includes *SAR Optimizer* (Abi-Zeid, et al., 2019), a search planning module involving simulation and optimization based on search theory (Stone et al., 2016). The output of *SAR Optimizer* is a search plan, namely the assignment of available search and rescue units (SRU) to rectangles, each enclosing a parallel search pattern. In the optimization module, the figure of merit to be maximized is the probability of success (POS), defined as the probability of finding the search object. The DSS evaluates multiple combinations of SRU and search rectangles to propose a best POS search plan in the planning time allowed.

The POS of a search plan is computed by simulating, at each time step, the positions of the SRUs and their proximity to the object's estimated position. However, a simulation approach is quite costly in terms of computation and time, in a context where a search plan must be produced within minutes, which constrains the number of search plans that can be assessed as potential solutions. This motivated us to explore whether and how machine learning (ML) can help make the POS computations faster and increase the number of candidate search plans evaluated. This study examines and compares the performances of four supervised ML algorithms to estimate the probability of success (POS) in search planning (Laperrière-Robillard et al. 2022).

2. Background

After being notified of an incident, the SMC creates, in the ASPT DSS, a SAR case containing all available information about the emergency, the characteristics of the vessel, the number of people involved, the last known point, possible sightings, relevant communications, etc. The next step is to perform a stochastic drift simulation based on Monte Carlo where 5,000 particles, equally likely to be the search object, are seeded using a bivariate Gaussian distribution with a standard deviation specified by the user. The particles are then moved by simulation in time and space according to a drift model that takes into account search object characteristics, surface currents, and winds. The result is a drift model providing the positions of the particles at each time step over a simulation horizon. The simulated trajectories of the particles represent equiprobable trajectories of the search object (Breivik and Allen, 2008). Subsequently, the SMC identifies available SRUs that can be tasked with performing the search operations. Using the drift model and the SRU information, *SAR Optimizer* recommends a search plan. Figure 1 provides a fictitious example of search plans (a parallel pattern and enclosing rectangle) for three SRUs. Since the optimization process terminates once the predefined time limit has been reached, there is no guarantee that all candidate search plans have been simulated and evaluated. The quality of the, possibly sub-optimal, recommended search plans depends not only on the total number of search patterns assessed but also on the sequence in which they were evaluated.

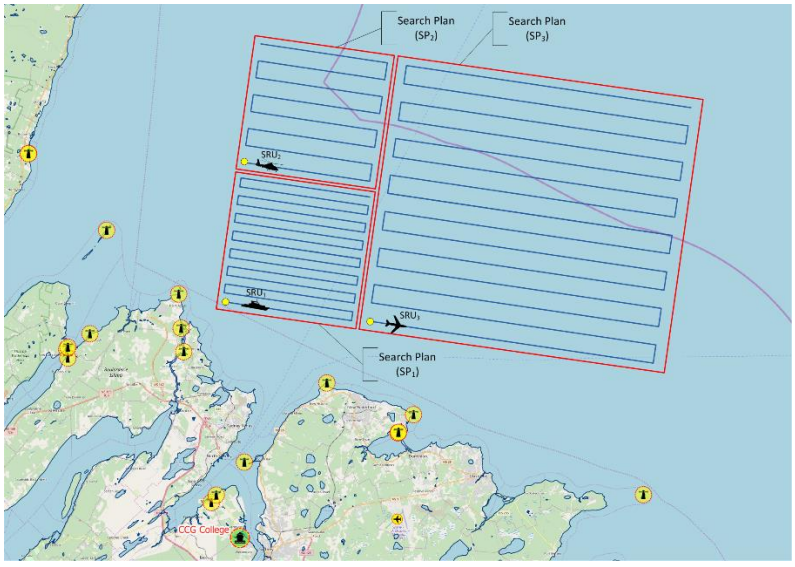


Figure 1: Example of a search plan with three SRUs conducting parallel search patterns in enclosed rectangles

Machine learning is a subfield of artificial intelligence with supervised learning as a particular case where a corpus of labeled learning examples is used to train a prediction model (Russell and Norvig, 2021). In supervised learning, the objective is to be able to predict a dependent variable (or label), here the POS, as a function of independent variables, referred to as learning features, here attributes of search plans. A dataset consisting of POS and attributes is used as learning examples. The learning phase, where a function transforming the attributes into a POS is followed by a prediction phase where this function is applied to obtain the value of a POS corresponding to unseen case attributes. A good model is one that is precise, meaning that it predicts correctly the POS of the training set, and that generalizes adequately, meaning that it predicts correctly the POS of unseen case attributes, evaluated using a test set. In many cases, fine-tuning of the hyperparameters of the ML algorithms is necessary. This consists of randomly picking a number of partitions of the training set k and repeating the training with each partition. The performance of the algorithm is obtained from the average performances over n sets of k partitions. Metrics to evaluate the performance of an algorithm include the mean absolute error (MAE) of the prediction (Kuhn and Johnson, 2013), computed here as the absolute difference between predicted and observed POS values.

3. Methods and Experiments

In order to evaluate an ML approach to predict a POS of a search plan, we followed a four-step process.

First, we generated the learning corpus by using *SAR Optimizer* to simulate and evaluate search plans. The drifting search object was a life raft and the SRU either a helicopter or a fixed-wing aircraft. The available search effort of a SRU is measured by the time spent searching. The experiments were conducted with two different effort levels, namely 3 and 6 hours. We assumed that a single SRU was on-scene searching. This setup resulted in 12 different scenarios (3 drifts × 2 SRUs × 2 effort levels). For each scenario, we generated, simulated, and evaluated the POS of almost 9,100 search plans. Therefore, our final corpus contained 12 sets of 9,100 evaluated search patterns.

We then applied four ML algorithms to each scenario, namely K-Nearest Neighbors (KNN), Support Vector Machines (SVM) Regression, Random Forest (RF), and Neural Networks (NN). We compared their prediction accuracy and the time they took to learn from data. Each model was trained using 70% of the dataset while keeping 30% for testing. This process was repeated 10 times with different partitions. The independent variables used to predict the POS included features of a search plan, namely the bearing, the area and the length and height of the enclosing rectangle, as well as the number and lengths of search and cross legs, and the starting coordinates of the search pattern.

Next, we tested how changing the size of the training set affects prediction accuracy for the ML model retained in the previous step. The goal was to find the best predictions while keeping training size small since generating the learning corpus is time consuming. In fact, it is not readily available from historical data since SAR incidents occur in different locations where the drifts have different characteristics. A smaller training set means faster data collection and training, which is important for real-life search missions. However, using too little data could lead to inaccurate predictions.

Finally, we selected the best ML model based on POS prediction quality, learning time, and training set size. The retained ML model was applied to predict the POS and rank candidate search plans in *SAR Optimizer* by decreasing POS (rectangle ordering heuristic). This allowed *SAR Optimizer* to evaluate, by simulation, the POS starting with the most promising search plans. In order to evaluate the benefits of using the ML model, we then compared the results of *SAR Optimizer* with and without ML based on the POS obtained. Hereafter, SAR Optimizer with the rectangle ordering heuristic is called *SAR Optimizer + ML*, and the standard version of SAR Optimizer is simply called *SAR Optimizer*.

4. Results

4.1 ML Algorithms comparison

The results of comparing the four ML models showed that although three out of four displayed relatively small differences in terms of POS prediction precision, the variability in terms of execution time was considerable. Although the SVM model had the shortest average learning time requiring only 0.37 minutes to train the model, its predictive performance was consistently inferior. The RF, NN and KNN models needed, on average, 51.9, 6.8 and 0.9 minutes respectively. Therefore, we retained the KNN algorithm as a candidate algorithm for predicting POS. Full results are available in (Laperrière-Robillard et al. 2022)

4.2 Prediction quality as a function of training size

Generating the learning set is an expensive operation. The average time needed to produce it for a scenario varies between 28 and 61 minutes when the training set size corresponds to 70% of the full dataset. This is obviously not acceptable since a search plan must be proposed in under 5 minutes. We therefore computed the average MAE for different training set sizes ranging from 455 search rectangles to 7,735. As expected, the prediction's quality improves (MAE decreases) as the size of the training set increases (Figure 2). However, we observed that using a training set of 455 rectangles resulted in an average POS estimation that deviated by approximately 0.012 from the actual ground-truth POS value. Considering that the highest POS values for search patterns range between 0.305 and 0.941, this margin of error is relatively small in practical terms, particularly in scenarios where the best POS exceeds 0.5.

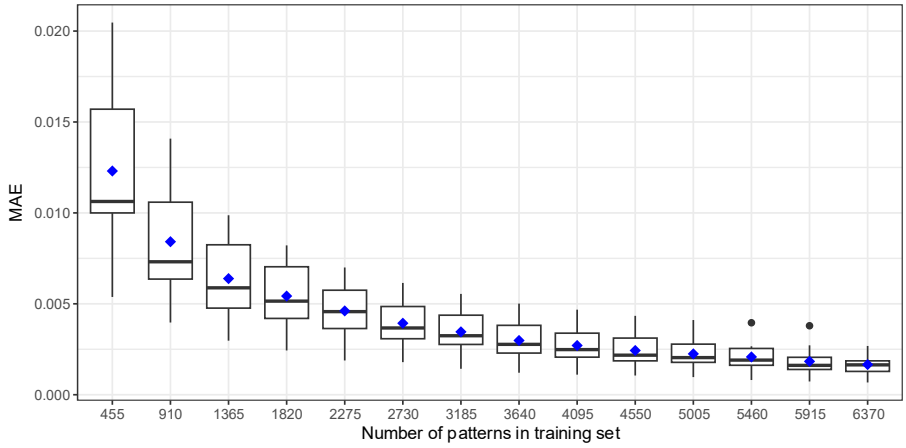


Figure 2: The mean average error of the predicted POS versus the number of plans in the training set **(adapted from Laperrière-Robillard et al., 2022)**

4.3 Comparing *SAR Optimizer* results with and without ML

Based on the previous results, we retained the KNN model with 455 search plans in the training set. We compared the performances of *SAR Optimizer* and *SAR Optimizer + ML* in terms of time needed to attain a best POS value. This comparison was computed over 30 runs representing learning with 30 different datasets of 455 patterns for 12 scenarios where we let *SAR Optimizer* run for 45 minutes. Each scenario is based on a drift (named A, B, or C) and involves a helicopter flying at 500 feet at 90 knots for 3 hours (case 1), a fixed wing flying at 1,000 feet at 120 knots for 6 hours (case 2), a helicopter flying at 90 knots for 3 hours (case 3), or a fixed wing flying at 1,000 feet at 120 knots for 6 hours. Table 1 shows that the total time to obtain the highest POS search plan was much higher without ML than with ML. This is explained by the fact that without ML, *SAR Optimizer* needed to evaluate a larger number of search plans before reaching the best one in the time allocated. Using the ML predicted POS as a heuristic to determine the order in which search plans were to be simulated proved very beneficial.

Table 1. Comparison of *SAR Optimizer* and *SAR Optimizer + ML* **(adapted from Laperrière-Robillard et al., 2022)**

	SAR Optimizer		***SAR Optimizer + ML***	
Scenario	Total time to best POS (min.)	Simulation rank of best plan POS	Average time to best POS (min.) with 95% CI	Median rank of best plan POS
A1	20.657	4,301	2.433 ± 0.149	17
A2	38.211	4,101	4.928 ± 0.378	24
A3	21.139	4,411	2.519 ± 0.146	29
A4	38.153	4,103	5.626 ± 0.485	72
B1	21.909	4,614	3.269 ± 0.517	112
B2	44.387	4,611	6.790 ± 0.985	130
B3	25.028	5,323	4.025 ± 0.812	201
B4	39.028	4,219	6.063 ± 0.991	90
C1	23.510	4,727	5.015 ± 1.279	340
C2	43.506	4,730	7.167 ± 1.135	154
C3	21.285	4,634	2.927 ± 0.355	80
C4	42.947	4,631	6.990 ± 1.362	154

5. Conclusions

In this paper, we set out to explore whether machine learning can improve *SAR Optimizer* in the Canadian maritime SAR planning DSS by developing higher POS search plans in less time. Our experiments showed that the K-Nearest Neighbors algorithm, trained on a relatively small training dataset, can provide accurate predictions within a reasonable timeframe. Given that *SAR Optimizer* can be stopped prematurely due to time constraints, we were able to improve the quality of the proposed search plans by guiding *SAR Optimizer* towards evaluating more promising search plans first, based on their ML-predicted POS.

Although training the ML model requires an initial set of simulated search models, we found that only a small subset is needed to effectively train the model. Our experimental results indicate that *SAR Optimizer + ML* outperforms the standard *SAR Optimizer*. Although we have specifically applied our approach to a maritime SAR decision support system, the method is generalizable and can be integrated into any simulation-based decision support system.

Our contribution consists of developing, testing and evaluating a new approach integrating ML to partially replace computationally expensive simulations in the Canadian SAR planning DSS. Future work includes extending our approach to multiple SRUs on scene.

References

Abi-Zeid, I., Morin, M. and Nilo, O. (2019). Decision Support for Planning Maritime Search and Rescue Operations in Canada. In Filipe, J., Smialek, M., Brodsky, A., Hammoudi, S. Eds Proceedings of International Conference on Enterprise Information Systems, Vol. 1, 316-327, Heraklion, Greece. https://doi.org/10.5220/0007730303280339

Fisheries and Oceans Canada (2009). Canadian Coast Guard Information Kit. 36 pages. DFO 2009-1563.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (Vol. 26, p. 13). New York: Springer.

Laperrière-Robillard, T., Morin, M., & Abi-Zeid, I. (2022). Supervised learning for maritime search operations: An artificial intelligence approach to search efficiency evaluation. Expert Systems with Applications, 206, 117857.

Russell, S., Norvig, P., 2021. Artificial Intelligence: A Modern Approach, Global Edition 4th, 4th ed. Pearson.

Stone, L. D., Royset, J. O., & Washburn, A. R. (2016). Optimal Search for Moving Targets (Vol. 237). Springer International Publishing. https://doi.org/10.1007/978-3-319-26899-6

Xu, X., Turner, C. A., & Santee, W. R. (2011). Survival time prediction in marine environments. Journal of Thermal Biology, 36(6), 340–345. https://doi.org/10.1016/j.jtherbio.2011.06.009

Acknowledgments

The authors wish the ASPT Working group and the SMCs for their support during the project. Special thanks to the project manager, the technical team from the Canadian Coast Guard College in Sydney, and to Oscar Nilo for their support.

Funding: The SAR Optimizer decision support system was developed under Canadian Government, contract number FP802-150046 to Neosoft Technologies. This research was funded by Natural Sciences and Engineering Research Council of Canada [grant number RGPIN-2021-03495, DGECR-2021-00189] and by Fonds de recherche du Québec – Nature et technologies [grant number 2017-B3-199085].

A comparison between DSS and ML models for churn prediction

Boris Delibašić [1*], Sandro Radovanović [1], Marko Bohanec [2], and Milija Suknović[1]

[1] University of Belgrade, Faculty of Organizational Sciences, Belgrade, Serbia
[2] Jožef Stefan Institute, Department of Knowledge Technologies, Ljubljana, Slovenia
[*] boris.delibasic@fon.bg.ac.rs

Abstract: This paper compares the accuracy and convenience of a classical machine learning algorithm, a decision tree, and a classical decision support system model, built by the DEX (Decision EXpert) multicriteria decision modelling method for categorical data, on a churn prediction data set. Decision support systems (DSS) are a technology from the 1960s that was predominantly overruled by machine learning (ML) in the 2010s due to the explosion of big data, and their cost effectiveness. Here we discuss the similar and different aspects of the two technologies, and demonstrate the performance of these different, yet intertwined technologies. We show that our proposed DSS model outperforms the ML model.

Keywords: Churn Prediction; DSS; Multi-Criteria Models, DEX; DIDEX; Decision Tree; Machine Learning

1. Introduction

Machine Learning (ML) and Decision Support Systems (DSS) are technologies that are built for the same purpose, to support decision-making. While ML systems are fuelled by data, it is the expert's knowledge that builds DSS. The problem this paper is addressing is which of these systems is more suitable for customer churn prediction.

Customer churn is a global problem occurring in many domains, from university education students, clients in telecommunication companies, to employees in companies worldwide. Both DSS and ML have been used for customer churn prediction with reported advantage of ML use over the DSS technology.

Although it is not easy to compare different approaches, as usually a multidimensional perspective on a problem boils down to comparing a single criterion (like accuracy, or generalization error). Here, we adopt a DSS perspective for comparing the models using intrinsic criteria proposed by (Bohanec, 2021). We also chose to compare an ML model (decision tree), a DSS model developed by experts (DEX), and a DSS model generated by ML (DIDEX).

The DSS criteria that will support our discussion, as proposed by (Bohanec, 2021), are:

1. Correctness: a model offered should be accurate but also aligned with the problem that should be solved.
2. Completeness: a model should be able to work on all possible input value ranges.
3. Consistency: a model should be internally consistent and should provide consistent solutions.
4. Comprehensibility: a model should be understandable and transparent to the decision makers.
5. Convenience: a model should be easy to use and possess convenient properties, like sensitivity analysis and counterfactual reasoning.

We will show that the five criteria are to some extent present in the ML model, where the DSS model possesses them *per se*. The remainder of this short paper is structured as follows: In Section 2 we show similar background research, in Section 3 we demonstrate the experiment, and we make our conclusions in Section 4.

2. Background research

Customer churn prediction is a critical challenge in various industries, particularly telecommunications. ML techniques have been widely applied to address this issue. Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and Neural Networks are commonly used for churn prediction (De et al., 2021; Raj et al., 2024). These methods have shown promising results in terms of accuracy, precision, recall, and F1 score (Raj et al., 2024). While hybrid and ensemble methods have improved model performance, authors propose future research directions to include exploring deep learning techniques for enhanced prediction accuracy and personalization (Raj et al., 2024). Britto & Gobinath (2020) claim that predictive analytics in customer churn prediction offer more accurate outcomes compared to other approaches.

Within the domain of DSS, several studies have explored the use of rough set theory (RST) for customer churn prediction in the telecommunications industry. Amin et al. (2014, 2015, 2017) consistently found that RST-based approaches, particularly those using genetic algorithms for rule generation, outperformed other rule generation methods such as exhaustive, covering, and LEM2 algorithms. These studies demonstrated that RST can effectively classify and predict customer churn, providing valuable insights for strategic decision-making (Amin et al., 2015, 2017). The researchers also compared RST as a one-class and multi-class classifier, with the latter showing superior performance in binary and multi-class classification problems (Amin et al., 2014). Furthermore, attribute-level analysis using RST was found to be beneficial in developing customer retention policies (Amin et al., 2017). Overall, these studies highlight the potential of RST as an efficient, rule-based approach for customer churn prediction in the telecommunications sector, offering a globally optimal solution when compared to other state-of-the-art methods.

Fuzzy set theory has been applied to customer churn prediction in various industries. In retail banking, fuzzy c-means clustering was used to develop churn prediction models, outperforming classical methods (Popovic & Basic, 2008; Popovic & Basic, 2009). In the finance sector, a three-phase framework incorporating fuzzy inference systems was proposed to predict churn among high-value customers (Safinejad et al., 2018). For telecommunications companies, a model combining fuzzy logic and neural networks was developed to improve churn prediction accuracy (Papa et al., 2021). This method utilized normalized data to form membership function parameters and employed neural network algorithms, demonstrating the potential of fuzzy neural networks in customer churn forecasting. A DIDEX model has been applied to public policy decision-making in the Republic of Serbia (Delibašić et al, 2023). While rule-based methods excel in interpretability, some struggle to provide prediction probabilities, which are valuable for customer evaluation (Huang et al., 2011). Overall, rule-based systems offer a balance between accuracy and explainability in churn prediction.

Based on the previous work, we can conclude that while ML algorithms excel in accuracy, rule-based algorithms excel in interpretability. In this short paper we will demonstrate that DSS models can be both accurate and interpretable.

3. The experiment

For the experiment in this short paper a churn dataset[18] was used. The dataset contained 3333 rows, 20 attributes from which one attribute was the binary outcome churn attribute. The experiment was conducted in Orange software[19], and DEXi software[20]. The goal was to build a prediction model that can "*Identify intention of clients to churn in future, so that there is a possibility to react and possibly prevent churn*".

The experiment had the following consecution:

1. An ML model, namely a decision tree, was built on the churn dataset in Orange.
2. Two DSS models were developed in DEXi for the same purpose, one "handcrafted" by an expert

[18] https://github.com/albayraktaroglu/Datasets/blob/master/churn.csv

[19] https://orangedatamining.com/

[20] https://kt.ijs.si/MarkoBohanec/dexi.html

and the other build on the churn dataset using the method DIDEX

3. All DSS model was evaluated on the churn dataset.
4. A what-if analysis on the churn dataset was performed with the DSS model.
5. The "best" model was chosen according to the criteria proposed above and the results discussed.

3.1 The ML model

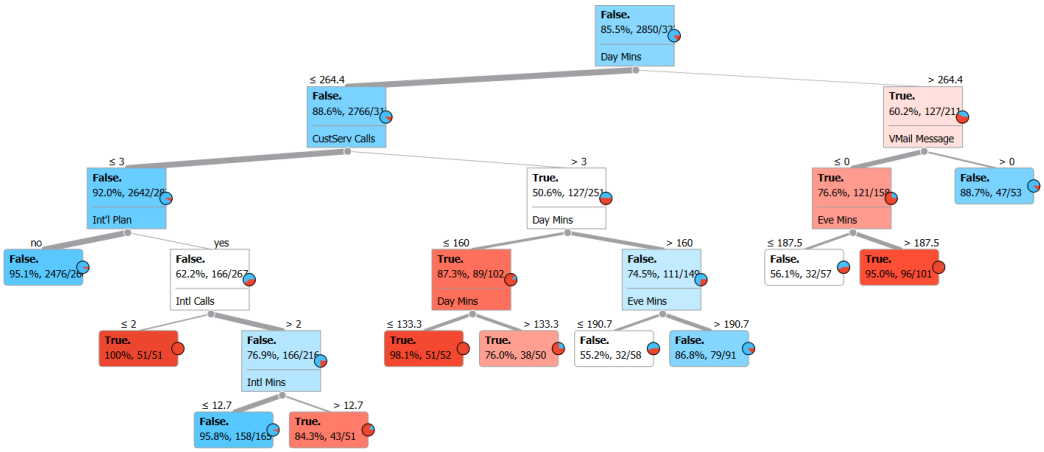


Figure 2: The churn prediction decision tree machine learned model

The decision tree, shown in Figure 1, has been grown in Orange software with limiting the depth of the decision tree to 6, as the generalization error was lowest with this depth. The decision tree model achieved the following performance: Accuracy 90.7%, Recall 56.7%, Precision 73.3%, and F1 63.9%. The tree was built using 10-fold cross validation. The following attributes were chosen as the most important from the 18 available in the dataset (without considering the *Phone number* attribute as the unique client id):

1. *Day Minutes* (how many minutes during a day a client made in a period)
2. *Customer Service Calls* (how many times a client called the customer service)
3. *Voice Mail Messages* (how many times a client left a voice mail message)
4. *Eve Minutes* (how many minutes during an evening a client made in a period)
5. *International Plan* (whether a client used the international plan package)
6. *International Calls* (how many international calls a client made in a period)
7. *International Minutes* (how many international minutes a client made in a period)

3.2 The First DSS model

Two DSS models were developed. The first was built in DEXi software, i.e. it was handcrafted by an expert in DEX methodology (shown in Figure 2). Six attributes were used for building the model. These were the 6 attributes identified in 3.1, excluding attribute 7. Attributes were discretized using the bounds in Figure 1. The first DEX model has a flat attribute structure, meaning that all attributes were on the same level of hierarchy w.r.t. to the root attribute [Churn].

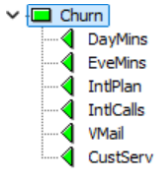


Figure 3: The expert modeled DSS using DEX methodology

96 rules were identified combining all categories in the six attributes and making a prediction for the churn outcome. A part of the decision rules is shown in Figure 3. Row 1 is read like this: If *Day Minutes* are Very Low (1), and *Eve Minutes* High (2), and *International Plan* is "yes", *International Calls* is Low, *Voice Mail* is "No", and *Customer Service Calls* is "Low" then [*Churn*] is "yes".

The 96 rules have been applied to the whole dataset, i.e. to the 3333 rows and the following performance of the DSS model were achieved: Accuracy 74.71%, Recall 62.94%, Precision 31.4%, and F1 41.9%.

	DayMins	EveMins	IntlPlan	IntlCalls	VMail	CustServ	Churn
1	1	2	yes	1	no	1	yes
2	1	2	yes	1	no	2	yes
3	1	2	yes	1	yes	1	yes
4	1	2	yes	1	yes	2	yes
5	1	2	yes	2	no	1	no
6	1	2	yes	2	no	2	yes
7	1	2	yes	2	yes	1	no
8	1	2	yes	2	yes	2	yes
9	1	2	no	1	no	1	no
10	1	2	no	1	no	2	yes
11	1	2	no	1	yes	1	no
12	1	2	no	1	yes	2	yes
13	1	2	no	2	no	1	no
14	1	2	no	2	no	2	yes
15	1	2	no	2	yes	1	no
16	1	2	no	2	yes	2	yes
17	1	1	yes	1	no	1	yes
18	1	1	yes	1	no	2	yes
19	1	1	yes	1	yes	1	yes
20	1	1	yes	1	yes	2	yes
21	1	1	yes	2	no	1	no
22	1	1	yes	2	no	2	yes
23	1	1	yes	2	yes	1	no
24	1	1	yes	2	yes	2	yes
25	1	1	no	1	no	1	no
26	1	1	no	1	no	2	yes
27	1	1	no	1	yes	1	no
28	1	1	no	1	yes	2	yes
29	1	1	no	2	no	1	no
30	1	1	no	2	no	2	yes
31	1	1	no	2	yes	1	no
32	1	1	no	2	yes	2	yes

Figure 4: A snippet of expert-modeled decision rules using the DEX method

A what-if analysis was performed (called ± Analysis in DEXi), a task that is uncommon usingML models, but intrinsically supported in DSS models.

In Figure 4 two clients (rows 6, and 3328) were chosen for the analysis. They are both identified as churners. The analysis reveals insights what intervention could be done with this client in order that the client changes his/her mind. In the figure an adverse decision (no) can be seen five times, three times with client 6, and two times with client 3328. Client 6 would have a different churn outcome if: The number of *Day Minutes* would drop from category 3 (high) to category 2 (low), *Eve Minutes* would drop from category 2 (high) to category 1 (low), and if *Voice Mail* would change from category no to category yes. Client 3328 would change the outcome in two cases: if *Day Minutes* would increase from category 1 (very low) to 2 (low), and if *Customer Service calls* would change from category 2 (high) to category 1 (low).

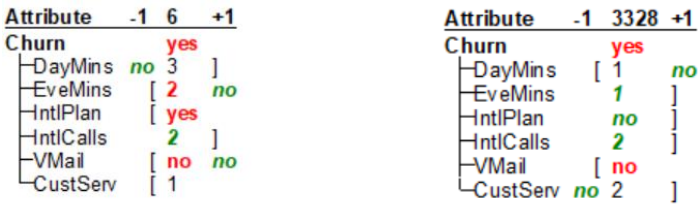


Figure 5: ±1 Analysis for clients 6 and 3328

3.3 The Second DSS model

The second DSS model was extracted using the DIDEX method (Radovanović et al. 2023), which automatically builds a DSS model from data. The difference with the DEX method is that in DEX method data does not exist, and experts create the attribute hierarchy and the decision rules by themselves. DIDEX extracts the attribute hierarchy and decision rules automatically from data. Only the names of newly constructed attributes are proposed by the user. The second DSS model is shown in Figure 5. After being learned from data, the model was read into DEXi software.

The attributes *International Plan* and *Voice Mail Plan* are grouped under a newly established feature [*Plans*], *International Calls* and *Customer Service Calls* are grouped under a new feature [*Calls*]. *Day Minutes* and *Evening Minutes* are grouped under a new feature [*Minutes*], where the features [*Calls*] and [*Minutes*] are grouped in a new feature [*Minutes & Calls*]. Finally, the decision on [*Churn*] is made based on features [*Plans*] and [*Minutes & Calls*].

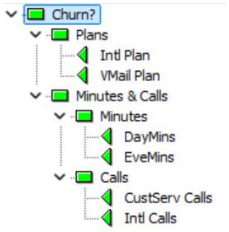


Figure 6: The expert modeled DSS using DEX method

A total of 22 decision rules were extracted using DIDEX method. From left to right, as shown in Figure 6, these rules are used to generate decision outcomes for the features [*Churn*], [*Plans*], [*Minutes & Calls*], [*Minutes*], and [*Calls*].

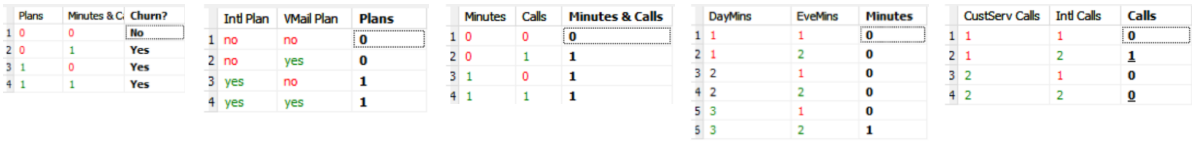


Figure 7: The extracted rules using DIDEX method

The 22 extracted rules were again applied to the whole dataset, and the following performance of the DSS model were achieved: Accuracy 64.66%, Recall 79.09%, Precision 26.18%, and F1 39.34%.

The what-if analysis was then performed on the client 3333 (Figure 7). This time a comparison was made between the first DSS (DEX expert modelled) model, and the second DSS (DIDEX extracted) model.

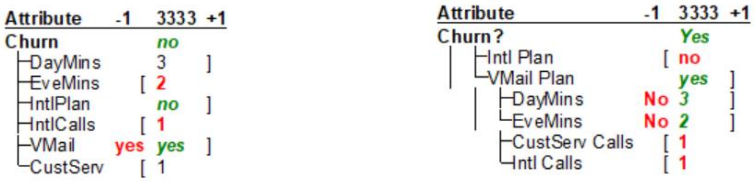


Figure 8: ±1 Sensitivity Analysis for clients 6 and 3328

It can be noticed that the two DSS models produce different churn predictions and have different recommendations on what to do if the churn decision should be different. While the first model thinks that the client 3333 will not churn and would do this in case where he/she would stop using voice mail, the second model thinks that the costumer would leave, and this outcome would have changed if either *Day Minutes* or *Eve Minutes* were reduced.

3.4 Which model to chose

It can be noticed that the ML model had the highest accuracy, precision and F1 value, and the poorest recall. The DSS DEX model achieved lower values from the ML model for all performance measures but recall, which was higher. The DSS DIDEX model had the lowest accuracy, precision, and F1 values, but the highest recall. The question is which model would be the most adequate.

According to Taskin (2023) in the case of telecom churn a false negative cost between 5 and 7 times more than a false positive, so the ranking of the models is:

1. DSS DEX model
2. DSS DIDEX model
3. ML model

This solution is stable on the whole interval [5,7]. The confusion matrices for the three models are shown in Table 1.

Table 6. Confusion matrices for DEX DSS model, DIDEX DSS model, and ML model

DEX	No	Yes	**DIDEX**	No	Yes	**ML**	No	Yes
No	2186	664	No	1765	1085	No	2750	100
Yes	179	304	Yes	173	310	Yes	209	274

While this result was highly unexpected, it was shown on this example of customer churn that a fully crafted DSS model can outperform, in some cases, ML models. In this example the recall of the DSS DEX model was the most adequate for tackling the costs of false classification. Having in mind the other DSS properties being satisfied, and the what-if analysis possibilities of the DSS models, the DSS model produced by DEX would be the most adequate choice in this situation.

4. Conclusions

In the 2010s, ML outperformed and mainly replaced DSSs due to a large availability of big data, availability of bigger processing power, easier integration into business processes, among others. ML does not require deep modeling from domain experts and easily overcomes their objective limitations.

Off-the-shelf ML algorithms most often optimize for accuracy, which may not be aligned with business goals. Even though users can influence this through decision thresholds or cost matrices, one needs to test whether these adjustments achieve the desired outcomes. Therefore, the property Correctness can easily be dissatisfied. In addition, if the available data is not representative of the entire population but is instead a biased sample (due to representation or sampling bias), the resulting model will inherit and potentially amplify those biases, leading to unfair or unreliable predictions which

endangers Consistency. This property is recently handled according to “responsible AI systems” that are fair, and transparent. The first-generation ML algorithms would easily fail the consistency test of DSSs. Comprehensibility is a feature that is recently tackled by explainable AI systems, where previously ML systems were often regarded as black-boxes, and this was acceptable due to their high accuracy performances. The property Convenience of the DSS is just partially satisfied by ML systems, as ML systems do have a lot of convenient properties, which made them the dominant AI technology. Easy integration into business processes, and automation of those, by lowering the costs of business processes significantly, made ML models the dominant technology. Still, intrinsic properties of DSS systems, like sensitivity analysis, are not easily carried out in ML systems.

We demonstrated in this short paper the strength of DSS systems, which, although lost their previous glory, still can produce valuable models for the decision-making process and in-depth decision analysis and can be a very respectable alternative to ML and AI systems.

References

Amin, A., Khan, C., Ali, I., & Anwar, S. (2014). Customer churn prediction in telecommunication industry: With and without counter-example. In Nature-Inspired Computation and Machine Learning: 13th Mexican International Conference on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, Mexico, November 16-22, 2014. Proceedings, Part II 13 (pp. 206-218). Springer International Publishing.

Amin, A., Shehzad, S., Khan, C., Ali, I., & Anwar, S. (2015). Churn prediction in telecommunication industry using rough set approach. New trends in computational collective intelligence, 83-95.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A., & Huang, K. (2017). Customer churn prediction in the telecommunication sector using a rough set approach. Neurocomputing, 237, 242-254.

Bohanec, M. (2021). From data and models to decision support systems: Lessons and advice for the future. In *EURO Working Group on DSS: A tour of the DSS developments over the last 30 years* (pp. 191-211). Cham: Springer International Publishing.

Britto, J. Gobinath R (2020).“A Detailed Review For Marketing Decision Making Support System In A Customer Churn Prediction,”. *Int. J. Sci. Technol. Res*, *9*(4), 3698-3702.

De, S., Prabu, P., & Paulose, J. (2021, September). Effective ML techniques to predict customer churn. In 2021 Third international conference on inventive research in computing applications (ICIRCA) (pp. 895-902). IEEE.

Delibašić, B., Radovanović, S., & Vukanović, S. (2023). A Decision Support System for Internal Migration Policy-Making. https://ipsitransactions.org/journals/papers/tir/2023jul/p7.pdf

Huang, Y., Huang, B., & Kechadi, M. T. (2011, May). A rule-based method for customer churn prediction in telecommunication services. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 411-422). Berlin, Heidelberg: Springer Berlin Heidelberg.

Papa, A., Shemet, Y., Yarovyi, A. (2021). Analysis of fuzzy logic methods for forecasting customer churn. Technology Audit and Production Reserves, 1 (2 (57)), 12–14. doi: http://doi.org/ 10.15587/2706-5448.2021.225285

Popović, D., & Bašić, B. D. (2008). Churn prediction model in retail banking using fuzzy c-means clustering. In Proceedings of the Conference on Data Mining and Data Warehouses, Ljubljana, Slovenia.

Popović, D., & Bašić, B. D. (2009). Churn prediction model in retail banking using fuzzy C-means algorithm. Informatica, 33(2).

Raj, R., Gupta, S., Mishra, R., & Malik, M. (2024, April). Efficacy of customer churn prediction system. In 2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT) (pp. 546-553). IEEE.

Radovanović, S., Bohanec, M., & Delibašić, B. (2023). Extracting decision models for ski injury prediction from data. *International Transactions in Operational Research*, *30*(6), 3429-3454.

Safinejad, F., Noughabi, E. A. Z., & Far, B. H. (2018). A fuzzy dynamic model for customer churn prediction in retail banking industry. Applications of Data Management and Analysis: Case Studies in Social Networks and Beyond, 85-101.

Taskin, N. (2023). Customer Churn Prediction Model In Telecommunication Sector Using Machinelearning Technique. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1774181&dswid=-3518

MCDM, Optimization, and Advanced Analytics

Towards Elevating DEMATEL with Spectral Analysis

Pavlos Delias[1*] and Kerasia Kalkitsa[1]

[1] Democritus University of Thrace, Kavala, Greece
[*] pdelias@af.duth.gr

Abstract: In this work we introduce a novel spectral analysis extension to the Decision-Making Trial and Evaluation Laboratory (DEMATEL) method, significantly enhancing its analytical capabilities for decision support. Although DEMATEL has proven valuable for mapping influence relationships in complex systems, its traditional analysis focuses primarily on direct and total effects through centrality measures. We demonstrate that the total relation matrix, under practical conditions of irreducible normalized direct influence, is diagonalizable with distinct eigenvalues. This mathematical property enables deeper insights into system dynamics, structural analysis, and intervention and control through spectral analysis. The dominant eigenvalue and its corresponding eigenvector reveal fundamental system characteristics including influence propagation patterns, stability thresholds, and potential cascade effects. This theoretical advancement provides practitioners with new tools for identifying optimal intervention points and assessing system-wide impacts. We illustrate these enhanced analytical capabilities through a case study of food safety governance factors in rural China, where spectral analysis reveals previously undetectable patterns of influence amplification and system convergence properties. The proposed extension transforms DEMATEL from a static influence mapping tool into a framework capable of quantifying dynamic system behavior, while maintaining its computational tractability and practical applicability.

Keywords: DEMATEL; spectral analysis; complex systems; decision support

1. Introduction

As systems grow in complexity, traditional analytical methods often fall short in capturing the cause-effect relationships, the subtle interplay between components and the cascading effects of interventions. Among the various methodologies developed to address this challenge, the Decision-Making Trial and Evaluation Laboratory (DEMATEL) method (Gabus & Fontela, 1972) has emerged as a particularly powerful tool, owing to its ability to not only map direct relationships but also quantify indirect influences propagating through the system (Quezada et al., 2018). This capability has made DEMATEL increasingly popular across diverse fields, from supply chain management to sustainability assessment, where understanding the ripple effects of decisions is crucial for effective system intervention (Si et al., 2018).

Although DEMATEL has become a popular methodology in complex system analysis (Chen, 2021), its analytical depth has remained relatively unchanged since its introduction. The method's standard procedure - constructing a direct influence matrix, applying normalization, and deriving the total relation matrix - provides valuable insights into system structure and component relationships (Sorooshian et al., 2023; Taherdoost & Madanchian, 2023). Current applications primarily focus on analyzing centrality measures and impact-relation maps derived from the total relation matrix. However, we have yet to fully leverage the mathematical properties of this matrix, particularly its diagonalizability, for more sophisticated system analysis beyond traditional DEMATEL outputs . This gap suggests a significant opportunity for methodological advancement.

This paper introduces a significant enhancement to DEMATEL through spectral analysis of the total relation matrix. By establishing that this matrix is diagonalizable under practical conditions of irreducible normalized direct influence, we unlock access to its eigenvalue spectrum. The maximum eigenvalue, in particular, serves as a powerful analytical tool that reveals fundamental system properties previously inaccessible through standard DEMATEL analysis. This spectral extension enables several crucial insights like the system dynamics and the influence propagation, the system's stability characteristics, the understanding of cascade effects, the detection of optimal leverage points for system intervention, and the intervention planning. These insights significantly expand DEMATEL's analytical capabilities while maintaining its computational tractability and practical applicability. The transition

from standard DEMATEL to spectral analysis follows a natural mathematical progression. The proposed approach complements traditional DEMATEL outputs by adding a new analytical dimension: on top of centrality measures and impact-relation maps that reveal the static distribution of influence, our spectral analysis extension uncovers the system's dynamic potential, stability boundaries, and critical intervention thresholds—fundamentally expanding DEMATEL's capabilities for complex system analysis.

2. Background

2.1 DEMATEL

The Decision-Making Trial and Evaluation Laboratory (DEMATEL) method is a structural modeling approach that analyzes complex causal relationships between components in a system. Originally developed by the Geneva Research Centre of the Battelle Memorial Institute (Gabus & Fontela, 1972), DEMATEL transforms qualitative assessments into quantifiable systemic relationships. The method begins with expert evaluations of direct influences between system factors, recorded in a direct influence matrix $\boldsymbol{A} = [a_{ij}]$, where a_{ij} represents the degree to which factor i influences factor j on a scale typically ranging from 0 (no influence) to 4 (very high influence).

The core DEMATEL procedure involves normalizing the direct influence matrix $\boldsymbol{A}$ into matrix $\boldsymbol{D} = \boldsymbol{A}/s$, where $s = \max\{\text{maximum row sum}, \text{maximum column sum}\}$, followed by calculating the total relation matrix $\boldsymbol{T} = \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{D})^{-1}$, where $\boldsymbol{I}$ is the identity matrix. From matrix $\boldsymbol{T}$, row sums (r) represent the total influence that factor i exerts on others (dispatch), column sums (c) indicate the total influence received by factor j (receipt), while $(r + c)$ and $(r - c)$ values respectively reveal the prominence (total involvement) and net effect (overall influence) of each factor in the system. This basic framework has been extended to include threshold values for significance (Li & Tzeng, 2009), helping practitioners focus on the most important relationships. These outputs enable decision-makers to visualize and understand the interdependencies between factors, identifying those with the greatest system-wide impact.

2.2 Algebraic Foundations

Matrix analysis provides essential tools for understanding complex systems through their mathematical properties. At its foundation are eigenvalues and eigenvectors: for a matrix A, if there exists a non-zero vector v and a scalar λ satisfying $Av = \lambda v$, then λ is called an eigenvalue and v is its corresponding eigenvector. A matrix is diagonalizable if it has a complete set of linearly independent eigenvectors, allowing it to be decomposed as $A = PDP^{-1}$, where D is a diagonal matrix of eigenvalues and P contains the corresponding eigenvectors. Among the eigenvalues, the spectral radius ρ(A) - the largest absolute value among all eigenvalues - is particularly significant for understanding system behavior.

When analyzing influence relationships, certain matrix properties become crucial. A nonnegative matrix (where all elements $a_{ij} \geq 0$) is called irreducible if its associated directed graph is strongly connected, meaning there exists a path between any two elements in the system. For such matrices, the Perron-Frobenius theorem establishes that the spectral radius is itself an eigenvalue, and its corresponding eigenvector has strictly positive components. These fundamental concepts from matrix theory (Meyer, 2023) provide the mathematical foundation for analyzing complex systems. These properties provide a rigorous foundation for analyzing interconnected systems, particularly when studying influence patterns.

3. Methodology

3.1 Theoretical Framework

Building upon the established DEMATEL methodology, we develop a spectral analysis approach to extract additional insights from the Total Relation Matrix $\boldsymbol{T}$. This section presents the theoretical

foundations that enable this extension.

3.1.1. Spectral Properties of the Total Relation Matrix

Our analysis begins with a fundamental result about the eigenstructure of $\boldsymbol{T}$:

Theorem 1: *If the normalized direct influence matrix $\boldsymbol{D}$ is irreducible, then the Total Relation Matrix $\boldsymbol{T}$ is diagonalizable with distinct eigenvalues.*

The proof relies on several key properties: $\boldsymbol{D}$ is irreducible, meaning its associated directed graph is strongly connected. By construction, $\boldsymbol{D}$ has also zero diagonal, and it is nonnegative, with all entries $d_{\text{ij}} \geq 0$. Under these conditions:

1. By Gershgorin's theorem (Varga, 2004) and the normalization of $\boldsymbol{D}$, all eigenvalues μ of $\boldsymbol{D}$ satisfy $|\mu| < 1$.
2. For any eigenvalue λ of $\boldsymbol{T}$ with eigenvector $\boldsymbol{v}$: From $\boldsymbol{Tv} = \lambda\boldsymbol{v}$, we obtain $\boldsymbol{D}(\boldsymbol{I}-\boldsymbol{D})^{-1}\boldsymbol{v} = \lambda\boldsymbol{v}$. This yields $\boldsymbol{Dv} = \lambda\boldsymbol{v} - \lambda\boldsymbol{Dv}$. Therefore, $(1+\lambda)\boldsymbol{Dv} = \lambda\boldsymbol{v}$.
3. This establishes a one-to-one correspondence: If $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{T}$ with eigenvalue λ, then $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{D}$ with eigenvalue $\mu = \lambda/(1+\lambda)$. Conversely, if μ is an eigenvalue of $\boldsymbol{D}$, then $\lambda = \mu/(1-\mu)$ is an eigenvalue of $\boldsymbol{T}$.

The irreducibility of $\boldsymbol{D}$, combined with its zero-diagonal structure, ensures distinct eigenvalues through Perron-Frobenius theory and the properties of strongly connected graphs ■

3.1.2. Spectral Decomposition

Under the conditions above, $\boldsymbol{T}$ admits the decomposition $\boldsymbol{T} = \boldsymbol{P\Lambda P^{-1}}$ where:

- $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ contains the distinct eigenvalues
- $\boldsymbol{P}$ contains the corresponding eigenvectors

This decomposition, fundamental to numerical matrix analysis (Golub & Van Loan, 2013), enables efficient computation of system properties through the eigenstructure. The dominant eigenvalue λ_{max} carries particular significance for system analysis, as we demonstrate in the following section.

3.2 Interpretations

The dominant eigenvalue λ_{max} of the total relation matrix $\boldsymbol{T}$ reveals fundamental insights into the system's structural characteristics and influence dynamics. Specifically, λ_{max} quantifies the maximum level of long-term influence that can propagate through the system, where larger values indicate stronger overall interdependencies between factors. The relationship between λ_{max} and the sum of all eigenvalues ($l = \lambda_{\text{max}}/\sum\lambda_i$) represents the system's concentration of influence in its primary mode, with values closer to 1 indicating a more dominated system. The corresponding eigenvectors further enrich this analysis by revealing influence clusters - groups of factors that exhibit similar influence patterns and collectively drive system behavior. These eigenvectors effectively decompose the complex network of relationships into distinct patterns of co-influence.

Then, the convergence properties illuminated by λ_{max} provide critical insights into system stability and behavior dynamics. The magnitude of λ_{max} (also called the spectral radius) directly corresponds to the convergence rate of indirect effects, where higher values indicate slower convergence and thus necessitate more iterations to capture all significant indirect influences. This relationship makes λ_{max} an effective measure of system complexity and coupling intensity. For instance, in systems with high λ_{max} values, indirect effects persist longer and potentially amplify through multiple paths, indicating a more intricately coupled system. This characteristic is particularly relevant when analyzing systems where cascade effects are of concern, as it helps quantify the potential for influence amplification through indirect ways.

The spectral analysis framework enables robust comparative analyses across different systems or temporal states. Changes in λ_{max} serve as reliable indicators of structural shifts in system interdependencies, providing a quantitative basis for monitoring system evolution. The sensitivity

analysis, performed through the examination of partial derivatives $\partial\lambda_{\max}/\partial a_{ij}$, identifies which relationships most significantly impact system coupling. This mathematical framework allows for systematic comparison of different system states or configurations, enabling researchers to quantify and validate structural changes in the system's interdependency patterns over time or across different contexts.

From an intervention perspective, $\lambda_{\max}$ serves as a crucial indicator for risk assessment and strategy development. Systems characterized by higher $\lambda_{\max}$ values demonstrate increased susceptibility to cascade effects, necessitating careful consideration of indirect influences in intervention planning. The components of the dominant eigenvector provide additional strategic guidance by identifying the most influential system elements - those with larger eigenvector values represent more significant *leverage points* for system modification. This understanding enables more targeted and effective risk mitigation strategies, allowing practitioners to focus interventions where they will have the most substantial impact while considering the potential for unintended consequences through indirect effects. The ways that the spectral analysis can augment the analysis potentials of DEMATEL are summarized and illustrated in Table 1.

Table 7. The contribution potentials of spectral analysis of the total relation matrix

<table>
<tr><th>System Dynamics & Stability</th><th>Structural Analysis & Pattern Detection</th><th>Intervention & Control Analysis</th></tr>
<tr><td>Maximum level of long-term influence that can propagate through the system</td><td>Strength of overall interdependencies between factors</td><td rowspan="2">Risk assessment and mitigation strategies:
- Systems with higher $\lambda_{\max}$ require more careful intervention planning.
- Greater attention to indirect effects in highly coupled systems.</td></tr>
<tr><td>Overall strength of the network’s total influence considering all direct/indirect effects</td><td>Ratio of $\lambda_{\max}$ to sum of all eigenvalues reveals primary influence mode dominance</td></tr>
<tr><td>Maximum amplification of influence possible through indirect effects</td><td>Eigenvectors reveal influence clusters (co-driving factors)</td><td rowspan="2">Sensitivity analysis through:
- Partial derivatives for relationship impact assessment
- Eigenvector $\boldsymbol{v}$ analysis for identifying leverage points</td></tr>
<tr><td>System’s stability and convergence characteristics</td><td rowspan="3">Comparative analysis capabilities:
- System comparison across different instances
- Temporal comparison of same system
- Validation of structural changes</td></tr>
<tr><td>Potential for cascade effects</td><td rowspan="2">Factor importance weighting and strategic intervention point identification</td></tr>
<tr><td>System complexity and coupling degree measurement</td></tr>
</table>

4. Case Study

4.1 Original Study

The original study of (Xie et al., 2023) examines the complex system of factors influencing farmers' participation in food safety governance in rural China. The researchers identified 20 key influencing factors across four dimensions: family characteristics (including education level, village cadre status, household income, family structure, eating habits, and self-supply of food), participant characteristics (victim experience, political trust, risk perception, media attention, government supervision, and village committee promotion), participation process (perception of effectiveness, cost perception, and government response), and participation environment (participation atmosphere, rural informatization degree, publicity of government information, participation channels, and incentive mechanisms). Using

the DEMATEL methodology alongside ISM and MICMAC analyses, they surveyed 20 experts in public management and food safety governance to evaluate the interrelationships and relative importance of these factors.

Through their DEMATEL analysis, the researchers evaluated both the influence degree (comprehensive influence of one factor on others), affected degree (how much a factor is influenced by others), center degree (overall importance in the system), and cause degree (net influence) of each factor. Their results were presented in a comprehensive influence matrix and summarized in tables showing these four key metrics for each of the 20 factors. The findings revealed that education level (a1) and village cadre status (a2) had the highest influence degrees (4.17 and 4.25 respectively) and cause degrees (both 1.69), while risk perception (a9) had the highest center degree (7.21) but a negative cause degree (-0.82), indicating it was more affected by other factors than influential. Family eating habits (a5) showed the lowest cause degree (-1.42), suggesting it was the most heavily influenced by other factors in the system.

4.2 Spectral Analysis for the System Dynamics & Stability

Maximum Level of Long-term Influence Propagation

The spectral analysis of the diagonalizable total relation matrix $\boldsymbol{T}$ reveals a maximum eigenvalue ($\lambda_{\max}$) of 3.27. This value indicates potential for long-term influence propagation within the system. When interpreting this metric for decision-making purposes, we observe that any $\lambda_{\max}$ exceeding 1 suggests that initial changes in the system can amplify over time. This characteristic demands careful consideration when designing interventions, as decision-makers must ensure high-quality implementations at critical entry points while allowing sufficient time for effects to mature through the system.

Our food security case study exemplifies this principle, where the impact of the primary influence drivers (education and village cadre status) grows systematically through rural networks, indicating that well-executed food safety education programs could create sustainable improvements in system resilience.

Overall Strength of the Network's Total Influence

The spectral analysis reveals several key indicators of network strength: $\lambda_{\max}$ approximates the Frobenius norm $\|\boldsymbol{T}\|_F$ (3.28), accompanied by a high $\lambda_{\max}/\lambda_2$ ratio (~51), and uniform eigenvector components ranging from 0.165 to 0.284. Together, these metrics indicate a robust, single-pattern influence structure. For decision-makers, this coherence suggests that system responses will be more predictable, favoring coordinated intervention strategies over fragmented approaches. The clear dominance of a single influence pathway provides an opportunity to align actions for maximum impact.

In our food security study, this unified influence pattern supports the implementation of integrated policies. The combination of education and cadre training programs demonstrates how interventions can achieve scalable impacts through predictable propagation pathways.

Maximum Amplification of Influence Through Indirect Effects

The spectral analysis yields crucial insights into influence amplification: with $\lambda_{\max} = 3.27$ and $\rho(\boldsymbol{D}) = 0.766$, we calculate a maximum amplification factor of 4.27 times the direct effects via $\boldsymbol{T} = \boldsymbol{D}(\boldsymbol{I} - \boldsymbol{D})^{-1}$. This amplification is stabilized by rapid secondary mode decay ($\lambda_2/\lambda_{\max} \approx 0.02$). This significant amplification potential emphasizes the importance of monitoring indirect effects in decision-making contexts, as minor interventions can escalate substantially throughout the system. To maintain control over this amplification potential, decision-makers should implement and test interventions incrementally.

Our case study demonstrates this principle, showing how a modest educational initiative could potentially amplify 4.27-fold across rural food security networks, supporting the study's emphasis on informatization as a high-leverage enhancement mechanism.

System's Stability and Convergence Characteristics

The spectral analysis reveals a high condition number ($\kappa = \lambda_{\max}/\lambda_{\min} \approx 1637$) and a convergence rate of approximately 3.93, derived from $\ln(\lambda_2/\lambda_{\max})$. These metrics indicate that while the system shows numerical sensitivity, it demonstrates rapid pattern alignment. For decision-makers, the high condition number necessitates precise data collection and analysis, but the fast convergence rate ensures predictable outcomes, supporting robust policy design that can be validated through multiple methods.

For the food safety governance case study, this means that interventions will follow predictable patterns of influence, so our focus should be on the dominant influence paths identified in the study. Then, because of the high κ, multiple measurement approaches are needed for reliable assessment of relationships.

5. Conclusions

The spectral analysis extension of DEMATEL we introduced in this work represents a significant advancement in this decision support methodology. Traditional DEMATEL analysis identifies influence patterns and centrality measures, but our approach unlocks deeper insights into system dynamics, stability characteristics, and intervention potential through eigenstructure examination. The food safety governance case study demonstrates these enhanced analytical capabilities: our method revealed critical patterns of influence amplification, system convergence properties, and coherent influence structures that standard DEMATEL analysis could not detect. These mathematical properties empower decision-makers to assess system stability, anticipate cascade effects, and identify strategic intervention points with unprecedented clarity. This methodological enhancement transforms DEMATEL from a static influence mapping tool into a dynamic framework capable of predicting system behavior and guiding targeted interventions with scientific rigor.

References

Chen, J.-K. (2021). Improved DEMATEL-ISM integration approach for complex systems. *PLOS ONE*, *16*(7), e0254694. https://doi.org/10.1371/journal.pone.0254694

Gabus, A., & Fontela, E. (1972). *World problems, an invitation to further thought within the framework of DEMATEL*. Battelle-Geneva Research Center.

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations* (Fourth edition). The Johns Hopkins University Press.

Li, C.-W., & Tzeng, G.-H. (2009). Identification of a threshold value for the DEMATEL method using the maximum mean de-entropy algorithm to find critical services provided by a semiconductor intellectual property mall. *Expert Systems with Applications*, *36*(6), 9891–9898. https://doi.org/10.1016/j.eswa.2009.01.073

Meyer, C. D. (2023). *Matrix analysis and applied linear algebra: Study and solutions guide* (Second edition). Siam, Society for Industrial and Applied Mathematics.

Quezada, L. E., López-Ospina, H. A., Palominos, P. I., & Oddershede, A. M. (2018). Identifying causal relationships in strategy maps using ANP and DEMATEL. *Computers & Industrial Engineering*, *118*, 170–179. https://doi.org/10.1016/j.cie.2018.02.020

Si, S.-L., You, X.-Y., Liu, H.-C., & Zhang, P. (2018). DEMATEL Technique: A Systematic Review of the State-of-the-Art Literature on Methodologies and Applications. *Mathematical Problems in Engineering*, *2018*, 1–33. https://doi.org/10.1155/2018/3696457

Sorooshian, S., Jamali, S. M., & Ebrahim, N. A. (2023). Performance of the decision-making trial and evaluation laboratory. *AIMS Mathematics*, *8*(3), 7490–7514. https://doi.org/10.3934/math.2023376

Taherdoost, H., & Madanchian, M. (2023). Understanding Applications and Best Practices of DEMATEL: A Method for Prioritizing Key Factors in Multi-Criteria Decision-Making. *Journal of Management Science & Engineering Research*, *6*(2), 17–23. https://doi.org/10.30564/jmser.v6i2.5634

Varga, R. S. (2004). *Geršgorin and His Circles* (Vol. 36). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-17798-9

Xie, J.-H., Tian, F.-J., Li, X.-Y., Chen, Y.-Q., & Li, S.-Y. (2023). A study on the influencing factors and related paths of farmer's participation in food safety governance—Based on DEMATEL-ISM-MICMAC model. *Scientific Reports*, *13*(1), 11372. https://doi.org/10.1038/s41598-023-38585-w

Predicting NBA Longevity: The Role of Physical Attributes and Early Career Performance

Petar Graovac[1*], Ilija Savić[1] and Milan Radojičić[1]

[1] University of Belgrade, Faculty of Organizational Sciences, Jove Ilića 154, 11000 Belgrade, Serbia
[*] petar.graovac@fon.bg.ac.rs

Abstract: Basketball analytics has transformed talent evaluation, influencing scouting, drafting, and player development decisions in the NBA. This study applies logistic regression with Lasso and Ridge regularization, to predict whether a player will reach their fifth NBA season taking their position into account. Using data from 618 college players who debuted in the NBA between 2010 and 2019, physical attributes, college performance metrics, and early NBA career statistics were analysed. Findings indicate that physical traits alone offer little predictive power for guards and forwards, but for centers they play a more significant role. College and NBA performance metrics significantly enhance prediction accuracy. The model using college and first-year NBA data achieves over 0.7 AUC across all positions, making it the earliest reliable point for predicting long-term NBA success. The most effective model integrates college statistics and both first and second year NBA performance data, achieving an AUC of 0.86 for forwards with Lasso regularization. These insights assist NBA teams in refining their draft strategies, improving player evaluation, and identifying long-term prospects. Future research can expand this approach by incorporating additional predictive techniques and more complex models, external factors such as injuries and team fit, alternative statistical indicators and the use of counterfactual examples.

Keywords: Machine learning; Logistic Regression; Sport Analytics; Player Performance Evaluation

1. Introduction

Basketball is more than a sport, it is a powerful economic, social, and commercial force that shapes global communities (Szymanski, 2010). Since professional leagues began, teams have used sports analytics to change how they pick players, plan strategies, and evaluate performance. Their aim is to secure more victories, attain higher levels of success, and build stronger organizations both on the court and off it (Steinberg, 2015).

Since its inception, the NBA has transformed basketball by drafting talented players and welcoming international stars alongside college prospects (Wilkinson, 2024). These changes have made the league more competitive, diverse, and dynamic and players are recruited and drafted through advances in sports science and analytics (Sarlis & Tjortjis, 2020). Today, it is the most famous basketball league and a top entertainment brand (Sun, 2015).

A growing body of sports analytics research seeks to understand how and why basketball players achieve long-term success (Karipidis et al., 2001). In today's data-driven NBA landscape, teams are increasingly looking beyond traditional scouting to predict a player's lasting prosperity (Moxley & Towne, 2015). Players secure a second contract by the time they reach the fifth season, after the first two guaranteed years and two additional years with team options (Zhou et al., 2023). Based on the analysis of 4,979 NBA players to date, the data revealed that the average career length in the league is 5.1 seasons, with 42.6% of players reaching their fifth year (Kubatko et al., 2007). Also, the salary for a player's second NBA contract is significantly higher than that of their first (Kuehn & Rebessi, 2023). This study uses machine learning to combine physical attributes with college, rookie, and sophomore performance metrics to predict which players will reach that fifth season mark taking their position into account.

Over the past decade, machine learning in sports prediction has been used to forecast winning teams (Teeter & Bergman, 2020) and manage player load and injuries (Robertson, 2014). Machine learning has also helped analyze how injuries affect basketball players and team performance (Sarlis et al., 2021).

In addition, these methods have been applied to predict individual player performance (Radovanović et al., 2014; Lee & Page, 2021), track player movement (Stephanson et al., 2021) and forecast game results in both the NBA and college sports (Thabtah, 2019; Ruiz & Perez-Cruz, 2015). Barnes (2008) further explored the factors that influence career longevity in the NBA.

The remainder of the paper is organized as follows. Section 2 explains the data and methodology used. Section 3 presents the findings and analysis, using logistic classification in multiple models analyzing player longevity in the NBA. Section 4 concludes the paper.

2. Data and Methodology

The Data and Methodology section includes a description of the data and the process of feature extraction from the basketball-reference database (Kubatko et al., 2007). The extracted data will then be utilized to develop a logistic regression model. For the analysis Google Collaboratory is used.

2.1 Data

The dataset contains 618 college players who made an NBA debut, with 360 having NBA Draft Combine measurements (body fat, hand size, standing reach, height, weight, and wingspan). Players are categorized into three positions: guard, forward and center. Data from players' college careers and their first two NBA seasons (from 2010 to 2019) were used to predict if their careers would last at least five years. College performance metrics from players' final season include basic (points, assists, rebounds, steals, blocks, and shooting percentages) (Barnes, 2008) and advanced stats (offensive rating, defensive rating, usage rate, box plus/minus, effective field goal percentage) (Kekez et al., 2021). Player performance from the first and second seasons was similarly evaluated using traditional and advanced indicators (added value over replacement player (VORP), player efficiency rating (PER), win share and usage rate). Players without NBA game appearances in the second season were assigned zeros for all statistics. For each year of analysis and player position, the dataset was split into training (80%) and test (20%) sets.

2.2 Logistic regression

Logistic regression is widely regarded as one of the most frequently used machine learning algorithms, with applications spanning numerous fields. Its primary advantage lies in its interpretability; the model's coefficients can be explained in terms of odds ratios and, by extension, probabilities. This clarity is particularly valuable in social science contexts, where each decision must be thoroughly justified. Essentially, logistic regression is a classification method that estimates the probability of a binary outcome y based on a set of independent variables X (Hastie, 2015). The formulation of this model is provided in equation (1).

$$\log\left(\frac{p}{1-p}\right) = \theta_0 + \theta_1 x_1 + \cdots + \theta_k x_k \tag{1}$$

where p represents the probability that player is going to play in fifth season (y = 1). Values of θ represent weights associated with independent features X. One wants to find the best values of θ such that the model provides the lowest possible error. Error is defined through loss function, called logistic loss (presented in the formula (2)), which must be minimized (Radovanović et al., 2021).

$$\min L\,(y, \hat{y}) = -\frac{1}{n}\sum_{i=1}^{n}[y \log(\hat{y}) + (1-y)\log(1-\hat{y})] \tag{2}$$

One challenge with using the logistic loss function is the risk of overfitting, fitting the training data too closely without retaining the ability to generalize to unseen examples, particularly when dealing

with many features. To address this, variations of logistic regression incorporate regularization into the logistic loss function. By adding a regularization term, the training process tolerates a slightly higher loss in exchange for better generalization to new data (James et al., 2013). In this paper both Lasso and Ridge regularization is used Lasso regularization incorporates the L1 norm, which drives some logistic regression coefficients to zero, effectively reducing the number of features needed to model the problem. Ridge regularization adds L2 norm which forces coefficients of logistic regression to be lower in general (Robert, 2011).

3. Analysis and Results

To predict whether a player's NBA career will last at least five years, four predictive models were developed. The first model depended exclusively on physical characteristics, meaning it used only draft combine measurements as its variables. The second model relied exclusively on college performance statistics, using seven variables for the basic and four for the advanced set. The third model integrated college statistics with data from the first NBA season by supplementing the college metrics with both basic and advanced performance data and total count of variables for basic is 17 and for advanced 8. The fourth model combined data from college and the first two NBA seasons, using 27 basic and 12 advanced variables. This method allowed us to evaluate how incorporating extra layers of performance data enhanced predictions of long-term NBA success across different positions.

Table 1 and Table 2 displays the study's outcomes using these models across different player positions. It is evident that as additional data regarding performance in the NBA is incorporated, the model's performance steadily improves.

Table 8. Performance of Lasso logistic regression models

	Lasso					
	Basic Variables (AUC)			Advanced Variables (AUC)		
	Center	Forward	Guard	Center	Forward	Guard
Model 1	0.67 ± 0.127	0.36 ± 0.043	0.52 ± 0.052	/	/	/
Model 2	0,54 ± 0,048	0,53 ±0,039	0,52 ± 0,034	0,68 ± 0,062	0,56 ± 0,023	0,58 ± 0,033
Model 3	0,73 ± 0,074	0,80 ± 0,038	0,72 ± 0,019	0,82 ± 0,072	0,71 ± 0,032	0,74 ± 0,042
Model 4	0,79 ± 0,077	0,86 ± 0,041	0,71 ± 0,027	0,74 ± 0,056	0,76 ± 0,040	0,73 ± 0,034

Table 9. Performance of Ridge logistic regression models

	Ridge					
	Basic Variables (AUC)			Advanced Variables (AUC)		
	Center	Forward	Guard	Center	Forward	Guard
Model 1	0,67 ± 0,118	0,36 ± 0,041	0,54 ± 0,047	/	/	/
Model 2	0,54 ± 0,053	0,55 ± 0,035	0,53 ± 0,032	0,68 ± 0,067	0,57 ± 0,027	0,58 ±0,030
Model 3	0,75 ± 0,084	0,77 ± 0,031	0,73 ± 0,02	0,74 ± 0,074	0,71 ± 0,033	0,73 ± 0,038
Model 4	0,79 ± 0,070	0,84 ± 0,036	0,74 ± 0,028	0,76 ± 0,064	0,76 ± 0,039	0,72 ± 0,034

Model 1, which relies solely on a player's physical attributes, struggles to differentiate between guards and forwards who have played at least five seasons in the NBA and those who have not, but it performs better in distinguishing such players among centers. Both Ridge and Lasso regularization methods yield similar AUC values In Model 2, even with Lasso and Ridge regularization applied to the basic variables, it remains challenging to accurately determine which players will make it to their fifth NBA season and which will not, however, when advanced variables are incorporated, results notably improve, reaching an AUC of 0.68 for centers. By including performance data from the first NBA season in Model 3 and from the first two NBA seasons in Model 4, the models demonstrate improved classification regardless of position. In Model 4, both Lasso and Ridge regularization yield strong AUC values of 0.86 and 0.84 for forwards, respectively. Generally, the overall AUC values for both Lasso and Ridge regularization remain similar when comparing basic to advanced variable models; however, a closer look at player positions reveals important nuances. In Model 3, for example, the impact of incorporating advanced variables is markedly position dependent. For centers, advanced variables push the AUC up from approximately 0.73 to 0.82 under Lasso. By contrast, forwards experience a drop from around 0.80 with basic variables to 0.71 with advanced ones while guards show only a modest improvement. Comparing the best configurations across models, Model 4 clearly outperforms Model 2 in both absolute terms and percentage gains. For instance, with Lasso regularization, the top AUC for forwards increases from roughly 0.53 in Model 2 to about 0.86 in Model 4, a gain of 61%. Likewise, the Ridge model experiences a similar enhancement. These improvements differ by player position. For centers, incorporating additional NBA performance data leads to substantial improvements, while guards show minimal differences between Models 3 and 4, with a significant boost compared to Model 2.

To determine how soon it can be predicted if a player reaches second NBA contract, it is crucial to establish what qualifies as good enough model performance. Nevertheless, all the models under consideration offer utility and surpass random (uninformed) decision making. According to a common rule of thumb, an AUC of 0.7 is often viewed as indicative of a good model (Radovanović et al., 2021).

Model 1 shows that relying solely on physical attributes offers little predictive value for guards and forwards, often only slightly better than random, while providing reasonable differentiation for centers. This indicates that although athleticism may influence draft stock, it is a weak predictor of long-term success for guards and forwards, where factors like skill, decision-making, and adaptability are likely more important. Consequently, teams should avoid overemphasizing physical traits, as this may result in overlooking players with high potential who lack standout physical profiles, even though strong physical attributes still play a significant role for centers.

Model 2, which utilizes college performance data, indicates that pre-NBA statistics possess moderate overall significance. For guards and forwards, both Lasso and Ridge deliver similar AUCs, basic variables yield around 0.52 - 0.55 and advanced variables about 0.56 - 0.58, indicating minimal gains from advanced analytics. While college statistics alone aren’t enough, they provide a solid starting point for further research. However, centers see a notable boost, with AUCs rising from roughly 0.54 with basic metrics to about 0.68 with advanced ones. This shows that deeper statistical analysis is crucial when evaluating prospects. Specifically, with effective field goal percentage (eFG%) emerging as the most important variable, it indicates that shooting efficiency is a critical factor in predicting long-term success for centers.

College statistics capture a player's foundational skills and efficiency in a structured setting, while first-year NBA data reflect their adaptation to tougher competition and faster gameplay. By merging these sources, Model 3 identifies key performance indicators that differentiate long-term successful players from those who struggle, aiding teams in refining scouting strategies and resource allocation. If a player shows indicators of long-term success, teams might invest more in their training, conditioning, and role development. Conversely, if a player appears unlikely to reach the five-year mark, teams may adjust their expectations, explore trade opportunities, or allocate resources differently. This model achieves an AUC above 0.7 across all positions and both Lasso and Ridge methods, marking a significant improvement over earlier models. It also represents the earliest point at which a player's

long-term NBA success can be reliably predicted, using only college and first-year NBA performance data. Forwards are particularly straightforward to predict, whereas centers benefit from advanced metrics, Lasso's AUC for centers reaches 0.82, notably outperforming Ridge at 0.74. Moreover, advanced variables like VORP consistently rank in the top two by influence, with PER and Box Plus-Minus (including college Box Plus-Minus) proving crucial for evaluating centers and forwards.

The last model offers a comprehensive view of a player's development over two seasons, outperforming earlier models by capturing both long-term consistency and short-term fluctuations. For forwards, the basic variable models deliver excellent AUCs under both Lasso and Ridge, indicating that foundational stats such as points, steals, assists, and FG% are highly effective. With Lasso regularization, the use of advanced variables results in higher AUC values for forwards and guards, while centers do not exhibit the same level of improvement. This suggests that while advanced metrics like second season PER and first season VORP enhance predictive accuracy for most positions, centers may rely more on traditional measures in this case. Overall, using advanced and basic metrics is crucial for differentiating players with genuine long-term potential from those who only display short-term promise.

4. Conclusion

This study highlights the use of machine learning in predicting long-term NBA career success. By analysing physical attributes, college performance, and early NBA statistics, it is demonstrated that data-driven approaches can enhance player evaluation and scouting strategies. The models built in this research progressively improved in predictive accuracy as more relevant performance data was incorporated across all positions.

Findings from Model 1 suggest that physical attributes alone have little predictive value for guards and forwards, supporting the notion that raw athleticism is not a key factor in determining career longevity. However, for centers, possessing extreme physical traits proves more beneficial, with an AUC value of 0.67. Model performance improves notably when moving from physical-only attributes to performance metrics, especially when advanced analytics are included. The third model, based on college and first-year NBA data, proves to be the earliest and most reliable stage for predicting long-term NBA success across all player positions. The results show that Model 4, which incorporates data from a player's college career and first two NBA seasons, delivers the highest classification performance across all positions. Forwards benefit the most from this approach, achieving AUC values up to 0.86 with Lasso regularization. Advanced statistics such as VORP, PER, and Box Plus-Minus consistently emerge as key indicators of success, underscoring the value of deeper performance metrics in scouting and player development. Furthermore, the findings reveal important position-based differences. Centers show strong improvements when advanced statistics are included early, whereas guards and forwards display more modest yet consistent gains. These findings indicate that predictive models must account for position-specific context, as a one-size-fits-all approach may not yield accurate results across all player roles.

Future research could refine these models by incorporating additional contextual factors, such as injuries, playing style fit within a team, and external influences like coaching and team stability. Expanding datasets beyond the NBA Draft Combine to include international and undrafted players could also provide further insights. Moreover, future work will explore the implementation of more complex models such as neural networks, random forest, and XGBoost algorithms. Another important direction for future research is the inclusion of counterfactual example analysis, which can enhance model interpretability and provide deeper insights into decision-making processes. Overall, this study demonstrates how advanced analytics can support talent identification and decision-making in professional basketball, enabling teams to maximize their chances of securing long-term contributors.

References

Barnes, J. C. (2008). Relationship of selected pre–NBA career variables to NBA players' career longevity. *The Sport Journal,(April-02)*.

Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity. *Monographs on statistics and applied probability*, *143*(143), 8.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, No. 1). New York: springer.

Karipidis, A., Fotinakis, P., Taxildaris, K., & Fatouros, J. (2001). Factors characterizing a successful performance in basketball. *Journal of human movement studies*, *41*(5), 385-398.

Kekez, I., Ćukušić, M., & Jadrić, M. (2021). Data mining approach for business value analysis in basketball. *Zbornik Veleučilišta u Rijeci*, *9*(1), 227-248.

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. *Journal of quantitative analysis in sports*, *3*(3).

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. Journal of quantitative analysis in sports, 3(3).

Kuehn, J., & Rebessi, F. (2023). The Importance of Team Fit for NBA Rookies' Career Earnings. *Journal of Sports Economics*, *24*(3), 285-309.

Lee, D. J., & Page, G. L. (2021). Big Data in Sports: Predictive Models for Basketball Player's Performance.

Moxley, J. H., & Towne, T. J. (2015). Predicting success in the national basketball association: Stability & potential. *Psychology of Sport and Exercise*, *16*, 128-136.

Radovanović, S., Delibašić, B., & Suknović, M. (2021). Predicting dropout in online learning environments. *Computer Science and Information Systems/ComSIS*, *18*(3), 957-978.

Radovanović, S., Radojičić, M., & Savić, G. (2014). Two-phased DEA-MLA approach for predicting efficiency of NBA players. *Yugoslav Journal of Operations Research*, *24*(3), 347-358.

Robert, T. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273-82.

Robertson, S. (2014). Improving load/injury predictive modelling in sport: The role of data analytics. *Journal of Science and Medicine in Sport*, *18*, e25-e26.

Ruiz, F. J., & Perez-Cruz, F. (2015). A generative model for predicting outcomes in college basketball. *Journal of Quantitative Analysis in Sports*, *11*(1), 39-52.

Sarlis, V., & Tjortjis, C. (2020). Sports analytics—Evaluation of basketball players and team performance. *Information Systems*, *93*, 101562.

Sarlis, V., Chatziilias, V., Tjortjis, C., & Mandalidis, D. (2021). A data science approach analysing the impact of injuries on basketball player and team performance. Information Systems, 99, 101750.

Steinberg, L. (2015). Changing the game: the rise of sports analytics. *Forbes. Retrieved March*, *14*, 2017.

Stephanos, D. K., Husari, G., Bennett, B. T., & Stephanos, E. (2021, April). Machine learning predictive analytics for player movement prediction in NBA: applications, opportunities, and challenges. In *Proceedings of the 2021 ACM Southeast Conference* (pp. 2-8).

Sun, Z. (2015). Brief probe into the brand and marketing strategy of NBA. *Asian Social Science*, *11*(16), 183.

Szymanski, S. (2010). The political economy of sport. In *The comparative economics of sport* (pp. 79-86). London: Palgrave Macmillan UK.

Teeter, A., & Bergman, M. (2020). Applying the data: Predictive analytics in sport. *Access*: Interdisciplinary Journal of Student Research and Scholarship*, *4*(1), 4.

Thabtah, F., Zhang, L., & Abdelhamid, N. (2019). NBA game result prediction using feature analysis and machine learning. *Annals of Data Science*, *6*(1), 103-116.

Wilkinson, R. (2024). *THE NBA AS A MODEL OF INTERNATIONAL BUSINESS DEVELOPMENT: LESSONS AND STRATEGIES* (Doctoral dissertation, University of Oregon).

Zhou, Q., Hu, D., & Shi, K. (2023). The persistent effects of early career contracts: evidence from NBA drafts. *Applied Economics*, *55*(58), 6876-6891.

Decision Support System in project portfolio: the Shapley values and knapsack problem combined

Carlos Jose de Lima[1] and Maisa Mendonça Silva[1*]

[1] Universidade Federal de Pernambuco, Recife - PE, Brazil
[*] maisa.ufpe@yahoo.com.br

Abstract: This paper introduces a mathematical model designed to optimize project portfolio selection by integrating Shapley values and the knapsack problem. These techniques are implemented into a decision support system (DSS), offering alternatives to expert choices; the DSS is available for access here. In the context of increasingly complex decision-making and abundant information, the efficient allocation of limited resources to maximize returns is fundamental. Shapley values are utilized to assess the individual contribution of each project to the overall portfolio utility, while the knapsack problem optimizes project selection within resource constraints. Results indicate that the combined techniques lead to improved decision-making and maximized financial returns compared to expert opinions applied to a case study in construction company in Brazil. Furthermore, this approach provides a novel perspective on project utility elicitation, enhancing the accuracy and efficiency of portfolio selection.

Keywords: Shapley values; knapsack problem; utility function; portfolio management; operations research.

1. Introduction

The management of project portfolios and corporate investments poses a challenge for organizations, not only because they compete for shared resources (Dreyer et al., 2022). A major challenge is aligning business strategy with portfolio execution, balancing short-term and long-term objectives, and managing available resources to deliver feasible results (Project Management Institute [PMI], 2017). Therefore, understanding that viewing the composition of a project portfolio as a competition for available resources allows for a parallel with Game Theory in the pursuit of equilibrium in cooperative games (Gibbons, 1992).

The approach outlined here aims to combine existing techniques to establish a method that not only applies mathematical models through Operational Research techniques that were combined into a technology product, a Decision Support System (DSS) that was developed. It also systematically utilizes an organization's existing data to optimize portfolio management processes. The criteria for portfolio selection have undergone significant changes over time, and related studies have introduced multicriteria approaches, proposing different formulations based on mathematical models, quantitative techniques, heuristics, and various other methods applied in diverse approaches (Aouni et al., 2014).

Achieving balance in this prioritization and selection process is not a simple task, as the complexity of organizations combined with changes in the external environment further increases the difficulty. Additionally, legal aspects and shifts in the consumption patterns of a product or service are challenging to predict with great accuracy (Jugend & Figueiredo, 2017). The proposed mathematical model aims to enhance the efficiency and applicability of portfolio management through an approach focused on the optimization of individual projects. The combined application of the Shapley method with the knapsack problem seeks to represent a cooperative game called budget games, focused on budgetary constraints, aiming to highlight the computational complexity of Shapley values in order to find a better solution than some algorithms.

2. Combined mathematical model

The first part of the model involves using the Shapley value method (Shapley, 1953) from Game Theory, enabling the characteristics of the projects to be evaluated in the form of a coalition. A coalition is understood as the union formed by players in pursuit of a common objective; with this goal, the Shapley value theory establishes a fair distribution of gains for each player based on their individual contribution to the coalition. Thus, the use of the Shapley value function measures the contribution of each player, calculated according to Equation 1:

$$\varphi[i,v] = \sum_{S \in C} \frac{(s\ -\ 1)!\,(n\ -\ s)!}{n!}(v(S)\ -\ v(S\backslash i)) \quad (1)$$

Where:

$\boldsymbol{i}$ is the player for whom we want to calculate the Shapley value;
$\boldsymbol{s}$ is the number of members in the coalition;
$\boldsymbol{S}$ represents a coalition;
$\boldsymbol{C}$ is the set of all possible coalitions;
$\boldsymbol{S\backslash i}$ is the coalition excluding player $\boldsymbol{i}$;
$\boldsymbol{n}$ is the total number of players in the coalition;
$\boldsymbol{v(s)}$ the coalition value.

The Shapley values for the existing indicators are combined to obtain a single utility value for each project. The combined value (u_i) is calculated as the sum of the Shapley values, according to Equation 2:

$$u_i = \varphi[i,v_1] + \varphi[i,v_2] +, \dots, + \varphi[i,v_n] \quad (2)$$

Where:

$\boldsymbol{u_i}$ represents the utility of an item $\boldsymbol{i}$;

$\boldsymbol{\varphi[i,v_n]}$ is the contribution of player $\boldsymbol{i}$, according to the game $\boldsymbol{v}$;

After obtaining the information about the marginal contribution of each item according to the established coalitions and the calculated Shapley values, the formulation of the knapsack problem will be presented. This problem involves selecting existing items (n), and for each item (j), considering the knapsack's capacity (b), the value (p), and the weight (a) of the item, to select the items that will compose the knapsack without exceeding its capacity (Arenales, 2011), as shown in Equation 3:

$$max \sum_{j=1}^{n} p_j x_j \quad (3)$$

Subject to:

$$\sum_{j=1}^{n} a_j x_j \ \leq\ b$$

$$x_j\ \in \{0,1\} \text{ para } j = 1,2, \dots, n.$$

Where:

$\boldsymbol{n}$ represents the total number of available items.;

$\boldsymbol{p_j}$ corresponds to the value associated with the item $\boldsymbol{j}$;

$\boldsymbol{a_j}$ represents the weight of the item $\boldsymbol{j}$;

$\boldsymbol{b}$ is the maximum capacity of the knapsack;

$\boldsymbol{x_j}$ corresponds to a binary variable that indicates whether item $\boldsymbol{j}$ is selected $(\boldsymbol{x_j = 1})$ or not $(\boldsymbol{x_j = 0})$;

In an analogy with portfolio management, the values obtained using the Shapley method represent the intrinsic utility of each Project, considering its contribution to other projects in the portfolio. This approach allows for an evaluation of each project's relative values among the available projects. The results of Shapley values serve as inputs to a value associated with an item. The knapsack problem serves as an optimization mechanism, selecting the combination of projects that maximize the total utility of the portfolio while adhering to budget constraints. The costs associated with executing each project are treated as weights in the knapsack problem, while the available budget for the portfolio corresponds to the total capacity of the knapsack.

By combining these techniques, the model aims to optimize the allocation of limited resources among competing projects, aiming the highest possible financial return for the organization. This maximization of total utility aligns with principles from Game Theory, particularly the concept of Nash's bargaining game, which seeks an equilibrium among the involved agents to achieve the best collective outcome while considering individual utilities (Nash, 1950). Integrating Shapley values with the knapsack problem offers a more robust and efficient approach to portfolio selection, addressing both the individual value of projects and the constraints on resources.

2.1 How the DSS was implemented

This section presents the Decision Support System (DSS) implementation and how it works. It uses the inputted data to calculate the Shapley values for two financial indicators. The marginal contributions obtained from the Shapley values serve as way to maximize the utility of a project according to a portfolio year. With the previous stage done, the DSS uses the budget provided by the user as a limit of knapsack capacity. From there it is possible to applying knapsack problem maximizing the function of the budget of a year.

The idea behind of a DDS is a system that automates the combination of the two techniques, displaying on-screen suggestions about the items, recommending which ones should be selected and which should not, based on the input data. The DSIS was developed using the Python programming language (Deitel & Deitel, 2020), employing computational resources to streamline the decision-making process, as advocated by Arenales (2011). At the end of the process, the tool provides an ordered list, indicating which items should be selected and the recommended prioritization order, Figure 1 describes how the DSS combines the techniques in the model proposed.

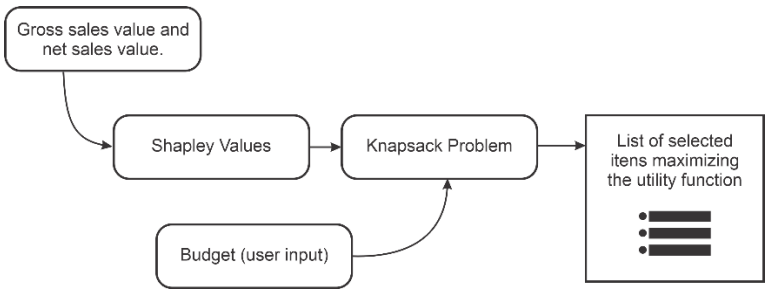


Figure 1: DSS and its model.

The proposed Decision Support System (DSS) combines Shapley values with the knapsack problem to facilitate project portfolio selection. The process starts by collecting input data on candidate projects, which includes information related to gross and net sales values. Shapley values are calculated to measure each project's marginal contribution to the overall portfolio value, reflecting the project's utility. These Shapley

values are then used as inputs in the knapsack problem and the budget for the portfolio is provided as another input. The knapsack problem determines the combination of projects that maximizes total portfolio utility while adhering to budget constraints. After performing the knapsack, a list of projects that should be selected to form an optimized portfolio is shown, assisting decision-makers in choosing projects that align best with the organization's objectives.

3. Case Study

In a construction company in Brazil a total of forty listed projects were subjected to a prior analysis determined they would be beneficial and feasible to develop based on legal and technical consideration, while promoting the expected return on the financial resources to be invested. With this stage completed, the projects are added to a portfolio list, from which some are selected to be part of the portfolio for a specific year, coinciding with the project's market launch. However, the market intelligence analysis is limited to gross and net sales values and project costs, as it pertains to distinct factors for each project, including macroeconomic and microeconomic aspects. The company did not share this data, considering it sensitive. When examining publications on the topic, it is noted that significant attention is given to local conditions, including business practices, cultural norms, and existing laws, which both shape and are shaped by real estate development. These elements can be regarded as structural factors but also as potential constraints (Boanada-Fuchs & Boanada-Fuchs, 2022).

The information obtained from the company reveals that there is an annual meeting to determine which projects will compose the company's portfolio for the following year. The meeting is the responsibility of an internal committee composed of directors and members of the new business team, and decisions are made based on financial criteria that ensure the company remains capable of pursuing new ventures, positions the brand positively in the market, and allows for sustainable growth. Having resources is crucial due to the high investment value of the projects, and considering the competition in the real estate market, there is a direct relationship between the company's financial condition and its ability to develop projects. However, projects differ in location, design, timeline, and pricing of the products offered (Fan et. al., 2022).

Regarding the decision-making process for portfolio composition, the company emphasizes that it is based on decisions driven by financial criteria (gross and net sales values) and expert opinions. No other tools are used beyond the analyses presented, and due to the large volume of information, extensive discussions take place until a consensus is reached among participants. At this point, it can be stated that proposing a mathematical solution to find the optimal strategy for decision-makers to accept or adjust the approach, in a context where there are stopping criteria and the capture of market contexts, is a critical factor for the success of the strategy (Abdelaziz & Mallek, 2018).

Annually, the Brazilian construction company typically considers between six and ten projects for its portfolio composition. Table 1 presents the number of projects selected in the years 2016, 2017, 2018, and 2019.

Table 10. Number of projects selected by year

Year	Number of projects
2016	7
2017	9
2018	11
2019	5

The company also informed that the available budgets for each year have a margin for upward adjustments, not exceeding 2.5% (five percent) of the total amount projected for that year. The projected budgets for each year are exhibited in millions of Reais on Table 2.

Table 2. Title of Example Table

Year	Avaliable Budget R$ Million
2016	155
2017	372
2018	465
2019	432

The Decision Support System (DSS) simulates the project selection process carried out by experts. To do this, the projects chosen by a committee of experts each year serve as input data for the DSS. This committee, equipped with knowledge and experience, selects the most relevant and appropriate projects from the set of projects available for each year, respecting the defined budget. The DSS, in turn, processes the same projects available to the committee and generates its selection, respecting the same budget value.. For example, in 2016, out of the ten available projects, the experts selected seven, and the tool was applied considering the same ten projects, performing its own selection.

Return on Investment (ROI) aims to identify the percentage of return on a specific amount invested, serving as a way to express the financial performance of a given investment (Brigham, 2010). In the case study, among other aspects, the ROI of the projects selected by the experts was compared with those chosen by the developed model. The ROI is determined by Equation 4.

$$ROI = \frac{Net\ Profit}{Total\ Investment} \times 100 \quad (4)$$

Where:

Net Profit: Total Revenue Generated from the Investment - Total Cost of the Investment
Total Investment: Total Amount Invested

Figure 2 illustrates the percentage increase in Return on Investment (ROI) when comparing the project selections made by the proposed model to those chosen by experts, for each year from 2016 to 2019. The graph demonstrates that the model consistently outperformed expert selections across all years. Specifically, the percentage increase in ROI was highest in 2016, at 4.46%, and lowest in 2018, at 3.63%. This data indicates that the model, implemented through the decision support system, resulted in a higher ROI for the company compared to the experts' selections.

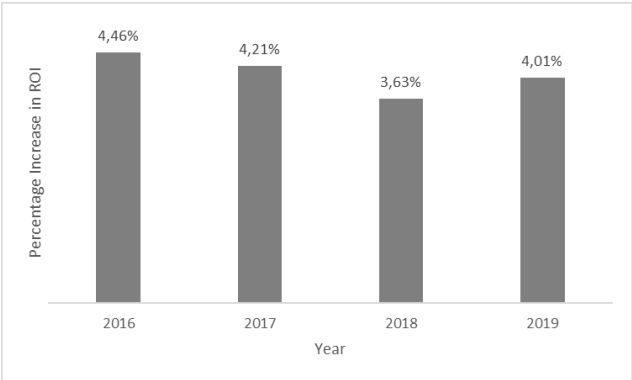


Figure 2: Percentage Increase in ROI achieved by the proposed model compared to expert selection, by year.

4. Conclusions

The decision support system combining Shapley values and Knapsack problem to maximize the utility function proved itself as an approach allowing a process improvement on portfolio selection due to the facts of best financial gains, and new possibilities composing a portfolio in contrast of expert opinions.

The proposed model, combining the two techniques, offers a novel perspective on the application of the knapsack problem combined with Shapley values aiming to elicit project utility in a way that contributes to preferences in the decision-making process. This provides a more dynamic view with coalitions in an interactive and cooperative process, promoting a distribution to maximize available resources, as seen in the return on investment in the applied case studied. It also allows for the selection of some projects that were not chosen by decision-makers, opening avenues for new portfolio compositions and significantly contributing to decision analysis.

This paper presents a mathematical model for optimizing project portfolio selection, combining Shapley values and the knapsack problem. The research explores the application of a project utility-focused objective function, diverging from traditional economic metric maximization. The proposed approach aims to offer a new perspective on project utility assessment and portfolio selection, particularly in contexts with limited resources. The investigation contributes to the field by providing an alternative methodology for portfolio composition, which can be applied in case studies such as the one presented, within a construction company in Brazil.

This paper shows impacts on both economic and social perspectives submitting the utility function in combined Operations Research techniques combined. From an economic standpoint, by aiming to optimize existing resources, that is, investing in a way that maximizes returns and minimizes risks, there is an expectation that organizations' strategic objectives will be well achieved. Thus, organizations would have a strategic tool to better conduct their actions and projects. Regarding social impacts, since resources are well managed, it is expected that there will be maintenance of jobs, the company's financial health, and the value chain.

Acknowledgements

We would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the National Council for Scientific and Technological Development (CNPq). for their financial support.

References

Abdelaziz, F. B., & Mallek, R. S. (2008). Multi-criteria optimal stopping methods applied to the portfolio optimisation problem. Annals of Operations Research, 267, 29–46.

Aouni, B., Colapinto, C., & La Torre, D. (2014). Financial portfolio management through the goal programming model: Current state-of-the-art. European Journal of Operational Research, 234(2), 536–545.

Arenales, M. (2011). Pesquisa operacional. Elsevier.

Brigham, E. F., & Ehrhardt, M. C. (2010). Administração financeira: Teoria e prática. Cengage Learning.

Boanada-Fuchs, A., & Boanada Fuchs, V. (2022). The role of real estate developers in urban development. Geoforum, 134, 173–177.

Deitel, P., & Deitel, H. (2020). Intro to Python for computer science and data science. Pearson.

Dreyer, S., Egger, A., Püschel, L., & Röglinger, M. (2022). Prioritising smart factory investments - A project portfolio selection approach. International Journal of Production Research, 60(3), 999–1015.

Fan, Y., Leung, C. K. Y., & Yang, Z. (2022). Financial conditions, local competition, and local market leaders: The case of real estate developers. Pacific Economic Review, 27, 131–193.

Gibbons, R. (1992). Game theory for applied economists. Princeton University Press.

Jugend, D., & Figueiredo, J. (2017). Integrating environmental sustainability and project portfolio management: Case study in an energy firm. Gestão & Produção, 24, 526–537.

Nash, J. F., Jr. (1950). The bargaining problem. Econometrica, 18, 155–162.

Project Management Institute. (2017). The standard for portfolio management. Project Management Institute.

Shapley, L. S. (1953). A value for n-person games. The Rand Corporation, 307–317.

Dynamic Model Selection Framework for Weekly Capacity Forecasting in E-Commerce Logistics

Ahmet Çay[1,2,*], Barış Bayram[2] , Ayşe Dilara Türkmen[2] and Alaeddin Türkmen[2]

[1] Technical University of Munich, Munich, Germany
[2] Hepsijet, Istanbul, Turkey
[*] ahmet.cay@hepsijet.com

Abstract: Accurate parcel volume forecasting is critical for optimizing cross-dock (**xdock**) operations in e-commerce logistics. It ensures efficient resource allocation and minimizes delays. Traditional forecasting models struggle with dynamic demand fluctuations, necessitating an adaptive approach to model selection. This study introduces a dynamic weekly model selection framework that autonomously selects the most suitable forecasting model for each xdock based on historical performance. Our methodology integrates a diverse range of machine learning and statistical models, including XGBoost, CatBoost, LightGBM, Random Forest, Linear Regression, Multi-Layer Perceptron, Bayesian Regression, K-Nearest Neighbors, ARIMA, SARIMAX, and Prophet. These models were selected due to their proven effectiveness in time series forecasting and structured data analysis in logistics forecasting applications.

To enhance forecasting accuracy, we propose two model selection strategies: **(1)** a heuristic rolling-based approach that evaluates past forecasting errors and selects models based on recent performance trends, and **(2)** a meta-learning framework that learns from previous selection decisions to refine model selection in a dynamic manner. Experimental results on real-world parcel shipment data indicate that our framework significantly outperforms traditional static models in terms of forecasting accuracy, adaptability, and robustness to demand shifts.

By combining machine learning-based meta-learning with heuristic techniques, this research advances Decision Support Systems (**DSS**) for logistics, offering a scalable, data-driven methodology for improving operational efficiency in cross-dock facilities. Future directions include incorporating external factors such as seasonality and economic trends to further enhance model selection.

Keywords: Logistics forecasting; cross-dock optimization; model selection; machine learning; decision support systems

1. Introduction

Efficient parcel forecasting is critical for logistics companies operating cross-dock (**xdock**) facilities, as it directly impacts resource allocation, workforce planning, and overall operational efficiency. Traditional forecasting models often fail to adapt to dynamic demand fluctuations, making them unsuitable for real-world logistics applications. To address this issue, we propose an adaptive weekly model selection framework that dynamically identifies the best forecasting model for each xdock location based on historical performance.

Our approach leverages multiple machine learning and statistical models, including XGBoost, CatBoost, LightGBM, Random Forest, Linear Regression, Multi-Layer Perceptron, Bayesian Regression, K-Nearest Neighbors, ARIMA, SARIMAX, and Prophet, to predict daily parcel volumes. Instead of relying on a single static model, we introduce a meta-learning framework that evaluates historical prediction errors and selects the most suitable model weekly. Additionally, we incorporate a heuristic rolling-based method, which utilizes past forecast performance to guide model selection.

This research contributes to the field of Decision Support Systems (DSS) for logistics by providing a scalable, data-driven methodology for improving parcel volume forecasting at cross-dock facilities. Our results demonstrate that dynamically selecting the best-performing model enhances forecasting accuracy compared to a fixed-model approach. By integrating meta-learning and heuristic strategies, we enable logistics companies to reduce forecasting errors, optimize decision-making, and improve operational efficiency.

The remainder of this paper is structured as follows: Section 2 discusses related work in forecasting and model selection strategies. Section 3 details the proposed methodology, including the meta-learning approach and heuristic rolling-based selection. Section 4 presents experimental results evaluating the effectiveness of our framework. Finally, Section 5 concludes with insights and potential future directions for improving adaptive forecasting systems in logistics.

2. Related Work

Model selection is a critical aspect of machine learning, as the choice of an appropriate model significantly impacts the performance of a given task. Over the years, various approaches have been proposed to address the challenges associated with model selection, ranging from meta-learning to dynamic selection strategies. This section reviews key contributions in the field, highlighting the evolution of methodologies and their applications across different domains.

A comprehensive survey by Khan et al. (2020) provides an in-depth analysis of meta-learning for classifier selection, addressing the algorithm selection problem (ASP) in classification tasks. The survey categorizes existing methods into three dimensions: meta-features, meta-learners, and meta-targets, offering a structured framework for understanding the classifier selection process. The authors also conduct an extensive empirical evaluation of prominent methods, comparing their performance across 17 classification algorithms and 84 benchmark datasets. This work not only summarizes the state-of-the-art but also identifies gaps and future research directions, making it a foundational reference for model selection studies.

Meta-learning has also been explored in the context of recommender systems, where Luo et al. (2020) propose MetaSelector, a framework for user-level adaptive model selection. By training a model selector via meta-learning, their approach achieves significant improvements over single-model baselines, demonstrating the potential of meta-learning in personalized recommendation tasks. Jasemi and Kimiagari (2012) investigate model selection criteria for technical analysis, specifically focusing on moving average (MA) models. Their study reveals that certain factors, such as the choice of MA calculation technique, have negligible effects on performance, while others, like the combination of MA lengths, are critical. This work provides practical insights for constructing effective technical analysis models.

In the context of meta-model selection, Most and Will (2008) propose another automatic approach for selecting optimal meta-models based on the coefficient of prognosis, which evaluates model performance using an additional test dataset. Their method efficiently reduces the variable space and is validated on both weakly and highly nonlinear problems. This work demonstrates the importance of objective assessment metrics in model selection and highlights the potential for extending such approaches to high-dimensional problems. For specialized tasks such as white blood cell (WBC) image classification, Rivas-Posada and Chacon-Murguia (2023) introduce a meta-learning-based methodology to automatically select base models. Their approach leverages meta-knowledge acquired from prior models and employs ensemble learning to achieve state-of-the-art performance on multiple datasets. This study underscores the applicability of meta-learning in domains with imbalanced and limited data, offering a promising direction for medical image classification tasks.

Dynamic model selection strategies have also gained traction, particularly in time-sensitive applications like retail sales forecasting. Yu et al. (2022) propose a dynamic approach that combines demand pattern classification with out-of-sample performance to select forecasting models. Their method, validated on large-scale retail datasets, demonstrates the effectiveness of adapting model selection strategies to specific demand patterns. Similarly, Silva et al. (2021) address time series forecasting by introducing a dynamic predictor selection method based on recent temporal windows. Their approach, Dynamic Selection based on the Nearest Windows (DSNAW), outperforms existing methods by focusing on the most relevant temporal patterns, emphasizing the importance of adaptability in model selection. These

studies illustrate the diverse methodologies and applications of model selection, highlighting the importance of adaptability, objective evaluation, and domain-specific considerations. Building on these foundations, this paper aims to contribute to the field by proposing a novel approach to model selection that addresses existing limitations and leverages recent advancements in meta-learning and dynamic selection strategies (Sun, 2014).

3. Methodology

Our proposed approach, shown in Figure 1, consists of two core components: **(i) a heuristic rolling-based model selection method** and **(ii) a meta-learning framework for selecting the best forecasting model weekly**. These methods aim to improve the adaptability of parcel volume forecasting by dynamically choosing the most suitable model based on historical performance.

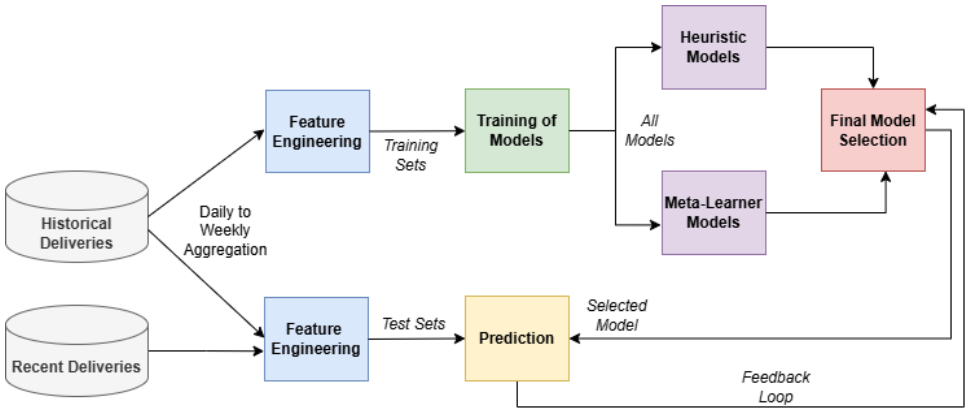


Figure 1. Model Selection Flowchart

The heuristic method selects models based on past prediction performance. Each week, for a given xdock, historical forecasting errors are analyzed over a fixed rolling window. The model that consistently achieves the lowest absolute percentage error (MAPE) within the rolling window is selected for the next forecasting cycle. This approach ensures that models are continually updated based on recent performance trends, allowing for quick adaptation to shifting demand patterns.

We implemented two different strategies for the heuristic selection:

1. **Weighted Loss-Based Selection**: We assign weights based on the errors of the most recent top-performing models and sum them for each unique model. The summed errors are then normalized by the number of models in that window for each model. This creates a loss-like function, and the model with the smallest loss is selected for forecasting.
2. **Model Frequency-Based Selection**: We count the occurrences of each unique model within the rolling window. If multiple models have equal counts, we apply the first approach (weighted loss-based selection) to determine the best model.

This combination of selection methods ensures that the approach is both adaptive and robust to variations in demand trends, optimizing the forecasting accuracy across different xdock locations.

To evaluate the performance of our model selection framework, we focus on forecasting accuracy and adaptability under varying conditions. The primary metric used is Mean Absolute Percentage Error (MAPE), which quantifies the average deviation between predicted and actual parcel volumes. In addition, we assess how frequently the selected model ranks among the top-performing models each week by measuring top-n accuracy, which reflects the framework's ability to consistently choose competitive models. Finally, we analyze how the heuristic-based selection method responds to different rolling window sizes, examining whether broader historical context improves adaptability and decision quality across fluctuating demand patterns.

The weekly model selection process is implemented as an automated pipeline that runs before each forecasting cycle. The system processes historical data, applies the heuristic and meta-learning methods, and selects the most appropriate forecasting model for each xdock. This adaptive pipeline ensures that the best model is chosen based on recent data trends, reducing forecasting errors and improving operational efficiency.

4. Experiments

To evaluate the effectiveness of our proposed weekly model selection framework, we conducted extensive experiments on real-world parcel shipment data across multiple xdock facilities. Our evaluation focuses on assessing the performance of different models, comparing the heuristic rolling-based method and the meta-learning approach.

4.1 Experimental Setup

Our initial dataset consisted of 38,143 rows, covering daily parcel volume predictions for 244 unique cross-dock (xdock) facilities. Each row included actual parcel volumes along with forecast results from multiple statistical and machine learning models. To improve data quality and consistency, we applied a filtering process based on data availability and continuity over a defined time period. Specifically, we retained only those xdocks with complete data from the beginning of January 2024 through mid-February 2025. After this refinement, the dataset was reduced to 5,553 weekly records representing 219 xdocks. During preprocessing, daily data was aggregated to weekly totals for both actual and predicted parcel volumes across all models. From these weekly aggregates, we calculated relative forecasting errors to evaluate model performance and identify the best-performing model for each xdock-week combination. To enrich the dataset, we also merged operational metadata from two additional sources: the Regional Management Center (BM), providing regional context, and the Transfer Center (TM), offering insight into parcel movement dynamics. These enriched features were subsequently used in the meta-learning component of our model selection framework.

It is important to note that all individual forecasting models were initially trained using time-series features such as lag variables, rolling statistics, and trend components. However, for the model selection dataset used in this study, we excluded these features, as our goal was not to re-train base models, but rather to evaluate and compare their precomputed weekly predictions. This ensures the meta-learning and heuristic selection strategies operate independently of the original model inputs while relying on historical forecasting performance for decision-making.

4.2 Evaluation Metrics

Forecasting accuracy in this study is primarily assessed using Mean Absolute Percentage Error (MAPE), which captures the average percentage difference between predicted and actual parcel volumes, providing a clear indication of overall predictive precision. In addition to MAPE, we evaluate the effectiveness of the model selection process through top-n accuracy, which reflects how often the chosen model ranks among the top-performing alternatives for a given week. This helps determine the consistency and competitiveness of the selection strategy. To further explore the adaptability of our heuristic-based method, we analyze its performance across different rolling window sizes, examining whether incorporating more historical data enhances its responsiveness to fluctuating demand conditions.

Table 1. Model Selection Results

Model	Window Size	Top-1 Acc. [%]	Top-2 Acc. [%]	Top-3 Acc. [%]	Predicted MAPE [%]	Actual MAPE [%]
Random Forest	-	22.48	37.41	50.71	9.42	3.06
XGBoost	-	20.50	36.15	53.05	9.34	3.06
LightGBM	-	19.78	36.51	49.10	9.75	3.06
Heuristic /w Weight	2/ 3/ 4	9.06/ 8.90/ 9.59	19.17/ 19.00/ 19.51	29.18/ 28.88/ 29.65	15.58/ 15.74/ 15.82	3.06
Heuristic /wo Weight	2/ 3/ 4	9.35/ 9.86/ 10.25	18.26/ 19.52/ 20.74	27.46/ 29.02/ 30.25	16.57/ 15.97/ 18.56	3.06

4.3 Results and Discussion

Our experimental results, summarized in Table 1, show that traditional machine learning models such as Random Forest, XGBoost, and LightGBM often rank among the top-performing models in terms of weekly forecasting accuracy. These models typically achieve lower Mean Absolute Percentage Error (MAPE) values compared to the heuristic-based selection strategies. However, while their predictive performance is strong in stable demand environments, they exhibit limitations in adapting to rapidly changing parcel volumes, which are common in e-commerce logistics.

In contrast, the heuristic rolling-based selection method demonstrates greater adaptability to dynamic fluctuations, despite showing lower average accuracy in static conditions. This is particularly evident in Figure 2, which presents the total predicted parcel volume across the entire logistics network over time. Although traditional models closely align with actual values during stable periods, the heuristic method better tracks rapid changes in demand. This suggests that heuristic strategies, especially those using weighted historical errors, offer a more resilient solution in volatile forecasting contexts.

The weighted heuristic variant, which prioritizes recent model performance, yields improved stability and responsiveness, particularly when larger rolling windows are used. This highlights the importance of incorporating sufficient historical context to inform model selection. On the other hand, the non-weighted heuristic approach shows similar trends but lacks robustness under more volatile conditions.

Interestingly, across all models—including heuristic-based methods—the predicted MAPE values remain higher than the actual observed MAPE. This discrepancy indicates that the selection framework tends to underestimate the potential accuracy of the best-performing models, possibly due to the limited scope of features or rigid selection logic. Introducing dynamic adjustments to the window size or integrating additional meta-features may further improve forecasting outcomes.

Lastly, our findings indicate that model performance varies significantly across different xdock locations and time periods. This variation underscores the need for a model selection mechanism that is both context-aware and scalable. In real-world logistics environments, the computational efficiency and automation of our framework make it suitable for large-scale deployment, supporting hundreds of facilities with minimal manual intervention.

5. Conclusions

Our experimental findings highlight that model performance varies across different xdock locations and time periods, underscoring the necessity of an adaptive selection mechanism. The heuristic rolling-based selection approach, particularly when weighted, exhibits stable performance in fluctuating

demand scenarios, while meta-learning effectively determines the best-performing model on a weekly basis, reducing forecasting errors and improving efficiency.

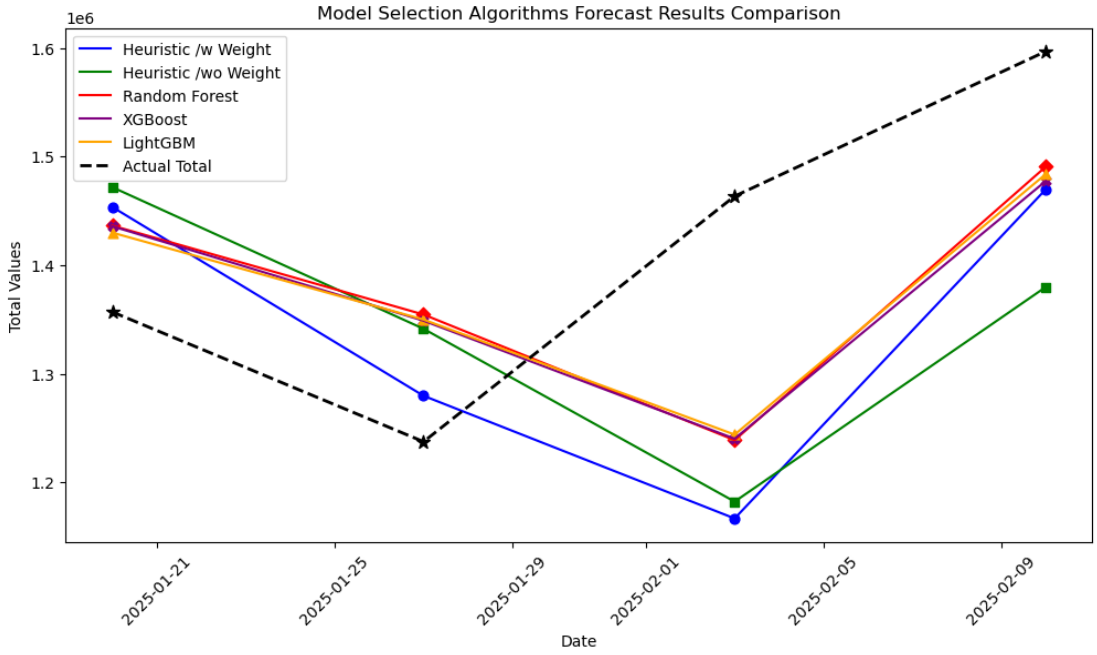


Figure 2. Forecasting Results of Model Selection Algorithms

Future research directions include developing hybrid strategies that combine machine learning with heuristic model selection to leverage the strengths of both approaches. Additionally, incorporating real-time external factors such as seasonality, economic trends, and operational constraints could further enhance forecasting performance. Finally, improving the interpretability of meta-learning decisions through explainability techniques could foster better adoption of automated model selection in logistics operations.

Our findings contribute to the Decision Support Systems (DSS) for logistics, offering a scalable, data-driven methodology to optimize cross-dock operations and improve forecasting accuracy.

References

Jasemi, M., & Kimiagari, A. M. (2012). An investigation of model selection criteria for technical analysis of moving average. *Journal of Industrial Engineering International*, *8*, 1-9.

Luo, M., Chen, F., Cheng, P., Dong, Z., He, X., Feng, J., & Li, Z. (2020, April). Metaselector: Meta-learning for recommendation with user-level adaptive model selection. In Proceedings of The Web Conference 2020 (pp. 2507-2513).

Khan, I., Zhang, X., Rehman, M., & Ali, R. (2020). A literature survey and empirical study of meta-learning for classifier selection. IEEE Access, 8, 10262-10281.

Most, T., & Will, J. (2008). Metamodel of Optimal Prognosis-an automatic approach for variable reduction and optimal metamodel selection. Proc. Weimarer Optimierungs-und Stochastiktage, 5, 20-21.

Rivas-Posada, E., & Chacon-Murguia, M. I. (2023). Automatic base-model selection for white blood cell image classification using meta-learning. Computers in Biology and Medicine, 163, 107200.

Silva, E. G., Neto, P. S. D. M., & Cavalcanti, G. D. (2021). A dynamic predictor selection method based on recent temporal windows for time series forecasting. IEEE Access, 9, 108466-108479.

Sun, Q. (2014). Meta-learning and the full model selection problem (Doctoral dissertation, University of Waikato).

Yu, M., Tian, X., & Tao, Y. (2022). Dynamic model selection based on demand pattern classification in retail sales forecasting. Mathematics, 10(17), 3179.

MCDM Model for Portfolio Selection in the Real Estate Market

Guilherme Freire de Mariz[1*], João Batista Sarmento dos Santos Neto[1] , Kassia Tonheiro Rodrigues[1] and Carolina Lino Martins[1]

[1] Universidade Federal de Mato Grosso do Sul - UFMS, Campo Grande - MS, Brazil
[*] guilherme.freire@ufms.br

Abstract: Real estate investment decisions require advanced approaches that can maintain a balance between several variables to optimize profits and reduce risks. Therefore, a multicriteria decision-making model for selecting project portfolios in the real estate market is proposed in this work. To achieve this, the study used a benefit-to-cost ratio-based approach for portfolio selection based on the FITradeoff method, which integrates important investment factors such as buildable area, land availability, internal rate of return, and housing availability while taking budgetary constraints into account. A numerical application employing actual data was carried out to choose plots in upscale residential condominiums for investment in Brazil to test the suggested approach. The outcomes showed how the model defined the best compromise solution for an investment portfolio and provided a systematic and organized approach to decision-making. By giving investors a clear framework that optimizes profits while effectively allocating scarce resources, this study confirms the usefulness of using multicriteria approaches when making real estate investment decisions.

Keywords: Portfolio selection; FITradeoff; benefit-to-cost; multicriteria analysis; plots; real estate market

1. Introduction

The selection of plots, that is, a small area to build a house, for real estate investment is a complex process that involves multiple criteria and conflicting objectives. Multi-criteria analysis methods have proven to be essential tools for handling this complexity, allowing for the structured integration of technical and economic factors. According to de Almeida et al. (2015), these methodologies combine qualitative and quantitative aspects, providing more well-founded and consistent decisions.

Land selection in the real estate sector is a classic multicriteria analysis problem, involving factors such as area, utilization coefficient, cost, and financial return. These factors are evaluated simultaneously, often under conditions of uncertainty and market competition. Multicriteria analysis has been used to prioritize technical and market criteria in real estate projects, contributing to more structured and well-founded decisions (Costa & de Almeida, 2021).

In the real estate sector, criteria such as buildable area, average selling price per square meter, cash exposure, and supply and demand in the region directly influence financial returns and project feasibility. These variables often present opposing objectives, such as maximizing returns while minimizing costs and risks, requiring tools that effectively balance these demands (Gomes, 2009).

The FITradeoff method for portfolio selection stands out by integrating these criteria flexibly, using partial information, and reducing the cognitive load in decision-making (de Almeida et al., 2016). Its application in the real estate market enables the identification of alternatives that optimize financial returns while respecting budgetary and risk constraints (Frej et al., 2021).

Therefore, this study aims to select a portfolio of plots for the construction of houses intended for sale, considering a budget of 30 million reais, covering both land acquisition and construction costs. To solve the problem, the research used a benefit-to-cost ratio-based approach for portfolio selection under multiple criteria with incomplete preference information, based on the Flexible and Interactive Tradeoff (FITradeoff) which starts from the principles of the traditional tradeoff procedure but works as a flexible and interactive way of eliciting the criteria weights, seeking only partial information about the decision maker's preferences (Frej et al., 2021). The analysis uses the FITradeoff method to assess criteria such as buildable area, land availability, house availability, internal rate of return, and investment margin. This approach seeks to identify the most advantageous alternatives to maximize financial returns and minimize risks, ensuring strategic and effective decision-making in the real estate market.

2. Benefit-to-cost ratio-based approach for portfolio selection with incomplete preference information

There are several approaches to solving portfolio selection problems. The traditional one is the knapsack problem, which is solved combinatorially by a mathematical programming model that looks for the optimal project combination while taking organizational constraints into account (Clemen & Smith, 2009).

The benefit-to-cost ratio of the projects is a different approach to solving portfolio selection problems. This heuristic method involves ranking projects according to their benefit-to-cost ratio, which is determined by dividing a project's particular value that indicates its benefit by a cost (Frej et al., 2021). Projects that fit within the allocated budget are chosen to be included in the portfolio once the ranking has been established. Without requiring an extensive computational evaluation of the combinatorial optimization model, the benefit-to-cost ratio technique has the advantage of solving the portfolio selection problem more realistically (Frej et al., 2021).

In this context, there is the FITradeoff method, which uses a flexible and interactive process to obtain the DM preferences, with a benefit-to-cost ratio approach (Frej et al., 2021). The benefit of each project *b*(*pi*) is the weighted sum of the project's performance about the criteria *vj*(*xij*) and the scaling constants of the criteria *kj*, as shown in Equation 1.

$$b\left(p_i\right) = \sum_{j=1}^{m} k_j v_j\left(x_{ij}\right) \quad (1)$$

The first step consists of establishing an ordering of criteria according to the decision maker's preferences. The second step is to elicit the decision maker's preferences regarding the consequences of the criteria. As this is a portfolio problem that takes into account the benefit-to-cost ratio analysis of each project, it is necessary to divide the aggregate performance of the project by its cost to find the BCR ratio per investment of each alternative, to maximize the benefit divided by the cost. Based on the values found by the ratio between benefit and cost of each project, FITradeOff will organize them in a ranking, from the highest value to the lowest value. Using the ranking, the method will form the subgroup until the sum of the project costs reaches the budget determined by the company, to maximize the benefit (Frej et al. 2021). For detailed information about the method, see Frej et al. (2021).

3. Decision Model Description

To test the proposed model, a numerical application employing real data was carried out in a real estate company located in a city in the Midwest of Brazil. The company needs to choose plots in upscale residential condominiums for investment. Figure 1 shows the decision model that was structured in stages to select a portfolio of plots in high-end condominiums, considering economic, technical, and

market criteria. Initially, the main condominiums were identified, and the available plots were surveyed through real estate listings. Duplicate listings were eliminated to ensure data consistency.

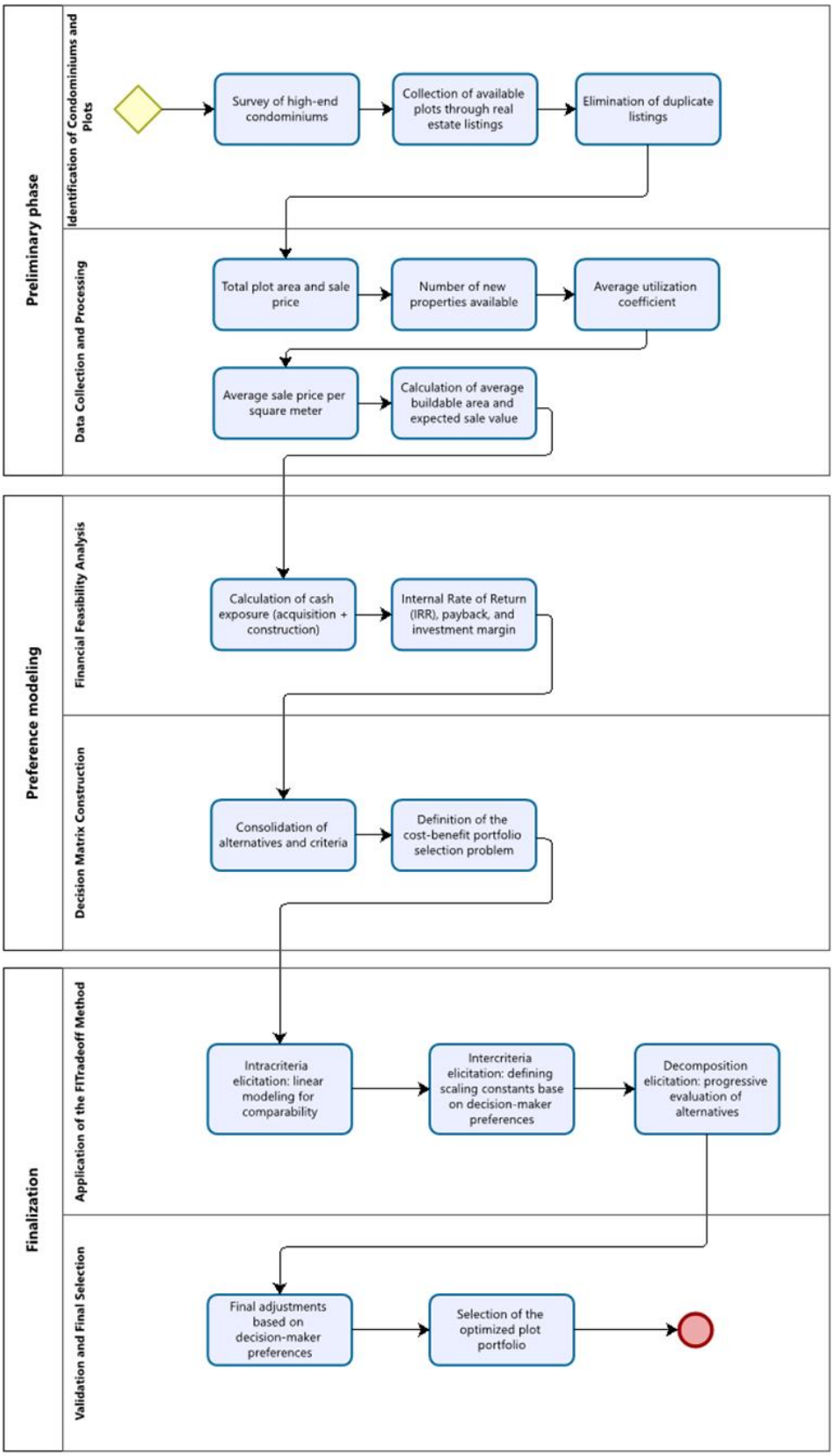


Figure 1: Decision Model for Land Portfolio Selection

For the application of the FITradeoff, the decision-maker (DM) was introduced to the family of criteria, the evaluation scale, and the set of alternatives to provide an overview of the key elements involved in the selection process. To validate the family of criteria, the DM was asked whether any criteria should be added or removed. Since the DM agreed with the predefined set, no modifications were made. Figure 2 shows the criteria of the model.

C(i)	Criteria	Function	Definition
C1	Construction Area (Sq. m)	Maximization	Square meters of each plot
C2	Available Plots (unit)	Minimization	Number of plots available for sale in the condominium
C3	Available Houses (unit)	Minimization	Number of houses available for sale in the condominium
C4	Internal Rate of Return (IRR)	Maximization	Expected return of this investment
C5	Cash Exposure Restriction	Minimization	Amount available to start an investment

Figure 2: Criteria of the model

For each plot, information such as total area, sale price, number of new properties available in the condominium, average utilization coefficient (ratio between built area and total area), and average sale price per square meter was collected. With this data, the average construction area and expected sale value were calculated, essential for assessing the financial potential of each plot.

The financial feasibility study included an analysis of indicators such as cash exposure (total acquisition and construction cost), internal rate of return (IRR), payback period, and investment margin. Cash exposure was used as the main constraint: the total cost of the selected portfolio could not exceed 30 million Brazilian Reais.

The collected data was organized into a decision matrix, consolidating alternatives and criteria. The company's initial portfolio consisted of 47 projects to choose from in upscale residential condominiums for investment in different locations of a city in the Midwest of Brazil, denoted A = {Alt.1, …, Alt.47}, as outlined in Fig. 3. The locations were defined according to their potential to bring greater benefits through previous studies carried out by the company. Also, the perspectives of some projects were obtained from data provided by the company and from interviews with experts.

In the Decision Support System (DSS), the benefit-to-cost ratio portfolio selection problem was defined, ensuring that the analysis considered maximizing returns concerning total costs. The criteria were modeled as linear functions through intra-criteria elicitation, ensuring comparability and consistency of scales among alternatives.

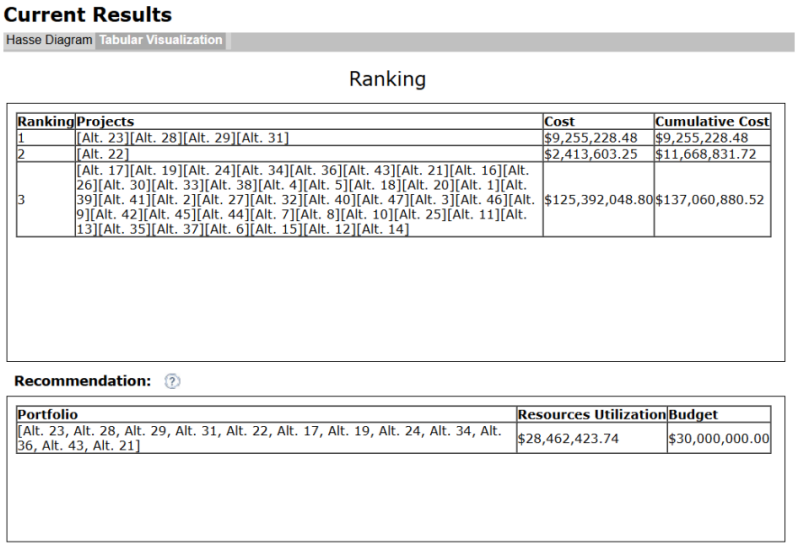
In the inter-criteria elicitation stage, scale constants were defined to reflect the decision-maker's preferences regarding the importance of the criteria. Decomposition-based elicitation allowed for a gradual analysis of alternatives, reducing the complexity of the evaluation.

The final solution was validated to ensure alignment with the decision-maker's objectives. This structured process integrated the defined criteria and adhered to financial constraints, resulting in an optimized selection of the plot portfolio.

3.1. Discussion of Results

The results revealed differences between the scenarios before and after the elicitation by decomposition, highlighting how adjustments in the model influenced the selection of the land portfolio.

In the initial scenario, the portfolio included the alternatives [Alt. 23, Alt. 28, Alt. 29, Alt. 31, Alt. 22, Alt. 17, Alt. 19, Alt. 24, Alt. 34, Alt. 36, Alt. 43, Alt. 21], with a total cost of R$28,462,423.74, respecting the limit of R$30,000,000.00. The solution prioritized plots with a high internal rate of return (IRR), ranging from 12.06% to 33.77%, and with significant buildable areas, from 269.76 m² to 434.93 m². It also minimized the availability of plots and houses, reducing competition within the condominiums. The selected portfolio used R$28,462,423.74, remaining R$1,537,576.26 of the budget unallocated.



Current Results

Hasse Diagram | Tabular Visualization

Ranking

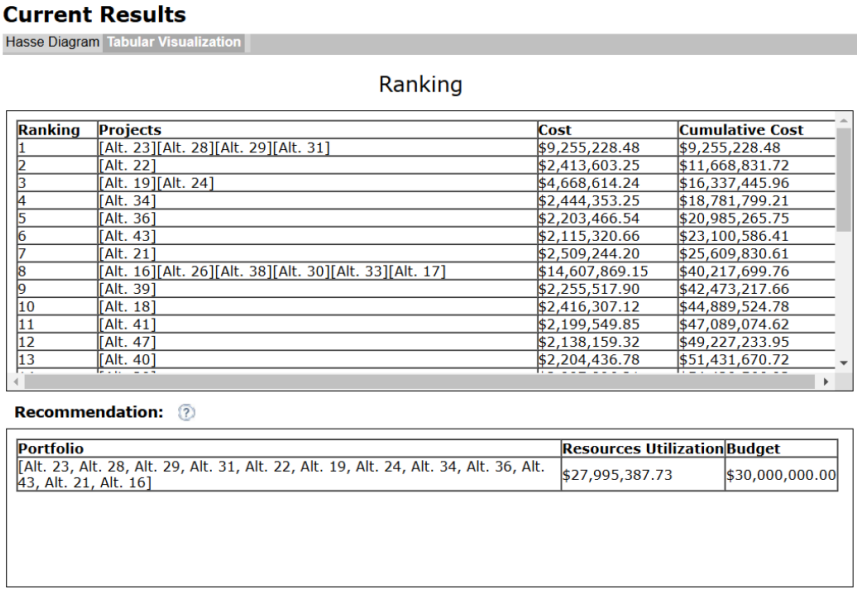
Ranking	Projects	Cost	Cumulative Cost
1	[Alt. 23][Alt. 28][Alt. 29][Alt. 31]	$9,255,228.48	$9,255,228.48
2	[Alt. 22]	$2,413,603.25	$11,668,831.72
3	[Alt. 17][Alt. 19][Alt. 24][Alt. 34][Alt. 36][Alt. 43][Alt. 21][Alt. 16][Alt. 26][Alt. 30][Alt. 33][Alt. 38][Alt. 4][Alt. 5][Alt. 18][Alt. 20][Alt. 1][Alt. 39][Alt. 41][Alt. 2][Alt. 27][Alt. 32][Alt. 40][Alt. 47][Alt. 3][Alt. 46][Alt. 9][Alt. 42][Alt. 45][Alt. 44][Alt. 7][Alt. 8][Alt. 10][Alt. 25][Alt. 11][Alt. 13][Alt. 35][Alt. 37][Alt. 6][Alt. 15][Alt. 12][Alt. 14]	$125,392,048.80	$137,060,880.52

Recommendation:

Portfolio	Resources Utilization	Budget
[Alt. 23, Alt. 28, Alt. 29, Alt. 31, Alt. 22, Alt. 17, Alt. 19, Alt. 24, Alt. 34, Alt. 36, Alt. 43, Alt. 21]	$28,462,423.74	$30,000,000.00

Figure 3: Initial Results

After the decomposition, the portfolio was adjusted to include the alternatives [Alt. 23, Alt. 28, Alt. 29, Alt. 31, Alt. 22, Alt. 19, Alt. 24, Alt. 34, Alt. 36, Alt. 43, Alt. 21, Alt. 16], with a total cost of R$27,995,387.73, increasing the unused budget margin to R$2,004,612.27. The inclusion of Alt. 16 diversified the portfolio while maintaining a high financial return and construction utilization.



Current Results

Hasse Diagram | Tabular Visualization

Ranking

Ranking	Projects	Cost	Cumulative Cost
1	[Alt. 23][Alt. 28][Alt. 29][Alt. 31]	$9,255,228.48	$9,255,228.48
2	[Alt. 22]	$2,413,603.25	$11,668,831.72
3	[Alt. 19][Alt. 24]	$4,668,614.24	$16,337,445.96
4	[Alt. 34]	$2,444,353.25	$18,781,799.21
5	[Alt. 36]	$2,203,466.54	$20,985,265.75
6	[Alt. 43]	$2,115,320.66	$23,100,586.41
7	[Alt. 21]	$2,509,244.20	$25,609,830.61
8	[Alt. 16][Alt. 26][Alt. 38][Alt. 30][Alt. 33][Alt. 17]	$14,607,869.15	$40,217,699.76
9	[Alt. 39]	$2,255,517.90	$42,473,217.66
10	[Alt. 18]	$2,416,307.12	$44,889,524.78
11	[Alt. 41]	$2,199,549.85	$47,089,074.62
12	[Alt. 47]	$2,138,159.32	$49,227,233.95
13	[Alt. 40]	$2,204,436.78	$51,431,670.72

Recommendation:

Portfolio	Resources Utilization	Budget
[Alt. 23, Alt. 28, Alt. 29, Alt. 31, Alt. 22, Alt. 19, Alt. 24, Alt. 34, Alt. 36, Alt. 43, Alt. 21, Alt. 16]	$27,995,387.73	$30,000,000.00

Figure 4: Results after elicitation by decomposition

In the first scenario, there was better budget utilization, with a smaller unused margin. In the second, the increase in the margin indicated a more conservative approach, ensuring diversification and alignment with the decision-maker's priorities.

Both scenarios maintained a high internal rate of return and maximized the buildable area while minimizing competition for land and houses. While the first scenario demonstrated greater financial efficiency, the second prioritized robustness, adding flexibility to the portfolio.

Decomposition refined the initial solution, generating a diversified portfolio aligned with the decision-maker's preferences. The FITradeoff method proved its ability to balance financial return, risk, and financial constraints, leading to robust and consistent solutions.

4. Conclusions

This study applied the FITradeoff multi-criteria method to select a portfolio of plots in high-end gated communities, considering criteria such as internal rate of return, buildable area, availability of plots and houses, and cash exposure, within a budget of 30 million reais. The results showed that FITradeoff efficiently organized the decision-maker's preferences. In the initial scenario, the portfolio cost R$28,462,423.74, while after decomposition, the cost was R$27,995,387.73. Both portfolios included plots with high financial returns and low competition, meeting the defined objectives.

The application of the model underscored the significant role that multi-criteria decision-making (MCDM) methods play in enhancing decision processes within the real estate sector. By incorporating structured evaluation frameworks, these methods facilitate a more transparent, systematic, and objective approach to selecting properties or investment opportunities. FITradeoff, integrated with the Decision Support System (DSS), proved to be flexible and effective in handling multiple criteria and budget constraints.

Overall, the findings reinforce the potential of MCDM approaches, particularly FITradeoff, in improving decision quality in the real estate sector. By fostering a structured yet flexible decision-making environment, these tools contribute to more informed and efficient choices, ultimately benefiting investors, developers, and policymakers involved in complex real estate transactions.

It is suggested to apply Value-Focused Thinking (VFT) to better structure the model's criteria and objectives. VFT can enhance the definition of the decision-maker's preferences, making the selection of plots more aligned with investment strategies.

References

Costa, A. P. C. S., & de Almeida, A. T. (2021). Decision support for sustainable urban planning: A case study using the FITradeoff method. Sustainable Cities and Society, 75, 103209.

Clemen, R.T. & Smith, J.E. (2009). On the Choice of Baselines in Multiattribute Portfolio Analysis: A Cautionary Note. Decision Analysis 6:256–262. https://doi.org/10.1287/deca.1090.0158

de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., & Garcez, T. V. (2015). Multicriteria and multiobjective models for risk, reliability, and maintenance decision analysis (Vol. 231). Springer International Publishing, New York, NY.

de Almeida, A. T., de Almeida, J. A., Costa, A. P. C. S., & Almeida-Filho, A. T. (2016). A new method for elicitation of criteria weights in additive models: Flexible and interactive tradeoff. European Journal of Operational Research, 250(1), 179–191.

Frej, E. A., Ekel, P., & de Almeida, A. T. (2021). A benefit-to-cost ratio-based approach for portfolio selection under multiple criteria with incomplete preference information. Information Sciences, 545, 487-498.

Gomes, L. F. A. M. (2009). An application of the TODIM method to the multicriteria rental evaluation of residential properties. European Journal of Operational Research, 193(1), 204-211.

Multicriteria Model for the Choice of Pour-on Acaricides in an Agribusiness Retail Company

Arthur Obici Pepino[1*], Carolina Lino Martins[1], João Batista Sarmento dos Santos Neto[1], Kassia Tonheiro Rodrigues[1], Nadya Kalache[1]

[1] Federal University of Mato Grosso do Sul (UFMS), Campo Grande-MS, Brazil
[*] arthurobicip@gmail.com

Abstract: This research proposes a multicriteria model using the FITradeoff method to select the most appropriate pour-on acaricide among the products of an agribusiness retail company, considering price, product, and supplier characteristics. The innovation of the study lies in the application of this model to pour-on acaricides, an area little explored in previous research. The methodology includes collecting information about FITradeoff, meetings with the decision-maker, and analysis of criteria about suppliers and products. The result revealed that a single product stands out for its profitability and unique characteristics. However, the product may vary depending on the context of each company, leading to the conclusion that the model allows for a specific decision, with potential for new studies in other regions of the country.

Keywords: MDCM/A; FITradeoff; Retail; Agribusiness; Acaricides

1. Introduction

Agribusiness has a significant participation in Brazil's GDP. According to recent data, the sector represents approximately 25-28% of the Brazilian Gross Domestic Product (IBGE, 2024). Within this context, the Brazilian cattle herd reached 234.4 million heads in 2022, an increase of 4.3% compared to 2021. The slaughter of cattle in Brazil in 2023 was 34.06 million head, a 13.7% increase compared to 2022 (IBGE, 2024). This demonstrates the significance of cattle within the country and highlights the vast market for veterinary treatment.

Cattle raising can be affected by ticks, which cause losses such as reduced body weight and lower leather quality. Furthermore, ectoparasites like ticks spread within herds and can even cause cattle mortality. Thus, controlling these parasites is necessary—despite the associated veterinary costs—to maintain productivity and avoid a decline in global food production (Gonzaga et al., 2023).

Among tick treatment methods are impregnated ear tags, injectables, sprays, pour-on solutions, and others. The pour-on method, as indicated by its name, requires application along the dorsal region of the animal in direct contact with the tick infestation area (Molento, 2020). This article focuses exclusively on pour-on application products.

Currently, the Brazilian veterinary market has approximately 250 registered products for cattle tick treatment. This high number results from the country's strategy of using synthetic acaricides for tick control, explaining the extensive catalog of available products (Klafke, 2024). Given this situation, retail companies dealing with these products need to focus on a specific product for commercialization. Within this context, Multicriteria Decision Making/Analysis (MCDM/A) can be used to identify the best product to sell.

Selecting a single option or ranking options (such as suppliers or products) is a widely discussed topic in academic literature. Among all methodologies, MCDM/A is classified as the most popular due to its adaptability across different fields and its ease of application and comprehension by decision-makers. MCDM/A considers the relationship between various common criteria among options and uses them to identify the most relevant option compared to others (Montes & Morais, 2023).

Among the various MCDM/A methods and tools, the FITradeoff method for choice problematic stands out for this study because it employs compensatory rationality based on a flexible and interactive process. Additionally, this method features a free online Decision Support System (DSS). Thus, this study aims to determine the best pour-on acaricide for a retail company in the agribusiness sector using the FITradeoff method, based on criteria established by an animal health sector's decision-maker.

2. Problem Modeling and Application

For the problem modeling stage in this paper, the decision-making process framework proposed by de Almeida et al. (2015) was used. Figure 1 shows the decision-making process framework.

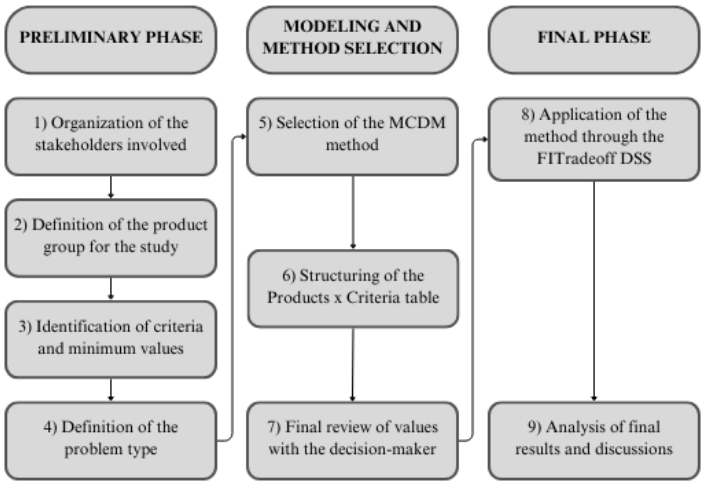


Figure 1: Decision-making Process Framework (de Almeida et al., 2015)

The first phase of the framework is the "Preliminary Phase," which begins with Step 1, organizing the actors in the decision-making process. It is necessary to focus on the decision-maker, who will be responsible for determining the criteria, the group of products to be analyzed, and preferences regarding the model. Following this, in Step 2, the products or groups of products in the portfolio that will be chosen for analysis must be identified, in this case, pour-on tick control products.

In Step 3, the relevant criteria for this product group were defined, as well as the rules limiting whether a product should be considered. Three criteria contain restrictions that exclude products that do not meet the rule: price, profit margin, and sales volume. The price and sales volume criteria had their limits set by excluding values significantly deviating from the linear progression trend (removing prices above R$ 1,000.00 and products with sales volumes below 100 liters). The profit margin criterion was defined by the decision-maker, stating that products with a margin below 15% should be automatically disregarded.

Table 1 presents the criteria set by the decision-maker and their role in the decision-making process.

Table 1. Criteria Description

C(i)	Criteria	Function	Type	Scale
C1	Cost per Liter	Minimization	Natural	R$ per Liter
C2	Payment Method	Maximization	Descriptive	3-choice Likert
C3	Profit Margin	Maximization	Natural	Percentage (%)
C4	Added Value	Maximization	Descriptive	4-choice Likert

C5	Quantity Sold in Liters	Maximization	Natural	Units per Liter
C6	Supplier Accessibility	Maximization	Descriptive	3-choice Likert
C7	Supplier After-Sales Support	Maximization	Descriptive	3-choice Likert
C8	Concentration	Maximization	Natural	Active In. grams

Tables 2, 3, and 4 outline how the Likert classification scale was used to quantify the descriptive criteria, including "Payment Method," "Added Value," "Supplier Accessibility," and "Supplier After-Sales Support."

"Payment Method" was described by the decision-maker using a table that details the payment conditions offered by the laboratories producing the pour-on tick control products in the company's current portfolio. The metric follows the logic that the greater the number of installments or the longer the payment intervals, the better the criterion score.

Table 2. Payment Method Criteria Scale

Grade	**Payment Method**				
1	28 days	30 days			
2	56 days	60 days	30/60 days	60/90 days	65/91 days
3	90 to 180 days	30/60/90 days	30/60/90/120 days		

Similarly, "Added Value" was also described by the decision-maker, considering that the product's effectiveness and duration determine its value; the more effective it is, the higher its score.

Table 3. Added Value Criteria Scale

Grade	**Added Value**	
1	Flies and Ticks	
2	Ticks (Medium Duration)	Flies, Ticks and Larvae
3	Ticks (High Duration and Prevents Reproduction)	
4	Ticks (Greater Duration and Prevents Reproduction)	

Finally, the "Supplier Accessibility" criterion accounts for response time and product lead-time (the period from order request to product delivery in-store), while the "Supplier After-Sales Support" criterion includes the supplier’s responsiveness to issues such as freight or product expiration dates. Both criteria follow a basic classification of bad, regular, or good.

Table 4. Supplier Accessibility and After-Sales Support Criteria Scale

Grade	**Supplier Accessibility/After-Sales Support**
1	Bad
2	Regular
3	Good

In Step 4, the problem to be addressed is decided: whether it involves product selection, product ranking, or grouping products into equivalence classes. Given the decision-makers' purchasing method, where purchases are always made for a single product when needed, the adopted problem is selection, as it limits the outcome to a single pour-on tick control product.

Due to the nature of the problem, where the decision-makers preference criteria exhibit compensatory rationality (i.e., a low score on one criterion can be offset by a high score on another), Step 5 opens the opportunity to use the FITradeoff selection method. FITradeoff was chosen for its consistency and flexibility, as well as its availability as a free Decision Support System (DSS) (de Almeida et al., 2016).

The FITradeoff method is based on analyzing the decision-maker's preferences, which are allocated into inequalities to form a linear programming model. During the process, the decision-maker responds to queries that help evaluate trade-offs and establish direct preferences between different alternatives. The obtained information is used to refine the scale constants space, improving selection possibilities (de Almeida et al., 2016).

Following the decision on the methodology to be used and the benefits of applying FITradeoff in this study, Step 6 defined, through Table 5, the decision matrix and their direct criteria. The "ID" refers to the product (followed by a number if there are multiple packages of the same product). Additionally, values highlighted in red indicate those reviewed with the decision-maker that will not be considered in the FITradeoff DSS execution, as they do not meet the criteria constraints. As a result, out of the 34 products currently in the portfolio, only 23 will be included in the analysis.

Table 5. Decision Matrix of the Problem

ID	C1	C2	C3	C4	C5	C6	C7	C8
A1	144,19	2	22,67%	3	145	2	1	2,5
A2	118,71	2	22,30%	3	740	2	1	2,5
B1	50,53	1	29,49%	2	557	2	1	5
B2	50,53	1	26,82%	2	652,5	2	1	5
C1	48,81	1	26,62%	2	582	2	1	5
C2	42,44	1	27,89%	2	3.445	2	1	5
D1	88,24	3	26,33%	1	34	2	2	1
D2	75,52	3	16,62%	1	415	2	2	1
E1	39,83	2	23,85%	1	1.081	2	1	5
E2	37,81	2	23,68%	1	3.676	2	1	5
E3	37,14	2	21,18%	1	3.625	2	1	5
F1	39,26	2	29,51%	2	123	2	2	5
F2	47,38	2	26,33%	2	855	2	2	5
G1	47,80	3	15,70%	2	1.188	2	1	6
G2	47,30	3	22,62%	2	1.485	2	1	6
H1	127,60	1	8,99%	2	580	2	2	1
I1	1.207,90	3	20,13%	4	24	2	2	2,5
I2	3.411,60	3	24,58%	4	2,75	2	2	2,5
I3	2.421,50	3	20,07%	4	95	2	2	2,5
J1	269,91	2	31,55%	3	5	2	2	3
K1	50,85	2	21,07%	3	295	2	1	2,5
L1	45,98	1	24,43%	3	53	2	2	0,5
L2	36,60	1	16,56%	3	2145	2	2	0,5
M1	192,44	3	14,32%	3	54	2	1	6
M2	38,48	3	15,17%	3	220	2	1	6
N1	36,40	1	23,71%	2	21	2	2	5
N2	36,40	1	13,46%	2	6.895	2	2	5
O1	39,00	3	21,70%	2	269	1	1	5
O2	34,08	3	26,61%	2	1.175	1	1	5
P1	113,52	1	15,88%	3	17	2	1	2,5
P2	98,58	1	12,18%	3	315	2	1	2,5

Q1	380,73	1	26,79%	2	1.304	2	2	5
R1	206,17	3	16,38%	3	109	2	2	2,5
R2	192,31	3	18,28%	3	220	2	2	2,5

Using a base table for importation, the refined data from Table 5 were imported into the FITradeoff DSS. With these values entered, the decision-maker, through their choices in cycle 0, defined the ranking preference order for the eight criteria, from most to least preferable, as follows: 1) Profit Margin; 2) Price per Liter; 3) Product Added Value; 4) Sales Volume (in Liters); 5) Payment Terms; 6) Supplier Accessibility; 7) Concentration; 8) Supplier After-Sales Support.

3. Results

Based on the criteria established for the analysis, it is observed that the most relevant factors for the decision-maker are directly related to the financial performance of the products, particularly profit margin and price. These aspects reflect the priority given to the economic return of choices, demonstrating that decisions are largely guided by the pursuit of better business opportunities.

The criteria considered less important include aspects related to the after-sales support provided by suppliers and the concentration of the active ingredients. This suggests that, for the decision-maker, technical issues or specific product details carry less weight compared to the business generation potential.

Figure 2 below visually demonstrates how the criteria defined by the decision-maker relate to each other, with some having a greater impact than others.

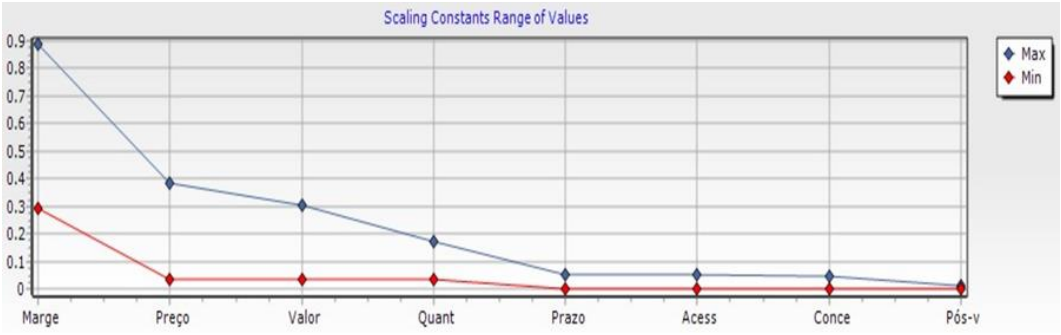


Figure 2: Scaling Constants

From Figure 2, it is possible to infer that the greatest variation among the criteria is in profit margin, with its fluctuation measured at 0.6 points, followed by a relatively smaller variation of 0.35 in price per liter, 0.25 in the product's added value, and 0.15 in the sales quantity per liter. Thus, it is evident that the product’s profit margin has a more significant influence compared to other criteria.

Five out of the 23 products are the best options for selection. Since the optimal choice among the five alternatives has not yet been determined, it is necessary to apply decomposition-based elicitation, which, similar to the questionnaire format used in cycle 0, refines the result only for the optimal alternatives. The presented and selected choices, the cycles, and the optimal alternatives are shown in Table 6 below.

Table 6. Decomposition Elicitation Cycles

Cycles	Choice A	Choice B	Selec.	Alternatives
0	Prior Definition of Criteria			B1-C2-F1-M2-O2
1	27 for C3 and 1 for C7	15 for C3 and 3 for C7	A	B1-C2-F1-M2-O2
2	27 for C3 and 381 for C1	15 for C3 and 34 for C1	A	B1-C2-F1
3	207 for C1 and 1 for C4	381 for C1 and 3 for C4	A	B1-F1
4	2 for C4 and 109 for C5	1 for C4 and 3,7K for C5	A	F1

In the fourth cycle of the elicitation process, FITradeoff returns the optimal alternative for selection. This alternative, along with its criteria and the maximum and minimum global values, can be seen in Table 7.

Table 7. Product Chosen as the Best Alternative

ID	C1	C2	C3	C4	C5	C6	C7	C8	Max Overall Value	Min Overall Value
F1	39,26	2	29,51	2	123	2	2	5	1.00	0.83

In summary, Table 6 exemplifies the decomposition-based elicitation process within FITradeoff's Decision Support System (DSS), determining that in the fourth cycle of interactions, by selecting consequence A as the most influential, the best option among the 23 evaluated is returned. With the final result, Table 7 reveals the most suitable pour-on acaricide for the company to continue with and even expand future negotiations, identified as product F1.

Thus, the study reaches the optimal result at the end of Stage 9 and conveys this information to the decision-maker, who will analyze it and depending on the current market situation, proceed with the purchase or consider the result for a future opportunity.

4. Conclusions

The selection of pour-on acaricides by the decision-maker was addressed through the application of a multicriteria model based on the FITradeoff choice method. This method allowed the selection of the best option among the available alternatives based on the decision-maker's preferences, utilizing the Decomposition-Based Elicitation process. The approach was efficient in reducing cognitive overload and facilitating the analysis of options.

The results confirmed that the model met expectations, enabling a direct selection. This decision allows the company to focus on marketing a targeted product within the pour-on acaricide group while also reducing costs associated with maintaining products in the portfolio that do not add significant value to the business.

The FITradeoff method stood out for its flexibility and interactivity. Additionally, its internal Decision Support System (DSS) further enhanced the process by ensuring consistency and ease of application.

For future applications, it is suggested that decision-makers and analysts adjust the criteria according to the characteristics of the products being evaluated. Furthermore, multicriteria methodologies can be explored in other categories of animal health products, including within the acaricide group, such as ear tags, sprays, and sprayers, further refining the portfolio of an agribusiness retail company.

References

de Almeida, A. T., Cavalcante, C. A. V., Alencar, M. H., Ferreira, R. J. P., de Almeida-Filho, A. T., & Garcez, T. V. (2015). Multicriteria and multiobjective models for risk, reliability, and maintenance decision analysis (Vol. 231). Springer International Publishing, New York, NY.

de Almeida, A. T., de Almeida, J. A., Costa, A. P. C. S., & de Almeida-Filho, A. T. (2016). A new method for elicitation of criteria weights in additive models: Flexible and interactive tradeoff. European Journal of Operational Research, 250(1), 179-191.

Gonzaga, M., Silva, R., Oliveira, L., & Santos, P. (2023). Essential oils and isolated compounds for tick control: advances beyond the laboratory. Parasites & Vectors, 16(415). https://doi.org/10.1186/s13071-023-05969-w

Instituto Brasileiro de Geografia e Estatística – IBGE. (2024). Em 2023, abate de bovinos cresce e o de frangos e suínos atinge recordes. IBGE. https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/39453-em-2023-abate-de-bovinos-cresce-e-o-de-frangos-e-suinos-atingem-recordes

Klafke, G. M., Martins, J. R., & Andreotti, R. (2024). Brazil's battle against Rhipicephalus (Boophilus)

microplus ticks: current strategies and future directions. Brazilian Journal of Veterinary Parasitology, 33(2). https://doi.org/10.1590/S1984-29612024026

Ministério da Agricultura e Pecuária – MAPA. (2023). Rebanho bovino brasileiro alcançou recorde de 234,4 milhões de animais em 2022. Ministério da Agricultura e Pecuária. https://www.gov.br/agricultura/pt-br/assuntos/noticias/rebanho-bovino-brasileiro-alcancou-recorde-de-234-4-milhoes-de-animais-em-2022

Molento, M. (2020). Avaliação seletiva de bovinos para o controle do carrapato. Ministério da Agricultura, Pecuária e Abastecimento. https://www.gov.br/agricultura/pt-br/assuntos/producao-animal/arquivos-publicacoes-bem-estar-animal/CARRAPATOS2.pdf

Montes, M., & Morais, D. C. (2023). Multicriteria negotiation model for selecting sustainable suppliers' problem in the agribusiness. Production Journal, 33(1). https://doi.org/10.1590/0103-6513.20220090

Environment, Agriculture, and Natural Resource Management

Architecting Context-Aware Decision Support Systems for Wildfire Management: An Event-Driven Approach

Angela-Maria Despotopoulou[1], Filippos Gkountoumas-Zrakić[1]; Konstantinos Christidis[1], Babis Magoutas[1]

[1] Frontier Innovations, 5 Laskaratou str., Athens, Greece
{angelina.despotopoulou; filippos.gkountoumas; kostas.christidis; b.magoutas}@frontier-innovations.com

Abstract: The management of wildfires relies on timely and informed decision-making, necessitating systems capable of integrating diverse data sources and dynamically adapting to evolving scenarios. This work presents an event-driven architectural framework for a software platform designed to support wildfire detection and response through the integration of remote sensing technologies and other data modalities. The architecture is event-driven, enabling the system to process real-time, heterogeneous data from various sources, including satellite-based fire detection systems, meteorological stations, environmental sensors, and AI-enhanced UAV imagery. Events such as temperature anomalies, fire detections, or environmental changes trigger the dynamic execution of response workflows. Remote sensing inputs are pivotal, offering high-resolution spatial and temporal information that drives fire propagation simulations and scenario-based analysis. The design of the architecture has been co-developed with domain experts, ensuring alignment with operational needs. These experts, including firefighters, environmental engineers, local authorities, and wildlife administrators, contributed to defining event patterns, simulation scenarios, and response workflows, ensuring the system is tailored to real-world wildfire management challenges. Outputs generated by the system, such as analytics, risk assessments, and detected situational hazards (e.g., endangered wildlife or infrastructure), are disseminated through user-centric interfaces. These include map-based dashboards, mobile applications, augmented reality devices, and text messaging systems, ensuring accessibility across diverse operational contexts. The architectural framework was validated within the TREEADS project, funded by the European Commission's Horizon 2020 Programme. In particular, two field trials with distinct scenarios took place in Samaria Gorge, Crete, Greece, and the Sorrento Peninsula, Italy.

Keywords: software architecture; decision support systems; process management; wildfire management system; context-aware wildfire detection

1. Motivation and Short Introduction

During the last two decades, an undeniable fact observed by the scientific community and the general public is the increasing frequency and intensity of wildfires across the European Union. In 2023, an area around twice the size of Luxembourg was burnt in the EU, amounting to more than half a million (504,002) hectares, according to the Advance report on Forest Fires in Europe, Middle East and North Africa 2023 *(San-MiguelAyanz et al., 2024)*. By the end of the year, the extent of the burnt area mapped by EFFIS reached 5,040.02 square kilometres, trailing 2017 (9,884.27 square kilometres), 2022 (8.372,12 square kilometres) and 2007 (5,883.88 square kilometres), the three worst years this century *(European Forest Fire Information System, n.d.) (European Commission, April 2024)*.

Even though the notion of reinforcing human fire-fighting efforts with technologically advanced hardware and software systems is not a recent development, several of those suffer from critical

limitations and unmet needs that significantly hinder their effectiveness. One of their commonest downsides is their exploitation of **fragmented data sources**, a fact that understandably leads to inability in providing to their end-users a comprehensive view of wildfire situations *(Arana-Pulido et al., 2018)*. Lack of real-time data ingestion and analysis, as well as the inability to process diverse data types – satellite imagery, IoT sensor readings, and meteorological information to provide some examples – imposes limits on the accuracy and timeliness of situational assessments. Moreover, these systems typically lack **scalability**, which has the effect of placing a ceiling in their ambitiousness: the increasing frequency, magnitude and severity of modern wildfire events renders them redundant frustratingly fast *(Saleh et al., 2023)*. Another infamous challenge is the significant gap in the **interoperability** of tools and platforms used by different agencies, even within the same statal jurisdiction, complicating coordination and collaboration during emergencies, despite good will *(Johnston et al., 2024)*. These are the gaps that inspired and driven the development of the platform illustrated in the present paper, aiming at proposing a solution with increased wildfire crisis management capabilities.

Efforts to build decision support systems (DSS) for wildfire detection and management have become increasingly prominent due to the rising frequency and severity of wildfires globally. To provide some examples, a DSS developed by the National Technical University of Athens integrates GIS technologies with semi-automatic satellite image processing, fuel maps, socio-economic risk modelling, and probabilistic fire models. Its key differentiator is the seamless integration of diverse data sources under a unified user interface, enhancing real-time fire management capabilities. However, the complexity of integrating such a range of technologies requires significant technical expertise *(Bonazountas et al., 2007)*. Similarly, the People&Fire webGIS tool simulates hazard and risk scenarios with a focus on land use transformation and incorporates social vulnerability into risk assessments. This human-centred approach offers a holistic perspective on wildfire impacts, although it depends heavily on access to high-quality land use and socio-economic data, which may be lacking in some regions (*Mileu et al., 2024*). Other innovative DSS include the AEGIS platform in Greece, a web-based GIS that supports early warning and real-time response by integrating weather data and fire simulations. Its standout feature is the real-time adaptability, though it relies on accurate user input, which can affect reliability (*Kalabokidis et al., 2016*). The Wildland Fire Decision Support System (WFDSS) in the U.S. integrates fire modelling and risk analysis to support large fire management, but its complexity and training requirements can hinder widespread use (*O'Connor et al., 2016*). The Risk Management Assistance program by the USDA Forest Service further emphasizes co-production with fire managers, increasing relevance but requiring ongoing adaptation (*Calkin et al., 2021*).

Aiming at addressing the uncertainty and socio-economic dimensions inherent in wildfire management, and to minimize the impact of wildfires through timely and informed decision-making, the present paper introduces an architectural framework for a software platform for Detection of Emerging Fire-related Situations and Response Process Management. The Decision Support System can be semantically divided into two groups of microservices, formulating the Event-Driven Situation Detector (EDSD) and Workflow Response Engine (WRE) *(European Commission, August 2024)*.

The **Event-Driven Situation Detector (EDSD)** whose purpose is to collect real-time and heterogenous outputs of modules developed under the auspices of the EU-funded TREEADS project *(Horizon 2020 programme, grant agreement number (101036926))*, as well as open data from external parties (e.g. weather stations and satellite systems owned and managed by scientific organisations), to support detection of crisisrelated situations that require specific response actions. EDSD handles the ingestion, processing and storage of data, as well as the maintenance of the relationships (that we will be calling "Rules") between conditions on the field and proposed actions from the part of the fire management stakeholders. EDSD can be also configured to host a scalable number of analytics and machine learning algorithms, extending to those produced by parties outside its development team.

The **Wildfire Response Engine** (**WRE**) handles the execution of fire response workflows given the datadriven assessments provided by EDSD. The WRE exploits the information stored by experts in the Rules Engine and the evaluations of the Analytics algorithms, to provide field actions, primarily recommendations for mitigating the situation directed to various applications addressed directly to end-users.

2. Methodology

The presented solution is the result of a multi-step process, following industrial-grade quality guidelines when it comes to software systems. The chosen methodology ensured a comprehensive, usercentred approach to developing a robust and effective wildfire crisis management decision support system.

To begin with, the development team analysed case studies and research papers on applications and systems addressing crisis management, operational decision-support and wildfire management situations. The next important step has been to identify and come into contact with key stakeholders (naturally including local authorities, professional firefighters, foresters, wildlife administrators, and environmental engineers). To this end, the team received considerable support from several partners within the TREEADS project, and most of all the leaders of the pilot tasks. Those groups of specialists, who in fact represent the target end-users of the platform, provided invaluable support by suggesting data sources, meteorological models, geographical information, business processes and common operational procedures (such as response scenarios to wildfirerelated conditions) and, overall, kindly helped the assemblage of requirements and user needs. Of great value has been, additionally, the contribution of the TREEADS project partners who developed the mathematical models that produced simulation scenarios (for fire propagation, smoke propagation and crowd distribution) which have been leveraged by the decision-support platform and, in particular, its analytics components. Armed with those inputs, the development team proceeded to collect samples from all identified usable data sources, as well as define interfaces and data models for exchange of information internally and with other TREEADS services (both data sources and data sinks). Interoperability has been a primary concern and has been tackled by using standardized data formats (such as JSON), interface design paradigms (such as REpresentational State Transfer, i.e. REST) and messaging protocols (such as Message Queuing Telemetry Transport, i.e. MQTT).

Subsequently followed the definition, design and documentation of the architectural structure of the platform, including the data ingestion components, the data management components, databases for storage and logging, algorithms for processing and decision support, integration of the aforementioned simulations, a state machine model to represent emergency level transitions and mechanisms to redirect outputs to user interfaces. Similarity metrics research and development as well as its validation has been a very important parallel process. What is more, expandability with third-party modules has been tested by integrating a deep learning wildfire progression module. During all those process steps, domain experts have been constantly consulted in defining event patterns and critical situation mitigation workflows.

3. Approach to System Architecture Design

The architectural paradigm that characterizes the synthesized component of EDSD and WRE, both internally and while interfacing with its data sources and data sinks, is known as an “event-driven architecture". The aspect that characterises this paradigm and sets it apart from its alternatives is the fact that the services communicate (1) asynchronously (2) via events. Asynchronous denotes that the downstream service can process the communication on its own schedule, without blocking the upstream service. Events, as opposed to requests, are significant occurrences or changes in the system's environment that trigger immediate processing and responses. Furthermore, event producers and

consumers are completely decoupled; meaning that producers should not be expected to be aware of or care which system components are consuming their events and how the consumers leverage those events in their service and to what end.

To capture and process events in our architectural approach, we employ Apache Kafka, a distributed event streaming platform capable of handling real-time data feeds. In practice, we assume two deployed infrastructures. For communication with other services of the TREEADS platform we leverage the global Event Streaming Platform, a backbone component of the TREEADS integrated proposition. For internal communication among the subcomponents of EDSD and WRE, we have deployed a second "local" Apache Kafka instance so as to ensure low-latency and high-throughput event processing. This dual Kafka setup allows for efficient segregation of global and local event traffic, optimizing performance and reliability. Additional messages, particularly IoT sensor readings from the field, are received via the TREEADS Sensors Data
Service, employing the MQTT protocol. Using MQTT allows EDSD to efficiently ingest a high volume of sensor data with minimal bandwidth, ensuring reliable and timely delivery of critical information. All messages exchanged through the aforementioned brokers, are formatted following the widely used JSON standard to facilitate interoperability.

The Decision Support System comprising EDSD, WRE and their shared components and messaging infrastructure has been architected as a collection of software microservices. Ahead of the pilot validation cycles those microservices have been deployed on virtual hosts, which are, in essence, cloud servers instantiated on a public cloud provider, named Hetzner Cloud. Sophisticated firewall rules have been established on each virtual server to meticulously regulate inbound and outbound network traffic, ensuring access is exclusively granted to authorized users and the applications that act as data sources or data sinks, including the TREEADS Holistic platform. Moreover, all storage units, whether ephemeral or persistent volumes, are encrypted to effectively mitigate the risk of data exposure. This deployment paradigm has been selected as we think it enhances system flexibility, allowing for independent scaling and updating of individual microservices without disrupting the entire system. Furthermore, the use of a public cloud infrastructure offers high availability and disaster recovery capabilities, safeguarding the system's resilience in case of hardware failures or other unexpected events. To ensure compliance with data protection regulations, comprehensive logging and monitoring mechanisms have been integrated, enabling real-time detection and response to any security incidents.

The key security features of the present deployment can be summarised as follows: (1) Distributed environment, which enables remote access and security-by-design. Deploying the software components and services on different VMs, makes it easier to horizontally scale the platform by introducing additional resources where necessary. (2) Secure communication among the deployed software components achieved by encrypted communications over TLS/HTTPS protocols. (3) Encryption-at-rest for ensuring data is inaccessible to malicious parties on the event of misplacement of the physical disk units that host the virtual servers. (4) Isolation and protection from unauthorized access hosts thanks to established firewall policies and rules.

While the event-driven microservices architecture of EDSD and WRE enables modularity and scalability, it introduces significant operational complexities. Each microservice runs in an isolated Docker container, demanding precise orchestration of lifecycle events, resource limits, and network bindings. Container sprawl and version drift across environments add to the DevOps burden, particularly when scaling horizontally on Hetzner Cloud's virtualized infrastructure. The use of dual Kafka clusters, local and global, alongside MQTTbased ingestion channels, introduces measurable inter-service network overhead, requiring careful tuning of throughput, buffering, and retention settings. The dual security mandate - encryption-in-transit and encryptionat-rest - further adds to this complexity, especially when balancing performance with cryptographic overhead and managing certificates across services. TLS encryption across internal and external endpoints, while essential, introduces processing

overhead and requires robust certificate management. Additionally, firewall enforcement across distributed VMs must be tightly managed to avoid accidental exposure of ports or services. To address observability, the architecture includes a dedicated monitoring component collection, as will also be mentioned below, to aggregate logs, analyse performance metrics, and visualises system health, ensuring traceability and operational awareness across the full deployment surface.

Figure 1 below shows the Reference Architecture of the Framework, depicting the microservices group that synthesise EDSD and WRE with distinct colours. Those will be subsequently analysed in detail.

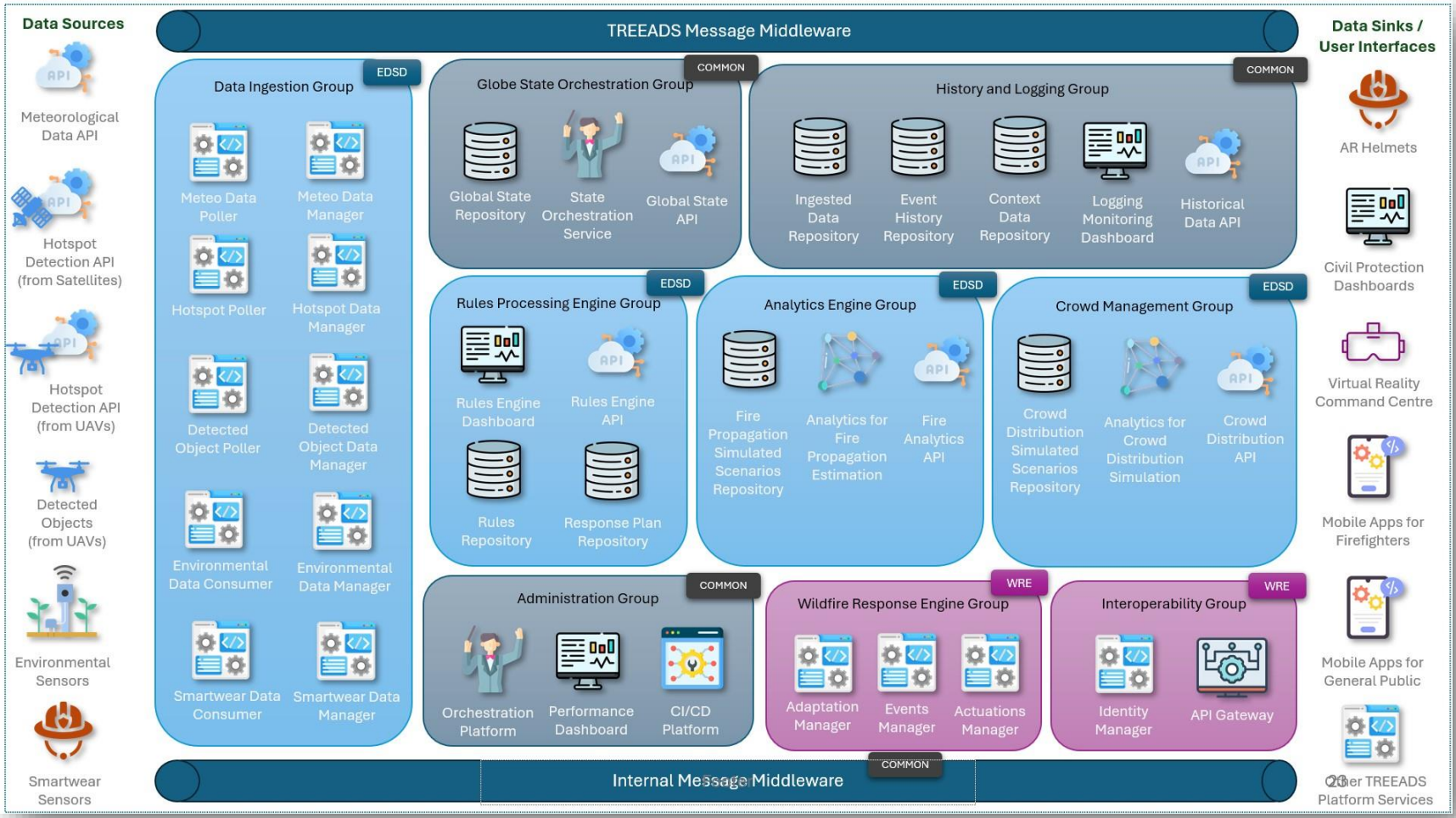


Figure 1: Reference Architecture for the Framework combining EDSD and WRE

4. Event-Driven Situation Detector (EDSD) Components and Interfaces

The present section revolves around the microservice groups that compose the Event-Driven Situation Detector (EDSD) whose purpose is to collect real-time and heterogenous outputs of other TREEADS modules, as well as open data from external parties, to support detection of crisis-related situations that require specific response actions. EDSD handles the ingestion, processing and storage of data, as well as the maintenance of the relationships (that we will be calling "Rules") between conditions on the field and proposed actions from the part of the fire management stakeholders. EDSD can be configured to host a scalable number of analytics and machine learning algorithms, extending to those produced by parties outside its development team. The management of proposed fire mitigation plans is handled by the WRE component that will be analysed in a subsequent sub-section. The principal microservice groups that compose EDSD are the following:

The **Data Ingestion Group** is responsible to ingest data from heterogeneous sources and then proceed to apply staple engineering procedures such as cleaning, data model transformation, cataloguing and, eventually, redirection to appropriate storages. This choreography is concluded by emitting a new event towards the internal message bus announcing the ingestion of new data from the outer environment. The microservices performing the ingestion – a collection of API pollers and message broker consumers (Kafka and MQTT) – are autonomous and decoupled from each other. Each is paired with a microservice performing the data engineering tasks integrating gracefully the data into EDSD (data model transformation, data storage, data cataloguing etc.). This convention presents the following advantages:

- **Scalability**: Microservices can be scaled independently, allowing frictionless addition and removal of data ingesting services adapted to specific data source specifications, as indicated by evolving business requirements.
- **Flexibility**: Each microservice can be developed, deployed, and updated independently, facilitating continuous integration, testing and deployment.
- **Resilience**: The failure of a particular microservice does not necessarily affect the availability and performance of the entire web of services, asserting robustness for the composed system.

On the other hand, managing a large number of microservices as opposed to a monolith software piece poses orchestration challenges that call for organisational discipline and well-chosen monitoring tools.

The **Rules Processing Engine Group** is tasked with defining and maintaining the set of rules that determine automated responses to events produced by the Data Ingestion Group as information arrives from the systems environment. In other words, it evaluates incoming data against predefined criteria (conditions) to trigger appropriate actions. This ensures, evidently, that the system can dynamically adapt to changing conditions. The criteria are expected to be procured via a user interface by domain experts (i.e. foresters, local authorities, emergency managers). This dashboard, depicted in Figure 2, allows the pairing environmental stimuli, such as elevated temperature, confirmation of fire as opposed to mere suspicion, animals in danger, etc.) to a variety of response plans that may include simple alerts, suggestions on evacuation, evaluations of wildfire severity, likelihood of a fire propagation scenario to evolve etc. Apart from the Rules and Response Plan repositories and the dashboard, the Rules Engine also exposes an Application Programming Interface (API) to allow CRUD operations to other services.

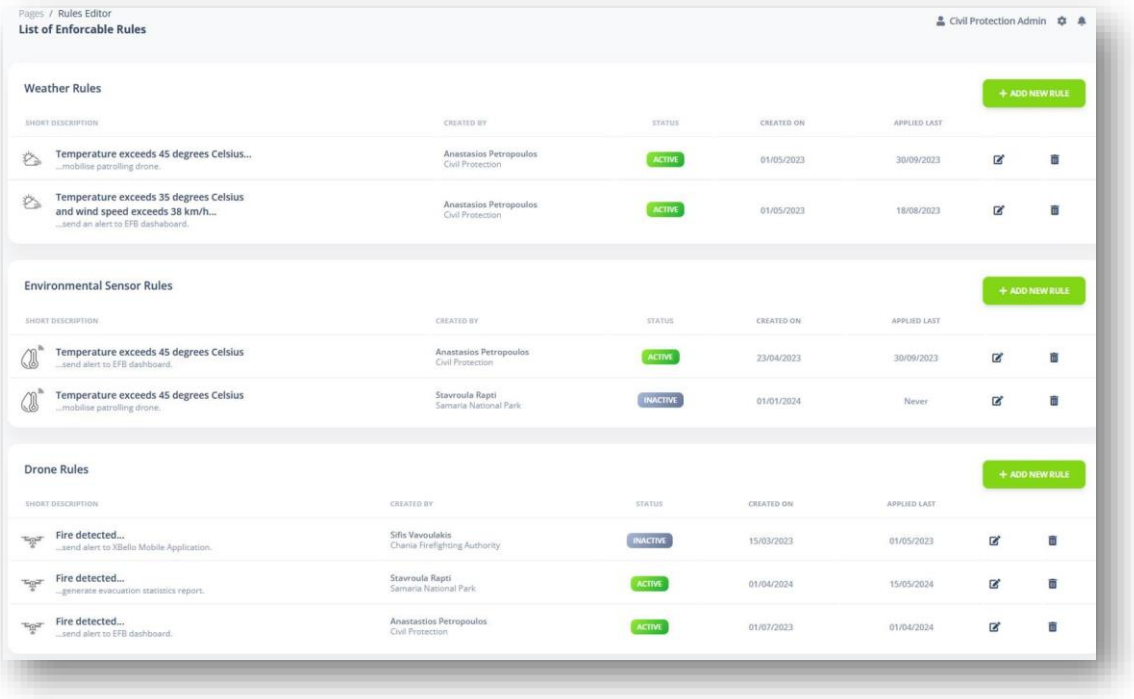


Figure 2: The editor of EDSD Rules Engine for defining response actions to detected events

The **Analytics Engine Group** has been built with analytics algorithm(s) as its core. It is responsible to juxtapose the environment conditions assimilated and announced by the Data Ingestion Group with preprepared fire propagation scenarios. By applying similarity metrics (e.g. Cosine Similarity, Manhattan Distance, Hamming Distance, Pearson Correlation), the group detects the simulation scenario that most closely matches the current real-time input, in an attempt to forecast the movement of fire and trigger the most appropriate response plan should the similarity level exceeds a threshold provided by domain experts. The subcomponents of this group apart from the similarity calculator itself are a repository for the simulated scenarios and an API to request and retrieve a new evaluation. Analogous is the architectural logic for the **Crowd Management Group**, with the only difference that this time the algorithm estimates the scenario under which people in the affected area will be notified on the danger and evacuated. Alternatively, the statistical estimation of their number based on past experience and historical data. It is noteworthy that EDSD can similarly be scaled to include in the future any number of analytics and/or machine learning algorithms that might enrich its business value proposition. All components of this group have been implemented using the Python programming language, as opposed to most of the other microservices that are employing the Java Spring Framework.

Those four groups comprise officially the Event-Driven Situation Detector (EDSD), however their functionality is extended towards field actuation with the WildFire Response Engine (WRE) that will be discussed a little further below.

5. Components common to EDSD and WRE

The architecture proposition of our decision-support system includes three groups of microservices that are common to EDSD and WRE and equally handle needs of both. Two perform critical functions to serve the logic around event management and, in particular, they

handle global state management, all-purpose storage of data (a data lake of sorts) and, finally, logging of events and system messages. The third is auxiliary and serves the development and maintenance teams of the platform. Analytically, the three groups common to both EDSD and WRE are the following:

The **Global State Orchestration Group** is responsible to assess at any given time the current level of emergency. This function has been built around a so-called state machine, i.e. a computational model used to track and manage the different states of emergency in the wildfire detection and management system. It transitions between predefined states (e.g., Monitoring, High Risk, Fire Suspected, Fire Confirmed, Fire Extinguished) based on incoming events and data. Depending on the current state, EDSD and WRE give emphasis to different prompts and perform different actions. To give an example, simple alerts for high temperature and low humidity make sense during the Monitor phase while alerts for possibly trapped animals get circulated only when a wildfire has been officially confirmed. Figure 3 illustrates an example of messages received by the end-user as the wildfire situation progresses through the various phases. The group of microservices that are correlated to this functionality comprises the State Orchestration Service, one that coordinates and announces the present Global State, a repository to maintain a historical record and an API publicising this information to other microservices.

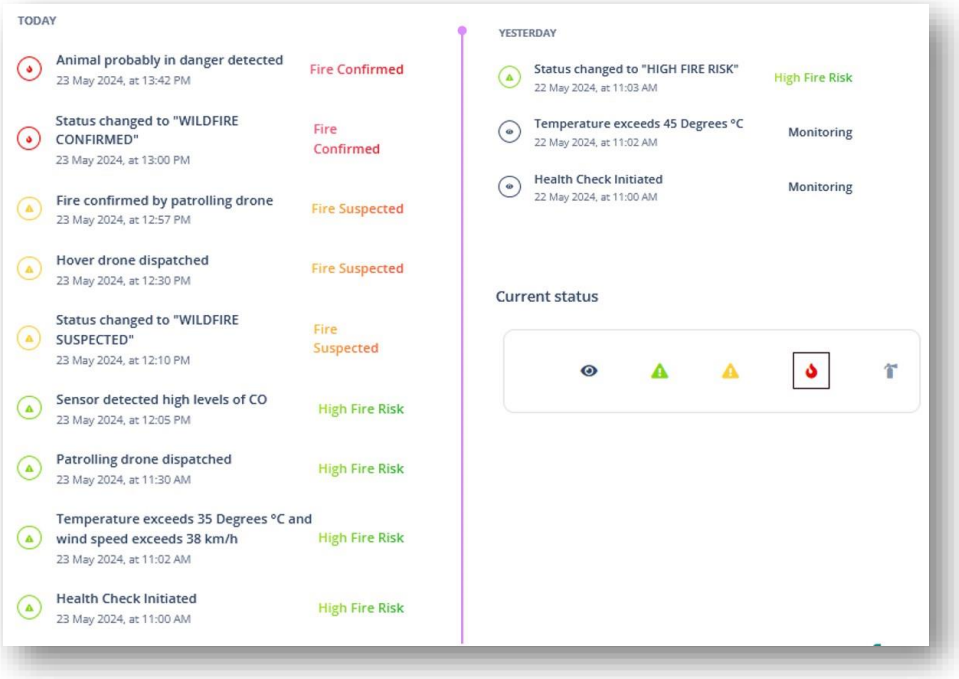


Figure 3: An example of messages received by the end-user as the situation progresses through the various phases

The **History and Logging Group** serves as a centralised global repository. It can be perceived as a data lake comprised by several smaller databases. One maintains processed ingested data by the Data Ingestion
Group of EDSD. Another database logs all generated events within the EDSD/WRE platform, partly for administration reasons and partly to be consulted historically. There is also a repository for historical context data, for example meteorological data of previous years, procured via open data APIs from

government sources (e.g. the repository of "Meteo.gr", a website maintained by the National Observatory of Athens). The aforementioned data is searchable and maintainable through a dedicated API. It is also monitorable via a dedicated dashboard leveraging selected parts of the open-source community version of the ELK technology stack (the user interface itself is based on Kibana).

The **Administration Group** is addressed to the development team and supports all platform maintenance tasks. It is not specific to wildfire management purposes, but it employs software industry partners that have become the de facto standard in recent years; an orchestration platform, dashboards for performance monitoring and metrics, as well as a platform for performing Continuous Integration and Continuous Delivery/Deployment tasks (including unit, functionality, integration and performance testing). The latter is composed of a set of highly popular software products: a code repository (GitHub), an automation server (Jenkins), an artifact repository (JFrog Artifactory), and the cloud deployment environment. It is selfexplainable that this group of services, while auxiliary, plays a critical role in maintaining the overall health and robustness of the system, enabling the development team to focus on building and improving features.

6. Wildfire Response Engine (WRE) components and interfaces

Having described the groups of microservices that synthesise EDSD and those common between EDSD and WRE, let us analyse similarly discuss the groups that comprise the latter. The Wildfire Response Engine (WRE) handles the execution of fire response workflows given the data-driven assessments provided by
EDSD. The WRE exploits the information stored by experts in the Rules Engine and the evaluations of the Analytics algorithms, to provide actuations on the field, primarily recommendations for mitigating courses of action directed to various applications addressed directly to end-users. Examples of such interfaces are user dashboards exposed via the TREEADS Holistic platform, as well as mobile devices such as portable command centres and AR-enhanced Helmets. It has been deemed important to interface with those frontends in a way generic enough to assert interoperability. The principal microservice groups that compose WRE are the following:

The **Wildfire Response Engine Group** is responsible for triggering the (semi)automated execution of fire response workflows depending on the current Global State and the assessments produced and announced via events by EDSD. The response plan workflows, as mentioned before, have been pre-prepared by domain experts and stored in the Rule Engine. WRE undertakes the task of adaption to the diagnosed situation following those Rules through a dedicated microservice, while another manages the actuations per se (e.g. sends an e-mail, redirects messages with recommendations to user interfaces etc.). A third microservice abstracts event management for all purposes within the EDSD and WRE combined ecosystem. It is evident that the business value of this group is in its ability to "trigger the action that is supposed to happen on the field" following EDSD's "diagnosis of the situation using data from the field".

Since EDSD and WRE are practically backend or middleware services, for WRE outputs to become perceivable by human actors it is essential to redirect them towards a data sink, or – in simpler terms – a user interface. Those interfaces may vary greatly from dashboards to mobile devices and wearables. Those interfaces must also be notified dependably and timely each time new information is being addressed to them. A third problem is that some user interfaces require trigger requests from data on their own initiative, without knowing the entire inner architecture of EDSD/WRE. Those problems are addressed by the **Interoperability Group** of microservices, the second and last of WRE. This group is responsible for redirecting

events to the TREEADS Event Streaming Platform announcing to user interfaces that new recommendations or alerts have been generated by WRE. Then it exposes an API that can be used by the interfaces to a) retrieve more information corresponding to the identifiers of events they have received and b) to request historical data on demand without an event prompt. In case of (a), the event messages also contain information on which user roles the message is being addressed to and for how long it remains valid (allowing thus user interfaces to ignore messages irrelevant to them). This approach abstracts the data flow of outgoing messages and decouples
WRE from the particular technical specifications that user interfaces might present. Finally, this group, since it communicates with the external environment of the EDSD/WRE platform, it also encompasses a simple identity manager.

7. Platform Validation in TREEADS Pilot Campaigns

Up to this point of the present document, the design of the system has been analysed in detail. The present chapter, as the next logical step, illustrates how EDSD and WRE have been incorporated into TREEADS pilot scenarios for validation. This process took take place during the autumn of 2024 at two pilot sites, the Samaria Gorge in Crete, Greece and the Sorrento Peninsula in Campania, Italy. It is self-evident that the scenarios described in the next two sections have been co-created and policed over in various iterations with the teams conducting the respective pilot activities, including the thresholds to be exceeded before a notification was supposed to be sent to end-users, the data analytics, statistical and AI models applicable to each situation as well as the recommended courses of action stemming from their respective stimuli.

Here follows a summarisation of the services EDSD/WRE demonstrated in the context of **the Samaria Gorge pilot scenario**:

- When meteorological conditions indicated high fire risk (e.g. high temperature and low humidity), alerts were directed to end users to remain vigilant. Also, recommendations were directed to the drone pilots to commence patrolling, actively searching for fire outbreaks. Meteorological conditions were being collected from weather stations and environmental sensors deployed in strategic places in the Gorge.
- When a patrolling drone indicated suspicion for fire or an environmental sensor indicates high concentration of smoke, alerts were directed to end users through the user interfaces accompanied with the location of the incident, along with the recommendation to deploy the second type of drone with onboard image processing capabilities.
- Through its analytics capabilities EDSD compared current meteorological conditions to meteorological models that corresponded to simulated past fire incidents. As the same meteorological models had been used as the basis for preparation of a variety of simulations by other research teams (fire propagation, smoke propagation and crowd evacuation scenarios), if a meteorological model displayed high similarity to that of a simulated past incident, WRE indicated this, so that the users could consult the simulations through the visualisation capabilities of their user interfaces (an example of such an interface offering simulation comparison is depicted in Figure 4 below).
- In addition, the EDSD/WRE platform, according to the time of day and month of the fire incident redirected to the end users an estimation based on statistical data from previous years of how many people are expected to be present in the Gorge before and after the ignition point (the trail in the Gorge is almost linear).

- As a fire incident was evolving, when a patrolling drone reported the presence of vehicles, people or animals in peril, this information was redirected by EDSD/WRE towards the user interfaces through alert messages.
- As a fire incident was evolving and firefighters with AR-Helmets had already been deployed on the field, their IoT sensors feed EDSD with additional measurements. The helmets also received textual alerts similar to those received by the dashboards and mobile devices. Such messages directed firefighters towards the source of smoke, or towards trapped people and vehicles.
- All throughout this exercise, the system reported the transition from one emergency level to the next.

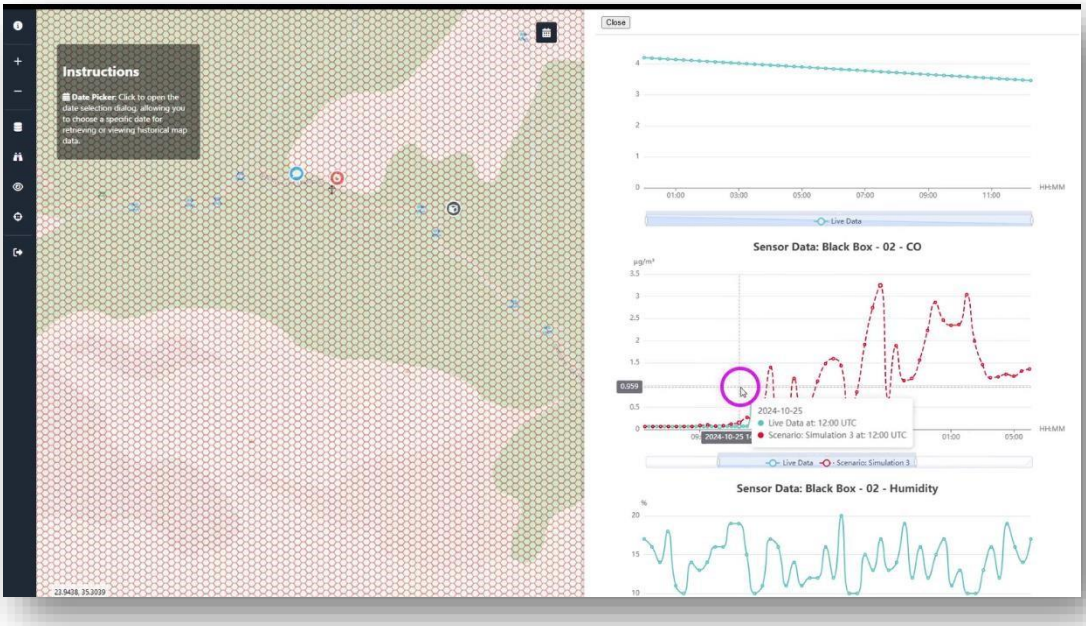


Figure 4: Visualisation tool comparing the actual situation to propagation simulations built leveraging past events

The scenario of the **Sorrento Peninsula pilot** leveraged the functionalities of EDSD and WRE in a different way than that of the Greek pilot from an end-user's perspective; however, on the level of technical implementation, similarities were significant. In addition, the methodology followed to receive expert opinion had been the same: with the assistance of the pilot leaders, experts from Sorrento had been consulted regarding their business needs, they have been approached for translating said business needs into use case scenarios that in turn were used as basis to define rules in the Rules Engine. Also, they had been asked to provide datasets including geographical locations of interest and, finally, they have been requested to select simulation scenarios to form the basis of the field exercise.

Characteristic user personas for the pilot scenario were the Field Exercise Responsible, the Municipal Command, the members of the Firefighting Team, the members of Volunteer Teams and the Director of Firefighting operations. These people had access to devices supporting navigation through the Internet via a browser. Their activities were carried out through dashboards whose primary features are maps; in the context of TREEADS those was be provided by an Open Street Maps-based visualisation tool (a sample is depicted in Figure 5),

and the messaging system of the TREEADS Holistic Platform. EDSD and WRE was acting as a back-end service providing insights regarding the evolution of a wildfire incident.

Volunteer team bases (e.g., squares) and Firefighting Team bases (e.g., triangles), together with Municipal Command bases (e.g., stars) were depicted on the map. Across the area under surveillance, a grid of environmental sensors was deployed, also visible as dots on the same map. It is noteworthy that the sensors formulating the grid were this time the only data sources for EDSD; as opposed to the Greek pilot, there were no data streams stemming from UAVs, satellites, weather stations nor smart wearables.

A wildfire event was demonstrated with real actors moving around the designated area. The propagation of fire was reported via the sensors, who changed colour on the map upon fire exposure and again each time the fire proximity was reported as getting worse. Simultaneously, the firefighting teams' locations were also being reported by the actual people in real time through mobile applications exposing the same map (the little squares and triangles were moving around). The area encompassing the sensors who reported fire was highlighted with a bright colour in the shades of orange. Sensors also report with a colour change (black) that the fire near them has been extinguished and thus the area was now considered "burnt" (a screenshot of the grid of sensors during the incident is depicted in Figure 6). All system messages were additionally propagated via e-mail.

In the context of the scenario just described for the Italian pilot the role of the EDSD/WRE platform was assigned to do the following:

- To ingest in real-time information from the environmental sensors regarding the temperature and smoke concentration, thus gaining a hint on the progress of the wildfire. Since during the field exercises no actual fire was commenced, the messages supposedly coming from sensors were artificially generated following fire propagation simulations - that constituted the output of work of a separate team by the University of Salamanca - and placed in an incoming stream from Apache Kafka.
- To follow the progression from emergency-state to emergency-state, as usual, through its State Machine.
- To consult its stored Rules on how to react each time a sensor reports fire ignition or extinguishment. More or less the response action will always be to report this to the user interfaces.
- To report periodically (per those rules) fire progression and the state of each individual sensor by placing text messages into Apache Kafka. Those reports will be then asynchronously received by the TREEADS dashboards for visualisation on maps.

Both the pilot field exercises demonstrated the platform's real-time analytics, emergency escalation detection, and support for tactical decisions. Stakeholders, including firefighters and municipal leaders, completed postexercise questionnaires regarding the demonstrated technologies as a whole, reporting 92% overall satisfaction. Key KPIs for EDSD/WRE showed a 78% improvement in situational awareness, a 70% reduction in intra-team communication delays and a 65% reduction in emergency response time compared to baseline drills. Participants noted the clarity of alerts, ease of accessing simulations, and effective coordination support as standout features. They particularly appreciated how the system filtered complex information into actionable insights, reducing cognitive load during high-pressure situations. Importantly, stakeholders highlighted that EDSD/WRE did not attempt to replace human decision-making but rather complemented it, thus respecting the value of on-the-ground experience and incorporating field wisdom into its recommendations, rather than relying solely on sterile, automated data outputs. These results, we hope, validate EDSD/WRE as a powerful decision support asset for wildfire risk and incident management across Europe.

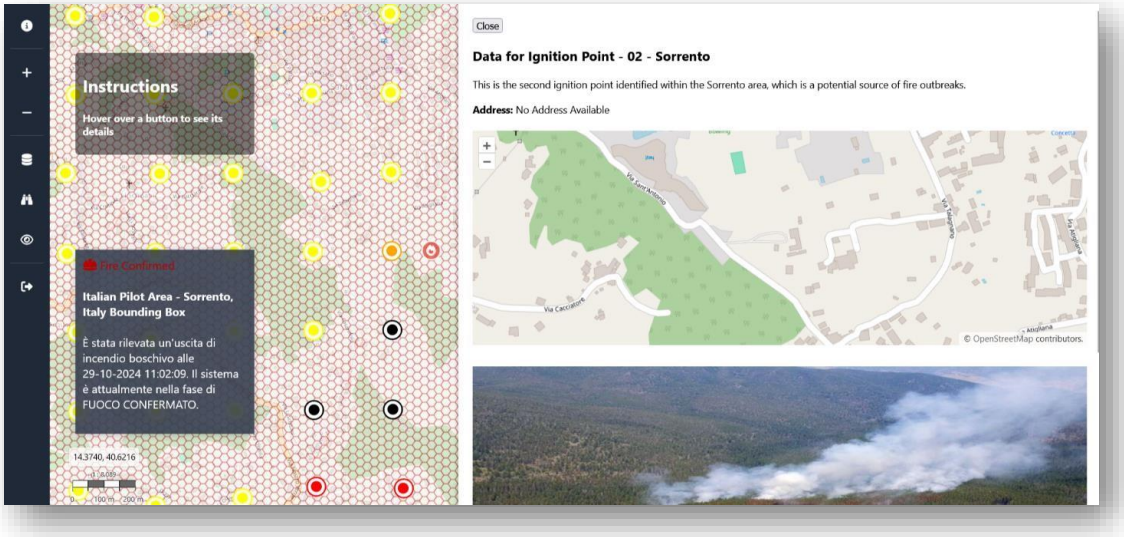


Figure 5: Map Visible to End-Users for the Italian Pilot – Fire Outbreak

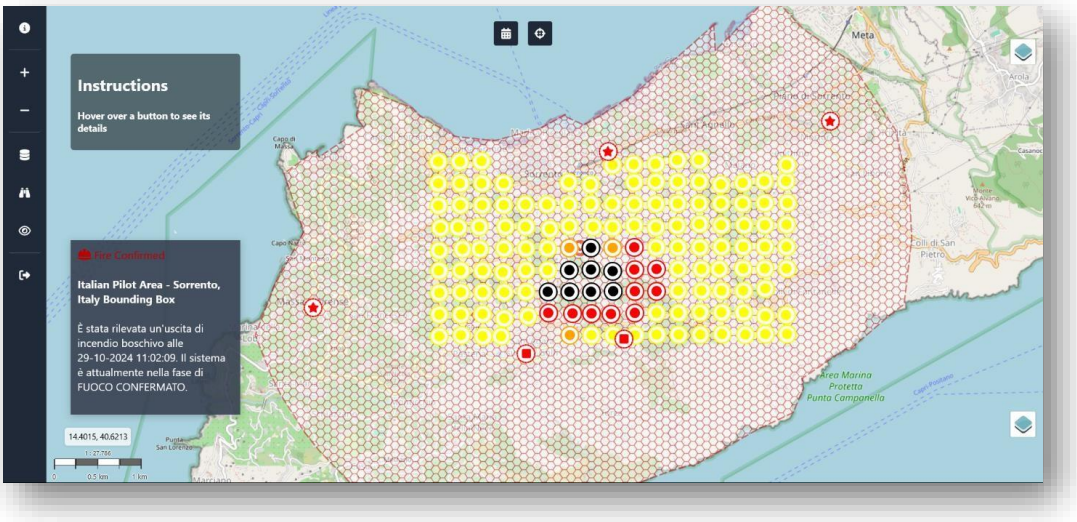


Figure 6: Map Visible to End-Users for the Italian Pilot – Areas Visibly Burnt

8. Lessons Learned and Challenges Discussion

The architectural framework was validated within the TREEADS project, funded by the European Commission's Horizon 2020 Programme. In particular, two field trials with distinct scenarios took place, as mentioned before, in Samaria Gorge, Crete, Greece, and the Sorrento Peninsula, Italy in October 2024.

The development of our analytics-enabled fire detection and mitigation system and its subsequent adaptation to those specific pilot demonstrations has been a complex yet rewarding journey. Throughout this process, we encountered numerous challenges that tested our technical capabilities, project management skills, and ability to adapt to evolving requirements.

However, these challenges also provided invaluable lessons that eventually strengthened the system's robustness, scalability, and effectiveness.

8.1. Integration of Heterogeneous Data Sources

One of the foremost challenges was the integration of diverse data sources, including IoT sensors, satellite imagery, and UAV-based visual data. Each data source presented unique challenges in terms of data formats, update frequencies, and varying reliability. Initially, the task of harmonizing these disparate data streams into a cohesive, real-time processing system proved daunting. However, we overcame this by developing a flexible data ingestion architecture (the Data Ingestion Group mentioned in the EDSD section above) that utilized distinct microservices for each data source, allowing for modular and scalable integration. This approach not only facilitated seamless data consolidation but also enabled easy adaptation to incorporate new data sources as they became available, ensuring the system's continued relevance and accuracy.

8.2. Managing Asynchronous Data Flows

One of the challenges we encountered was managing asynchronous data flows from various sources, such as IoT sensors, satellite feeds, and user inputs, which often arrive at different times and intervals. This variability can lead to difficulties in synchronizing data and ensuring that analytics are based on the most current and comprehensive dataset available. We successfully addressed this challenge by employing an eventdriven system architecture, which allowed the system to react to new data events in real-time. This architecture facilitated the immediate processing and integration of data as it became available, regardless of the source or timing. By leveraging event-driven techniques, we ensured that the system could handle dynamic and sporadic data streams efficiently, providing up-to-date insights and maintaining the integrity of the analytics pipeline. This approach not only improved the system's responsiveness but also enhanced its capacity to support timely and accurate decision-making in critical wildfire situations.

8.3. Interoperability with a variety of User Interfaces

Another significant challenge was ensuring the system's interoperability with a range of diverse user interfaces, whose technical specifications where initially unknown, as it was essential that the system's outputs be accessible and usable across various platforms and devices. These interfaces included map-equipped dashboards, mobile applications, augmented reality (AR)-enabled helmets, and basic text messaging systems. The diversity of these interfaces required the system to be highly adaptable, with a smartly designed API and flexible data output formats that could accommodate different display and interaction technologies. By prioritizing a design that was agnostic to specific devices or platforms, we created a system capable of delivering critical information effectively, regardless of the end user's hardware or software environment.

8.4. Scalability and Futureproofing

Ensuring the scalability and futureproofing of the system was another major challenge. As data volumes and the complexity of analysis increase over time, the system needed to be capable of scaling up without performance degradation. Additionally, it had to be adaptable to future technological advancements and the incorporation of new functionalities. We addressed this by adopting a microservices architecture and cloudbased infrastructure, which provided the necessary flexibility and scalability. This architecture allows for the addition of new features or the integration of advanced AI models without significant reworking of the system’s core structure. It also facilitates the easy scaling of computational resources in response to increasing data loads, ensuring that the system remains efficient and responsive as demands grow.

9. Future Research Directions

To maximize the impact and sustainability of the Decision Support Tool delineated in the present paper, ongoing development and refinement are essential. This may include expanding the range of data sources, incorporating new Analytics and AI algorithms, and experimenting with a greater variety of intended user interfaces. Experimentation with the use of edge computing to enhance real-time data processing could contribute towards reducing latency and addressing unstable Internet connectivity. Feedback and usability testing from a bigger variety of stakeholder perspectives will undeniably reveal actionable insights. The investigation of incorporation of various standards into the system's design, as well as contributing feedback to their maintenance teams, is a method to build synergies around this software product. Additionally, exploring opportunities for interoperability with other wildfire crisis management systems at the national and European levels equally seems a promising research direction. Finally, a daring endeavour could also entail expansion of the solution towards managing different environmental hazards, such as floods and landslides.

Acknowledgements

This work has been carried out under the TREEADS project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101036926. The authors acknowledge valuable help and contributions from all partners of the project. Moreover, the content reflects only the authors' view, and the European Commission is not responsible for any use that may be made of the information it contains. The Dashboards in Figures 4, 5 and 6 have been designed and implemented by Engineers for Business (EFB) based in Thessaloniki, Greece.

References

Arana-Pulido, V., Cabrera-Almeida, F., Perez-Mato, J., Dorta-Naranjo, B. P., Hernandez-Rodriguez, S., & Jimenez-Yguacel, E. (2018). Challenges of an Autonomous Wildfire Geolocation System Based on

Synthetic Vision Technology. Sensors (Basel, Switzerland), 18(11), 3631. https://doi.org/10.3390/s18113631

Bonazountas, M., Kallidromitou, D., Kassomenos, P., & Passas, N. (2007). A decision support system for managing forest fire casualties. Journal of environmental management, 84(4), 412–418. https://doi.org/10.1016/j.jenvman.2006.06.016

Calkin, D. E., O'Connor, C. D., Thompson, M. P., & Stratton, R. D. (2021). Strategic Wildfire Response Decision Support and the Risk Management Assistance Program. Forests, 12(10), 1407.

https://doi.org/10.3390/f12101407

European Commission. (2024, April 10). Wildfires in 2023 among the worst in the EU this century. EU Science Hub. Joint Research Centre. https://joint-research-centre.ec.europa.eu/jrc-news-andupdates/wildfires-2023-among-worst-eu-century-2024-04-10_en

European Commission. (2024, August 31). TREEADS Deliverable 5.5: A decision support tool for wildfire response (Horizon 2020 Project No. 101036926).

European Commission. (n.d.). TREEADS: A holistic fire management ecosystem for prevention, detection, and restoration of environmental disasters (Horizon 2020 grant agreement No. 101036926).

TREEADS. https://treeads-project.eu/

European Forest Fire Information System. (n.d.). EFFIS official website. Copernicus Emergency Management Services. https://forest-fire.emergency.copernicus.eu/

Johnston, F., Jones, C., Li, F., Stehr, A., Torres-Vázquez, M. Á., Turco, M., & Veraverbeke, S. (2024). Interdisciplinary solutions and collaborations for wildfire management. iScience, 27(8), 110438. https://doi.org/10.1016/j.isci.2024.110438

Kalabokidis, K., Ager, A., Finney, M., Athanasis, N., Palaiologou, P., & Vasilakos, C. (2016). AEGIS: A wildfire prevention and management information system. Natural Hazards and Earth System Sciences, 16(3), 643–661. https://doi.org/10.5194/nhess-16-643-2016

Mileu, N., Zêzere, J. L., & Bergonse, R. (2024). People&Fire webGIS tool for wildfire risk assessment. MethodsX, 12, 102709. https://doi.org/10.1016/j.mex.2024.102709

O'Connor, C. D., Thompson, M. P., & Rodríguez y Silva, F. (2016). Getting Ahead of the Wildfire Problem: Quantifying and Mapping Management Challenges and Opportunities. Geosciences, 6(3), 35. https://doi.org/10.3390/geosciences6030035

Saleh, A., Zulkifley, M. A., Harun, H. H., Gaudreault, F., Davison, I., & Spraggon, M. (2023). Forest fire surveillance systems: A review of deep learning methods. Heliyon, 10(1), e23127. https://doi.org/10.1016/j.heliyon.2023.e23127

San-Miguel-Ayanz, J., Durrant, T., Boca, R., Maianti, P., Liberta`, G., Oom, D., Branco, A., De Rigo, D., Suarez Moreno, M., Ferrari, D., Roglia, E., Scionti, N., & Broglia, M. (2024). Advance report on forest fires in Europe, Middle East, and North Africa 2023 (JRC137375). Publications Office of the European Union. https://doi.org/10.2760/74873

Decision support system to evaluate the socio-economic impact of the National Agrarian Reform Program in Brazil

Aime Maria Sebold de Almeida[1], João Batista Sarmento dos Santos Neto[1], Carolina Lino Martins[1], Mario de Araujo Carvalho[1], Kassia Tonheiro Rodrigues[1]*

[1] Federal University of Mato Grosso do Sul (UFMS), Campo Grande, Brazil
* kassia.rodrigues@ufms.br

Abstract: This paper aims to develop a data-driven decision support system (DSS) to evaluate the socio-economic impact of the Brazilian National Agrarian Reform Program (PNRA) on resettled families. The DSS integrates a comprehensive questionnaire based on literature review and expert consultations, covering quality of life, organizational management, and post-settlement perceptions. A pilot field test was conducted using a custom system designed to streamline data collection. Data was processed through Power BI to generate three analytical dashboards: respondent demographic profiles, organizational management metrics, and settler perceptions of program impacts. Based on feedback from experts and field technicians, the questionnaire, software interface, and dashboard visualizations were refined to enhance result interpretation. The proposed DSS demonstrated effectiveness in supporting evidence-based decision making regarding PNRA implementation, providing tools that facilitate rigorous evaluation of the program's socio-economic outcomes. This approach has potential to inform strategic improvements to the program based on empirical data rather than non-systematic evidence.

Keywords: Decision support system, agrarian reform evaluation, data visualization, socio-economic impact assessment, evidence-based policy.

1. Introduction

An agrarian reform settlement consists of a group of agricultural units, referred to as plots or lots. In Brazil, these settlements are established by the National Institute for Colonization and Agrarian Reform (INCRA) on rural land, as defined by the Ministry of Agrarian Development. Each plot is allocated to a family of farmers or rural workers who lack the financial means to purchase their own land (Brasil, 2020).

Settlers perceive an improvement in their quality of life within these settlements compared to their previous living conditions. It is assumed that the previous level was very low, similar to the reality of the majority of poor Brazilians in the agrarian environment. Thus, the establishment of settlements is regarded as a strategy for the social integration of this population (Albuquerque et al., 2005).

Sousa et al. (2010) emphasize that agrarian reform must go beyond land distribution and ensure conditions for production, income generation, and access to fundamental rights such as healthcare, education and sanitation. To assess its actual impact, it is essential to directly engage with settlers and analyze their experiences and challenges. Persistent difficulties may jeopardize the economic viability of the plots and the long-term settlement of families, requiring a multidimensional and strategic approach.

The decision-making process is essential in addressing these challenges. Due to its multidisciplinary nature, decision-making in this context requires the analysis of multiple criteria (Kunsch et al. 2009; Katsikopoulos, 2011). The increasing amount of information further complicates the selection among alternatives (Gomes et al., 2004). Therefore, decisions must align policies with the actual needs of families, fostering social inclusion and improving quality of life.

Effective organizational decision-making depends on proper information management (Clericuzi et al., 2006). The ability to make data-driven decisions provides a competitive advantage because it ensures objectivity and accuracy (Souza, 2023). Recent technological advancements have facilitated the adoption of Business Intelligence (BI) tools, including information dashboards and data visualization systems (Dover, 2004; Gitlow, 2005; Paine, 2004). BI tools enable the detailed analysis of stakeholder behavior, market trends, and external factors. Negash and Gray (2008) define BI as a data-driven decision support system that utilizes analytical tools such as dashboards. The systems enhance information accessibility by structuring and organizing large volumes of data.

Given the complexity of decision-making in agrarian reform settlements—where economic, social, and environmental factors intersect—traditional assessment methods often fail to capture the full spectrum of challenges faced by settlers. Decision Support Systems (DSS) offer a structured framework for analyzing these multidimensional problems, seamlessly integrating data-driven insights to inform and enhance policy-making processes.

Despite the potential benefits, few studies in the literature explore the application of data-driven decision support systems using BI in the context of agrarian reform. This gap hinders the development of effective data-driven strategies and highlights the need for an integrated approach that combines qualitative and quantitative assessments to support policy decisions. Therefore, this study aims to develop a data-driven decision support system to evaluate the perception of beneficiaries of the National Agrarian Reform Program (PNRA) regarding the socioeconomic impacts of the program. This approach enhances decision-making by providing policymakers with a comprehensive and dynamic analysis of settlement conditions. The proposed model includes: (i) a semi-structured questionnaire; (ii) TerraSurvey software for data collection; (iii) an interactive dashboard. This approach will make it possible to understand the results and challenges faced by families, facilitating data analysis and promoting more agile and assertive decisions.

2. Methodological Procedures

This research included the following phases: Questionnaire development, software development, dashboard development, testing and validation.

2.1. Questionnaire development

The questionnaire, based on exploratory research and a literature review, sought to capture PNRA beneficiaries' perceptions of the socioeconomic impacts of the program. It addressed questions related to quality of life, social perceptions, and economic impacts.

A mixed approach was adopted, with quantitative questions allowing an objective assessment of indicators such as increase in family income, length of stay in the settlement, and number of family members, and qualitative questions to capture the settlers' subjective perceptions of the program's impacts and challenges. The questionnaire was validated by experts and program technicians who suggested adjustments to ensure clarity, accessibility, and alignment. The experts provided insight into the suitability of the questions for the research objectives and suggested adjustments to the wording and scope of the questions to ensure that the questions were understandable and accessible to the beneficiaries. The technicians provided feedback on the practical application to the target audience to ensure that the questions were aligned with the specificities of the interviewees.

2.2.TerraSurvey Software Development

The TerraSurvey software, developed as a Progressive Web App (PWA) using the JavaScript framework called Vue.js, allows data to be collected and managed in a flexible and accessible way. TerraSurvey was built based on the open source tool ToFormy (Carvalho, 2024), using Pinia as a state management library, RxJS to facilitate the use of reactive programming, while Dexie.js was used to implement a local database.

The compilation and distribution process was managed by Vite, an application building tool that provides a fast and efficient development experience. The questionnaire digitization process was structured as a series of technical steps that transformed physical forms into interactive software components. Using the SurveyJS component, it was possible to create a dynamic question flow that adapts to the user's previous answers (SurveyJS, 2024).

TerraSurvey's interface was designed with simplicity and accessibility in mind, allowing users with limited technology skills to use it efficiently. Using the Vuetify framework already present in ToFormy, the application has a responsive design that adapts to any screen size, from smartphones to desktop computers. Simplified menus, intuitive icons, and a clean layout were deliberate choices to minimize the learning curve for settlers and maximize the efficiency of data collection.

The software offers support for two languages - English and Portuguese - making it easy to use in different cultural and geographical contexts, options for configuring themes, and is available as a PWA application on the website: https://terrasurveyapp.web.app.

2.3. Dashboard using PowerBI

TerraSurvey data was integrated into Power BI. The data was imported using files in common formats, Excel and CSV. The data then went through a cleansing and organization process. This process included eliminating duplicates, handling missing values, standardizing responses, and categorizing qualitative variables. To address this challenge, we created a specialized column that transforms qualitative data into numerical values, allowing for seamless integration with quantitative metrics. This methodological approach substantially enhanced cross-referencing capabilities and analytical precision. To mitigate the inherent subjectivity of qualitative responses, we deployed artificial intelligence to systematically categorize textual inputs based on semantic patterns and contextual relevance. This AI-driven classification system enabled more nuanced grouping of questions into precisely defined categories, thus reducing human bias and enhancing the reliability of subsequent analyses.

The next step was to transform them into charts and interactive reports in Power BI to provide a clear and intuitive visualization of the results. The key metrics were organized into key categories that grouped together sections of the questionnaire that dealt with related topics. The process resulted in the creation of three main dashboards: (i) interviewed data analysis, which provides an overview of the demographic and socioeconomic profile of the settlers; (ii) organizational management profile, which analyzes responses related to the management of activities, use of resources, and marketing practices; and (iii) interviewees' post-settlement perceptions, which provides insights into the perceived impacts, challenges, and improvements observed in their lives.

The dashboard, developed in Power BI, provides interactivity, allowing data to be filtered by various criteria and segmented analysis, as well as real-time analysis with periodic data updates.

3. Results

Figure 1 shows the flow of the data-driven decision support system developed in this study, from the creation of a questionnaire to the provision of relevant information in dashboards for decision-makers. Technicians and analysts are responsible for the development, integration, questionnaire application, and data analysis phases, and stakeholders are the decision makers who use the reports and dashboards

to support strategic decisions. The system integrates data collection software with Power BI and data analysis to facilitate information-based decisions.

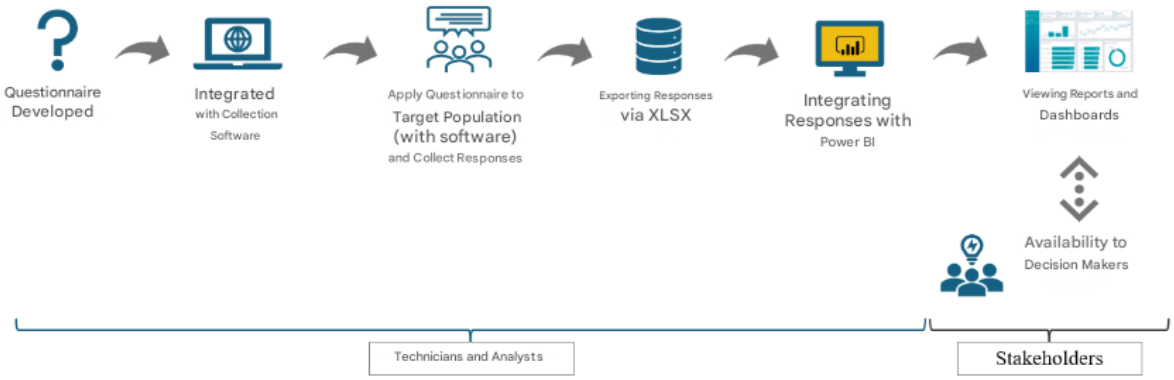


Figure 1: Representation of the DSS architecture

The questionnaire developed was divided into sections to collect essential information about the respondents and their characteristics. The first section, "Interviewee Data," covers age, family composition, and length of residence. The second, "Interviewee Profile," covers land ownership, education, work, and income, and analyzes legal security and management capacity.

The third section, "Property Information," examines land regularization and economic activities. The fourth section, "Rural Organization, Technical Assistance and Training," examines participation in organizations and management training. The fifth section, "Property Performance and Management," analyzes financial control, infrastructure, and strategies for overcoming deficiencies. The sixth section, "Commercialization, Credit, Rural Insurance and Environmental Regularization," examines financial and operational management. The seventh section, "Impacts of Land Regularization," examines investments in infrastructure and basic services and improvements in quality of life. The final section, "Explore in Interview (optional)," allows for discursive and optional responses to provide an open space for respondents to share experiences and perceptions not covered in the previous sections.

The software developed supports the collection of data through questionnaires, allowing for easy customization as needed, and the intuitive interface makes it easy to manage surveys (Figure 2). The initial menu offers options such as "Surveys" where the user can create, view, edit, archive, and delete records. The data (responses) can be viewed in a spreadsheet and exported in XLSX format.

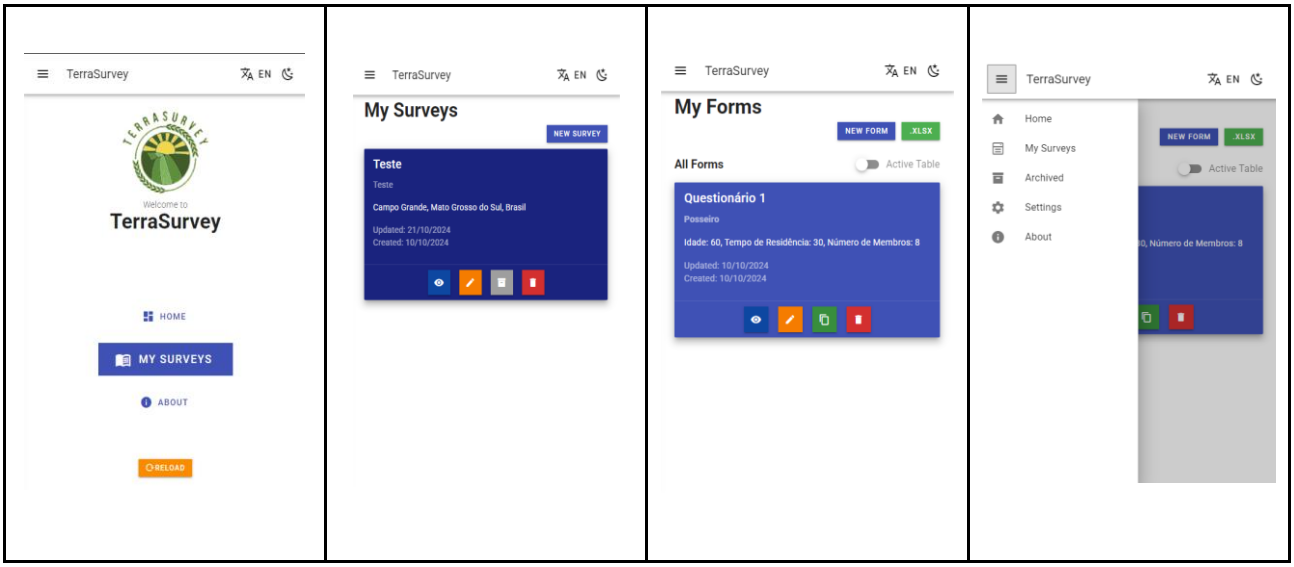


Figure 2: Main screens of the application.

The dashboard organizes the data through the use of graphics, icons, and strategic colors, transforming a complex set of data into an intuitive and accessible interface and providing insights that facilitate decision making.

The first panel presents demographic data of the interviewees, reflecting the answers from sections 1 and 2 of the questionnaire (Figure 3) such as age group, education and family composition. The aim of this panel is to provide a basis for understanding the socio-economic profile of the settlers, making it easier to identify trends and patterns between the groups.

Figure 3: Dashboard 1 - Interviewee profile

The second panel presents the interviewees' organizational management profile, providing an analysis of their management and operating practices in the settlement (Figure 4). This panel summarizes the responses to sections 4, 5 and 6 of the questionnaire, which covers topics such as financial control, production planning, marketing and access to resources and credit. This visualization makes it possible to identify patterns of management behavior, making it easier to understand which practices are most widely adopted and which areas have the greatest gaps.

Figure 4: Dashboard 2 - Organization and Management Profile

The third panel assesses the impact on the beneficiaries' quality of life after settlement, the advantages and disadvantages of participating in the land reform program, and the challenges that remain (Figure 5).

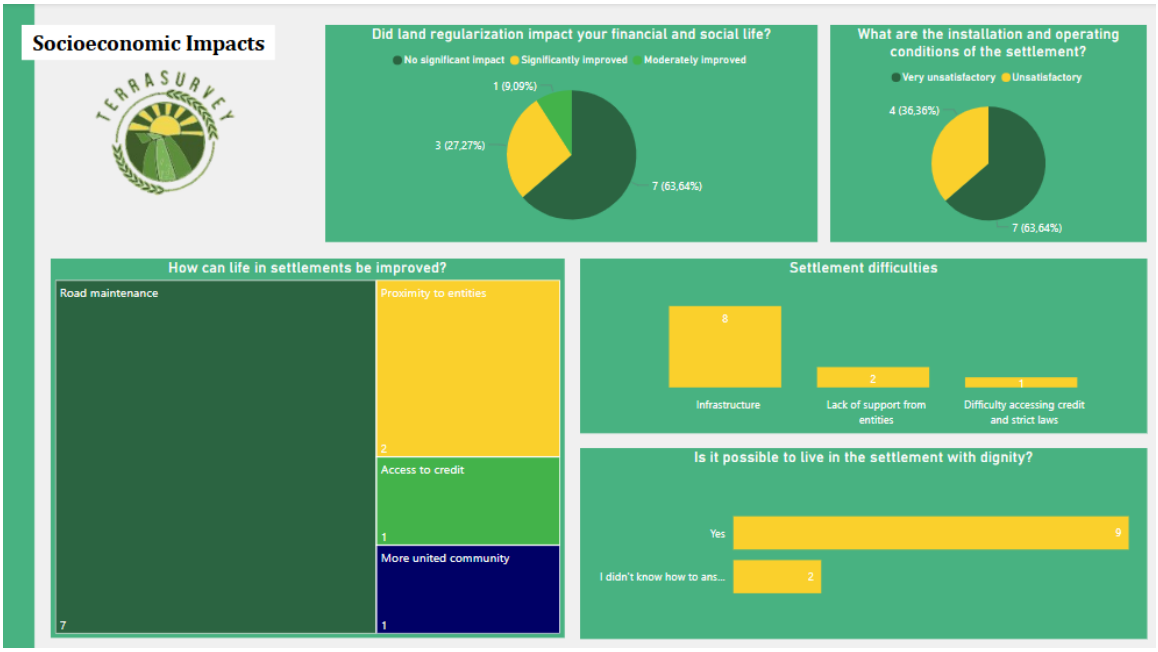


Figure 5: Dashboard 3 - Impacts on Post-Settlement Life

3.1. Testing and Validation

To validate and illustrate the use of the tool, data was collected from 11 PNRA beneficiaries in Santarém, Pará, Brazil. Since many of the beneficiaries are illiterate, technicians administered the questionnaires to them using their own devices.

After the application, the technicians provided feedback on the difficulties encountered during the application. Based on this, confusing questions were revised or removed, and the language was

simplified. The dashboard interface was also adjusted to ensure the integrity of the data and make it easier for managers and researchers to interpret.

3.2. Discussion of results

The results obtained reflect the proposed objectives by allowing the structured collection of information through a questionnaire, the operationalization of the data by the software and its visual and interactive presentation on the dashboard.

This approach confirms the relevance of integrated systems to support data-based decisions, as pointed out by Vinke (1992), who emphasizes the need to consider multiple criteria. The system proved capable of aggregating different perspectives and variables, providing input for more informed and strategic decisions.

Implementation occurs within a sensitive context that directly impacts the quality of life of people seeking better conditions for their survival and transforming their reality. This challenge requires an attentive and empathetic view, taking into account the diversity of stakeholders - beneficiaries, public managers and social organizations - and the need to handle data in a careful and responsible manner.

The analytical environment presents significant complexity, with each decision potentially creating substantial impact. Aligning effective strategies and clear action plans in a dynamic and multifactorial scenario makes the process even more challenging. This reinforces the importance of tools that combine social sensitivity with technical and technological skills, allowing careful analysis of data.

The system's design aligns with the best practices in literature. Aguarón et al. (2003) point out that the increasing complexity of decision problems requires the use of more flexible and open approaches that use software tools as technological support. Previous studies suggest that interactive dashboards facilitate the interpretation of information and promote greater managerial involvement. Resina & Javier (2018) emphasize the importance of easy-to-interpret visualizations that allow decisions to be made based on data, pointing out that dashboards capture key business indicators.

Despite achieving its objectives, the system has some limitations. The accuracy of the analysis depends on the quality of the information provided by the beneficiaries. The usability of the software and the dashboard requires adequate user training. In addition, the need to manually update the dashboard can cause delays that affect the agility of strategic decisions. An automated update mechanism should be explored to minimize data latency and improve responsiveness. Additionally, implementing error detection and validation procedures can enhance data reliability.

Business Intelligence fundamentally relies on organizational data for decision-making processes. The data warehouse transforms data into information, which is then transformed into knowledge through analytical tools. However, feedback latency can significantly impact operational responsiveness (Negash and Gray, 2008). The update process and data transmission timeline to decision-makers require optimization. Furthermore, accurate data collection remains critical for meaningful visualization. Future enhancements could include machine learning models to detect inconsistencies or anomalies in responses, thereby improving data integrity.

To provide a clearer understanding of the system's workflow, Figure 6 presents a visual summary of its key components and processes. This approach enables a comprehensive understanding of how the collected data is processed, analyzed, and strategically transformed into actionable insights, thereby providing a foundation for data-driven decision-making processes across organizational levels.

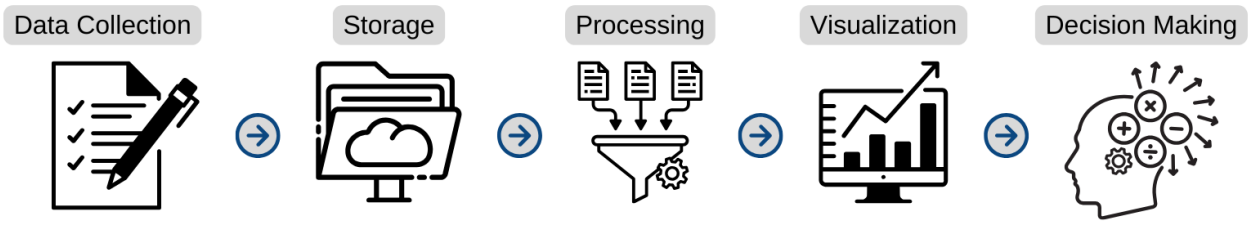


Figure 6: Representation of the DSS workflow focused on data processing and analysis

The findings have significant implications both practically and theoretically. From a practical standpoint, the system provides a tool for public managers and policy makers to make decisions based on more accurate information. Theoretically, it reinforces the value of technology integration in assessing the socioeconomic impact of social programs. Future research should pursue a multidimensional agenda encompassing these key directions: the integration of AI-driven analytics and predictive modeling to enhance decision-making processes, particularly in forecasting long-term settlement sustainability; the evaluation of visualization effectiveness across diverse user profiles, especially those with limited data analysis experience, through usability studies aimed at optimizing dashboard design for maximum accessibility; and the measurement of the proposed tool's actual impact on decision quality, investigating whether the system demonstrably improves policy formulation and implementation outcomes.

4. Conclusions

This study presented a data-driven decision support system to assess the perceptions of PNRA beneficiaries. The integrated solution comprises a validated questionnaire, the TerraSurvey mobile data collection platform, and an interactive analytics dashboard that together form a cohesive framework for program assessment. The DSS integrates qualitative and quantitative assessments to evaluate the socioeconomic impacts of the program. Unlike conventional evaluation models, which often rely on isolated economic indicators, the proposed DSS incorporates stakeholder perceptions through a semi-structured survey. This approach enhances decision-making by providing policymakers with a comprehensive and dynamic analysis of settlement conditions. By adopting a multidimensional perspective, this study moves beyond traditional impact evaluations, offering a more nuanced and data-driven approach to policy assessment.

The system enables real-time monitoring and facilitates multidimensional analysis of socioeconomic impacts while significantly enhancing accessibility and transparency of beneficiary outcomes. Managers and researchers thus have a robust and efficient tool to evaluate the program, identify challenges, and direct public policies more effectively.

Our contributions extend beyond tool development to include a validated methodological approach for gathering reliable beneficiary data and a structured framework for assessing critical program outcomes such as land tenure security, income stability, and quality of life improvements. This approach addresses previous limitations in monitoring systems while establishing replicable protocols for long-term impact assessment.

It is recommended that future research implement an automated flow in the dashboard directly from the TerraSurvey database. This integration will streamline access to information and allow for faster strategic decisions. The backend of the software needs to be developed, including the implementation of a database to store questionnaire responses in the cloud and APIs for easy access to the data. Although the system performs well within the current scope, its scalability in larger and more complex deployments requires further assessment. Future studies also should explore how the system handles increased data volume, multiple user interactions, and integration with governmental data

infrastructures. Potential bottlenecks, such as processing delays and storage limitations, should also be considered to enhance robustness.

Acknowledgments

We would like to thank the Ministry of Agrarian Development and Family Farming (MDA), the National Institute of Agrarian Reform (INCRA).

References

K.N. Flam, S.B. Rem, G.A. Tim, A. Moss, A. Afzu, Example 1 of Reference for the ICDSST 2019 Short Paper Template*, EPRI TR-100382*, Electric Power Research Institute, CA., USA, 2002.

Aguarón, J., Escobar, M. T., & Moreno-Jiménez, J. M. (2003). Consistency stability intervals for a judgement in AHP decision support systems. European Journal of Operational Research, 145(2), 382-393.

Albuquerque, F. J. B. D., Coelho, J. A. P. D. M., & Vasconcelos, T. C. (2004). As políticas públicas e os projetos de assentamento. Estudos de psicologia (Natal), 9, 81-88.

Brasil. Instituto Nacional de Colonização e Reforma Agrária (INCRA). (n.d.). Assentamentos. Retrieved from https://www.gov.br/incra/pt-br/assuntos/reforma-agraria/assentamentos

Clericuzi, A. Z., Almeida, A. T. de, & Costa, A. P. C. S. (2006). Aspectos relevantes dos SAD nas organizações: um estudo exploratório. Production, 16(1), 8–23. https://www.scielo.br/j/prod/a/BtjLCsGVhgJ3VCrJ766xnwP/?lang=pt#

Dover, C. (2004). How dashboards can change your culture. Strategic Finance, 86(4), 43-48.

Gitlow, H. (2005). Organizational dashboards: Steering an organization towards its mission. Quality Engineering, 17(3), 345-357.

Gomes, L. F. A. M., Araya, M. C. G., & Carignano, C. (2004). Tomada de decisões em cenários complexos. São Paulo: Pioneira Thomson Learning.

Kunsch, P. L., Kavathatzopoulos, I., & Rauschmayer, F. (2009). Modelling complex ethical decision problems with operations research. Omega-International Journal of Management Science, 1100-1108.

Moreno-Jiménez, J. M., Aguarón-Joven, J., Escobar-Urmeneta, M. T., & Turón-Lanuza, A. (1999). Multicriteria procedural rationality on SISDEMA. European Journal of Operational Research, 119(2), 388-403.

Negash, S., & Gray, P. (2008). Business intelligence. In F. Burstein & C. W. Holsapple (Eds.), Handbook on Decision Support Systems 2 (pp. 175-193). Springer-Verlag Berlin Heidelberg.

Sousa, J. M. M. de, Loreto, M. das D. S. de, Cunha, B. G., & Locatel, D. C. (2010). A reforma agrária e a qualidade de vida das famílias assentadas em Sergipe. UNIARA. Retrieved from https://www.uniara.com.br/legado/nupedor/nupedor_2010/00%20textos/sessao_5A/05A-07.pdf

Blockchain, Supply Chain, and Logistics

Understanding the determinates of blockchain technology adoption in the supply chain: A multi-method approach using PLS-SEM and fsQCA

Xiaotian Xie[1*]

[1]Newcastle University, Newcastle, UK
* xiaotian.xie@newcastle.ac.uk

Abstract: The recent interest in Blockchain technology has led many companies to leverage it to enhance supply chain performance. However, in the literature, there are few deep insights regarding how organisations build blockchain technology capability. This study aimed to identify the determinants of blockchain adoption in the supply chain and verify the influence of the combination of various factors on adoption intention. To fill the gap, the study draws on the Technology-Organization-Environment (TOE) framework, Diffusion of Innovation (DOI) theory, and the resource-based view (RBV) and proposes a blockchain adoption conceptual model. Then, this study uses a mixed-methods approach; we integrate Partial Least Squares Structural Equation Modelling (PLS-SEM) to assess hierarchical relationships and fuzzy-set Qualitative Comparative Analysis (fsQCA) to explore configurational effects. This study provides supply chain managers with actionable strategies to enhance blockchain implementation, supporting more resilient and efficient supply chain operations.

Keywords: blockchain; supply chain management; fsQCA; PLS-SEM; Digital transformation

1. Introduction

Emergent technologies, mainly associated with industry 4.0 and digital service, are expected to have a disruptive impact on the supply chain and force firms to develop new business strategy models. One of the most promising of these technologies is blockchain (Maull et al., 2017; Queiroz et al., 2019). Blockchain is a decentralised, distributive, and advanced technology that we can describe as a digital shared ledger that is disturbed over the network (Dutta et al., 2020). Every node in this peer-to-peer network maintains an upgraded copy of a database with all transaction history (Bahga & Madisetti, 2016; Scott et al., 2017). Once the records are added, they cannot be edited by a single actor but are verified and managed using automated and shared governance protocols (Christidis & Devetsikiotis, 2016). Blockchain technology (BCT) has raised the significant interest of supply chain managers because of its unique nature of disintermediation, transparency, confidentiality, security and automation (Wang et al., 2021). Many organisations have started to explore the application of BCT in the supply chain, ranging from product provenance and traceability (Jabbar et al., 2021), smart contacts (Chang & Chen, 2020), supplier trust (Brookbanks & Parry, 2022), cross-border trade (Chang et al., 2020) and supply chain sustainability (Saberi et al., 2019). Utilising blockchain in the supply chain can improve supply chain transparency and traceability, improve supply chain efficiency, and reduce risks and administrative costs (Gökalp et al., 2022; Rashid et al., 2022). Based on the recent survey conducted by Deloitte with data collected from more than 1000 participants around the world, the results show that blockchain will become one of the top five strategic priorities for the next two years (Pawczuk et al., 2019).

To date, blockchain is still in the early stage of development, and the studies in the literature that unify blockchain with the supply chain are in their infancy (Queiroz et al., 2019; Schmidt & Wagner, 2019; Wang et al., 2021). Thus, many scholars call for empirical research to investigate the full

implications of BCT in the supply chain (Pournader et al., 2020; Rejeb et al., 2021; Sheel & Nath, 2019). Besides, although several studies investigate the impact of blockchain on the supply chain, most researchers focus on conceptualising and examining blockchain possibilities and challenges for supply chains and providing a solution for particular use cases. This is reinforced by several literature reviews of BCT in the supply chain (Dutta et al., 2020; Moosavi et al., 2021; Queiroz et al., 2019; Rejeb et al., 2021). Furthermore, only 8% of the 26,000 blockchain projects started in 2016 were still actively developed in 2017 (Browne, 2017). In other words, the issue of how to guarantee that blockchain provides value to their companies and supply networks remains unanswered. Therefore, more in-depth research is required to identify the factors that impact blockchain adoption in the supply chain areas. To accomplish the research aim, this study aimed to achieve the following objectives: (1) use the partial least squares structural equation model (PLS-SEM) method to identify the determinates of blockchain adoption in the supply chain; (2) combine the fuzzy-set qualitative comparative analysis (fsQCA) approach and the PLS-SEM method to explore the synergistic effect among these determinants.

This study investigates these research questions from the lens of the technology-organisation-environment (TOE) framework, diffusion of innovation theory (DOI) and resource-based view (RBV) theory. In this background, this study builds upon Supply chain management (SCM) literature that uses the BCT, to identify the determinates of blockchain adoption in the supply chain. Then an integration of the PLS-SE and fsQCA methods was used to test the established theoretical mode.

2. Theoretical model

BCT is regarded as a capability that requires specific resources for its development. Drawing on the RBV, researchers have examined the necessary resources and capabilities for establishing and sustaining blockchain-based systems within supply chains. For example, Tsolakis et al. (2021) investigated the role of data interoperability as a resource in the fish supply chain, showing that blockchain-driven food supply chain management relies on effective data interoperability. Wamba and Queiroz (2022) investigated the relationship between top management support, technology competence, and blockchain adoption, finding that both top management support and technology competence positively influence adoption. Overall, RBV can help evaluate how resources are organized for BCT adoption and how they are integrated with organizational processes to create a BCT-driven supply chain capability.

Further, the TOE framework and DOI can explain critical factors for BCT adoption in the supply chain. According to current research, the TOE framework helps conceptualize how inter- and intra-organizational factors influence the effective deployment of blockchain in supply chain operations. Among the technological considerations are issues like technology immaturity, security (Xu et al., 2022) and standardization (Kurpjuweit et al., 2021); organisational: top management support (Deng et al., 2022), lack of technical competence and technology awareness (Samad et al., 2022); environmental: regulatory support, rivalry pressure (Chittipaka et al., 2022) and labour market availability (Xie et al., 2022). The TOE framework thus serves as a useful guide for understanding the diffusion of BCT and analysing how technology, organizational structures, and environmental forces interact in the adoption process.

Moreover, DOI was developed by Rogers (2010) to describe the process through which new ideas, practices or technologies are spread into a social system. DOI has been utilized to understand BCT formalization and diffusion in the supply chain and to explore how it spreads or results in rejection (Helliar et al., 2020). Based on DOI, Agi and Jha (2022) explored factors affecting blockchain adoption based on three characteristics of innovation—relative advantage, compatibility, and complexity—and concluded that compatibility and complexity are particularly influential, while relative advantage exerts a primary influence on supply chain adoption. DOI theory evaluates organizational innovation adoption by considering individual, internal, and external dimensions but does not extensively address environmental factors (Jayashree et al., 2021; Lu et al., 2021). Laaraj et al. (2022) propose a model that integrates DOI with the TOE framework, thereby encompassing the technological, organizational, and environmental contexts to broaden the understanding of factors affecting innovation adoption. Their findings highlight the significant impact of limited knowledge and ethical barriers on the intention to

adopt BCT. Overall, DOI theory can elucidate how blockchain is formalized and disseminated in supply chains, allowing researchers to analyse barriers and drivers of diffusion across different stages of the innovation-decision process. When coupled with the TOE framework, DOI provides deeper insights into the organizational, external, and technology-specific elements that influence BCT adoption within supply chains.

Based on the above analysis, a conceptual model for this study is proposed. It considers the factors of the organisation itself, external environment, technology, governance and people. These determinants do not operate in isolation; they interact and converge to shape a company's strategic direction. The framework includes 17 constructs, with blockchain adoption as the dependent variable. Figure 1 shows the proposed model.

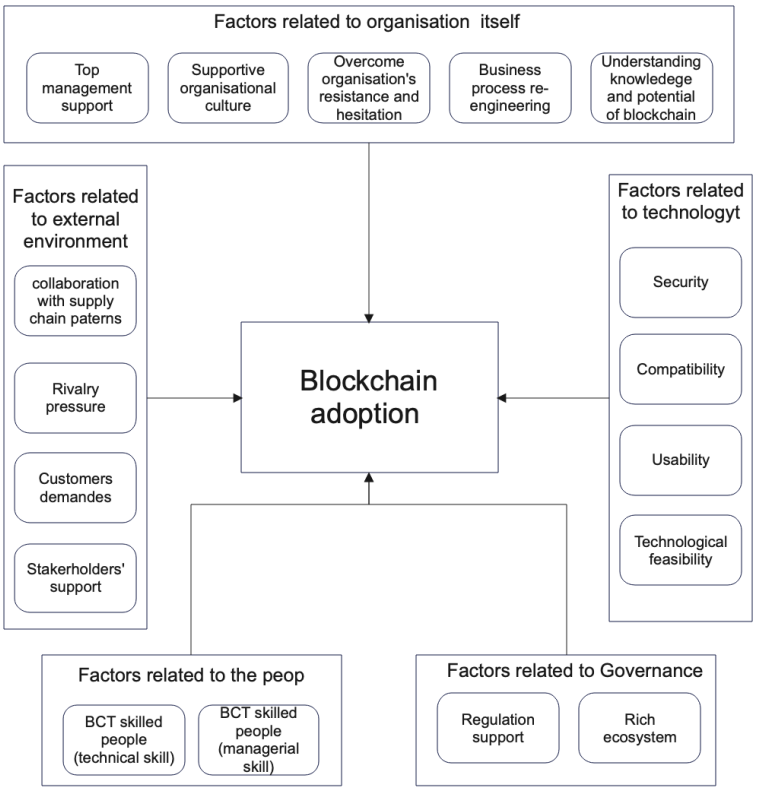


Figure 9: Theoretical framework

3. Research methodology

This study employed a mixed-methods approach to identify factors influencing blockchain adoption in the supply chain and to explore the synergistic and alternative interactions among these multiple tiers of factors (see Figure 2). To develop a framework for measuring BCT adoption, the research began with a literature review on blockchain technology and supply chain management to pinpoint the resources needed to construct a blockchain system. Subsequently, the linear relationships and configurational effects among variables were examined using two methods: PLS-SEM and fsQCA.

PLS-SEM is well-suited for assessing hierarchically structured data at multiple levels and can yield valid, reliable findings in complex models (Henseler and Sarstedt, 2013). Given the constructs identified in this study and the need to analyse interrelationships among various dependent and independent variables simultaneously, PLS-SEM was deemed an appropriate choice. Following this, the data were also analysed using fsQCA, which is a comprehensive technique that offers two main advantages. First, it recognizes the possibility of asymmetric causal relationships and allows for multiple solutions leading to the same outcome (Kaya et al., 2020). Second, fsQCA accommodates causal complexity, meaning that not all conditions need to be present to produce a particular result, and different combinations of

conditions can yield the same outcome (Cheng & Chong, 2022). By employing fsQCA alongside PLS-SEM, this study gained a more nuanced interpretation of its research questions

.

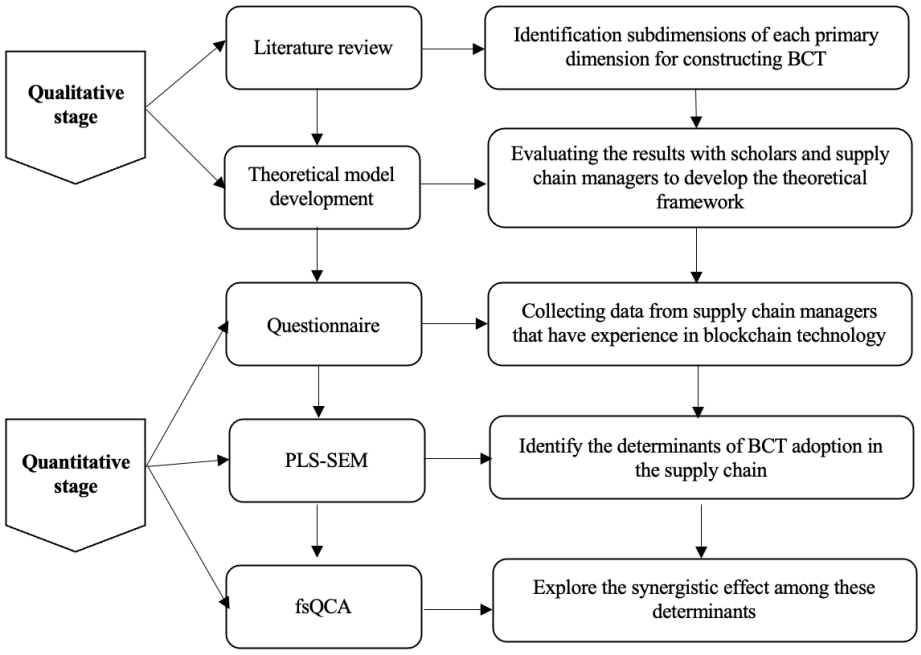


Figure 2: Research methodology framework

3.1 Empirical data and collection plan

This study employed both purposive and snowball sampling to identify potential respondents who would be willing and qualified to participate. Initially, purposive sampling was used to select individuals with the requisite knowledge and experience to complete the questionnaire. Two criteria guided this selection: (1) respondents had to have at least one year of professional experience related to blockchain technology and supply chain management, ensuring sufficient expertise, and (2) respondents needed to be at the middle or senior management level, ensuring a comprehensive understanding of supply chain operations. Following this, snowball sampling was utilized to expand the participant pool. Individuals who had already taken part in the study were invited to recommend other professionals who met the inclusion criteria. This approach not only increased the sample size but also helped reach a more diverse and informed group of respondents across different organizations and industries. The study aims to collect responses from approximately 150 to 200 participants, ensuring a robust dataset for both PLS-SEM and fsQCA analyses. Data will be collected primarily through an online questionnaire, which allows for efficient distribution and broad geographic reach. The questionnaire will include closed-ended items designed to measure the constructs identified in the research framework.

4. Intended contribution

This ongoing research is expected to contribute significantly to both theoretical and practical understanding of BCT adoption within supply chain management. From a theoretical perspective, this research aims to deepen insights into the complex, multi-level interactions among factors influencing BCT adoption. Through a comprehensive analysis of the PLS-SEM and fsQCA results, it offers deeper insights into the adoption process. Furthermore, this study extends the traditional TOE framework by incorporating additional dimensions—namely, people and governance—to offer a more holistic and integrated model of BCT adoption.

From a managerial perspective, the research seeks to provide practical guidance for organizations exploring or implementing blockchain solutions in their supply chains. By identifying key enablers and barriers to adoption, the study is expected to support supply chain managers in making more informed, strategic decisions. The findings aim to enhance organizational readiness, reduce implementation risks, and facilitate more effective adoption strategies aligned with both technological and governance structures.

Acknowledgements

The authors gratefully acknowledge the funding contribution of the Engineering and Physical Sciences Research Council (UK) to the Next Stage Digital Economy Centre in the Decentralized Digital Economy (DECaDE) [Grant Reference EP/T022485/1].

References

Agi, M. A., & Jha, A. K. (2022). Blockchain technology in the supply chain: An integrated theoretical perspective of organizational adoption. *International Journal of Production Economics*, *247*, 108458.

Bahga, A., & Madisetti, V. K. (2016). Blockchain platform for industrial internet of things. *Journal of Software Engineering and Applications*, *9*(10), 533-546.

Brookbanks, M., & Parry, G. (2022). The impact of a blockchain platform on trust in established relationships: a case study of wine supply chains. *Supply Chain Management: An International Journal*.

Browne, R. (2017). There were more than 26,000 new blockchain projects last year-only 8% are still active. *November*, *9*, 2017.

Chang, S. E., & Chen, Y. (2020). When blockchain meets supply chain: A systematic literature review on current development and potential applications. *Ieee Access*, *8*, 62478-62494.

Chang, Y., Iakovou, E., & Shi, W. (2020). Blockchain in global supply chains and cross border trade: a critical synthesis of the state-of-the-art, challenges and opportunities. *International Journal of Production Research*, *58*(7), 2082-2099.

Cheng, M., & Chong, H.-Y. (2022). Understanding the determinants of blockchain adoption in the engineering-construction industry: multi-stakeholders' analyses. *Ieee Access*, *10*, 108307-108319.

Chittipaka, V., Kumar, S., Sivarajah, U., Bowden, J. L.-H., & Baral, M. M. (2022). Blockchain Technology for Supply Chains operating in emerging markets: an empirical examination of technology-organization-environment (TOE) framework. *Annals of Operations Research*, 1-28.

Christidis, K., & Devetsikiotis, M. (2016). Blockchains and smart contracts for the internet of things. *Ieee Access*, *4*, 2292-2303.

Deng, N., Shi, Y., Wang, J., & Gaur, J. (2022). Testing the adoption of blockchain technology in supply chain management among MSMEs in China. *Annals of Operations Research*, 1-20.

Dutta, P., Choi, T.-M., Somani, S., & Butala, R. (2020). Blockchain technology in supply chain operations: Applications, challenges and research opportunities. *Transportation Research Part E: Logistics and Transportation Review*, *142*.

Gökalp, E., Gökalp, M. O., & Çoban, S. (2022). Blockchain-based supply chain management: understanding the determinants of adoption in the context of organizations. *Information Systems Management*, *39*(2), 100-121.

Helliar, C. V., Crawford, L., Rocca, L., Teodori, C., & Veneziani, M. (2020). Permissionless and permissioned blockchain diffusion. *International Journal of Information Management*, *54*, 102136.

Jabbar, S., Lloyd, H., Hammoudeh, M., Adebisi, B., & Raza, U. (2021). Blockchain-enabled supply chain: analysis, challenges, and future directions. *Multimedia systems*, *27*(4), 787-806.

Jayashree, S., Reza, M. N. H., Malarvizhi, C. A. N., & Mohiuddin, M. (2021). Industry 4.0 implementation and Triple Bottom Line sustainability: An empirical study on small and medium manufacturing firms. *Heliyon*, *7*(8), e07753.

Kaya, B., Abubakar, A. M., Behravesh, E., Yildiz, H., & Mert, I. S. (2020). Antecedents of innovative performance: Findings from PLS-SEM and fuzzy sets (fsQCA). *Journal of Business Research*, *114*, 278-289.

Kurpjuweit, S., Schmidt, C. G., Klöckner, M., & Wagner, S. M. (2021). Blockchain in additive manufacturing and its impact on supply chains. *Journal of Business Logistics*, *42*(1), 46-70.

Laaraj, M., Nakara, W. A., & Fosso Wamba, S. (2022). Blockchain diffusion: the role of consulting firms. *Production Planning & Control*, 1-13.

Lu, L., Liang, C., Gu, D., Ma, Y., Xie, Y., & Zhao, S. (2021). What advantages of blockchain affect its adoption in the elderly care industry? A study based on the technology–organisation–environment framework. *Technology in Society*, *67*, 101786.

Maull, R., Godsiff, P., Mulligan, C., Brown, A., & Kewell, B. (2017). Distributed ledger technology: Applications and implications. *Strategic Change*, *26*(5), 481-489.

Moosavi, J., Naeni, L. M., Fathollahi-Fard, A. M., & Fiore, U. (2021). Blockchain in supply chain management: a review, bibliometric, and network analysis. *Environmental Science and Pollution Research*, 1-15.

Pawczuk, L., Massey, R., & Holdowsky, J. (2019). Deloitte's 2019 Global Blockchain Survey-Blockchain gets down to business. *Deloitte Insights*, *6*.

Pournader, M., Shi, Y., Seuring, S., & Koh, S. L. (2020). Blockchain applications in supply chains, transport and logistics: a systematic review of the literature. *International Journal of Production Research*, *58*(7), 2063-2081.

Queiroz, M. M., Telles, R., & Bonilla, S. H. (2019). Blockchain and supply chain management integration: a systematic review of the literature. *Supply Chain Management: An International Journal*.

Rashid, A., Ali, S. B., Rasheed, R., Amirah, N. A., & Ngah, A. H. (2022). A paradigm of blockchain and supply chain performance: a mediated model using structural equation modeling. *Kybernetes*(ahead-of-print).

Rejeb, A., Rejeb, K., Simske, S., & Treiblmaier, H. (2021). Blockchain technologies in logistics and supply chain management: a bibliometric review. *Logistics*, *5*(4), 72.

Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.

Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain technology and its relationships to sustainable supply chain management. *International Journal of Production Research*, *57*(7), 2117-2135.

Samad, T. A., Sharma, R., Ganguly, K. K., Wamba, S. F., & Jain, G. (2022). Enablers to the adoption of blockchain technology in logistics supply chains: evidence from an emerging economy. *Annals of Operations Research*, 1-41.

Schmidt, C. G., & Wagner, S. M. (2019). Blockchain and supply chain relations: A transaction cost theory perspective. *Journal of Purchasing and Supply Management*, *25*(4), 100552.

Scott, B., Loonam, J., & Kumar, V. (2017). Exploring the rise of blockchain technology: Towards distributed collaborative organizations. *Strategic Change*, *26*(5), 423-428.

Sheel, A., & Nath, V. (2019). Effect of blockchain technology adoption on supply chain adaptability, agility, alignment and performance. *Management research review*.

Tsolakis, N., Niedenzu, D., Simonetto, M., Dora, M., & Kumar, M. (2021). Supply network design to address United Nations Sustainable Development Goals: A case study of blockchain implementation in Thai fish industry. *Journal of Business Research*, *131*, 495-519.

Wamba, S. F., & Queiroz, M. M. (2022). Industry 4.0 and the supply chain digitalisation: a blockchain diffusion perspective. *Production Planning & Control*, *33*(2-3), 193-210.

Wang, Y., Chen, C. H., & Zghari-Sales, A. (2021). Designing a blockchain enabled supply chain. *International Journal of Production Research*, *59*(5), 1450-1475.

Xie, S., Gong, Y., Kunc, M., Wen, Z., & Brown, S. (2022). The application of blockchain technology in the recycling chain: a state-of-the-art literature review and conceptual framework. *International Journal of Production Research*, 1-27.

Xu, X., Tatge, L., Xu, X., & Liu, Y. (2022). Blockchain applications in the supply chain management in German automotive industry. *Production Planning & Control*, 1-15.

Integrating Expert Systems in AI-Based Decision Support: A DEX Approach to Risk Assessment

Rok Drnovšek[1], Marija Milavec Kapun[2], Vladislav Rajkovič[3], and Uroš Rajkovič[4*]

[1] University Medical Centre Ljubljana
[2] University of Ljubljana, Faculty of Health Sciences
[3] University of Maribor, Faculty of Organizational Sciences
[4]University of Maribor
[*]Corresponding Author: uros.rajkovic@gmail.com

Abstract: Decision Support Systems have the potential to enhance risk management, particularly in healthcare, where uncertainty and complexity require structured decision-making approaches. This paper presents a multi-criteria decision support model based on the DEX methodology, an expert system approach that relies on symbolic AI principles. The model structures expert knowledge into a hierarchical set of criteria, enabling transparent and explainable risk assessment.

A key challenge in risk assessment is handling uncertainty, which arises both in probability estimation and consequence evaluation. To address this, we integrate fuzzy logic into the decision model, allowing for gradual and probabilistic categorizations instead of rigid classifications. The model was tested in a clinical setting to evaluate its applicability compared to traditional Risk Matrices. Results demonstrate that the DEX-based approach improves consistency, transparency, and adaptability in risk evaluation.

By leveraging expert systems within AI-based decision support, this research highlights the importance of structured expert knowledge in risk assessment. The findings suggest that symbolic AI methods remain highly relevant for decision-making in complex domains where explainability is crucial, such as healthcare.

Keywords: Decision Support Systems; Expert Systems; AI in Healthcare; DEX Methodology; Risk Assessment; Fuzzy Logic

1. Introduction

Risk management is an essential component of healthcare, where decision-making occurs in complex and uncertain environments with potentially severe consequences. Systematic risk assessment is crucial for ensuring patient safety, optimising resource allocation, and improving overall healthcare outcomes. However, conventional risk assessment tools, such as Risk Matrices, are often criticised for their subjectivity, lack of transparency, and inability to handle uncertainty effectively (Ferdous et al., 2011; Heikkilä et al., 2024). These limitations necessitate the development of more structured and knowledge-driven decision support systems (DSS) to improve risk evaluation in clinical practice.

DSS have the potential to enhance risk assessment by providing structured frameworks for decision-making, yet their application in healthcare remains underdeveloped. Traditional DSS approaches often rely on numerical scoring methods, which may not adequately capture expert-driven qualitative knowledge that is essential in complex risk evaluations. Expert systems, a subset of artificial

intelligence (AI), offer an alternative by structuring expert knowledge into decision-making models. DEX methodology is a well-established multi-criteria decision analysis (MCDA) approach, which represents expert knowledge using hierarchically organised qualitative criteria and rule-based decision-making (Bohanec et al., 2013; Mihelčič & Bohanec, 2016).

This paper presents a DEX-based decision support model for healthcare risk assessment, designed to improve existing risk evaluation methods by:

- Structuring expert knowledge into a hierarchical set of criteria, allowing for a more systematic and transparent assessment of risk factors.
- Addressing uncertainty in risk assessment by integrating fuzzy logic, which enables the representation of probabilistic uncertainty in risk estimation (Chen et al., 2014; Chicco et al., 2023).
- Comparing the proposed model with conventional risk assessment tools, such as Risk Matrices, through practical application in a clinical setting.

The study employs an expert-driven methodology for developing and validating the proposed model, using the DEXi software, which supports hierarchical decision models and facilitates rule-based decision-making (Boshkoska et al., 2020). The model was tested in a healthcare environment, where experts assessed multiple clinical risks, including patient falls, pressure injuries, workplace injuries, and burnout among healthcare professionals.

Findings demonstrate that the DEX-based approach enhances decision-making by providing a more structured, transparent, and adaptable framework for risk assessment compared to traditional methods. Unlike conventional risk matrices, which rely on rigid categorisations, the proposed model enables a more flexible and evidence-driven evaluation by incorporating symbolic AI techniques (Zadeh, 1996; Özkan & Türkşen, 2014). The results highlight the continued relevance of expert systems within AI-driven decision support and suggest their broader applicability in risk-sensitive domains beyond healthcare.

2. Methodology

The development of a structured and transparent risk assessment model requires a decision-support approach that effectively integrates expert knowledge while addressing uncertainty in risk evaluation. This study utilises the DEX methodology, a multi-criteria decision analysis approach, to construct a hierarchical risk assessment model. The methodology consists of the following key steps: model development, uncertainty management, and validation in a healthcare setting.

2.1 Development of the DEX-Based Risk Assessment Model

The proposed model was built using the DEX methodology, which structures expert knowledge into qualitative hierarchical tree of criteria (Bohanec et al., 2013). The model consists of basic and aggregated criteria, where:

- Basic criteria represent independent attributes of the risk being assessed, and
- Aggregated criteria are hierarchical groupings that integrate multiple lower-level criteria using predefined decision rules.

The evaluation process follows a rule-based system, where decision rules are manually defined by with specialised domain expertise. These rules determine how different combinations of basic criteria contribute to higher-level risk assessments, ultimately leading to a final risk classification (Mihelčič & Bohanec, 2016).

To ensure model consistency, we used DEXi software (Boshkoska et al., 2020), which provides a structured environment for model construction and evaluation. The software supports hierarchical decision models and allows for easy integration of expert-defined rules.

2.2 Addressing Uncertainty in Risk Assessment

A key challenge in healthcare risk assessment is uncertainty, which arises from limited data availability, subjective expert judgments, and variability in risk consequences (Ferdous et al., 2011). To address this issue, we integrated fuzzy logic into the model, allowing for a more nuanced risk classification.

Uncertainty in probability estimation: Instead of using rigid probability categories, the model allows for graded probability levels, reducing the impact of subjective bias.

Uncertainty in consequence evaluation: Traditional methods often oversimplify consequences by categorising them into a few rigid risk levels. To overcome this, we use fuzzy membership functions to represent different severity levels, ensuring a more flexible classification (Chicco et al., 2023).

2.3 Validation in a Healthcare Setting

To assess the practical applicability of the model, we conducted a comparative evaluation with traditional Risk Matrices in a clinical setting. The model was tested in a large healthcare institution in Slovenia, where healthcare professionals assessed real-world risks, including patient falls, pressure injuries, workplace injuries, and burnout among healthcare professionals.

Participants evaluated risks using both the traditional Risk Matrix and the proposed DEX-based model, allowing for a direct comparison of assessment consistency, usability, and decision transparency.

Feedback from healthcare professionals was collected through structured interviews and think-aloud protocols, ensuring qualitative validation of the model's usability. The results were analysed to determine whether the DEX approach provided more structured, transparent, and reliable risk assessments compared to conventional methods.

3. Implementation: A Case Study

The developed DEX-based risk assessment model was implemented and tested in a clinical healthcare setting to evaluate its applicability in real-world risk assessment. The goal of the case study was to compare the DEX approach with traditional risk assessment methods, focusing on consistency, transparency, and usability.

3.1 Application of the Model in Healthcare Risk Assessment

The proposed model was applied to assess various clinical risks that are commonly encountered in healthcare environments. The selection of risks was based on expert consultations and an analysis of historical risk reports within the chosen institution. The following four risk categories were assessed:

- Patient falls – a common patient safety concern with varying degrees of severity (Heikkilä et al., 2024).
- Pressure injuries – preventable complications that require structured monitoring and intervention.
- Workplace injuries – occupational hazards affecting healthcare professionals.
- Burnout among healthcare professionals – a long-term risk with both individual and organisational consequences.

For each risk category, the DEX model was used to classify risks based on hierarchically structured criteria, and its outputs were compared to assessments conducted using a traditional Risk Matrix.

3.2 Evaluation Process and Comparison with Risk Matrices

To ensure a systematic comparison, healthcare professionals were asked to assess the selected risks using: (1) a traditional Risk Matrix, where risks are classified based on a predefined likelihood-consequence grid, and (2) the DEX risk assessment model, where risks are evaluated through structured criteria and expert-defined rules.

Each participant was provided with historical data on risk occurrences and was asked to rate the risks independently using both methods. The following aspects were evaluated:

- Consistency of risk classification – whether similar risks were assigned similar severity levels.
- Handling of uncertainty – how well each method accounted for incomplete or ambiguous data (Ferdous et al., 2011).
- Transparency and usability – participants' perceptions of how well the model supported decision-making.

3.3 Experts' Feedback and Usability Assessment

To validate the model's usability, structured interviews were conducted with the participants. The think-aloud method was used during the assessment process to capture real-time feedback on the decision-making experience.

Experts valued the structured nature of the DEX model, which provided a more transparent justification for risk classifications compared to the traditional Risk Matrix.

The model's ability to integrate multiple criteria was identified as a major advantage, particularly in cases where traditional methods failed to differentiate risks with similar likelihood scores but different contextual factors.

Uncertainty handling was perceived as more intuitive in the DEX model, as it allowed for gradual classifications rather than rigid category boundaries (Chicco et al., 2023).

Overall, the case study confirmed that the DEX-based approach enhances risk assessment by providing a more systematic approach and evidence-based evaluation. The results suggest that expert systems remain a relevant AI-driven solution for structured decision support in healthcare.

4. Discussion

The application of the DEX-based risk assessment model in a clinical setting provided important insights into its effectiveness and advantages over traditional risk assessment methods. The comparison with the Risk Matrix revealed that the DEX model ensures greater consistency in risk classification, as risks with similar characteristics were assigned more uniform assessments. Unlike the Risk Matrix, where subjective interpretation often led to variations, the rule-based structure of the DEX model provided a more systematic approach to decision-making.

One of the key benefits observed was the enhanced transparency of risk evaluation. The hierarchical organisation of qualitative criteria allowed experts to trace how risk levels were determined, whereas the Risk Matrix lacked explicit reasoning behind category assignments. Additionally, the DEX model facilitated a more nuanced risk differentiation, considering multiple influencing factors rather than relying solely on a likelihood-consequence pairing.

A major limitation of traditional risk assessment methods is their inability to address uncertainty effectively (Ferdous et al., 2011). The DEX model mitigated this issue through fuzzy logic, which allowed for gradual classifications rather than rigid categories. This feature was particularly beneficial in assessing risks such as burnout, where data availability was inconsistent, and conventional methods struggled to differentiate between low- and high-risk cases.

Healthcare professionals who participated in the study provided positive feedback on the model's

usability. They highlighted that the structured nature of the model aligned well with clinical reasoning, making risk assessment more intuitive and justifiable. Unlike the Risk Matrix, which often led to simplified or ambiguous risk classifications, the DEX model enabled users to understand and explain the rationale behind each decision. However, some participants noted that the initial learning curve for using the model was steeper than that of the Risk Matrix, suggesting the need for training and support tools to facilitate adoption.

These findings reinforce the ongoing relevance of expert systems in AI-driven decision support, particularly in domains where explainability and structured reasoning are crucial (Chicco et al., 2023; Zadeh, 1996). Unlike black-box machine learning models, expert systems like DEX provide structured, transparent decision processes that align with expert knowledge and are adaptable to various clinical settings. The study suggests that hybrid AI approaches, where symbolic AI methods complement data-driven techniques, could further enhance decision support in healthcare.

5. Conclusions

This study presents a DEX-based decision support model for risk assessment in healthcare, addressing key limitations of traditional methods such as Risk Matrices. The results demonstrate that the structured, rule-based approach of the DEX methodology improves consistency, transparency, and adaptability in risk evaluation. Unlike conventional methods that rely on rigid probability-consequence pairings, the proposed model integrates hierarchical qualitative criteria, ensuring a more systematic and explainable risk classification.

A major advantage of the model is its ability to handle uncertainty, achieved by incorporating fuzzy logic, which allows for gradual classifications rather than binary risk categories. This feature proved particularly useful in assessing risks with incomplete data or subjective expert evaluations, such as burnout and workplace safety risks. Expert feedback further confirmed that the DEX model provides better justification for risk ratings, supporting more informed decision-making in healthcare risk management.

However, the study also identified practical challenges associated with implementing the model in real-world settings. While participants found the model intuitive and aligned with clinical reasoning, some noted that the initial learning curve was higher compared to simpler tools like the Risk Matrix. This suggests that training and user support tools will be essential for wider adoption.

The findings highlight the continued relevance of expert systems in AI-driven decision support, particularly in fields where explainability and structured decision-making are critical. Future research should explore integration with data-driven approaches, such as machine learning models, to enhance the adaptability of decision rules while maintaining the interpretability of expert-driven assessments. Additionally, testing the model across different healthcare institutions and various risk domains would provide further validation and insights into its scalability.

In conclusion, the DEX-based risk assessment model represents a promising approach for improving decision support in healthcare, offering a transparent, structured, and adaptable framework for risk evaluation. The study reinforces the importance of symbolic AI methods in risk-sensitive environments and provides a foundation for further research into hybrid AI solutions that balance expert knowledge with data-driven insights.

References

Bohanec, M., Žnidaršič, M., Rajkovič, V., Bratko, I., & Zupan, B. (2013). DEX methodology: Three decades of qualitative multi-attribute modeling. Informatica, 37(1), 49–54.

Boshkoska, B. M., Miljković, D., Valmarska, A., Gatsios, D., Rigas, G., Konitsiotis, S., Tsiouris, K. M., Fotiadis, D. I., & Bohanec, M. (2020). Decision support for medication change of Parkinson's disease patients. Computer Methods and Programs in Biomedicine, 196, 105552. https://doi.org/10.1016/j.cmpb.2020.105552

Chen, S. Y., Ross, B. H., & Murphy, G. L. (2014). Decision making under uncertain categorization. Frontiers in Psychology, 5(AUG), 89567. https://doi.org/10.3389/FPSYG.2014.00991/ABSTRACT

Chicco, D., Spolaor, S., & Nobile, M. S. (2023). Ten quick tips for fuzzy logic modeling of biomedical systems. PLOS Computational Biology, 19(12). https://doi.org/10.1371/JOURNAL.PCBI.1011700

Ferdous, R., Khan, F., Sadiq, R., Amyotte, P., & Veitch, B. (2011). Fault and event tree analyses for process systems risk analysis: uncertainty handling formulations. Risk Analysis : An Official Publication of the Society for Risk Analysis, 31(1), 86–107. https://doi.org/10.1111/J.1539-6924.2010.01475.X

Heikkilä, A., Lehtonen, L., & Junttila, K. (2024). Consequences of Inpatient Falls in Acute Care—A Retrospective Register Study. Journal of Patient Safety. https://doi.org/10.1097/PTS.0000000000001230

Mihelčič, M., & Bohanec, M. (2017). Approximating incompletely defined utility functions of qualitative multi-criteria modeling method DEX. Central European Journal of Operations Research, 25(3), 627–649. https://doi.org/10.1007/s10100-016-0451-x

Özkan, İ., & Türksen, İ. B. (2014). Uncertainty and fuzzy decisions. V S. Banerjee & S. S. Erçetin (ur.), Chaos Theory in Politics (str. 17–27). Springer. https://doi.org/10.1007/978-94-017-8691-1_2

Zadeh, L. A. (1996). Fuzzy logic = computing with words. IEEE Transactions on Fuzzy Systems, 4(2), 103–111. https://doi.org/10.1109/91.493904

Public Sector, Education, and Administration

Pathways to Peace: Applying Decision EXpert Methodology in the Balkans

Sandro Radovanović[1*], Nemanja Džuverović[2], Goran Tepšić[2] and Đorđe Krivokapić[2]

[1] University of Belgrade, Faculty of Organizational Sciences, Belgrade, Serbia
[2] University of Belgrade, Faculty of Political Sciences, Belgrade, Serbia
[*] sandro.radovanovic@fon.bg.ac.rs

Abstract: This paper discusses the development and application of the Balkan Peace Index, a novel metric designed to measure peacefulness across seven Western Balkan countries. This index diverges from traditional peace indices by integrating both quantitative and qualitative data through a synthesis of Decision EXpert methodology and ethnographic approaches, in line with the "local turn" in peace and conflict studies. This method emphasizes local involvement and contextual knowledge, offering a culturally and contextually informed perspective of regional peace. The paper outlines the theoretical underpinnings and practical application of the index, highlighting its aim to provide actionable insights for policymakers and enrich peace research methodology. The index is presented transparently, with all data, models, and results publicly accessible through a dedicated web application, encouraging external validation and broad usage.

Keywords: Peace Index; Peace Classification Method; Decision EXpert; Sensitivity Analysis; Open Science

1. Introduction

This paper presents the experiences and design choices involved in the development and application of the Balkan Peace Index (BPI), a comprehensive metric designed to measure peacefulness across seven Western Balkan nations and territories: Albania, Bosnia and Herzegovina, Croatia, Kosovo, Montenegro, North Macedonia, and Serbia. In response to the local turn in critical peace and conflict studies (Mac Ginty & Richmond, 2013; Džuverović, 2021), which emphasizes the importance of local involvement in creating knowledge in post-conflict settings, the BPI combines decision support systems with ethnographic methods (Ljungkvist & Jarstad, 2021). This methodology diverges from traditional indices like the Global Peace Index, which predominantly utilize quantitative data and present scores as rankings. Instead, the BPI employs the Decision EXpert methodology (Bohanec, 2021), which leverages both quantitative and qualitative data to incorporate opinions and thoughts from local populations, aiming to offer a comprehensive, culturally, and contextually informed picture of peace in the region.

The BPI's methodology reflects a synthesis of globally recognized indexing techniques and in-depth, localized engagement strategies. By incorporating direct feedback from the communities it assesses, the index gains a unique depth, making it a highly relevant and practical tool for policymakers and researchers alike (Firchow, 2020; Bohanec, 2021). Fieldwork conducted by researchers fluent in the local languages and culturally attuned to the region ensures that the data collection processes capture not only the statistical dimensions of peace but also the subjective experiences and perceptions of the local populace. This comprehensive approach allows the BPI to address the complex dynamics of peace and conflict in the Balkans more effectively, presenting a balanced view that leverages both established research methodologies and innovative, participatory techniques.

In terms of decision support systems, the proposed BPI satisfies several critical criteria: it is *accurate*, as it is developed by local experts; it is *complete*, providing feedback on any given combination of input data; it is *consistent* in reasoning, reducing the possible logical fallacies human decision-makers may have (Maier et al., 2011); it is *comprehensive*, with seven domains measuring both positive and negative aspects of peace; and it is *convenient to use*, with developed web applications allowing users to experiment and see potential outcomes (Keeney, 2004; Jonassen, 2012).

A fundamental idea of this paper is to share the process of building the BPI, specifically the application of the Decision EXpert method to obtain and disseminate the peace index results. By doing so, we aim to facilitate additional validation of our findings, enable their integration into diverse research efforts, and provide a valuable resource for decision-makers and policymakers. This commitment to transparency and accessibility underscores our belief in the collaborative nature of scientific inquiry, encouraging the broader research community to scrutinize, build upon, and apply our work in various contexts.

2. Background

Over the past two decades, peacebuilding has become a vital component of global governance, attracting substantial investments from intergovernmental organizations, non-governmental entities, and various donors. These entities annually invest billions of dollars and deploy extensive personnel to support peace initiatives. Despite these efforts, and a significant expenditure of $6.45 billion by the United Nations on peacekeeping in 2012 alone, the results have often been mixed (Caplan, 2020). According to recent studies (Gledhill et al., 2021; de Coning, 2022), one significant challenge hindering successful peacebuilding is the lack of effective tools to measure progress toward sustained peace. Current assessments tend to be ad-hoc and subjective, lacking clear criteria and often resulting in selective interpretations. This gap highlights the need for strategic, comprehensive evaluations to guide peacebuilding efforts effectively, which the BPI aims to address.

However, the most challenging issue in developing the BPI is defining peace. This is inherently complex due to the multidimensional nature of peace. One compelling perspective is that peace extends beyond the absence of conflict to include active societal integration that prevents violence (Ide & Mello, 2022), as well as elements such as economic stability and political freedom (Mross et al., 2022; Hirblinger et al., 2023). Each of these dimensions uniquely contributes to the sense of peace and peacefulness experienced by individuals. In post-conflict scenarios, such as in the Balkans, the factors influencing peace and conflict are intricate and varied. For instance, while democracy is often seen as a goal in peacebuilding, its process can increase the likelihood of conflict, suggesting a need for a comprehensive understanding of peace dynamics that goes beyond traditional frameworks (Džuverović, 2014; Stewart, 2016). Thus, developing comprehensive peace indicators for the BPI is indeed a challenge. Additionally, the concept of peace is fundamentally contested, with major international actors like the UN, NATO, EU, OSCE, and the World Bank generally aligning with the liberal paradigm of peace (Acharya, 2014; Hauenstein & Joshi, 2020). This paradigm encompasses aspects such as security provision, reconstruction, strengthening of political institutions, economic and social development, and the protection of human rights.

In response to these complexities, our paper proposes a peace classification method that seeks to provide policymakers with a detailed understanding of peace on a continuum, rather than as a binary outcome, ranking, or numerical score. A peace classification system offers a nuanced understanding by categorizing different aspects or types of peace, such as the absence of violence (negative peace) and the presence of justice and equality (positive peace) (Mustafa et al., 2023; van Iterson Scholten, 2024). This approach reflects the complexity of peace more informatively than binary outcomes or numerical scores. Furthermore, peace classifications can provide actionable insights by indicating specific areas that need attention, guiding targeted policy interventions more effectively than rankings, which only show relative positions without explaining underlying reasons. They are also less prone to misinterpretation compared to rankings and scores, where minor differences might be viewed as significant disparities (Bohanec, 2021). Moreover, classification systems can be easily updated or revised to reflect new insights, maintaining their relevance and utility in policy discussions (Bui, 2000). This approach for the BPI aims to facilitate comparisons between countries, foster international dialogue, encourage cooperative strategies for achieving and sustaining peace, and consequently make them invaluable for policymakers.

Finally, we advocate for a transparent approach, planning to share our methodology and results openly through a website, encouraging public scrutiny and application (Gebru et al., 2022). This initiative aims to foster broader engagement and allow stakeholders to evaluate their own countries and

simulate potential outcomes, thus supporting informed decision-making and promoting a deeper, more contextual understanding of peace.

3. Methodology

In developing the BPI, our methodology integrates both qualitative and quantitative metrics, addressing the challenge of harmonizing diverse measures of peace into a coherent scoring system useful for informed decision-making. Unlike conventional indices that reduce complex assessments to a single numerical score, our approach categorizes Western Balkan countries into distinct, ordered classifications. This strategic deviation allows for an interpretation of results along the peace-violence continuum, offering policymakers not just a score but a contextualized understanding and specific, actionable recommendations. By looking beyond mere numbers to understand the implications, we ensure that comparisons reflect not just statistical differences but also the realities of each society. Additionally, our methodology is tailored to the unique historical and current dynamics of the Balkan region. With a past marked by intense ethnic conflicts and nationalism, notably during the breakup of Yugoslavia in the 1990s, the Balkans present complex challenges that require a sensitive and informed approach to peace measurement. This context demands a methodology that not only captures the current state of peace but also respects the deep-seated historical grievances and ongoing transitional challenges within the region. Therefore, we have employed the Decision Expert methodology (Bohanec, 2021) in constructing the Balkan Peace Index, ensuring that our approach is as comprehensive and region-specific as possible.

The Decision EXpert (DEX) methodology is a qualitative multi-criteria decision analysis tool selected to develop the BPI due to its ability to handle complex decisions involving multiple and potentially conflicting attributes (Bohanec, 2021). DEX simplifies decision-making through a hierarchical decomposition approach, breaking down complex problems into manageable subproblems. This method is particularly beneficial for categorizing countries along a peace/violence continuum by using a hierarchy of attributes that define peace and violence, thus facilitating a more nuanced understanding of these concepts. DEX utilizes IF-THEN decision rules to aggregate lower-level attributes into higher-level outcomes, ensuring a systematic approach to decision-making. The effectiveness of the DEX methodology in the Balkan Peace Index meets the '5C requirements'—Correctness, Completeness, Consistency, Comprehensiveness, and Convenience—which ensure the model's accuracy, relevance, and utility. The model's design also emphasizes traceability, allowing decision-makers to understand and verify the decision-making process from the basic attributes to the final classification along the peace/violence continuum. Moreover, the capacity for sensitivity analysis within the DEX framework offers a dynamic tool for assessing how changes in individual attributes affect the overall classification. This feature is crucial for enabling policymakers to make informed decisions and adjust strategies based on comprehensive and reliable data analysis, thereby enhancing the overall effectiveness of the peace/violence assessment process (Radovanović et al., 2023; Delibašić et al., 2023). The reason why the DEX methodology has advantage over other multi-criteria decision-making methods for the problem at hand is in easily handling both qualitative and quantitative data. Further, DEX uses a clear hierarchical structure, breaking down problems into smaller parts and using simple IF-THEN rules to make decisions. This makes the process transparent and easy to follow. More importantly, the method is user-friendly and visually intuitive, making it easier to communicate and involve stakeholders (Bohanec, 2021; Delibašić et al., 2023).

Formally, a DEX model M is a four-tuple $M = \{X, D, S, F\}$, where X is the set of attributes, S is the descendant function that determines the hierarchical structure of attributes, D is the set of value scales of attributes in X (domain of X), and F the set of aggregation functions. The set X contains n attributes: $X = \{x_1, x_2, \dots, x_n\}$. Attributes represent observable properties of the decision problem and decision alternatives. In DEX models, attributes are typically given unique and meaningful names. The structure is defined by the function $S : X -> 2^X$, which associates each $x \in X$ with a set of its descendants $S(x)$ in the hierarchy. The relationship between an attribute and its descendants indicates both dependence and influence. Attributes without parents are called roots and represent the main outputs of the model, while attributes without descendants, $S(x) = \emptyset$, are called basic attributes and represent model inputs.

Attributes with $S(x) \neq \emptyset$ are referred to as aggregate attributes and are considered partial, lower-level outputs of the model. DEX models are most often structured as trees, where all attributes, except a single root attribute, have exactly one parent. Each attribute $x \in X$ associated with a value scale $D_x \in D$, defined as an ordered set of symbolic (qualitative) values. The fourth component of the DEX model is $F = \{f_x, x \in X\}$, a set of aggregation functions used to evaluate aggregate attributes based on the values of their immediate descendants. Aggregation functions are represented by decision tables that define the function value for each combination of argument values, ensuring a systematic approach to decision-making in the form of IF-THEN rules. Alternatives $A = \{A_1, A_2, \ldots, A_q\}$ are considered external data objects processed by the model. Each alternative A_i is represented by a set of values $A_i = \{a_{(x,i)} \in D_x, \forall x \in X\}$, where each $a_{x,i}$ represents the value of A_i assigned to attribute X. Values of $a_{x,i}$ are propagated from the bottom of the hierarchy using aggregation function in F to the root of the hierarchy, which represents the outcome of the decision process. (Bohanec, 2021; Radovanović et al., 2023)

4. Results and Discussion

Rather than ranking countries, the BPI positions them along a peace continuum that includes five stages: Violent Conflict, Contested Peace, Polarised Peace, Stable Peace, and Consolidated Peace (Figure 1). This method assesses the overall quality of peace in each country based on inputs from local researchers and populations, ensuring a rich and detailed understanding of peace dynamics. For example, a survey conducted in Serbia involved 1,213 respondents, representing a balanced mix of gender, rural and urban dwellers, educational levels, and age groups.

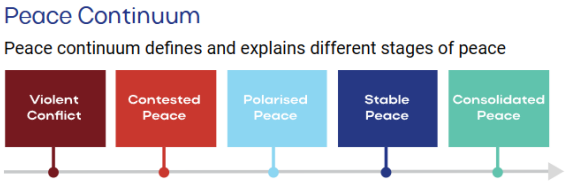


Figure 10. BPI Peace Continuum (MIND - Balkan Peace Index, 2024)

The BPI model, developed through extensive consultation with domain experts, interviews, and focus groups, comprises seven domains each containing multiple sub-indicators, as presented in Figure 3. The model distinguishes between *negative peace*, represented by the domains of Political Violence and Fighting Crime, and *positive peace*, which includes Regional and International Relations, State Capacity, Socio-Economic Development, Political Pluralism, and Environmental Sustainability. These domains were shaped by the insights of local researchers in the Balkans, reflecting a deep understanding of the regional factors that are crucial for maintaining peace.

From a global perspective, the region is seen as peaceful, free from armed conflicts for over two decades. However, European views highlight ongoing tensions and potential conflicts (Džuverović, 2021), influenced by the legacy of the 1990s wars and ongoing political and ethnic tensions. In 2022, Kosovo experienced a violent crisis, distinguishing it along with Bosnia and Herzegovina, which also faces political instability, as having *contested peace* (Newman & Visoka, 2024). Meanwhile, Serbia and Montenegro are categorized under *polarized peace*, North Macedonia under *stable peace*, and Albania and Croatia as *consolidated peace*.

The situation is very similar in 2023. Albania and Croatia are regarded to belong to a *consolidated peace* state as these countries are developing without major conflicts with their neighbours and on the worldwide level. North Macedonia is under the *stable peace*, as well as Montenegro. The improvement of Montenegro is attributed to improvements in political polarization. Kosovo and Bosnia and Herzegovina remained in *contested peace*, but Serbia joined them due to worsening the freedom of expression and media indicator.

To present a glimpse of what is possible to achieve with the DEX model for BPI, we present what-if analysis for Serbia in 2022 in Figure 2. On the left panel one can see what indicators lead to “left” on the peace continuum scale – leading to worse outcome. Those factors are indicators one needs to consider not to make country less safe. Similarly, on the right panel one can find indicators leading to “right” on the peace continuum – leading to more safe and secure country.

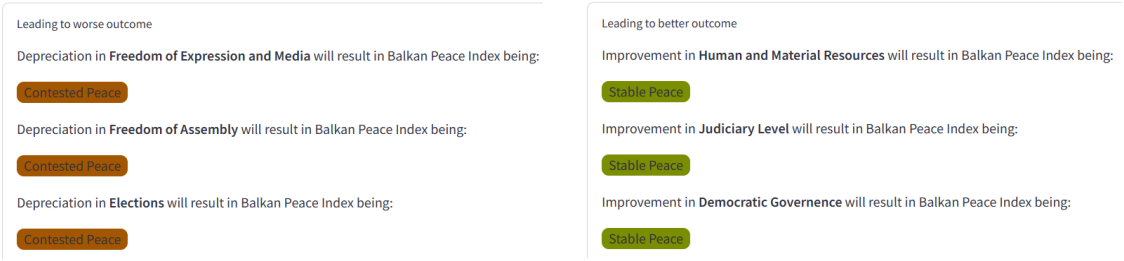


Figure 11. What-if analysis for Serbia in 2022

The BPI website (MIND Balkan Peace Index Web Application, 2024) offers a detailed analysis of peace in the Balkans, categorizing countries based on a range of domains like Political Violence, Fighting Crime, and Environmental Sustainability, among others. Political crises and ethnic tensions, especially in Kosovo and Bosnia and Herzegovina, contribute to ongoing instability, yet are unlikely to escalate into war due to international peacekeeping efforts. Challenges in fighting crime, particularly in terms of organized and state-sponsored crime, along with issues in environmental sustainability exacerbated by climate change, significantly impact the region's peace and development. Other domains such as Regional and International Relations and State Capacity show varying levels of success and challenge across the region, affecting overall peace quality. In addition, the data is made available open and free (Džuverović et al., 2025).

5. Conclusions

This paper aims to present the Balkan Peace Index and its development ideas and design choices. It aims to advance the measurement of peace beyond traditional numerical rankings by introducing a classification system that reflects the complexities of peace across a continuum. Grounded in the local turn in peace research, the Balkan Peace Index combines expert analysis with local insights and decision support systems to provide a detailed perspective on each country's peace status, offering actionable insights for improvement and identifying potential pitfalls. The index is designed to be transparent, with all data, models, and results made publicly available on a dedicated website (MIND Balkan Peace Index Web Application, 2024). This openness facilitates external scrutiny and allows users to evaluate countries, simulate potential outcomes, and perform sensitivity analysis. This will hopefully enhance the applicability of the developed index across different contexts.

However, the BPI faces several limitations that could affect its effectiveness and scope. Currently, the index relies on two years, which limits its ability to analyze trends over time and correlate events with changes in peace classifications. Future efforts will focus on developing a multi-year dataset to enrich the analysis and provide more robust insights. Additionally, the index's methodology is highly localized, tailored specifically to the Western Balkans, which restricts its generalizability to other regions. This localization is intentional, based on the theoretical approach that emphasizes region-specific, contextually appropriate methods for evaluating peace. Despite these limitations, there is optimism for positive community engagement and the potential for developing a more comprehensive Peace Index using similar methodologies.

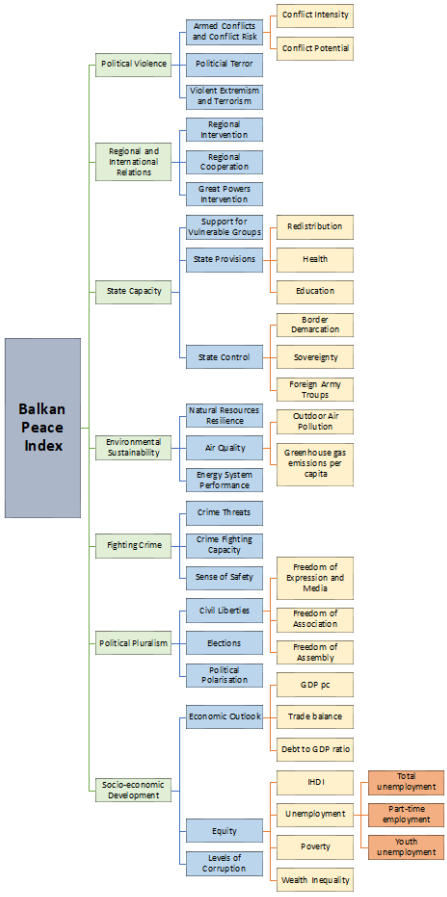


Figure 12. Balkan Peace Index

Acknowledgement

This research was supported by the Science Fund of the Republic of Serbia (grant number 7744512) Monitoring and Indexing Peace and Security in the Western Balkans – MIND.

References

1. Acharya, A. (2014). Global international relations (IR) and regional worlds: A new agenda for international studies. *International Studies Quarterly, 58*(4), 647-659.
2. Bohanec, M. (2021). From data and models to decision support systems: Lessons and advice for the future. In *EURO Working Group on DSS: A tour of the DSS developments over the last 30 years* (pp. 191-211). Cham: Springer International Publishing.
3. Bui, T. X. (2000). Decision support systems for sustainable development: An overview. *Decision Support Systems for Sustainable Development: a Resource Book of Methods and Applications*, 1-10.
4. Caplan, R. (2020). Measuring Peace: Principles, Practices, and Politics. *Ethnopolitics, 19*(3), 311-315.
5. de Coning, C. (2022). Insights from complexity theory for peace and conflict studies. In *The Palgrave Encyclopedia of Peace and Conflict Studies* (pp. 600-610). Cham: Springer International Publishing.
6. Delibašić, B., Radovanović, S., & Vukanović, S. (2023). A Decision Support System for Internal Migration Policy-Making.

7. Džuverović, N. (2014). Inequality–Conflict Research Beyond Neo-Liberal Discourse. *Peace Review, 26*(4), 547-555.
8. Džuverović, N. (2021). 'To romanticise or not to romanticise the local': local agency and peacebuilding in the Balkans. *Conflict, Security & Development, 21*(1), 21-41.
9. Džuverović, N., Radovanović, S., Tepšić, G., & Krivokapić, Đ. (2025). Balkan Peace Index decision EXpert model and data. *Data in Brief, 58*, 111181.
10. Firchow, P. (2020). World peace is local peace. *Ethics & International Affairs, 34*(1), 57-65.
11. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., & Crawford, K. (2022). Excerpt from datasheets for datasets. In *Ethics of Data and Analytics* (pp. 148-156). Auerbach Publications.
12. Gledhill, J., Caplan, R., & Meiske, M. (2021). Developing peace: the evolution of development goals and activities in United Nations peacekeeping. *Oxford Development Studies, 49*(3), 201-229.
13. Hauenstein, M., & Joshi, M. (2020). Remaining seized of the matter: Un resolutions and peace implementation. *International Studies Quarterly, 64*(4), 834-844.
14. Hirblinger, A. T., Hansen, J. M., Hoelscher, K., Kolås, Å., Lidén, K., & Martins, B. O. (2023). Digital peacebuilding: A framework for critical–reflexive engagement. *International Studies Perspectives, 24*(3), 265-284.
15. Ide, T., & Mello, P. A. (2022). QCA in international relations: A review of strengths, pitfalls, and empirical applications. *International Studies Review, 24*(1), viac008.
16. Jonassen, D. H. (2012). Designing for decision making. *Educational technology research and development, 60*, 341-359.
17. Keeney, R. L. (2004). Making better decision makers. *Decision analysis, 1*(4), 193-204.
18. Ljungkvist, K., & Jarstad, A. (2021). Revisiting the local turn in peacebuilding–through the emerging urban approach. *Third World Quarterly, 42*(10), 2209-2226.
19. Mac Ginty, R., & Richmond, O. P. (2013). The local turn in peace building: a critical agenda for peace. *Third World Quarterly, 34*(5), 763-783.
20. Maier, B., Shibles, W. A., Maier, B., & Shibles, W. A. (2011). Decision-Making: Fallacies and Other Mistakes. The Philosophy and Practice of Medicine and Bioethics: *A Naturalistic-Humanistic Approach*, 23-44.
21. *MIND - Balkan Peace Index*. (2025). MIND - Balkan Peace Index. https://bpi.mindproject.ac.rs/. Accessed on January 13, 2025.
22. *MIND Balkan Peace Index Web Application*. (2025). https://mind-bpi.streamlit.app/. Accessed on January 13, 2025.
23. Mross, K., Fiedler, C., & Grävingholt, J. (2022). Identifying pathways to peace: How international support can help prevent conflict recurrence. International Studies *Quarterly, 66*(1), sqab091.
24. Mustafa, G., Jamshed, U., Nawaz, S., Arslan, M., & Ahmad, T. (2023). Peace: A Conceptual Understanding. *Journal of Positive School Psychology*, 853-863.
25. Newman, E., & Visoka, G. (2024). NATO in Kosovo and the logic of successful security practices. *International Affairs, 100*(2), 631-653.
26. Radovanović, S., Bohanec, M., & Delibašić, B. (2023). Extracting decision models for ski injury prediction from data. *International Transactions in Operational Research, 30*(6), 3429-3454.
27. Stewart, F. (Ed.). (2016). *Horizontal inequalities and conflict: Understanding group violence in multiethnic societies*. Springer.
28. van Iterson Scholten, G. M. (2024). Peace beyond the Absence of War: Three Trends in the Study of Positive Peace.

Efficiency of Elementary Schools in Belgrade: A DEA and SHAP Analysis

Sandro Radovanović[1*], Boris Delibašić[1], Milija Suknović[1], Rôlin Gabriel Rasoanaivo[2], Elandaloussi Sidahmed[2,] Pascale Zaraté[2]

[1] University of Belgrade, Faculty of Organizational Sciences
[2] Université de Toulouse, CNRS, IRIT
*sandro.radovanovic@fon.bg.ac.rs

Abstract: This study investigates the efficiency of elementary schools in the Belgrade region in preparing students for middle school admission, highlighting the critical role of socio-economic factors. Using a modified output-oriented Data Envelopment Analysis model with municipality score constraints, we assess school performance by comparing grade point averages and student numbers (inputs) to standardized test scores (outputs). Our findings reveal stark differences in efficiency across municipalities, with the highest-performing areas, such as Stari Grad, Voždovac, and Zemun, benefiting from dynamic economies and diverse opportunities. In contrast, outer-circle municipalities like Sopot, Mladenovac, and Obrenovac face challenges tied to limited economic prospects. Leveraging Shapley Additive Explanations analysis, we uncover that unemployment is the most influential factor, with municipalities with lower unemployment achieving significantly lower efficiency scores. These findings not only demonstrate the interconnectedness of socio-economic conditions and education but also emphasize the need for targeted interventions to bridge performance gaps. By addressing unemployment and fostering economic growth, particularly in disadvantaged areas, we believe that policymakers can create a more equitable educational landscape.

Keywords: Data Envelopment Analysis; Shapley Additive Explanations; Educational Efficiency

1. Introduction

An important issue in education is understanding why some elementary schools consistently outperform others in preparing students for middle school admission. This disparity often stems from socio-economic factors, such as unemployment, economic opportunities, and access to resources, which influence both student performance and school efficiency (Ozturk, 2008; Hayat et al., 2022). In Serbia, this disparity is particularly evident, with schools in socio-economically advantaged areas achieving better outcomes compared to those in disadvantaged municipalities (Pešikan & Ivić, 2021).

Understanding the efficiency of elementary schools is critical because it directly impacts students' academic success and their future opportunities. By identifying the factors that influence efficiency, we can uncover actionable insights to reduce educational disparities, improve resource allocation, and enhance overall system performance, ultimately fostering a better education system. In the educational domain, Data Envelopment Analysis has been widely used as a powerful tool to evaluate school efficiency by comparing inputs, such as student performance metrics, to outputs, such as standardized test scores. For example, in the higher education one can find ranking of different education systems in (Sun et al., 2023), or (Lee & Johnes, 2022) who evaluated teaching efficiency to graduate employment. More often, data envelopment analysis was used for ranking of schools (Henriques et al., 2022; Chiariello et al., 2022), but also for discovering determinants of success (Kounetas et al., 2023). Applying DEA in the context of education allows for a structured assessment of how schools convert resources and academic preparation into successful outcomes, providing valuable benchmarks for policymakers to identify inefficiencies and drive improvements in the education system. By identifying the factors that influence efficiency, we can uncover actionable insights to reduce educational disparities, improve resource allocation, and enhance overall education system performance. Studying education efficiency is not only academically intriguing but also socially impactful, as it helps identify

what factors can hinder student success.

In Serbia, admission to middle schools, including gymnasiums and vocational schools, is determined by a standardized entrance exam known as the Final Exam, typically taken at the end of elementary education (8th grade). This exam assesses students' knowledge and skills in key subjects: Serbian language and literature (or the mother tongue), mathematics, and a third test which is selected in advance by the student and can be one of the following: history, geography, biology, physics, and chemistry. The results of the Final Exam are combined with students' grade point averages from the last three years of elementary education (6th, 7th, and 8th) to calculate their total score, which determines their eligibility for specific middle schools. Students rank their preferred schools and programs, and admission is competitive, with higher-ranking schools typically requiring better results due to limited capacity. The process ensures a merit-based transition from elementary to secondary education.

The key components of this approach include applying a data envelopment analysis model to measure the efficiency of elementary schools in transforming inputs, such as grade point averages and the number of students, into outputs, including standardized test scores. However, data envelopment analysis mathematical model should be adapted for the specificities of the educational system in Serbia. More specifically, the model is adapted so the efficiency scores are bounded for the average municipality efficiency scores are not overly high, nor low. This is one of the technical contributions of the paper. In addition to the data envelopment analysis, we wanted to inspect disparities in efficiency scores between municipalities by incorporating socio-economic factors like unemployment rates, average salaries, and active companies. To do so, we incorporate SHAP explanations to analyse the influence of socio-economic factors on municipal efficiency scores.

2. Methodology

Our research goal is to evaluate efficiency of schools regarding the admission to middle school and assess if there are differences in results that can be attributed to the socio-economic factors of the municipality.

The motivation for this analysis lies in understanding the efficiency of elementary schools in preparing students for this critical transition. Higher GPAs in 6th, 7th, and 8th grades are expected to correlate positively with higher Final Exam scores, as they reflect consistent academic preparation and mastery of curriculum content. Therefore, efficient schools should be those that maximize exam scores relative to their performance of students during the elementary education, and *vice-versa* inefficient schools should be those where exam scores are low compared to the grade average during the elementary education. The reason why we would like to assess socio-economic factors comes from the OECD's PISA 2022 assessment that indicates that socio-economically advantaged students (the top 25% in terms of socio-economic status) outperformed disadvantaged students (the bottom 25%) by a considerable margin (Organisation for Economic Co-operation and Development, 2022) in Serbia.

As a suitable methodology for efficiency analysis (Ji & Lee, 2010), we employ data envelopment analysis to assess the efficiency of each individual elementary school and inspect if there are differences in average scores within and between municipalities. This analysis aligns closely with the structure of Serbia's education system, where the transition to middle schools heavily depends on performance in the Final Exam, which includes the same subject areas as the outputs in the DEA model.

2.1 Design Choices

Data. To assess the efficacy of elementary schools, we utilized publicly available results from the reporting portal "My Middle School[21]," which represents the systematic and comprehensive collection of *ad-hoc* reports related to the school level performance, academic achievement, and standardized admission exam results.

[21] Web address: https://izvestavanje.mojasrednjaskola.gov.rs/

For the purposes of the research, we focused solely on the region of the capital city – Belgrade Region – and regular elementary schools. In other words, we removed schools aimed for the adult education, special needs education, and those for the education of the gifted. The former are schools aimed for students with development delays or those who failed to finish elementary school regularly due to various factors. These students often have lower academic performance, as teaching and grade evaluations take place after regular working hours. Upon completing elementary education, these individuals can continue their education by enrolling in secondary schools, including gymnasiums and vocational schools. However, the Adult Education Survey conducted in 2022 indicates that only 19.9% of adults engaged in formal or non-formal education or training, which is significantly below the European Union average of 45.1% (European Commission, 2024). It would be unfair to compare schools aimed at adult education or special needs education with regular schools because their objectives, target demographics, and evaluation metrics significantly differ. Their primary goals often centre on skill acquisition, individual progress, and reintegration into formal education or the workforce, rather than traditional academic achievement metrics.

On the other hand, schools for gifted students are designed to provide specialized, advanced instruction tailored to the exceptional abilities of their students in areas such as mathematics, science, or the arts. It is offered for students from the sixth grade of the elementary school and as such take only the most advanced students in the generation. These institutions often have access to additional resources, smaller class sizes, and customized learning environments that foster accelerated learning and innovative thinking. As a result, their performance metrics, such as academic achievement or standardized test scores, are naturally higher due to the exceptional aptitude of their student body. Including these schools in a comparison with regular elementary schools, which serve a broader and more diverse range of abilities, would create an uneven playing field and misrepresent the efficiency of regular schools by failing to account for these inherent advantages.

As a result, we obtained 176 elementary schools in Belgrade regions distributed over 17 municipalities of the Belgrade region.

Selection of input and output attributes. To assess the efficiency of the elementary school we selected the number of students and grade point averages at 6th, 7th, and 8th grade as inputs and average scores on mother tongue, mathematics, and the third test as outputs is highly relevant to the middle school admission process. In an ideal scenario, efficiency would depend solely on the selected inputs, reflecting the schools' ability to objectively assess students' knowledge throughout the elementary education and consequently translate knowledge into desired academic outcomes. However, we recognize that other external factors, such as socio-economic conditions, may also significantly influence school performance. That effects can be in a form of an exogenous influence (socio-economic factors influence inputs only), unobserved heterogeneity in the outcome (socio-economic factors influence outputs only), or even a confounder effect (socio-economic factors influence both inputs and outputs). To have a notion of account for these effects, we aim to explore how socio-economic factors could explain the observed differences in efficiency scores across schools. This approach will provide a more comprehensive understanding of the disparities in school efficiency and help identify areas for targeted interventions.

2.2 Data Envelopment Analysis Model

To calculate the efficiency of schools we adopted an output-oriented data envelopment analysis model with several adjustments. Since schools belong to a municipality who do have some notion of influence on the school organization and functioning through financial and organizational benefits, we want to ensure that average efficiency does not differ to much from other municipalities. Therefore, when observing the efficiency scores, we would like every municipality to have at most γ percentage difference of average efficiency scores compared to the remaining schools. One way to ensure that is to have a multi-step procedure where we would calculate the efficiency scores, observe municipality average efficiency scores, construct a constraint and repeat the data envelopment analysis with new set of constraints. In case of an infeasible solution, we would relax the constraints and try again. A more elegant solution would be to incorporate a set of constraints in the mathematical model directly. To do

that, we must introduce one mathematical model for calculation of efficiencies for every school instead of n mathematical models for each school separately, which is common in data envelopment analysis (Camanho et al., 2024).

We use the fact that the sum of linear functions is also a linear function (Aggarwal et al, 2020). Therefore, solving for n models can be posed as a single model. More importantly, the optimal solution of such formulation will be the same as the sum of n individual models. Consequently, efficiency scores for each school will remain the same with a caveat that this kind of formulation can result in a large number of variables one needs to optimize for, as well as in a large number of constraints. Effects of the model increase can be found in (Radovanović et al., 2022).

The model we employed is presented below:

$$\max \sum_{d=1}^{n} \sum_{r=1}^{s} u_{rd} y_{rd} \tag{1}$$

$$s.t.$$

$$\sum_{i=1}^{m} v_{id} x_{id} = 1, d = 1, \dots, n \tag{2}$$

$$\sum_{r=1}^{s} u_{rd} y_{rj} - \sum_{i=1}^{m} v_{id} x_{ij} \leq 0, \forall d, j = 1, \dots, n \tag{3}$$

$$\frac{1}{N_c} \sum_{d \in c} \sum_{r=1}^{s} u_{rd} y_{rd} - \frac{1}{N_{\rceil c}} \sum_{d \notin c} \sum_{r=1}^{s} u_{rd} y_{rd} \leq \gamma \tag{4}$$

$$\frac{1}{N_{\rceil c}} \sum_{d \notin c} \sum_{r=1}^{s} u_{rd} y_{rd} - \frac{1}{N_c} \sum_{d \in c} \sum_{r=1}^{s} u_{rd} y_{rd} \leq \gamma \tag{5}$$

$$v_{id} \geq \varepsilon, i = 1, \dots, m \ \wedge d = 1, \dots, n \tag{6}$$

$$u_{rd} \geq \varepsilon, r = 1, \dots, s \ \wedge d = 1, \dots, n \tag{7}$$

This DEA model optimizes the efficiency of decision-making units, in this case elementary school, while incorporating constraints to ensure fairness across groups or categories (in this case municipalities). Since the model is output-oriented one seeks to maximize the weighted sum of outputs presented in equation (1) ($u_{rd} y_{rd}$) for each school d, where u_{rd} represents the weights for outputs r, and y_{rd} is the output values. To ensure that efficiency can be interpreted as an efficiency (Ahn et al., 1988) the weighted sum of inputs ($v_{id} x_{id}$) for each school d must equal 1 (equation (2)), where v_{id} represents the weights for inputs i, and x_{id} is the input values. In addition, to ensure that no school can dominate others in terms of efficiency we must introduce a set of constraints presented in equation (3). For every school d, the weighted sum of outputs minus the weighted sum of inputs for all other schools must be less than or equal to 0. One can interpret this set of constraints also that efficiency can be at most one. In the context of school efficiency, every school is compared to every school in the Belgrade region to form an absolute frontier.

One of the contributions of this paper is introduction of statistical parity set of constraints presented in equations (4) and (5). These constraints enforce that the average efficiency of schools within a specific municipality c should not significantly deviate from those outside the municipality by a factor γ. More specifically, equation 4 ensures the average efficiency within the category does not exceed that of the rest (schools not in the municipality c) by more than a threshold (γ), while equation 5 ensures the reverse, maintaining fairness and comparability.

Finally, to ensure faster convergence of optimization procedure we normalized the data using L_∞

norm (Aggarwal et al, 2020). Also, we set γ to 0.1 and ε to 0.01 but also bounded the maximum of both u_{rd} and v_{id} so the contribution of a single or several outputs or inputs is limited. For the optimization of the model we used interior point method.

3. Results and Discussion

After conduction the data envelopment analysis, we have obtained that only two elementary schools were indeed efficient. The histogram presented in Figure 1 shows efficiency scores are ranging from 0.5959 to 1, with majority of the schools achieved relatively high efficiency scores (0.75–0.90).

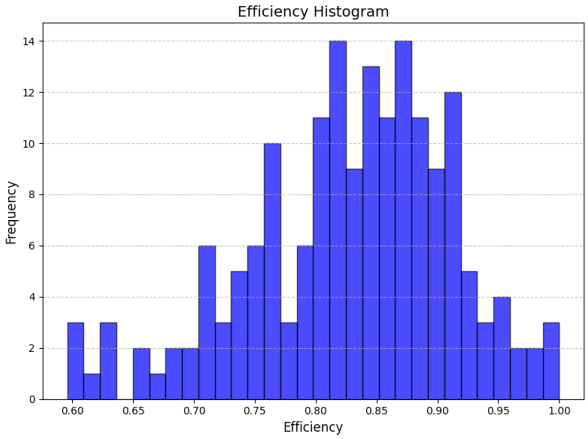


Figure 13: Efficiency scores histogram

If we are to compare the results on the test and efficiency, we can notice that higher test scores consistently correspond to greater efficiency (Figure 2), with strong clustering of high-efficiency observations in the upper-right quadrants of all plots. Similarly, there is a slight variation in test scores which correspond to low efficiency in the lower-left quadrants.

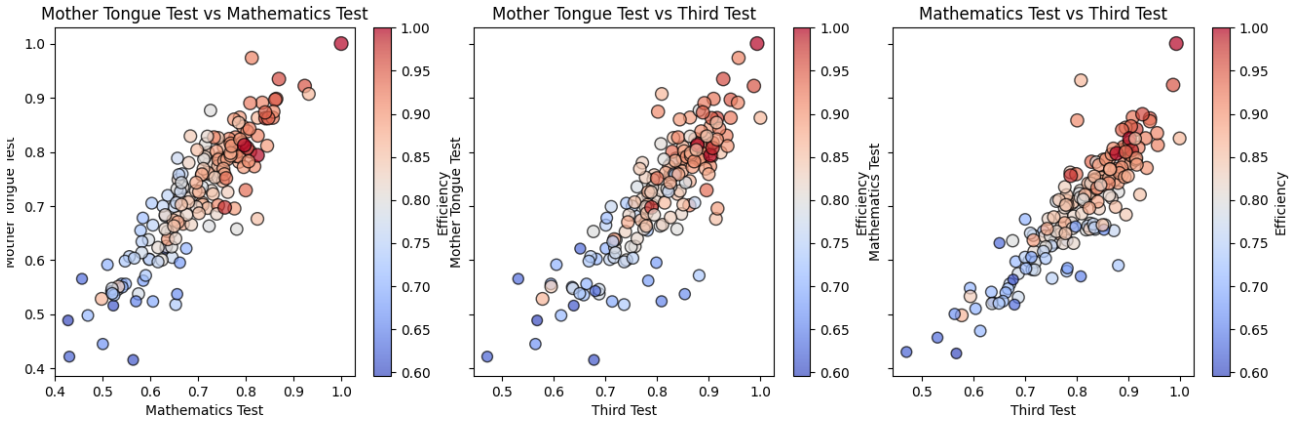


Figure 14: Correlation between efficiency scores and output

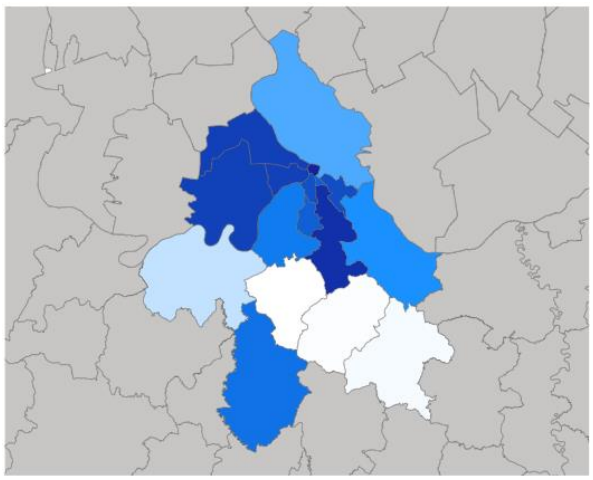


Figure 15: Average efficiency scores across municipalities

We can observe in Figure 3 that efficiency scores differ across municipalities. White represents lower efficiency scores, while darker blue color represents higher efficiency scores. Based on the average efficiency scores obtained, the top three municipalities are Stari Grad (0.8816), Voždovac (0.8722), and Zemun (0.8610). These municipalities more-or-less belong to the city center. Stari Grad is a hub for commerce, tourism, and culture, contributing to its high economic activity. Voždovac hosts significant industrial zones and residential neighbourhoods, supporting a diverse economy, and Zemun, known for its developed industries, boasts sectors ranging from manufacturing to services, enhancing its economic standing.

On the other side of the efficiency spectrum are Sopot (0.7304), Mladenovac (0.7321), and Obrenovac (0.7487). Sopot and Mladenovac are more rural, with economies primarily based on agriculture and small-scale industries, which may limit their economic opportunities.

To compare the efficiency scores with the socio-economic factors we need to use data from the Republic Statistical Office Overview of the state and development of municipalities[22] database for 2023. More specifically, we selected several socio-economic factors to explain efficiency scores as they capture key aspects of economic activity (e.g., active companies, entrepreneurs, and salaries), population demographics (e.g., ageing index and population estimates), and social welfare (e.g., child allowances and institutional care). Additionally, employment and unemployment metrics provide insights into labour market dynamics. Namely, alphabetically we selected Active companies per 1,000 inhabitants, Active entrepreneurs per 1,000 inhabitants, Ageing index, Number of beneficiaries of increased allowance for assistance and care of another person per 1,000 inhabitants, Number of beneficiaries of increased child allowance (aged 0-17) per 1,000 inhabitants, Number of beneficiaries of public residential institutions aged 65 and over per 1,000 inhabitants, Percentage of registered employees by municipalities of residence, and Registered unemployed per 1,000 inhabitants.

To inspect the association between efficiency scores and socio-economic attributes we employ Shapley value explanations (Lundberg, 2017). It is a model-agnostic technique based on cooperative game theory that explains the contribution of each feature to a model's predictions. It calculates the average marginal effect of a feature across all possible feature combinations. Global summary plot is presented in Figure 4.

[22] Web address: http://devinfo.stat.gov.rs/Opstine/libraries/aspx/home.aspx

Figure 16: SHAP Summary plot

The SHAP summary plot reveals that *Registered unemployed per 1,000 inhabitants* is the most influential factor, with higher unemployment significantly impacting the efficiency score. More specifically, municipalities with lower unemployment are performing better than municipalities with higher unemployment. An interpretation can be found in (Ananat et al., 2011) who concluded that living in a family with unemployed parents leads to increased stress and decreased well-being of children, which can adversely influence children's educational outcomes. What is more interesting is that unemployment is more important than employment (attribute *Registered employees by municipalities of residence comparing to population number*) as SHAP values are higher in intensity. Another interpretation is given by (Bordot, 2022), which puts unemployment into perspective of lack of opportunities of work in the vicinity of residence. Given that municipalities with higher unemployment in the Belgrade region are those at outer circle of Belgrade, this might suggest that lack of economic opportunities leads to lower education performance. This might be the case as other important factors by SHAP values are *Active companies* and *Average annual net salaries and wages*.

Features like child allowances, ageing index, and institutional care have smaller impacts, suggesting less direct influence on the outcome. Research indicates that socio-economic factors such as child allowances, ageing index, and institutional care have varying impacts on educational outcomes. As (Baker et al., 2024) stated, increased child benefits can improve children's mental health but may not significantly affect standardized test scores. The main reason for such results is that mental health is a prerequisite for good performance, but student must have capacity to reach the better score.

4. Conclusions

This study highlights the significant role of socio-economic factors in shaping the efficiency of elementary schools in the Belgrade region. Through an innovative data envelopment analysis model and SHAP analysis, we reveal that unemployment is the most influential factor, with municipalities with lower unemployment achieving lower efficiency. Economic drivers such as active companies and average salaries further underscore the connection between local economic conditions and educational outcomes. While factors like child allowances and institutional care have less direct impact, they remain important for addressing broader inequalities in efficiency scores. Another benefit one can have from the data envelopment analysis is the set of role models, or efficient schools that utilized the resources better. This can be used as a part of a discussion for the policy and decision making.

The sole fact that results highlight unemployment as a factor that significantly impacts educational outcomes, this suggests the need for targeted interventions in high-unemployment municipalities. Investing in mental health programs, tutoring, and skill-building initiatives can mitigate the adverse effects on students. Additionally, addressing regional disparities in economic opportunities through job creation and infrastructure improvements can enhance access to education and reduce inequality. Promoting holistic policies that integrate economic and educational support can improve overall student performance while fostering long-term social equity and workforce development.

In the future work, we plan to extend the DEA efficiency scores to the entire Serbia, not just Belgrade Region. In addition, we plan to implement preference elicitation methods that would guide the efficiency frontier for different region, municipalities, and school types. Preference elicitation would define input-output combinations that would consequently define groups of similar elementary schools that could be an input to DEA method or, more preferably, a *human-in-the-loop* procedure for DEA method. That way, a decision- or policy-maker would see the consequences of choices and correct the efficiency score calculation in real-time. Another line of research would be aimed at replacing linear input and output utility function DEA assumes with more complex utility functions. One approach worth considering is replacing the utility function with Gaussian processes or neural networks with autoencoder structures.

References

Aggarwal, C. C., Aggarwal, L. F., & Lagerstrom-Fife. (2020). *Linear algebra and optimization for machine learning* (Vol. 156). Cham: Springer International Publishing.

Ahn, T., Charnes, A., & Cooper, W. W. (1988). Efficiency characterizations in different DEA models. *Socio-Economic Planning Sciences*, *22*(6), 253-257.

Ananat, E. O., Gassman-Pines, A., Francis, D. V., & Gibson-Davis, C. M. (2011). *Children left behind: The effects of statewide job loss on student achievement* (No. w17104). National Bureau of Economic Research.

Baker, M., Koebel, K., & Stabile, M. (2024, May). The Impact of Family Tax Benefits on Children's Health and Educational Outcomes. In *AEA Papers and Proceedings* (Vol. 114, pp. 429-434). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.

Bordot, F. (2022). Artificial intelligence, robots and unemployment: Evidence from OECD countries. *Journal of Innovation Economics & Management*, (1), 117-138.

Camanho, A. S., Silva, M. C., Piran, F. S., & Lacerda, D. P. (2024). A literature review of economic efficiency assessments using Data Envelopment Analysis. *European Journal of Operational Research*, *315*(1), 1-18.

Chiariello, V., Rotondo, F., & Scalera, D. (2022). Efficiency in education: primary and secondary schools in Italian regions. *Regional Studies*, *56*(10), 1729-1743.

European Commission (2024). *Serbia: Adult education and training - funding*. Eurydice. Retrieved January 26, 2025, from https://eurydice.eacea.ec.europa.eu/national-education-systems/serbia/adult-education-and-training-funding.

Hayat, K., Yaqub, K., Aslam, M. A., & Shabbir, M. S. (2022). Impact of societal and economic development on academic performance: a literature review. *iRASD Journal of Economics*, *4*(1), 98-106.

Henriques, C. O., Chavez, J. M., Gouveia, M. C., & Marcenaro-Gutierrez, O. D. (2022). Efficiency of secondary schools in Ecuador: A value based DEA approach. *Socio-Economic Planning Sciences*, *82*, 101226.

Ji, Y. B., & Lee, C. (2010). Data envelopment analysis. *The Stata Journal*, *10*(2), 267-280.

Kounetas, K., Androulakis, G., Kaisari, M., & Manousakis, G. (2023). Educational reforms and secondary school's efficiency performance in Greece: a bootstrap DEA and multilevel approach. *Operational Research*, *23*(1), 9.

Lee, B. L., & Johnes, J. (2022). Using network DEA to inform policy: The case of the teaching quality of higher education in England. *Higher Education Quarterly*, *76*(2), 399-421.

Lundberg, S. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Organisation for Economic Co-operation and Development. (2022). *PISA 2022 results: Volume I and II: Country notes: Serbia*. Retrieved January 26, 2025, from https://www.oecd.org.

Ozturk, I. (2008). The role of education in economic development: a theoretical perspective. *Available at SSRN 1137541*.

Pešikan, A., & Ivić, I. (2021). The impact of specific social factors on changes in education in Serbia. *Center for Educational Policy Studies Journal*, *11*(2), 59-76.

Radovanović, S., Savić, G., Delibašić, B., & Suknović, M. (2022). FairDEA—Removing disparate impact from efficiency scores. *European Journal of Operational Research*, *301*(3), 1088-1098.

Sun, Y., Wang, D., Yang, F., & Ang, S. (2023). Efficiency evaluation of higher education systems in China: A double frontier parallel DEA model. *Computers & Industrial Engineering*, *176*, 108979.

Evaluation system of the regional directorates of national education through public primary school resources

Rôlin Gabriel Rasoanaivo[1,2*], Sidahmed Elandaloussi[1], Pascale Zaraté[1], Boris Delibašić[3] and Sandro Radovanović[3]

[1] IT Faculty (IRIT), Toulouse Capitole University, Toulouse, France
[2] Faculté des Sciences et Technologies, Université de Toamasina, Toamasina, Madagascar
[3] Faculty of Organizational Sciences, University of Belgrade, Belgrade, Serbia
[*] rolin-gabriel.rasoanaivo@ut-capitole.fr

Abstract: Effective resource allocation in the education sector requires a comprehensive understanding of the current situation. This study aims to support the Ministry of National Education in Madagascar in distributing resources more equitably among the 23 Regional Directorates of National Education (DREN). To achieve this, we developed a decision support system (DSS) using the multi-criteria methods Centroidous to assign importance to the criteria and combined compromise for ideal solution (CoCoFISo) to rank the DREN. The evaluation considered nine criteria related to the availability of human and material resources and their impact on student performance in examinations. The system analysed data from primary schools across Madagascar for the 2022–2023 academic year. The Centroidous method prioritise the criteria in divergent manners, thus resulting in fluctuations in the rankings of the DRENs according to the CoCoFISo method. This variation was observed to be contingent upon the criteria weights and the availability of resources in each primary school. Notably, the Analamanga DREN is in a leading position. The analysis revealed that the distribution of resources among the DRENs was found to be inequitable. Consequently, the findings of this study could assist the head of the Ministry of National Education in Madagascar in making more informed and equitable decisions regarding resource allocation.

Keywords: Resource allocation; Primary school; Decision support system; CoCoFISo; Centroidous.

1. Introduction

Primary education constitutes the foundational level of education for most pupils worldwide. The vast majority of a nation's population has previously undergone primary education prior to assuming various roles in the workforce. In certain countries, primary education is compulsory for all children as part of the national education policy (Khan, 1998; Southworth, 2003; Subiyakto & Mutiani, 2020; Vincent, 2021). In Madagascar, statistical data from the 2022-2023 school year pertaining to primary schools offers a comprehensive perspective. The nation currently boasts a total of 27,176 primary schools, encompassing 86,145 classrooms that collectively accommodate 4,108,804 pupils. This indicates a pupil-to-classroom ratio of 47. Among children aged 6 to 10, the net enrolment rate was 97.97%. The proportion of the student population enrolled in primary schools represents 70.81% of the total student population, including pre-school and high school students. The complete set of statistical information is contained in a document entitled "*Annuaire statistique 2022-2023*" (MEN, 2023) which is published annually by the Ministry of Education. This document contains a wealth of information on the static availability of resources for pre-schools, primary schools, secondary schools and high schools in Madagascar.

The present study focuses on analysing the resources available at these primary schools, given that the majority of the population attends them. The central research question that prompted our interest is as follows: would it be feasible to evaluate regional directorates of education using primary school data? If so, what evaluation method would allow all this information to be considered? The intention is to determine the rank of the regional directorates of national education based on the resources available to them individually. This is because the national education directorates, which are spread across all the

regions of Madagascar, aim to successfully implement national education policy.

Nine criteria were identified that illustrate the presence of human and material resources, as well as the success of pupils in examinations. These criteria appeared to be essential and interdependent for the operation and development of primary schools. The absence of material resources would prevent human resources from carrying out their tasks, and conversely, the absence of human resources would render the future of material resources uncertain. It was concluded that the presence of both human and material resources was necessary for the achievement of satisfactory results at the end of each school year.

Given the conflicting criteria, we found that this situation is part of a multi-criteria problem (Bouyssou & Perny, 1997). This is how we decided to apply multi-criteria decision-making methods to evaluate the regional directorates of national education in Madagascar. The literature contains a variety of multi-criteria decision-making methods applied in various sectors. The following are just a few examples. In the context of the efficiency of public procurement, the Technique of Order Preference Similarity to the Ideal Solution (TOPSIS) method was used to evaluate the member countries of the European Union (Milosavljević et al., 2021). The Combined Compromise For Ideal Solution Gray number (CoCoFISo-G) method was used to select organic suppliers (Sen & Toksoy, 2024). The Analytic Hierarchy Process (AHP) and VlšeKriterijuska Optimizacija I Komoromisno Rešenje (VIKOR) methods were used to guarantee the accuracy of a long-range rifle shot (Aleksić et al., 2024). The AHP and Combined Compromise Solution (CoCoSo) methods were combined to optimise the choice of the ideal waste management site (Yazdani et al., 2025). The Mean Weight and Combined Compromise For Ideal Solution (CoCoFISo) methods were used to assess Madagascar's employment regions (Nirinarivelo & Rasoanaivo, 2025). The Centroidous and the Multi-Objective Optimization on the basis of Simple Ratio Analysis (MOOSRA) methods have been used to evaluate the drones (Mohamed et al., 2025a).

Among these methods, the Centroidous (Vinogradova-Zinkevič, 2024) was selected for the prioritisation of criteria, and the CoCoFISo (Rasoanaivo et al., 2024) was employed for the evaluation of regional directorates of national education in the context of this study. This decision was made based on the effectiveness of Centroidous and CoCoFISo in solving multi-criteria problems (Chang, 2025; Fujita, 2025; Mohamed et al., 2025b; Rasoanaivo & Tata, 2024).

The subsequent section presents the algorithmic processes of these two methods. Following this, an analysis of the data is conducted to obtain the evaluation result. Finally, a discussion of the results and a proposal for future work is provided.

2. Research methodology

It is essential to recognize that the implementation of multi-criteria decision support methods relies on establishing of a hierarchical structure for the criteria, prior to the evaluation of the available alternatives. In this section, an exposition is provided on the algorithmic underpinnings of the Centroidous method and the CoCoFISo method, with the objective of elucidating their operational mechanisms. A significant benefit of the Centroidous method is that it avoids the need for the decision-maker to intervene during its use. The approach is founded on the premise of the performance matrix's availability. The CoCoFISo method has been found to be particularly compelling in that it has effectively addressed the limitations of the combined compromise solution method by enhancing the latter's algorithm. Both methods are based on the presence of a performance matrix of *n* criteria and *m* alternatives in **Formula 1** below.

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} \ldots & x_{1n} \\ x_{21} & x_{22} & x_{23} \ldots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \\ x_{m1} & x_{m2} & x_{m3} \ldots & x_{mn} \end{bmatrix} \quad (1)$$

2.1 Centroidous method

The Centroidous (Vinogradova-Zinkevič, 2024) method is an objective approach to evaluating the relative importance of criteria. It operates without the need for a decision-maker to establish a hierarchical structure for the criteria. The method is based on the availability of the performance matrix and aims to identify a compromise solution for the weights, based on a centre of gravity of the data for each criterion. The following **Formulas 2 to 6** describes the process.

- Performance matrix normalization using L_1norm

$$\widetilde{x_{ij}} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \tag{2}$$

- Centre of gravity obtained as average of each criterion i

$$c_i = \frac{1}{n}\sum_{j=1}^{n} \widetilde{x_{ij}} \tag{3}$$

- Euclidean distance between the alternative j and centre of gravity

$$d_j = \sqrt{\sum_{i=1}^{m}\left(\widetilde{x_{ij}} - c_i\right)^2} \tag{4}$$

- Euclidean distance normalization

$$\widetilde{d_j} = \frac{j^{\min(d_j)}}{d_j} \tag{5}$$

- Criteria weight

$$w_j = \frac{\widetilde{d_j}}{\sum_{j=1}^{n} \widetilde{d_j}} \tag{6}$$

2.2 CoCoFISo method

CoCoFISo (Rasoanaivo et al., 2024) is a recent multi-criteria decision-making method that has been developed as an improvement on the combined compromise solution method. It is based on the assembly of solutions derived from various multi-criteria methods, and its algorithm is based on four phases detailed in **Formulas 7 to 13** below.

- Performance matrix normalization using L_2 norm

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m}(x_{ij})^2}} \tag{7}$$

- Comparability sequence weighting using weighted sum and weighted product

$$S_i = \sum_{j=1}^{n}\left(w_j r_{ij}\right) \tag{8}$$

$$P_i = \sum_{j=1}^{n}(r_{ij})^{w_j} \tag{9}$$

- Deduction of aggregation strategies from comparability sequences

$$k_{ia} = \frac{P_i + S_i}{\sum_{i=1}^{m}(P_i + S_i)} \tag{10}$$

$$k_{ib} = \left(\frac{S_i + P_i}{1 + \frac{S_i}{1+S_i} + \frac{P_i}{1+P_i}}\right) \tag{11}$$

$$k_{ic} = \frac{\lambda(S_i) + (1-\lambda)(P_i)}{\left(\lambda \max_i S_i + (1-\lambda)\max_i P_i\right)} \quad (12)$$

$0 \leq \lambda \leq 1;\ \lambda$ is chosen by decision $-$ makers (usually $\lambda = 0.5$).

- Determining the final score

$$k_i = (k_{ia}k_{ib}k_{ic})^{\frac{1}{3}} + \frac{1}{3}(k_{ia} + k_{ib} + k_{ic}) \quad (13)$$

3. Data and results

In the context of an investigation into the decentralised national education departments present in each region of Madagascar, consultation was made of the information published by the Ministry of National Education on its website[23]. This consultation revealed a document entitled "Annuaire statistique 2022-2023" (MEN, 2023), published on an annual basis[24] by this ministry, which lists the resources available in each district and region of Madagascar, followed by school results. We have chosen to use data from the 2022-2023 school year because it is the most recent information available on this website. In other words, this information reflects the current state of resource availability in schools. Among the information presented in this document, the decision was thus taken to opt for data on primary schools and select nine criteria based on data availability. These criteria fall into three categories, including the availability of material resources, human resources and examination results for each educational establishment. **Table 1** below describes the different criteria we will be using.

Table 1. Description of evaluation criteria

Criteria group	Criretia	Description
Human resource	Student	This is the total number of students enrolled for an academic year in all the primairy schools in each decentralised directorate.
	Teacher	Teachers are distinguished by their employment contract. There are four categories of teacher: those who are civil servants, those on contract paid by the state, those on contract and temporary paid by parents, and those who are volunteers. Given this diversity, the objective of this study is to determine the percentage of teachers paid by the state budget. Therefore, the term "teacher" in this study refers to all civil servants and contract teachers paid by the state, represented as a percentage compared to the other categories.
	Staff	These are the administrative and technical staff responsible for managing primary schools.
Material resource	School	These are the numbers of primary schools operational at each decentralised directorate.
	Classroom	This figure denotes the number of classrooms designated for or allocated to primary schools for a single academic year.
	Board	The quantity of blackboards in each primary school classroom is a matter of interest.
	Seating	This is the total number of existing seats in all the bench tables, which is the capacity for students.

[23] https://www.education.gov.mg/

[24] https://www.education.gov.mg/ressources/annuaires-statistiques/

	Manual	These are textbooks in various subjects provided by the state to primary schools.
Admission	Success	This is the percentage of students admitted to higher classes. Originally, the data concerned the number of students repeating a year, but we have processed them in relation to the number of students enrolled in order to obtain this percentage.

There are twenty-three regional education directorates that require assessment. Each of these directorates is located in a specific region of Madagascar and is designated by the name of the region in which it is situated. **Figure 1** below presents a list of these 23 DRENs in alphabetical order. **Figure 2** below illustrates the twenty-three regions along with their respective criteria values. This figure represents the performance matrix. The screenshot of the system that has been developed is presented herewith.

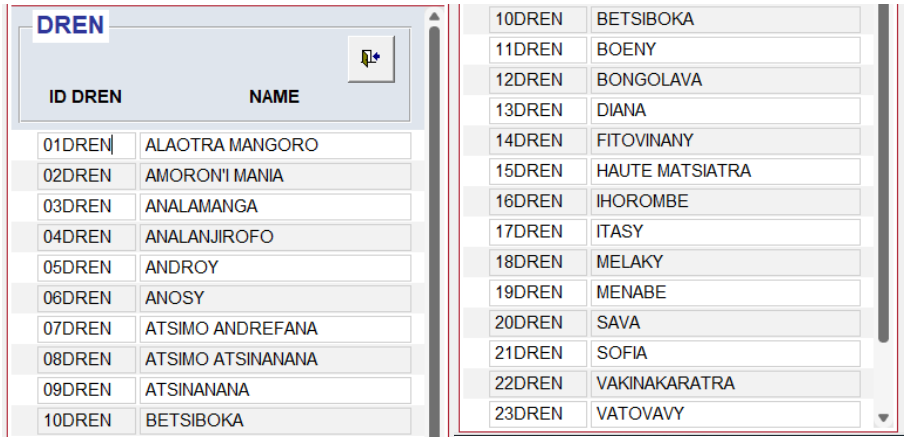


Figure 1. Regional directorates of national education (DREN) in Madagascar

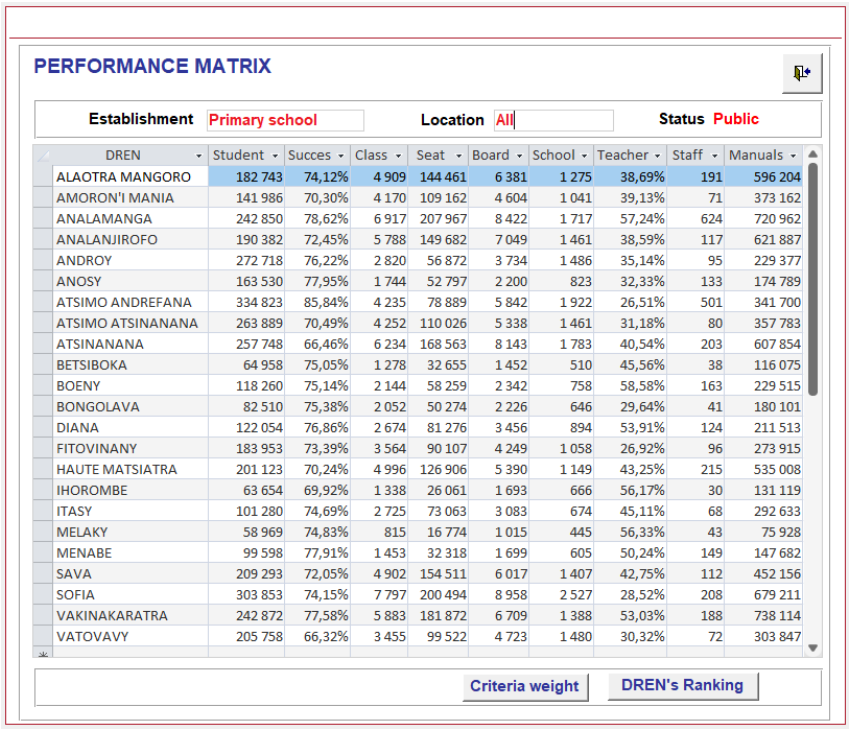


Figure 2. Performance matrix

Following the implementation of the Centroidous and CoCoFISo algorithms within the system, the

results can be viewed by clicking on one of the Criteria weight or DREN's ranking buttons displayed in **Figure 2** above. Consequently, **Figure 3** below displays the weights of the criteria calculated by the centroid method, and **Figure 4** shows the ranks of the DRENs according to the CoCoFISo method.

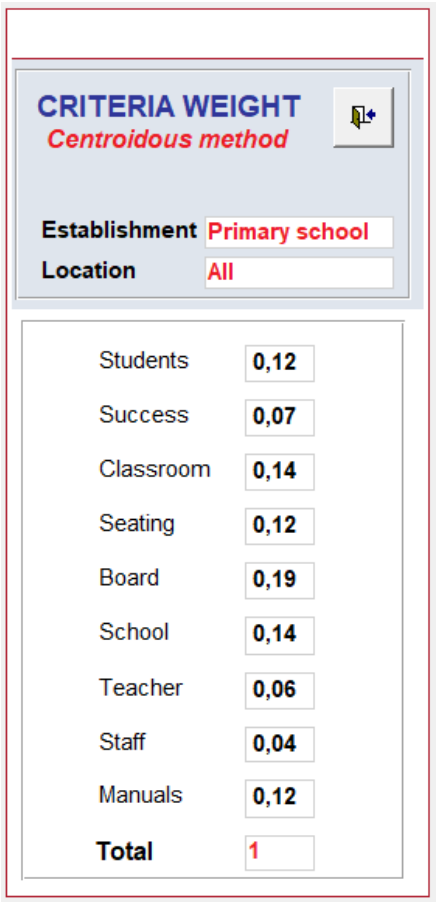
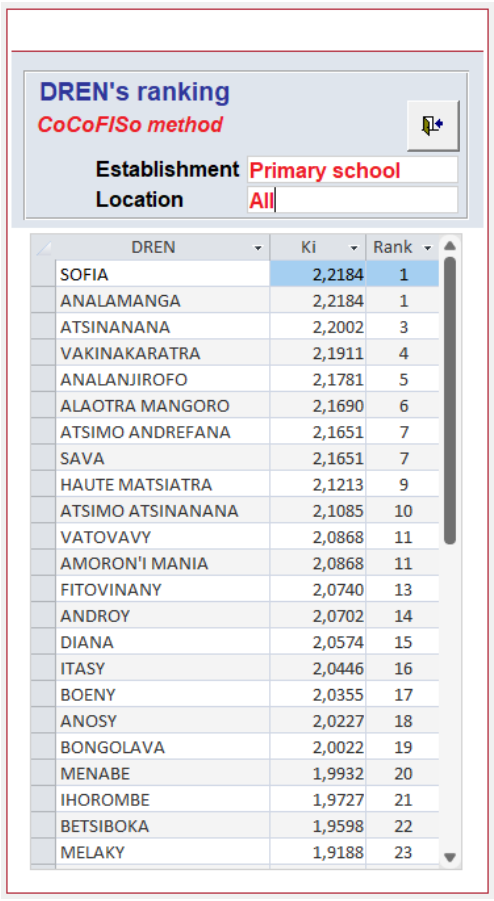


Figure 3. Criteria weight



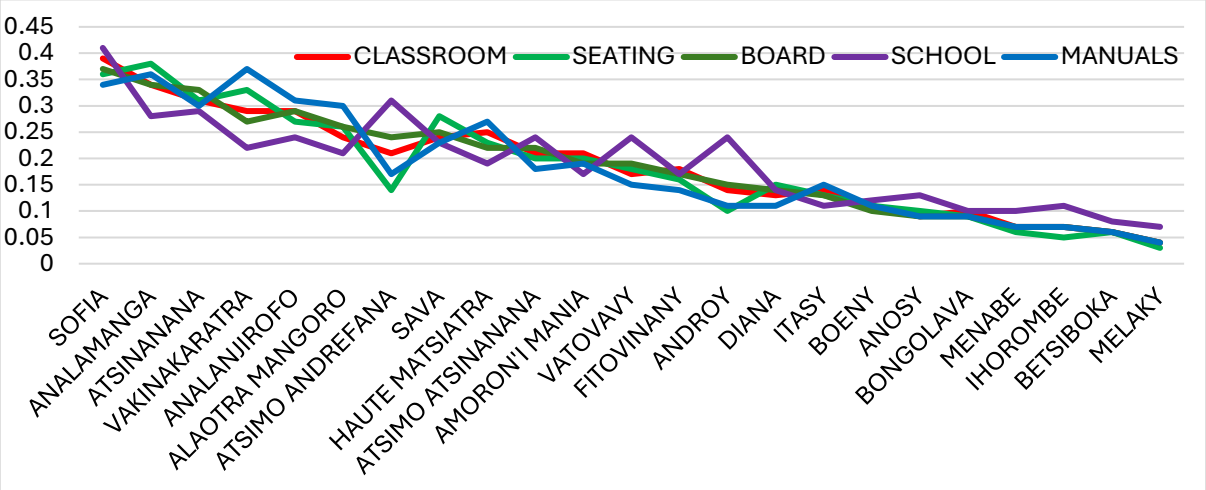
DREN	Ki	Rank
SOFIA	2,2184	1
ANALAMANGA	2,2184	1
ATSINANANA	2,2002	3
VAKINAKARATRA	2,1911	4
ANALANJIROFO	2,1781	5
ALAOTRA MANGORO	2,1690	6
ATSIMO ANDREFANA	2,1651	7
SAVA	2,1651	7
HAUTE MATSIATRA	2,1213	9
ATSIMO ATSINANANA	2,1085	10
VATOVAVY	2,0868	11
AMORON'I MANIA	2,0868	11
FITOVINANY	2,0740	13
ANDROY	2,0702	14
DIANA	2,0574	15
ITASY	2,0446	16
BOENY	2,0355	17
ANOSY	2,0227	18
BONGOLAVA	2,0022	19
MENABE	1,9932	20
IHOROMBE	1,9727	21
BETSIBOKA	1,9598	22
MELAKY	1,9188	23

Figure 4. DREN's ranking by primary school

Following a thorough analysis of the criteria, it is evident that those pertaining to material resources have been accorded a high priority. The *Board* criterion, with a weighting of 0.19, has been prioritised. The *Classroom* and *School* criteria have been allocated a second priority, with an identical weighting of 0.14. The *Seating* and *Manuals* criteria, which both occupy third priority, carry an identical weight of 0.12. At this point, we also identify the presence of the *Students* criterion in the human resources category, which also has a weight of 0.12. On the other hand, the other two human resources criteria have only minor weights, with *Teacher* at 0.06 and *Staff* at 0.04. It is also noteworthy that the *Success* criterion, with a weight of 0.07, holds greater precedence over the other two. As a result, the material resources criterion group has a total weight of 0.71, the human resources criterion group has a weight of 0.22 and admission to the examination has a weight of 0.07.

In accordance with the established criteria weightings (**Figure 3**) and performance matrix (**Figure 2**), the CoCoFISo method was applied to evaluate the twenty-three DRENs, as illustrated in **Figure 4** above. Consequently, the Sofia and Analamanga DRENs were positioned identically, at the top of the evaluation. Subsequently, DREN Atsinanana was placed third. The DREN Vakinakaratra, DREN Analanjirofo and DREN Alaotra Mangoro are positioned in fourth, fifth and sixth place respectively. DREN Atsimo Andrefana and DREN Sava obtained an identical evaluation and were both ranked seventh. DREN Haute Matsiatra and DREN Atsimo Atsinanana are positioned in ninth and tenth place respectively and son on. In contrast, the final five places in the ranking are occupied by the Bongolava DREN, the Menabe DREN, the Ihorombe DREN, the Betsiboka DREN and the Melaky DREN, which respectively occupy the nineteenth and twenty-third rankings.

This ranking is indicative of the availability of resources in the primary school for each DRENs. To illustrate the distribution of material resources among each DREN, reference was made to the normalised matrix obtained using the CoCoFISo method. This is because, due to the various values and units of measurement present in the initial data, the normalised matrix standardises the values of these data between 0 and 1 through its calculation method. **Figure 5** below shows this distribution according



to DREN rank.

Figure 5. Availability of material resources in primary schools for each DRENs

The findings reveal a consistent decrease in material resources from the first to the last rank, suggesting that the Sofia and Analamanga DRENs, occupying the top positions, possess a substantial amount of material resources in comparison to other DRENs.

Conversely, an analysis of the distribution of human resources at each primary school in the DRENs exposes significant variations across different DRENs. The subsequent **Figure 6** provides a visual representation of these variations.

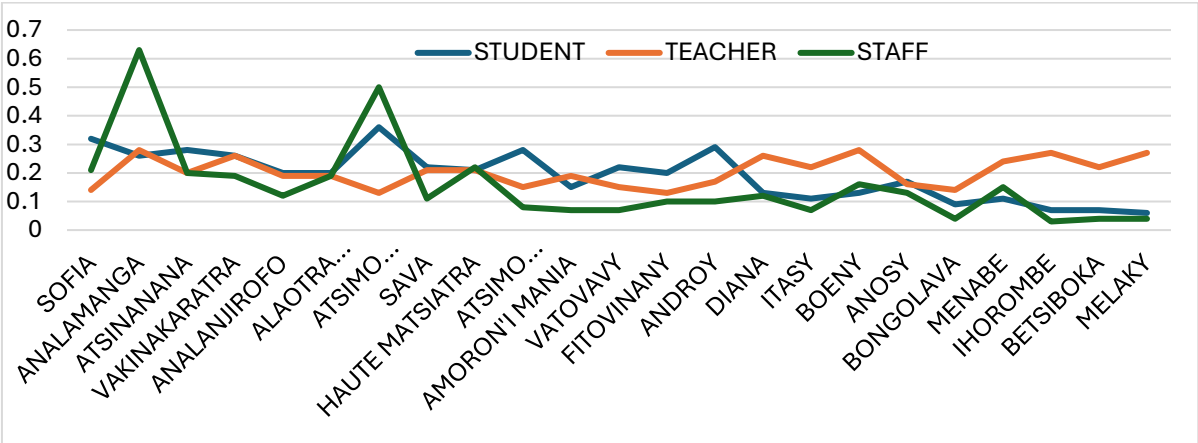


Figure 6. Availability of human resources in primary schools for each DRENs

The reliability of the DREN ranking method is evidenced by the initial weighting of the material resources criterion group at 0.71, in comparison to 0.22 for human resources. This results in a subsequent ranking of DRENs with primary schools that possess significant material resources, placing them higher in the evaluation. This is the reason why we find Melaky DREN at the bottom of our evaluation because these primary schools have low material resources compared to the others according to **Figure 5**.

4. Conclusions

The present study, which is focused on the evaluation of Madagascar's regional directorates of national education in terms of the availability of resources in primary schools, will indubitably be of

great assistance to Ministry of National Education officials when allocating resources. Multi-criteria decision-making methods were implemented based on the current state of human and material resources and examination results in primary schools located within the DREN. The Centroidous method was employed to establish the criteria weights, while the CoCoFISo approach was used to rank the DRENs. The findings demonstrated the efficacy of these two approaches in evaluating DRENs. To enhance the ease of application of these methods, these algorithms were integrated into a decision support system. The result discovered the disparity in the availability of resources between the primary schools in each DREN, in terms of both material and human resources. In this context, the equitable distribution of resources will be possible, thus ensuring that all Malagasy pupils in primary schools have an equal opportunity to access the same education. To illustrate this point, consider the distribution of resources, whether material or human, to primary schools. One approach would be to allocate resources based on the total number of pupils. However, a more comprehensive approach would involve the distribution of material resources to each DREN for primary schools, contingent on the total number of children aged 6 to 10 residing in the respective regions. This ensures that all children have access to educational opportunities. The allocation of human resources will then be based on the availability of material resources.

The present study focuses specifically on the case of public primary schools in Madagascar in order to evaluate the regional directorates of national education. In future research, it would be relevant to also consider the availability of resources in secondary schools and high schools when evaluating the DRENs. Above all, the overall evaluation of the DREN should consider the availability of resources in all three establishments. As soon as the DREN rankings for primary schools, secondary schools and high schools are available, this application will implement a method for aggregating the rankings to obtain an overall assessment of the DRENs. Another research direction could be to develop a decision support system predicting the distribution of resources within each DREN, for primary schools, based on the number of children aged 6 to 10 residing in the region in order to support the Ministry responsible for national education by proposing interventions on equitable education development across Madagascar.

Acknowledgements

The present study is indebted to the Madagascar's Ministry of National Education for its decision to make the statistical yearbooks available to the public. The acquisition of these data was instrumental in the successful execution of the research project.

References

Aleksić, A. R., Delibašić, B. V., Jokić, Ž. M., & Radovanović, M. R. (2024). Application of the AHP and VIKOR methods of individual decision making and the Borda method of group decision making when choosing the most efficient way of performing preparatory shooting at serial number one from a 12.7 mm long range rifle M-93. *Vojnotehnički Glasnik / Military Technical Courier*, *72*(3), Article 3. https://doi.org/10.5937/vojtehg72-52204

Bouyssou, D., & Perny, P. (1997). Aide multicritère à la décision et théorie du choix social. *Nouvelles de la Science et de la Technologie, tm*, *15*, 61-72.

Chang, H. (2025). Decision Algorithm for Digital Media and Intangible-Heritage Digitalization Using Picture Fuzzy Combined Compromise for Ideal Solution in Uncertain Environments. *Symmetry*, *17*(3), Article 3. https://doi.org/10.3390/sym17030443

Fujita, T. (2025). N-Superhypersoft Set and Bijective Superhypersoft Set. *Advancing Uncertain Combinatorics through Graphization, Hyperization, and Uncertainization: Fuzzy, Neutrosophic, Soft, Rough, and Beyond*, 138.

Khan, A. N. (1998). Obligation to Attend School: The English Law. *Australia & New Zealand Journal of Law & Education*, *3*, 74.

MEN, M. (2023). *Annuaire statistique national 2022-2023*. https://www.instat.mg/p/men-annuaire-statistique-de-leducation-2022-2023

Milosavljević, M., Radonovanović, S., & Delibašić, B. (2021). Evaluation of public procurement efficiency of the EU countries using preference learning topsis method. *Economic Computation and Economic Cybernetics Studies and Research*, *55*(3), 187-202. https://doi.org/10.24818/18423264/55.3.21.12

Mohamed, M., Mouty, A. M. A., Mohamed, K., & Smarandache, F. (2025a). *SuperHyperSoft-Driven Evaluation of Smart Transportation in Centroidous-Moosra: Real-World Insights for the UAV Era*. Infinite Study.

Mohamed, M., Mouty, A. M. A., Mohamed, K., & Smarandache, F. (2025b). *Superhypersoft-driven evaluation of smart transportation in centroidous-moosra: Real-world insights for the uav era*. Infinite Study. https://books.google.com/books?hl=fr&lr=&id=rLRCEQAAQBAJ&oi=fnd&pg=PA154&dq=Centroidous+method&ots=P99X3pbtpq&sig=UKvLIqGQWCGxm1doG5eJj6qGTYw

Nirinarivelo, H., & Rasoanaivo, R. G. (2025). Multi-criteria Evaluation of Madagascar's Regions in the Context of Employment Using the CoCoFISo Method. *Spectrum of Decision Making and Applications*, *2*(1), Article 1. https://doi.org/10.31181/sdmap21202514

Rasoanaivo, R. G., & Tata, J. A. (2024). A New Technique of Ranking Madagascar's Universities Using CoCoFISo Method in a Multi-Criteria Decision Support System: MadUrank. *International Journal of Scientific Research in Computer Science and Engineering*, *12*(4), 18-31.

Rasoanaivo, R. G., Yazdani, M., Zaraté, P., & Fateh, A. (2024). Combined compromise for ideal solution (CoCoFISo): A multi-criteria decision-making based on the CoCoSo method algorithm. *Expert Systems with Applications*, 124079. https://doi.org/10.1016/j.eswa.2024.124079

Sen, H., & Toksoy, M. S. (2024). Green Supplier Selection with CoCoFISo-G. *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, *32*, 257-265. https://doi.org/10.55549/epstem.1598450

Southworth, G. (2003). *Primary School Leadership in Context: Leading Small, Medium and Large Sized Schools*. Routledge. https://doi.org/10.4324/9780203711750

Subiyakto, B., & Mutiani. (2020). *Learning Motivation In Street Children (Case Study on Street Children Who Attend School in Public Elementary School (Sekolah Dasar Negeri/SDN) Mawar 2 Banjarmasin)*. 264-270. https://doi.org/10.2991/assehr.k.200803.033

Vincent, G. (2021). *L'École primaire française : Étude sociologique*. Presses universitaires de Lyon.

Vinogradova-Zinkevič, I. (2024). Centroidous Method for Determining Objective Weights. *Mathematics*, *12*(14), Article 14. https://doi.org/10.3390/math12142269

Yazdani, M., Ye, C., Shaayesteh, M. T., & Zaraté, P. (2025). Decision support system for waste management: Fuzzy group AHP-CoCoSo. *International Journal of Production Management and Engineering*, *13*(1), Article 1. https://doi.org/10.4995/ijpme.2025.21558

DSS Design, User Experience, and Business/Process Frameworks

The role of asset management maturity in competitiveness: guidelines for decision-making

Gabriel Herminio de Andrade Lima[1*], Ana Paula Cabral Seixas Costa[1]

[1] Center for Decision Systems and Information Development - CDSID, Universidade Federal de Pernambuco, Recife, PE, Brazil
[*] gabriel.herminio@ufpe.br

Abstract: Asset management (AM) has emerged as a strategic factor that enables the creation and delivery of value in asset-intensive industries, which has been consolidated as a locus of research. The current scenario of Industry 4.0 technologies has supported the progress of AM practices. Despite the wide range of benefits of AM practices, AM literature still needs to advance on the strategic, decision-making, and sustainability approaches, and mainly, on whether AM maturity affects business performance. Thus, many decisions are based on intuitions, which have demanded evidence to support AM decision-making. Therefore, this paper seeks to bring evidence that AM maturity, which is composed of AM capabilities, contributes to competitiveness. For this, the partial least squares structural equation modeling (PLS-SEM) methodology was applied to the database of AM maturity assessments from Brazilian enterprises. The results illuminate AM context describing some relationship between AM maturity and competitiveness, specifically the contributions to business performance and competitive priorities, which can assist in the adoption of Industry 4.0 technologies for improvements in AM capabilities. Therefore, the paper's findings provide underpinning theoretical and managerial insights into decision-making processes. Finally, it illustrates how managers can use the paper's findings to develop AM plans to improve business performance.

Keywords: Asset Management; Business performance; competitiveness; roadmap.

1. Introduction

Asset Management (AM) plays a role in enabling asset-intensive enterprises (manufacturing, infrastructure, transportation, and others) to deliver business value, which includes sustainable aspects, through the management of asset lifecycle while balancing risk, performance, and cost (El-Akruti et al., 2013; IAM, 2024; ISO, 2014; Sandu et al., 2023). So, enterprises that invest and develop AM initiatives may achieve the best organizational results by their assets.

AM involves activities and tools that come from different fields (El-Akruti et al., 2013) , e.g., maintenance, risk management, and performance assessment. Due to this interdisciplinary approach, many challenges arise, such as data availability and interoperability, the interplay processes, and the relationship between a company's assets and organizational structure process (Tscheikner-Gratl et al. 2019; Daulat et al.,2022; IAM 2008), which demand new solutions.

With advances in technologies, especially Industry 4.0 technologies, new tools, and methodological progress assist in developing AM capabilities and initiatives (Frank et al., 2019; Rampini & Re Cecconi, 2022). These technologies have allowed the collection and integration of data in real-time, and their analyses to support decision-making in the industry (Frank et al., 2019).

Therefore, many benefits are expected from applying AM practices such as improvements in financial performance, informed decision-making, the delivery of services and goods (Han et al., 2021; ISO, 2014; Maletič et al., 2020). Despite these benefits reported, there is a demand for empirical evidence that demonstrates if AM maturity would reflect in business performance (BP) (Lima et al., 2021).

In addition, there is also a gap in the empirical demonstration of the relationship between asset management strategies, BP, and competitive advantages (Gavrikova et al., 2020). Therefore, this can contribute to scenarios where AM decisions have been made based on intuition (Komonen et al., 2012; van Riel et al., 2014). Thus, decision-makers need evidence showing that investing in AM maturity leads to organizational results.

BP is a fundamental element in managing and monitoring business strategies. Literature has demonstrated that the measurement of firm performance is complex (Siemieniuch & Sinclair, 2002), evolving multiple factors such as people, processes, and organizational systems. In this context, there are also few studies investigating the relationship between organizational maturity and BP (Tarhan et al., 2016).

Intending to increase business competitiveness, organizations develop operational strategies and align organizational capabilities to attain competitive priorities, such as quality, flexibility, and cost (Boyer and Lewis, 2002). Precisely in the AM context, assets play a role in organizational competitiveness and growth, impacting the competitive priorities related to flexibility and delivery (Cai & Yang, 2014; Smith & Sharif, 2007). Therefore, managers need AM decision-making that explores the best results in competitive priorities, and consequently, in BP. In this scenario, the Industry 4.0 technologies adoption is valuable to AM capabilities, which must be used considering the relevance of AM capabilities to competitiveness.

Seeking evidence to clarify these gaps, Partial Least Squares Structural Equation Modeling (PLS-SEM) has been applied to data from maturity and BP assessments from Brazilian enterprises. Specifically, this paper uses the AM maturity model integrated into a decision support system by Lima & Costa (2023) – asset management maturity assessment model (AMAP). AMAP comprises ten core AM dimensions: AM Policy, AM Strategy and Objectives, AM Planning, Data and Information Management, Asset Information System, AM Leadership, Competence Management, Risk Assessment and Management, Asset Performance, and Asset Cost and Valuation. PLS-SEM methodology requires a solid description of the hypotheses that will be investigated.

1.1 Hypotheses development

There is evidence that AM capabilities affect BP (Han et al., 2021; ISO, 2014; Maletič et al., 2018, 2020), supporting the development of Hypothesis 1. In exploratory research, Lima et al. (2021) identify that AM capabilities impact competitive priorities, which demand yet empirical evidence. In this way, when an organization develops a specific AM dimension, it is expected a result or contribution to BP. Thereby, using some of these relationships and AM dimensions available in AMAP, Hypothesis 2 to Hypothesis 9 were developed.

$\boldsymbol{H_1}$: there is a positive relationship between AM Maturity and BP.

$\boldsymbol{H_2}$: there is a positive relationship between AM Policy and Quality.

$\boldsymbol{H_3}$: there is a positive relationship between AM Planning and Quality.

$\boldsymbol{H_4}$: there is a positive relationship between Data and Information Management and Quality H_{4a} and Financial H_{4b}.

$\boldsymbol{H_5}$: there is a positive relationship between the Asset Information System and Quality.

$\boldsymbol{H_6}$: there is a positive relationship between AM Leadership and Flexibility.

$\boldsymbol{H_7}$: there is a positive relationship between Competence Management and Flexibility.

$\boldsymbol{H_8}$: there is a positive relationship between Asset Performance and Customers' satisfaction H_{8a} and Quality H_{8b}.

$\boldsymbol{H_9}$: there is a positive relationship between Asset Costing and Valuation and Financial.

2. Materials and methods

Unlike other multivariate tools, structural equation modeling methods allow the inclusion of unobservable variables that must be measured indirectly by variables (Dash & Paul, 2021; Hair et al., 2014). Among them, PLS-SEM has been used to explore the relationship between multiple variables, being useful in applications that evolve samples of small size, formative indicators, and nonnormal data (Hair et al., 2017). Considering these characteristics and the aim of this paper to explain the impact of AM maturity on BP, PLS-SEM has been applied to data from asset maturity and performance assessment, using the SmartPLS tool. Specifically, the sample is composed of seventy Brazilian enterprises (Figure 1) that have used an AM maturity model – AMAP (Lima & Costa, 2024). Moreover, a questionnaire (Appendix 1) that assesses BP on the Likert scale is available in AMAP. Each question was assigned to competitive priorities. It is worth noting that BP can be measured by the composition of all questions.

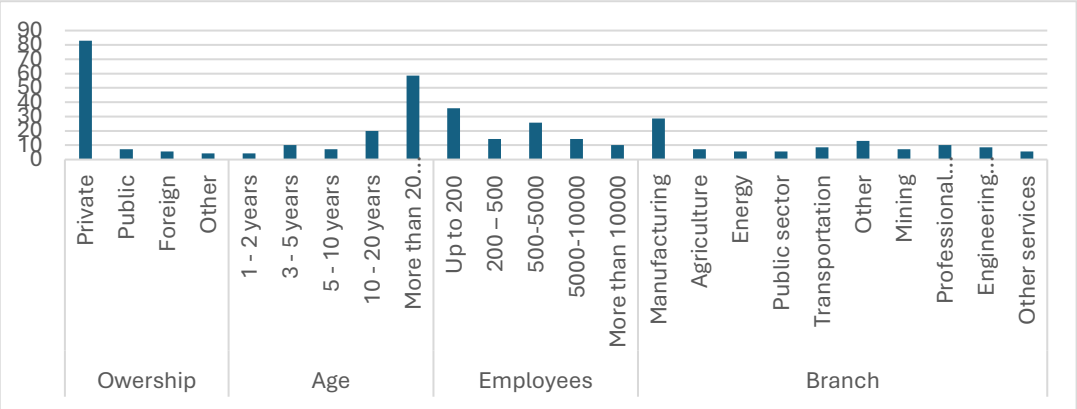


Figure 17: profiles of seventy Brazilian enterprises (%)

Initially, PLS-SEM approach requires the verification of the reliability and validity of constructs, using the confirmatory factor analysis. For this, the measuring of Cronbach's alpha (CA) and average variance extracted (AVE) was made to each competitive priority, identifying CA above 0.7 and AVE above 0.5 for all of them. Furthermore, the outer loading analyses were made, in which most indicators are above the threshold (0.7). Table 2 summarizes the reliability and validity of the constructs.

Table 2. Reliability and validity

Competitive Priorities	Indicator	Outer loading			
Flexibility	BP4	0.910	AVE	=	0.835
	BP8	0.917	AC	=	0.803
Quality	BP2	0.681	AVE	=	0.626
	BP3	0.886	AC	=	0.741
	BP10	0.793			
Costumers' Satisfaction	BP5	0.731	AVE	=	0.624
	BP10	0.816	AC	=	0.705
	BP13	0.818			
Financial	BP1	0.811	AVE	=	0.626
	BP5	0.736	AC	=	0.877
	BP12	0.894			
	BP13	0.808			
	BP7	0.595			
	BP14	0.867			

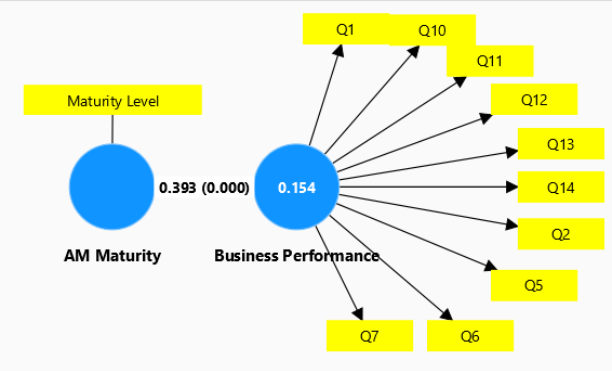
Note: p-value < 0.05.

As mentioned, all questions also assess BP. However, questions BP3, BP4, BP8, and BP9 which obtained outer loading below the threshold, were eliminated to increase AVE from 0.429 to 0.505 in BP.

Several small models were established based on developed hypotheses, setting up the structural models. To proceed in evaluation, discriminant validity was performed using the Fornell-Larcker criterion, which demonstrated differences among constructs. Thereby, the analyses of structural models can be made.

2.1 AM maturity and BP

Figure 2: Structural model H_1.

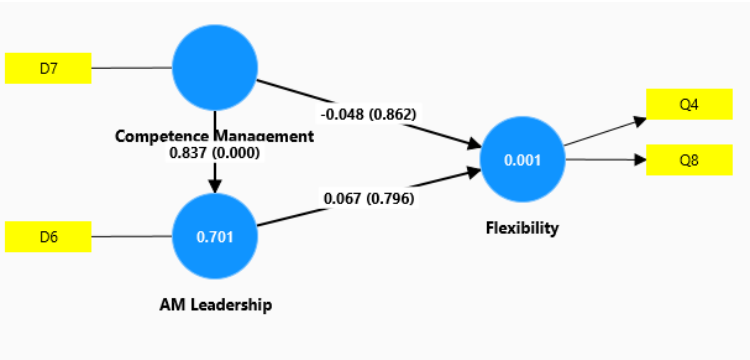


Exploring the effects of AM maturity on BP (Figure 2), which is reflected in its capabilities, found a path coefficient (β) of 0.393, which demonstrates a positive impact of AM maturity on BP.

2.2 AM Leadership, Competence Management and Flexibility

Considering the impact of Leadership & People competencies on Quality, it did not find evidence that AM Leadership (β=0.067, p-value > 0.05) and Competence Management (β=-0.048, p-value > 0.05) affect flexibility (Figure 3). On the other hand, it was possible to observe a relationship between Leadership and Competence Management (β=0.837, p-value > 0.05).

Figure 3: Structural model H_6 and H_7.



2.3 Strategy & Planning and Quality

Bringing evidence to validate H_2 and H_3, the path coefficients of AM Policy and AM Planning with Quality were identified β=0.399 and β=0.365 respectively.

2.4 Asset Information System and Quality

Note also that Quality can be improved through asset information system capability (β=0.380, p-value < 0.05).

2.5 Data & Information Management, Quality, and Financial

Considering the contribution on quality, it was found evidence that developing governance and use of data can leverage quality (β=0.333, p-value<0.05). Similarly, the financial perspective is related to

Data and Information Management (β=0.283, p-value<0.05). Therefore, this dimension has a positive association with two competitive priorities.

2.6 Asset performance, Quality and Costumers' Satisfaction

Quality demonstrated fit with Asset performance capability (β=0.371, p-value<0.05). In parallel, the hypothesis related to costumers' satisfaction was also satisfied (β=0.288, p-value<0.05).

2.7 Asset Cost & Valuation and Financial

Developing the capability to measure asset cost and value may contribute to financial performance (β=0.276, p-value<0.05).

3. Managerial Implications

Initially, the relationship between AM Maturity and BP highlights the value of AM practices in BP, aiding AM decision-makers to validate AM actions based on evidence and value, which can be used to demonstrate the relevance of investment in AM areas to business. On the other hand, this finding demonstrates the potential contribution of AM maturity assessment as a managerial tool to foster the improvement of AM maturity, which includes the AMAP system.

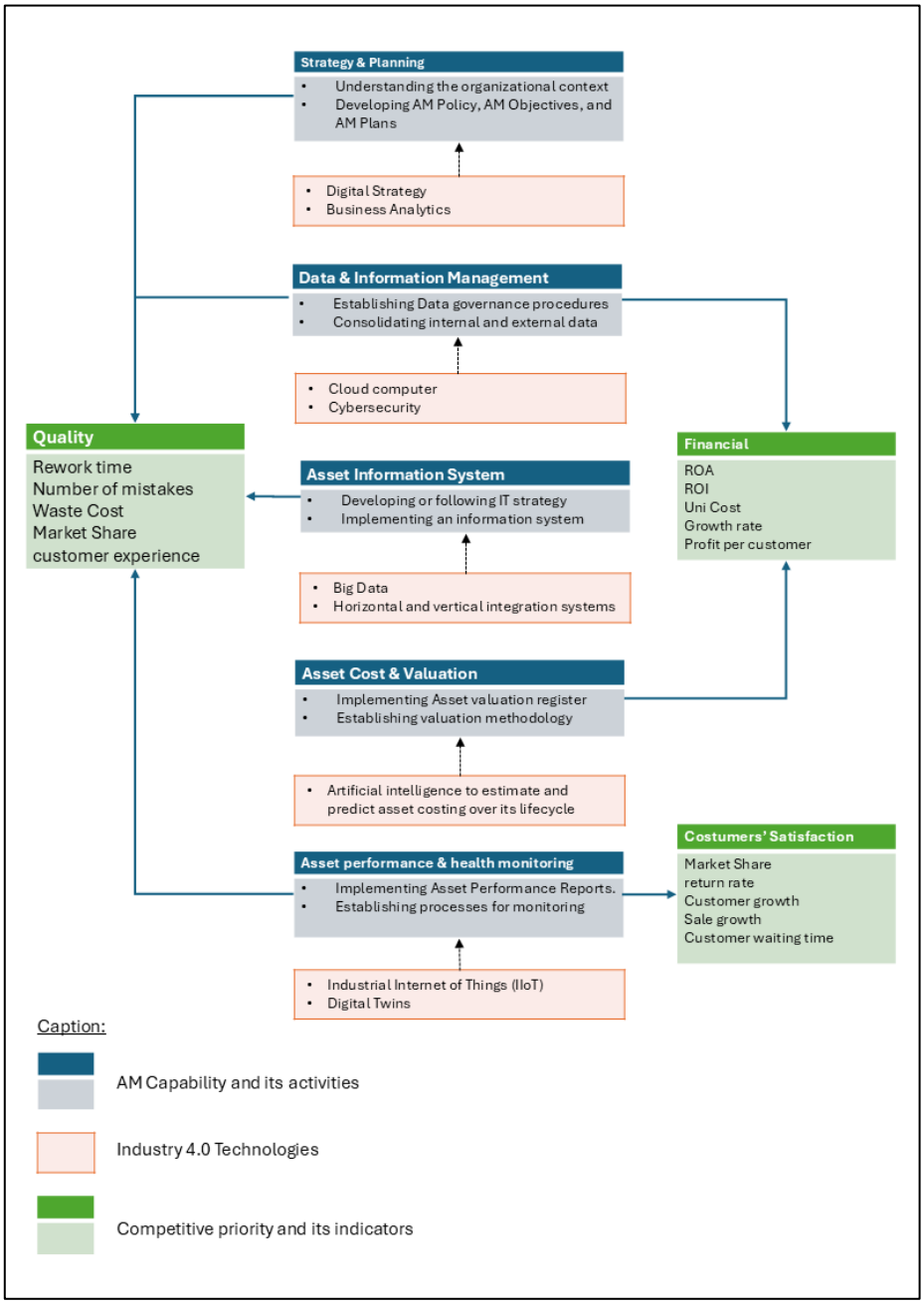
On the other hand, the findings summarized in Table 3 reveal that AM capabilities may assist organizations in articulating actions that impact their competitive priorities, which include Quality, Financial, and Costumer. For example, AM policy guarantees the development of AM actions aligned with organizational objectives, so the AM processes including manufacturing must follow the pattern established. In this sense, the effectiveness of the outcome of processes may deliver Quality.

Table 3. Hypotheses test results

H_1	AM Maturity → BP.	Support
H_2	AM Policy → Quality.	Support
H_3	AM Planning → Quality	Support
H_{4a}	Data and Information Management → Quality	Support
H_{4b}	Data and Information Management → Financial	Support
H_5	Asset Information System → Quality	Support
H_6	AM Leadership → Flexibility.	No support
H_7	Competence Management → Flexibility.	No support
H_8	Asset Performance → Customers' satisfaction	Support
H_8	Asset Performance → Quality	Support
H_9	Asset Costing and Valuation → Financial.	Support

In order to illustrate the applicability of the results, Figure 4 exemplifies the roadmap to implement actions in AM capabilities, which can be made using Industry 4.0 technologies, and its impact on competitive priorities. GFMAM (2024) and ISO (2014) have elaborated AM activities that compose AM capabilities, which may impact the competitive priorities. It was established indicators (Van Looy & Shafagatova, 2016) for each priority to monitor the effectiveness of AM actions.

Figure 4: Roadmap to implement actions in AM capabilities



4. Conclusions

Advancing in the value of AM practices, this paper has explored the importance of AM to competitiveness, bringing an overview of the relationship between AM capabilities and competitive priorities. Firstly, it was found that AM maturity affects BP, filling a relevant gap in AM literature. Secondly, it brought empirical evidence that AM capabilities contribute to competitiveness.

These findings allow leverage the decision-making in the AM context, drawing actions and programs that consider the relationships to improve in areas that are critical to competitiveness. So, this research begins with a landscape that can be adapted and expanded taking into account the objectives of the organization.

Considering this exploratory nature, it is fundamental to progress in other competitive priorities, by investigating new implications of AM capabilities on BP, as well as systemizing Industry 4.0

technologies that can contribute to an effective AM.

Acknowledgment

This study had partial support from FACEPE (Sciences and Technology Foundation of Pernambuco State - Brazil), the Coordination Unit for Improving the Qualifications of Higher Education Personnel – Brazil (CAPES) (grant 001) and the Brazilian Research Council (CNPq) (306860/2022-8) for all of which the authors are grateful

References

Cai, S., & Yang, Z. (2014). On the relationship between business environment and competitive priorities: The role of performance frontiers. *International Journal of Production Economics*, *151*, 131–145. https://doi.org/10.1016/j.ijpe.2014.02.005

Dash, G., & Paul, J. (2021). CB-SEM vs PLS-SEM methods for research in social sciences and technology forecasting. *Technological Forecasting and Social Change*, *173*. https://doi.org/10.1016/j.techfore.2021.121092

El-Akruti, K., Dwight, R., & Zhang, T. (2013). The strategic role of Engineering Asset Management. *International Journal of Production Economics*, *146*(1), 227–239. https://doi.org/10.1016/j.ijpe.2013.07.002

Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. *International Journal of Production Economics*, *210*, 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004

Gavrikova, E., Volkova, I., & Burda, Y. (2020). Strategic Aspects of Asset Management: An Overview of Current Research. *Sustainability*, *12*(15), 5955. https://doi.org/10.3390/su12155955

GFMAM. (2024). *The Asset Management Landscape* (Third Edition). Global Forum on Maintenance & Asset Management. https://gfmam.org/sites/default/files/2024-06/GFMAM_AM_Landscape_v3.0_English_2024.pdf

Hair, J. F. ., Hult, G. T. M. ., Ringle, C. M. ., & Sarstedt, Marko. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage.

Hair, J. F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. In *European Business Review* (Vol. 26, Issue 2, pp. 106–121). Emerald Group Publishing Ltd. https://doi.org/10.1108/EBR-10-2013-0128

Han, S., Li, C., Feng, W., Luo, Z., & Gupta, S. (2021). The effect of equipment management capability maturity on manufacturing performance. *Production Planning and Control*, *32*(16), 1352–1367. https://doi.org/10.1080/09537287.2020.1815246

IAM. (2024). *Asset Management - An anatomy*. https://theiam.org/media/5615/iam-anatomy-version-4-final.pdf

ISO. (2014). *ISO 55002:2018. Asset management — Management systems — Guidelines for the application of ISO 55001*. https://www.iso.org/standard/70402.html

Komonen, K., Kortelainen, H., & Räikkönen, M. (2012). Corporate Asset Management for Industrial Companies: An Integrated Business-Driven Approach. In *Asset Management* (pp. 47–63). Springer Netherlands. https://doi.org/10.1007/978-94-007-2724-3_4

Lima, E. S., McMahon, P., & Costa, A. P. C. S. (2021). Establishing the relationship between asset management and business performance. *International Journal of Production Economics*, *232*. https://doi.org/10.1016/j.ijpe.2020.107937

Lima, G. H. de A. L., & Costa, A. P. C. de S. (2023). Decision Support Procedure for Maturity Assessment in Asset Management. *Decision Support System in an Uncertain World: The Contribution of Digital Twins*.

Lima, G. H. de A. L., & Costa, A. P. C. S. (2024). Decision support system for maturity assessment in asset management. *24th International Conference on Group Decision and Negotiation & 10th International Conference on Decision Support System Technology*.

Maletič, D., Maletič, M., Al-Najjar, B., & Gomišček, B. (2018). Development of a Model Linking Physical Asset Management to Sustainability Performance: An Empirical Research. *Sustainability*, *10*(12), 4759. https://doi.org/10.3390/su10124759

Maletič, D., Maletič, M., Al-Najjar, B., & Gomišček, B. (2020). An analysis of physical asset management core practices and their influence on operational performance. *Sustainability (Switzerland)*, *12*(21), 1–20. https://doi.org/10.3390/su12219097

Rampini, L., & Re Cecconi, F. (2022). ARTIFICIAL INTELLIGENCE IN CONSTRUCTION ASSET MANAGEMENT: A REVIEW OF PRESENT STATUS, CHALLENGES AND FUTURE OPPORTUNITIES. *Journal of Information Technology in Construction*, *27*, 884–913. https://doi.org/10.36680/j.itcon.2022.043

Sandu, G., Varganova, O., & Samii, B. (2023). Managing physical assets: a systematic review and a sustainable perspective. *International Journal of Production Research*, *61*(19), 6652–6674. https://doi.org/10.1080/00207543.2022.2126019

Siemieniuch, C. E., & Sinclair, M. A. (2002). On complexity, process ownership and organisational learning in manufacturing organisations, from an ergonomics perspective. In *Applied Ergonomics* (Vol. 33).

Smith, R., & Sharif, N. (2007). Understanding and acquiring technology assets for global competition. *Technovation*, *27*(11), 643–649. https://doi.org/10.1016/j.technovation.2007.04.001

Tarhan, A., Turetken, O., & Reijers, H. A. (2016). Business process maturity models: A systematic literature review. *Information and Software Technology*, *75*, 122–134. https://doi.org/10.1016/j.infsof.2016.01.010

Van Looy, A., & Shafagatova, A. (2016). Business process performance measurement: a structured literature review of indicators, measures and metrics. *SpringerPlus*, *5*(1), 1797. https://doi.org/10.1186/s40064-016-3498-1

van Riel, W., Langeveld, J. G., Herder, P. M., & Clemens, F. H. L. R. (2014). Intuition and information in decision-making for sewer asset management. *Urban Water Journal*, *11*(6), 506–518. https://doi.org/10.1080/1573062X.2014.904903

Appendix 1 – BP questionnaire

Questions		Competitive priorities
BP1	Has Return on Assets increased above the industry average during the last 3 years?	Financial
BP2	Has the average lead time decreased during the last 3 years?	Quality Effectiveness
BP3	Has the percentage of internal scrap and rework decreased during the last 3 years?	Quality Environmental Productivity
BP4	Has the ability to introduce new products increased in the last 3 years?	Flexibility
BP5	Has market share increased during the last 3 years?	Costumers' Satisfaction Financial Effectiveness
BP6	Has on-time delivery performance improved during the last 3 years?	Effectiveness Productivity
BP7	Has the Unit cost of manufacturing decreased during the last 3 years?	Financial Productivity
BP8	Has flexibility to change product mix improved during the last 3 years?	Flexibility

BP9	Has the ability to handle emergencies increased during the last 3 years	Environmental
BP10	Has the improvement in product customization and reliability increased during the last 3 years?	Costumers' Satisfaction Quality
BP11	Has the consumption of resources decreased in the last 3 years?	Environmental Effectiveness Productivity
BP12	Has the return on investment increased above the industry average during the last 3 years?	Financial
BP13	Has sales growth increased above the industry average during the last 3 years?	Costumers' Satisfaction Financial
BP14	Has the growth in profit growth increased above industry average during the last 3 years?	Financial

A decision-support system applied to Law: Reasoning and explicability of the decision

Jérémy Bouche-Pillon [1*], Pascale Zaraté [1,2], Yannick Chevalier [1,3] and Nathalie Aussenac-Gilles [1]

[1] Université de Toulouse - CNRS - IRIT, France
[2] Université Toulouse Capitole, France
[3] Université Toulouse 3 Paul Sabatier, France
[*] jeremy.bouche-pillon@irit.fr

Abstract: The emergence of the digital transition brought an increasing need to control the processing of digital information, including in Law Enforcement Agencies (LEAs). At the EU level, in recent years, many regulations have emerged to control data processing and exchange. Texts other than the GDPR, such as the "Law Enforcement Directive (LED)", appeared to regulate specifically how Law Enforcement Agencies (LEAs) could process data. A formal representation of these regulations can be part of decision systems that support LEAs in processing data in compliance with the regulations. Although many new formalisms have emerged to represent legal norms and rules, few are provided with a reasoning mechanism. The explainability of the results of systems using these formalisms also remains a major issue. This paper describes a framework to operate formal rules from regulations and illustrate its integration in an interface to guide a user in its decision process in a situation of data processing by LEAs.

Keywords: decision-support; rule-based; decision trees; law compliance; explainability; user interface

1. Introduction

With the advent of the digital transition and the increase in data volumes in many domains, protecting the privacy of people's information and ensuring their lawful usage has become more and more complex. This has led to numerous new regulations, notably at the European level. Texts other than the GDPR (European Union, 2016), such as the "Law Enforcement Directive (LED)" (European Union, 2016), appeared to regulate specifically how Law Enforcement Agencies (LEAs) could process data. The increasing number of regulations makes it more and more challenging for law experts to work with so many sources. To assist them in their work, regulations started being formalized with purposes ranging from creating synthetic analyses to automatically reasoning over a set of rules. Although many formalisms have emerged in recent years to represent legal norms and rules, few are actually accompanied by a mechanism for reasoning on the rules. Another major issue when using them in a decision support system is the inability of such reasoning mechanisms to give a satisfying explanation of the decision. Moreover, the growing need to reason on increasingly complex information has led to store it no longer as simple "data" but as "knowledge", moving from relational databases to knowledge graphs (Lazarska, 2019; Medhi, 2017) along with all the semantic web technologies that come with it. Based on these observations, we propose a framework that supports the implementation of a decision support system to check the conformance of a situation to a set of rules taken from regulations. This system also explains the decision made, while integrating semantic web technologies. To illustrate the design and use of this framework, we selected a use case about checking the conformance of data processing in LEAs to several European regulations.

Part of this framework has been detailed in previous work (Bouche-Pillon et al., Dec. 2024). This paper specifically focuses on the following aspects of the framework: (i) The explicability of the results of

reasoning over the rules ; (ii) The integration of the decision-support framework in an interface and how a user interacts with it.

2. Decision-support framework

Given the lack of explicable reasoning frameworks over formal legal rules, this article proposes a decision support framework based on Marakas' model (Marakas, 2003), which will ensure both the operability of formal rules and the explicability of reasoning results. The framework, described in a previous paper (Bouche-Pillon et al., Dec. 2024), is illustrated in Figure 1.

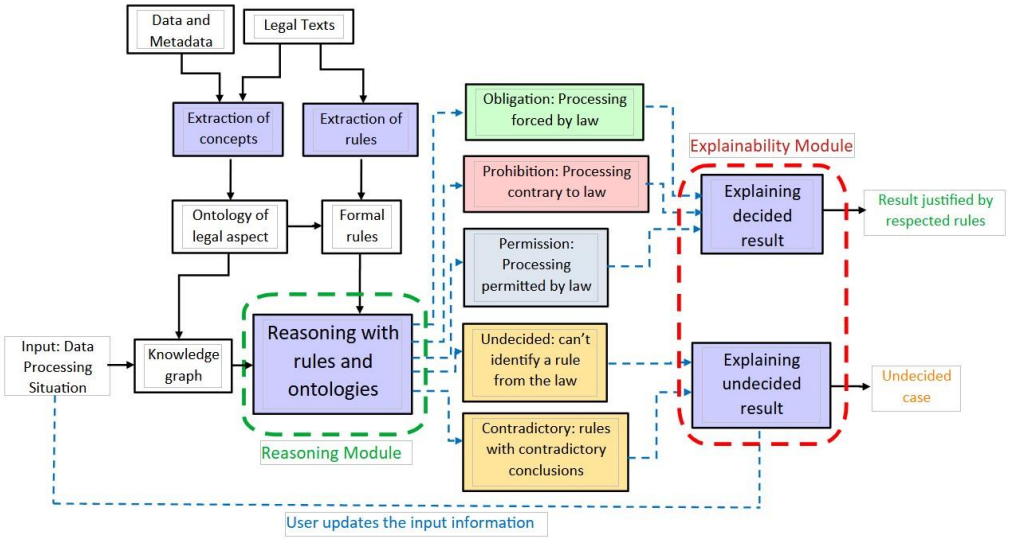


Figure 1: Framework architecture

The framework is composed of the following elements: (i) Formal rules are extracted from the regulations and expressed in SPARQL[25], using an ontology (Bouche-Pillon et al., July 2024) which serves as a model of the knowledge graph containing the data and metadata to reason about. (ii) The input is an action the legality of which is to be checked and the context in which it occurs. Users describe it through a form processed into TriG[26] and included in the knowledge base. (iii) A first module performs reasoning over the input context based on the formal rules and concludes one of the five possible outputs: "Obligation", "Prohibition", "Permission", "Undecided" and "Contradictory". The first three outputs are decided cases where coherent rules are respected. The last two outputs are undecided cases where either no rule is respected or several rules are respected but incoherent. (iv) A second module is in charge of justifying the outputs from the reasoning module, with a distinction between whether or not the result or reasoning is decided.

3. Formal Rules

To adapt the framework to the use case, it is necessary to formalize the relevant regulations. Users select the appropriate applicable regulation according to the context to be checked, from which a set of rules is extracted and then formalized in the framework. The rules obtained that way are "explicit" and can be automatically extracted. There are however other rules that are not explicitly present in regulations yet are essential to take into account. These "implicit" rules serve as "default cases" or priority rules defining when some rules should prevail on others.

Both the rules and the input contexts are represented as a graph using the concepts and properties of the ontology (Bouche-Pillon et al., July 2024). SPARQL is used as rule formalization language, since many

[25] https://www.w3.org/TR/sparql11-query

[26] https://www.w3.org/TR/rdf12-trig/

reasoning engines and endpoints such as GraphDB[27] or Apache Jena Fuseki[28] exist to reason directly over knowledge graph content. However, for clarity reasons, in this paper, the formalization will be presented in a First-Order Logic (FOL) form following a *Conditions* → *Effect* syntax. The SPARQL equivalents of the FOL expressions can be found on a Github repository[29].

3.1. Extract explicit rules

Explicit rules are those that can be directly extracted from the regulations in natural language. Although extraction is currently manual, rule extraction from regulations can be automated using LLMs and semantic analysis (Fawei, 2024; Ferraro et al., 2020; Recski et al., 2021), which we plan to do in future works. The formalization principle applied in this framework is adapted from Gandon et al.'s work (2017), where each rule extracted from the regulation does not indicate directly whether an action is permitted, prohibited or mandatory. Instead each rule is classified as either permission, obligation or prohibition and the reasoning aims at assessing the compliance of the input situation to each rule. Moreover, as explained in a previous paper (Bouche-Pillon et al., Dec. 2024), rules are split in two, with a first part indicating the applicability of the rule to the input situation and a second part indicating the respect of the rule by the situation. To illustrate this principle, the analysis of the article in Figure 2 would decompose as follows:

Article 10

Processing of special categories of personal data

Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be allowed only where strictly necessary, subject to appropriate safeguards for the rights and freedoms of the data subject, and only:

(a) where authorised by Union or Member State law;

(b) to protect the vital interests of the data subject or of another natural person; or

(c) where such processing relates to data which are manifestly made public by the data subject.

Figure 2: Example of law article: article 10 from 2016/680/UE directive (LED)

(i) The deontic class of this article is "Permission", as indicated by the terms "shall be allowed". (ii) The object of the rule relates to the processing of "sensitive personal data". (iii) The conditions of this rule are "the strict necessity" of the processing, the "safeguard of rights and freedom of the data subject" and a disjunction of conditions: "allowed by Union or Member State" OR "protection of vital interests" OR "the data are public". From this analysis, the "permission" aspect of the rule is expressed by the axiom *isPermision(LED10)* and the rest can be formalized as the following 2 rules, in FOL:

1) $Processing(action) \wedge InvolvesDataset(action, dataset) \wedge ContainsData(dataset, data) \wedge SensitivePersonalData(data) \rightarrow IsApplicable(LED10, situation)$

2) $IsApplicable(LED10, situation) \wedge Action(action) \wedge InvolvesDataset(action, dataset) \wedge Necessary(action) \wedge SafeguardRights(action) \wedge (AuthorizedLaw(action) \vee ProtectsVitalInterest(action) \vee (\forall data, SensitivePersonalData(data) \Rightarrow PublicData(data)) \rightarrow HasCompliance(LED10, situation)$

Since there is no universal quantifier in SPARQL, there is a slight difference between the FOL expression and its SPARQL equivalent: rather than expressing that "all personal data are public", the

[27] https://graphdb.ontotext.com

[28] https://jena.apache.org/documentation/fuseki2/

[29] https://github.com/JeremyBOUCHEPILLON/legalDataProcessing/blob/main/rules/

SPARQL expression states that “there is no public data that is not public”. However, it appears that only considering the explicit rules that can be extracted from the regulations does not provide a full cover of possible situations. A lot of implicit statements can be derived from the explicit ones and need to be identified.

3.2. Identify implicit rules and conflict solving rules

In addition to the explicit rules, it is necessary to make explicit, with the help of experts, rules that are either implicit or that allow to solve conflicts between other rules. Those conflicts can have several causes. For example one rule may be an exception to another, or a legal doctrine defining priority between two rules may have not been applied (lex superior, lex posterior. This can be illustrated with the article in Figure 2. Once analyzed, this article states that “the processing of sensitive personal data is permitted ONLY where some conditions are met”. Intuitively, this suggests that there is an implicit “default” rule stating that “the processing of sensitive personal data is forbidden” and the rule from the LED is an exception to this default rule. The formalization of “default” rule is based on the axiom *isProhibition(LED10_default)* and the FOL expression:

$$Processing(action) \wedge InvolvesDataset(action, dataset) \wedge ContainsData(dataset, data) \wedge SensitivePersonalData(data) \rightarrow HasCompliance(LED10\_default, situation)$$

And the rule that resolves the conflict states that if both rules are complied with, only the exception has to be considered (represented by negating the compliance with the default rule here):
$$HasCompliance(LED10\_default, situation) \wedge HasCompliance(LED10, situation) \rightarrow \neg HasCompliance(LED10\_default, situation)$$

4. Reasoning and explainability

The extraction of explicit rules and the identification of implicit, exceptions and priority solving rules generates a full rule base on which to reason. The reasoning itself is decomposed in two steps: (i) A first step where the input situation is checked against explicit and implicit rules. This is done by executing all the corresponding SPARQL requests and compiling the resulting generated knowledge in the situation knowledge graph. This step generates, for the input situation, a list of all applicable rules and a list of all respected rules. It can be noted that the list of “respected rules” is a sublist of the “applicable rules” list. (ii) The exception and priority rules are then applied to the “respected rules” list, to try solving potential conflicts in the respected rules by removing problematic rules from the list. The last step of the framework provides justifications to explain the result obtained from the reasoning.

4.1. General algorithm for explainability

The explanation given to the user will depend on the type of results from the reasoning phase, following the logic illustrated in Algorithm 1. In most cases, when all the respected rules are coherent, the explanation is straightforward and consists of informing the user of the regulation parts that support the result. However, there are two problematic cases. The first one is when there are still contradictory rules among the respected ones, even after the second step of reasoning where conflict solving rules were applied. This indicates that such rules are still missing in the rule base. The result returned to the user highlights the contradiction, indicating the conflicting rules and asking him which rule should prevail. The answer will be added as a new temporary conflict solving rule in the system, waiting for validation from an expert. The second problematic case is when not a single rule from the rule base has been respected, yet there are still applicable rules, indicating that the input situation is in the scope of application of the verified regulations. This specific case will trigger an interaction with the user to try to improve the results.

Algorithm 1: Explainability of the reasoning's results

Input: *The result from the reasoning :(listApplicableRules, listRespectedRules)*
Output: *Explained result in the form : (result, explanation)*
Initialization of variables: \
if *listApplicableRules is empty* ***then*** /*There are no applicable rules*/
    return *"The situation is outside the scope of applicability of the verified regulations"*
else if *listRespectedRules is empty* ***then*** /*None of the applicable rules are respected*/
    result ← undecided
    try to improve the result /*Involves interaction with the user*/
else
    if *all rules in listRespectedRules are Prohibition* ***then***
        return *(Prohibition, listRespectedRules)*
    else if *all rules in listRespectedRules are Permission* ***then***
        return *(Permission, listRespectedRules)*
    else if *all rules in listRespectedRules are either Obligation or Permission* ***then***
        return *(Obligation, listRespectedRules)*
    else */* There are still contradictions in the respected rules*/*
        return *(Contradictory, listRespectedRules)*
    end
end

4.2. Explicability of undecided cases

In the case where no rule from the ruleset is respected, the system will ask the user for complementary information regarding the input situation. In order to determine which information to request from the user, the idea is to check what part of the rules that were "the closest to being complied with" were not respected. To do this, the following procedure is applied: (i) Consider only the rules "applicable" to the situation. (ii) For each of these rules, generate a binary decision tree where each node is one of the conditions of the rule, the left edge of a node corresponds to "the condition of the node is respected" and the right edge to "the condition of the node is not respected. An example applied to article 10 of the LED is illustrated Figure 3. (iii) Confront the input situation to each tree and keep a track of the non respected conditions in them. (iv) sort the rules depending on how deep in their decision tree the verification went. (v) Use the non respected conditions in each rule to ask the user complementary information regarding the situation in input.

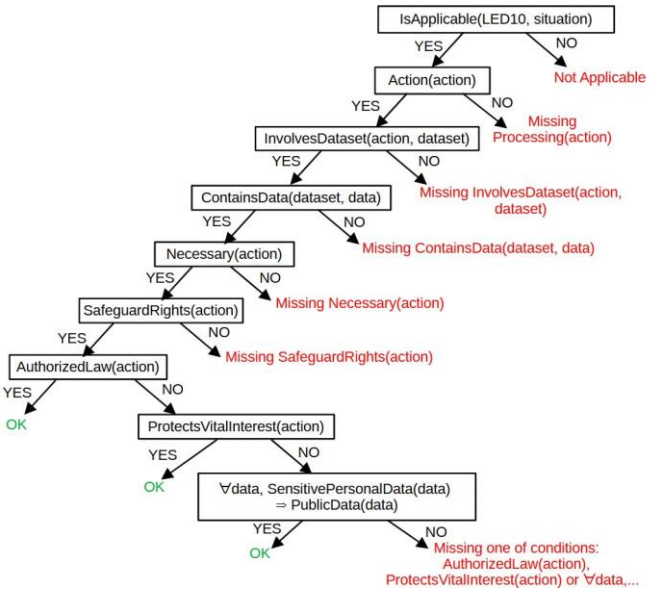


Figure 3: Decision tree corresponding to the Article 10 of the LED

If the user is able to give supplementary information, the system processes the updated input information from the beginning of the framework, hoping to obtain a decision. If after this new processing the reasoning fails to find respected rules again or if the user can't give supplementary information, the case is classified as "undecided". Currently, the explainability part of the framework is implemented up until the fourth step.

5. Integration in a decision process

The framework aims at being used in a decision process. A prototype of integration is currently being developed in Python, using Jena Fuseki for the storage and interfacing of the knowledge base. To illustrate its feasibility, it is tested in a use case of data processing in and between European Law Enforcement Agencies (LEAs). Whenever a data controller in a LEA is faced with a data processing situation, he has to take a decision while ensuring that this action is lawful. For example, in a situation where a data controller wants to know if he can transfer some data regarding a suspect in an investigation case from a storage bank to another, he can enter all the relevant information using an interface like the one in Figure 4a.

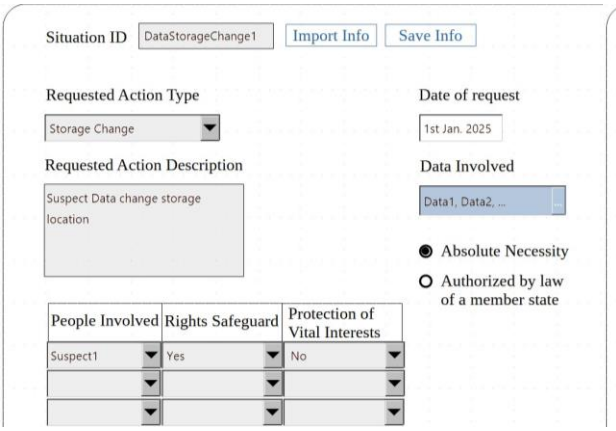


Figure 4a: Example of interface

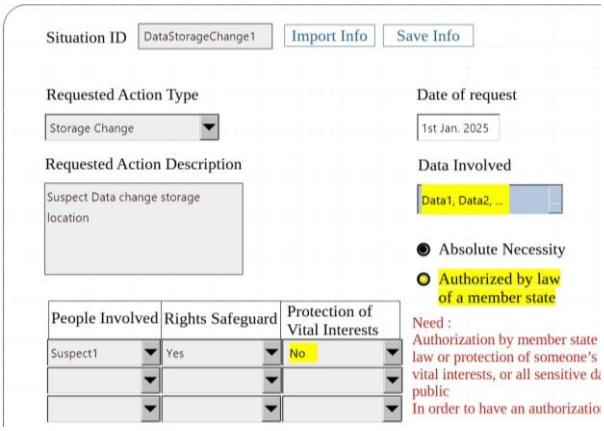


Figure 4b: Example of highlighted interface

These data are processed and added to the knowledge base of the framework as a new named graph. The reasoning is then applied to this new named graph, using the SPARQL rules and following the process detailed in section 4. The algorithm for explainability finally links the results from the reasoning to a justification that will both be returned to the data controller, or, in "undecided" cases, the data controller will be asked supplementary information, by highlighting in the interface the fields of the form that would need a different value and giving indications ("Need" text in red) on what changes would be required in order to increase the chances of getting a "decided" case, as illustrated Figure 4b.

6. Conclusions

This paper presented the reasoning and justification aspects of a framework for decision support systems as well as its integration in a decision process, illustrated by the use case of data processing in and between European Law Enforcement Agencies. Future works involve completing the implementation of the explainability part of the framework to test the completed version. It also involves extending the number of formal rules in the framework, by using state-of-the-art Natural Language Processing methods (Fawei, 2024; Ferraro et al., 2020; Recski et al., 2021) to new directives that have come into force in 2023 regarding the data processing by LEAs (European Union, 2023). Finally, while the formal language used in this study is SPARQL, other formal standards like LegalRuleML (Palmirani, 2011) could be tested to compare the results. A more complex case study will also be tested.

Acknowledgments

This work is partially funded by the H2020 project STARLIGHT ("Sustainable Autonomy and Resilience for LEAs using AI against High priority Threats") that received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101021797. We would also like to thank Ronan PONS, PhD student in Law, who assisted this work by providing his insight as a legal expert.

References

Bouché-Pillon, J., Aussenac-Gilles, N., Chevalier, Y., & Zaraté, P. (2024, July). An ontology for legal reasoning on data sharing and processing between law enforcement agencies. *3rd international workshop Knowledge Management and Process Mining for Law* (KM4LAW 2024), IAOA, Jul 2024, Enschede, Netherlands. ⟨hal-04654770⟩

Bouche-Pillon, J., Aussenac-Gilles, N., Chevalier, Y., & Zarate, P. (2024, December). Decision Support in Law: From Formalizing Rules to Reasoning with Justification. *JURIX 2024: The Thirty-seventh Annual Conference*, J. Savelka; J. Harasta; T. Novotna; J. Misek, Dec 2024, Brno, Czech Republic. ⟨10.3233/faia241231⟩.

European Union. (2016, May 04). *Regulation - 2016/679 - EN - gdpr - EUR-Lex.europa.eu.* https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

European Union. (2016, May 04). 2016/680 - EN - Law Enforcement Directive; LED - EUR-Lex. https://eur-lex.europa.eu/eli/dir/2016/680/oj/eng

European Union. (2023, May 22). Directive - 2023/977 - EN - EUR-Lex. EU. https://eur-lex.europa.eu/legal-content/fr/ALL/?uri=oj:JOL_2023_134_R_0001

Fawei, B. (2024). NLP-Based Rule Learning from Legal Text for Question Answering. *Asian Journal of Research in Computer Science*, 17(7), 31–40. https://doi.org/10.9734/ajrcos/2024/v17i7475

Ferraro, G., Lam, H. P., Tosatto, S. C., Olivieri, F., Islam, M. B., van Beest, N., & Governatori, G. (2020). Automatic extraction of legal norms: Evaluation of natural language processing tools. In *New Frontiers in*

Artificial Intelligence: JSAI-isAI International Workshops, JURISIN, AI-Biz, LENLS, Kansei-AI, Yokohama, Japan, November 10–12, 2019, Revised Selected Papers 10 (pp. 64-81). Springer International Publishing.

Gandon, F., Governatori, G., & Villata, S. (2017). Normative requirements as linked data. In *Legal Knowledge and Information Systems* (pp. 1-10). IOS Press.

Lazarska, M., & Siedlecka-Lamch, O. (2019, November). Comparative study of relational and graph databases. In *2019 IEEE 15th International Scientific Conference on Informatics* (pp. 000363-000370). IEEE.

Marakas, G.M. (2003). *Decision Support Systems in the 21st Century*. Prentice Hall

Medhi, S., & Baruah, H. K. (2017). Relational database and graph database: A comparative analysis. *Journal of process management and new technologies*, 5(2).

Palmirani, M., Governatori, G., Rotolo, A., Tabet, S., Boley, H., & Paschke, A. (2011, November). Legalruleml: Xml-based rules and norms. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web* (pp. 298-312). Berlin, Heidelberg: Springer Berlin Heidelberg.

Recski, G., Lellmann, B., Kovacs, A., & Hanbury, A. (2021, June). Explainable Rule Extraction via Semantic Graphs. In *ASAIL/LegalAIIA@ ICAIL* (pp. 24-35).

Assessing consumer perceptions and attitudes towards Cultured Meat using X (Twitter) data: implications for decision support

Guoste Pivoraite[1*], Shaofeng Liu[1], Saeyeon Roh[1] and Guoqing Zhao[2]

[1] Plymouth Business School, University of Plymouth, Plymouth, PL4 8AA, UK
[2] School of Management, Swansea University, Swansea, SA1 8EN, UK
guoste.pivoraite@plymouth.ac.uk

Abstract: Cultured meat (CM) is considered to have a potential to address environmental, sustainability, ethical concerns, and food security, however, consumer acceptance remains a challenge. This study empirically validates and extends the Cultured Meat Attitude and Perception Assessment (CAPA) model by analysing public discussions on X (Twitter) to understand consumer perceptions of CM. Using BERTopic for topic modelling and qualitative analysis, the study explored unprompted public narratives, mapping them to CAPA constructs to examine how public opinions align with or challenge existing theoretical concepts. The findings revealed a combination of cultural, political, ethical, and economic narratives shaping consumer behaviour and their decision-making processes related to CM acceptance. Existing constructs such as Information Influence, Product Attributes, Ethical Issues, and Social Influences were validated. However, the emergence of themes and sub-themes such as Conspiracy Theories, Political Resistance, Specific Health Risk Narratives, and Economic and Market Forces suggests the need to refine the model providing a more comprehensive framework for understanding consumer behaviour. This study contributes to knowledge by offering a deeper understanding of the factors influencing CM acceptance, validating and extending the CAPA model, and demonstrating the effectiveness of combining machine learning with qualitative analysis. It highlights the importance of transparent communication to counter misinformation and ideological resistance, as well as the need for strategic communication for product positioning and policy development. This approach advances academic understanding and provides actionable insights to inform decision-making in the commercialisation of sustainable and ethical food innovations.

Keywords: Cultured meat; decision-making; opinion mining; social media analytics; natural language processing

1. Introduction

Cultured meat (CM), sometimes referred to as lab-grown, cell-based, or cultivated meat, represents a technological advancement within the alternative protein sector. By cultivating animal cells in vitro rather than relying on conventional livestock, CM aims to reduce greenhouse gas emissions, mitigate ethical concerns related to animal welfare, and address food security issues (Deliza et al., 2023). As of 2023, over 150 companies worldwide are actively engaged in developing CM products, reflecting a surge in interest and investment in this emerging sector (Wunsch, 2024). Market projections show the potential of CM within the global meat industry. Estimates suggest that by 2040, CM could constitute over 30% of the worldwide meat market, translating to a market value of approximately 1.8 billion U.S. dollars (Wunsch, 2024). Although plant-based alternatives have gained public attention, the significance of CM lies in its potential to more closely mimic the sensory and nutritional attributes of conventional meat (Lewisch & Riefler, 2023).

However, consumer acceptance remains a challenge, largely influenced by perceptions of unnaturalness, health risks, cultural beliefs, political debates, and ethical ambiguity (Hanan et al., 2024; Monaco et al., 2024). Existing research on consumer attitudes toward CM has primarily relied on structured surveys and controlled experiments, which may overlook the complexity and authenticity of

unprompted public opinions shared on social media platforms (Onwezen & Dagevos, 2024; Tsvakirai et al., 2024). Additionally, theoretical frameworks, such as the Cultured Meat Attitude and Perception Assessment (CAPA) model (Pivoraite et al., 2024), still lack empirical validation through real-world data.

Addressing this gap, the current study empirically validates and extends the CAPA framework by analysing public narratives about CM on X (Twitter). Unlike prior studies that often applied predefined categories or media frames (Pilařová et al., 2022; Kouarfaté & Durif, 2023), this research captures naturally occurring consumer discussions without imposing theoretical preconceptions. The study uses BERTopic topic modelling and qualitative analysis to map these narratives onto CAPA constructs, identifying both alignment and emerging themes. Preliminary analysis covers the first 50 out of 267 topics from a dataset of 13,362 tweets, collected between July and October 2024.

The findings contribute to a richer understanding of factors shaping CM acceptance and provide insights for strategic communication, market positioning, and policy-making to support sustainable and ethical food innovations.

2. Literature Review

2.1 Consumer perceptions and acceptance of Cultured Meat

Consumer acceptance of CM is shaped by multiple factors, including knowledge, cultural identity, political ideologies, and ethical beliefs. Studies indicate that ethical and environmental motivations positively influence willingness to try CM however, concerns about unnaturalness, safety, and ethical ambiguity pose challenges for widespread acceptance (Hanan et al., 2024; Monaco et al., 2024). In particular, perceptions of artificiality and health risks, such as associations with genetic manipulation or carcinogenic effects, amplify distrust and resistance (Siddiqui et al., 2022; Lewisch & Riefler, 2024). Cultural identity and political ideologies notably shape public perceptions too, with certain cultural groups viewing CM as a threat to traditional food practices and cultural norms (Mancini & Antonioli 2019; Tsvakirai, 2024). Additionally, consumer acceptance varies across regions due to social norms, religious beliefs, and political affiliations, showing the need to understand cultural narratives influencing public sentiment (Bryant, 2020; Xanthe Lin et al., 2025). This highlights the importance of exploring how unstructured public narratives, particularly on social media, shape perceptions of CM.

2.2 The CAPA model and theoretical constructs

The CAPA framework (Figure 1), proposed by Pivoraite et al. (2024), builds on traditional models like the Theory of Planned Behaviour (Ajzen, 1991) but extends beyond attitudes, subjective norms, and perceived behavioural control. CAPA integrates multidimensional knowledge (awareness, familiarity, comprehension), emotional responses (e.g., curiosity, distrust, neophobia), and a range of external influences (ethical, cultural, marketing, regulatory, etc.) to capture the complexity of consumer attitudes toward CM. CAPA is particularly relevant for emerging food technologies such as CM, because it considers emotional responses and contextual factors that shape consumer perceptions, beyond purely practical benefits (Pivoraite et al., 2024).

Despite its theoretical foundation, the CAPA model would benefit from empirical validation, such as using unstructured social media data to map public narratives to its constructs. This approach captures authentic, unprompted consumer narratives that are difficult to obtain through traditional methods, offering a better understanding of public perceptions (Laestadius & Caldwell, 2015; Pilařová et al., 2022; Kouarfate & Durif, 2023; Pivoraite et al., 2024). Some studies have explored consumer perceptions using social media analytics, primarily focusing on media framing rather than systematically mapping narratives to theoretical constructs (Pilařová et al., 2022; Kouarfate & Durif, 2023). This study addresses this gap by validating and potentially extending the CAPA model using opinions from X (Twitter). It employs Natural Language Processing (NLP) techniques, including

BERTopic (a machine learning model for topic discovery) to identify topics from unstructured social media data (Egger & Yu, 2022). In this preliminary stage, 50 out of 267 topics were analysed, with the remaining topics and sentiment analysis planned for subsequent stages. The sentiment analysis will include overall sentiment, theme-specific sentiment (based on the thematic analysis), negative sentiment clustering (to reveal potential new themes), and customer segmentation (e.g., sceptics, optimists, etc.). This approach is novel in CM studies that utilised social media data, which have primarily focused on media framing and thematic analysis rather than systematically linking thematic constructs to sentiment patterns (Laestadius & Caldwell, 2015; Pilařová et al., 2022; Kouarfate & Durif, 2023).

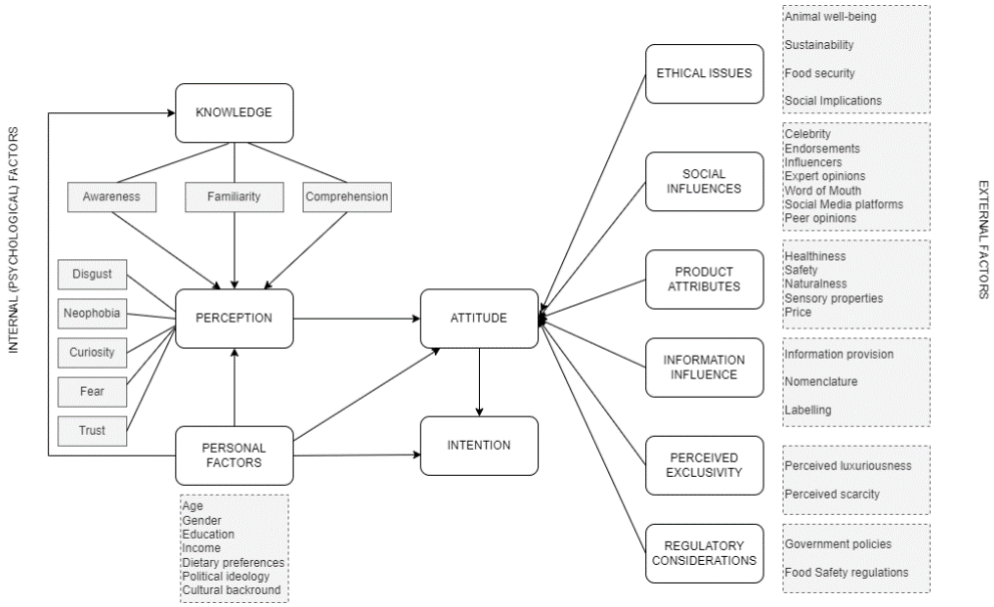


Figure 1. *Cultured Meat Attitude and Perception Assessment (CAPA) framework* (Pivoraite et al., 2024, p. 117)

By mapping public narratives to CAPA model constructs through thematic analysis, followed by sentiment analysis at the later stages, this study offers a deeper understanding of consumer views and their influence on CM acceptance, providing strategic insights for industry stakeholders in communication strategies, product positioning, and policy advocacy, ultimately supporting responsible commercialisation and sustainable adoption of CM.

3. Methodology

3.1 Methodological rationale

This study adopted a mixed-methods approach combining social media analytics, BERTopic topic modelling, and qualitative interpretation to explore consumer perceptions and attitudes toward cultured meat. X (Twitter) was chosen as the data source for its capacity to capture unprompted public narratives, providing a more authentic view of consumer perceptions compared to structured surveys (Specht et al., 2020). BERTopic was selected over traditional models like Latent Dirichlet Allocation (LDA) due to its strong performance with short social media texts, using contextual embeddings to better understand the meaning behind words (Egger & Yu, 2022). A qualitative interpretation was employed to map emergent themes to CAPA constructs, ensuring conceptual alignment and enabling theoretical validation and extension of the framework. Figure 2 provides an overview of the methodological workflow, illustrating the sequential process from data collection to thematic and sentiment analysis, integrated with the CAPA framework for contextualisation.

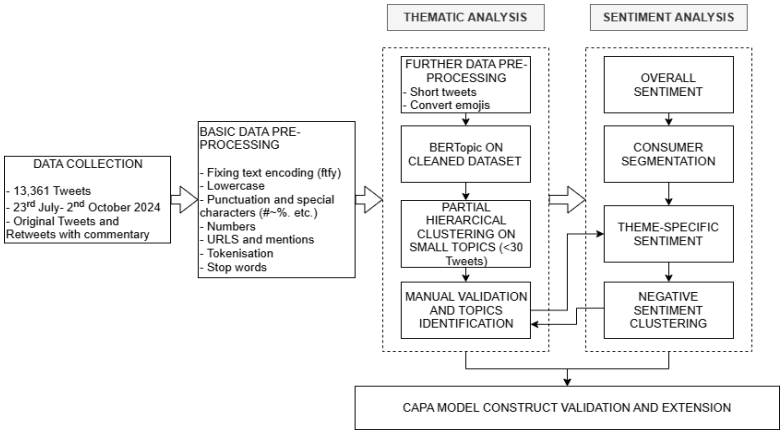


Figure 2. Methodology process

This approach helped to identify relevant themes while allowing flexibility in discovering new narratives. By combining social media analytics with some NLP techniques and manual qualitative analysis, the study captured genuine public conversations, leading to more meaningful findings (González-Rostani, Incio & Lezama, 2024; Mendonça & Figueira, 2024). This provides a foundation for understanding consumer perceptions of cultured meat and offers evidence that can inform both theory and practice (Mendonça & Figueira, 2024).

3.2 Data collection

X (Twitter) data was collected between July and October 2024. Tweets and retweets containing commentary were gathered through Twitter's API using keywords "lab grown meat," "lab meat," "cell cultured meat," "cultivated meat," "cultured meat," "in vitro meat," "factory grown meat," "cell-based meat," and "bioengineered meat." A total of 13,362 tweets were obtained.

3.3 Preprocessing and data cleaning

Data preprocessing followed established social media analytics protocols, including lowercasing, removal of URLs, punctuation, and numeric strings, as well as converting emojis to text for semantic consistency (Specht et al., 2020). Tokenisation and lemmatisation were applied to standardise word variants (e.g., "eats" to "eat"). Tweets with fewer than five words were discarded to maintain contextual integrity (Boot et al., 2019; Tommasel & Godoy, 2019). While X's (Twitter) Terms of Service acknowledge the public nature of posted content, allowing for its use in research (X, 2024), ethical standards necessitate protecting user privacy (Takats et al., 2022). To balance data utility with confidentiality, original tweet IDs were replaced with unique codes, enabling internal tracking without exposing user identities.

3.4 Topic modelling and thematic analysis

This study used BERTopic technique for topic modelling to analyse public opinions because of its effectiveness with short social media texts and its ability to capture different meanings (Egger & Yu, 2022) which grouped tweets into 267 distinct topics based on semantic similarity, identifying both more notable and more subtle narratives in this study. BERTopic assessed topic clarity using class-based TF-IDF (c-TF-IDF), an algorithm highlighting words uniquely representative of each topic. Resulting keywords were manually reviewed for coherence, improving topic interpretation and ensuring meaningfulness and consistency. The model automatically discovered patterns without requiring a predefined number of topics, allowing for a data-driven exploration of public discussions, aligned with previous research methodologies (An et al., 2023; Ju, 2024).

In this context, "topics" are machine-generated clusters of semantically related words and tweets. These clusters were represented by three elements- representation (a summary of the most relevant words that capture the essence of each topic), top words (the most frequent and contextually significant terms

within each cluster), and representative tweets (examples that best illustrate the narrative of each topic) which were identified by the model, with top 10 tweets automatically selected to reflect each topic's core idea. The first 50 topics analysed were chosen by their size, prioritising the largest topics to reflect dominant narratives, with analysis of the remaining 217 topics planned to refine or adjust these insights.

After topic modelling, a thematic analysis was conducted to organise the topics into broader themes and sub-themes. In this study, "themes" refer to broader interpretative categories that grouped related topics to provide a better understanding of public perceptions, while "sub-themes" captured specific aspects within each theme. For example, topics related to health risks, particularly cancer and tumor fears, as well as safety concerns, were grouped under theme "Health risks", with the sub-theme "Long-term health risks" and "Uncertainty". This categorisation helped to maintain the detailed focus of each topic while presenting a clearer, more organised view of public narratives.

This initial analysis covered 50 topics, providing basis for more detailed thematic and sentiment analyses planned for later stages. By starting with topic modelling and then applying thematic analysis, the study identified data-driven themes while also allowing for broader interpretation. This approach facilitated the systematic mapping of public opinions to the CAPA model constructs, supporting theoretical validation and the potential extension of the CAPA framework (Braun & Clarke, 2006; Nowell et al., 2017).

3.5 CAPA contextualisation and qualitative review

To contextualise the findings within the CAPA framework, a qualitative review of the BERTopic-generated topics was conducted. This process aimed to map public opinions to existing CAPA constructs while also identifying emerging narratives that could extend the model. The review involved manually examining topics to evaluate their contextual relevance and alignment with CAPA constructs. The review assessed the top words and representative tweets identified by the model. Top words were manually reviewed to confirm they accurately captured the essence of each topic, while representative tweets were evaluated for contextual accuracy and thematic relevance. Tweets were prioritised if they clearly expressed the main idea of the topic and related to a CAPA construct or emerging narrative. For example, tweets mentioning safety concerns or cancer risks were mapped to Product Attributes (Safety), while tweets linking cultured meat to conspiracy theories involving public figures were mapped to Information Influence (Misinformation and Mistrust). Decisions were made by examining the semantic context and emotional tone of tweets where ambiguous tweets were reviewed alongside others in the same topic to clarify the broader narrative. Topics that were too similar or unclear were combined or adjusted for clarity. Emerging narratives that did not fit existing constructs were documented as potential new themes, supporting the extension of the CAPA model.

This process was iterative and flexible, refining topics to fit the CAPA framework or identifying them as new themes when needed. By combining human judgement with machine-generated insights, the study improved the accuracy and consistency of the analysis, supporting both the validation and expansion of the CAPA framework.

4. Results and Discussion

4.1 Thematic analysis and CAPA mapping

Thematic analysis using BERTopic uncovered a range of narratives that were mapped to CAPA model constructs and revealed a combination of cultural, political, ethical, and economic factors influencing consumer perceptions of cultured meat. The initial 50 topics were grouped into broader themes corresponding to *Information influence, Product attributes, Ethical issues, Social influences,* and *Economic/market forces* (Table 1). This mapping facilitated a systematic examination of how public narratives align with or challenge the theoretical constructs within the CAPA model. The first 50 topics revealed one potential new construct called Economic/Market Forces, and six sub-themes not explicitly

present in the original CAPA model, all of which are summarised in Table 1. Only new insights were included in the table. Themes that reinforced existing CAPA constructs were excluded to focus on additions or refinements.

Table 1. Emerging constructs and sub-themes from initial 50 topics

Theme	CAPA construct	Sub-theme	Representative tweets
Conspiracy theories	Information influence (existing)	Misinformation and mistrust (new)	• "Stay minimum 300 meters away from any mRNA gene therapy products. Same goes for lab grown meat, bugs and UN/WEF orchestrated PsyOps propaganda." • "You need to add to that poll investigate the fda, cia, fbi, nsa, wef, who, and every other political organization! The FDA has been using chemicals that are poisoning our food. Get rid of lab grown meat and vegetables. Also convict Fauci and bill gates" • "If Bill Gates creating lab grown meat you guys need to look into this as something serious you can remotely control a human like this CIA Secret Pentagon project program funded Besides other plans to completely remove the Bible from society Ben all guns"
Insect and bug comparison	Product attributes (existing)	Unnaturalness and impurity (new)	• "What's on the buffet menu, is it fake lab meat and organic lettuce with a side of insects?" • "Lab meat is chemical carbslop that would be worse than eating rats and insects for dinner" • "We do not want toxic insects and lab meat! We want vegetables, fruits, real meat and fish!" • "lab grown meat = insects"
Health risks (cancer, tumor)	Product attributes- Safety (existing)	Long-term health risks and uncertainty (new)	• "It is a known fact that lab grown meat is cancer causing! NOT SAFE!" • "But lab grown meat causes cancer, wtf are we expecting from this fake food. More BS from the FDA!" • "Cultivated meat is the same as cancer, immortal/infinite cell growth" • "lab grown meat is literally just tumors" • "lab grown tumors! is the same mechanism to grow "meat" in a lab"
GMO and ethical concerns	Ethical issues (existing)	Artificiality and manipulation of nature (new)	• "GMOs create lesser quality but larger quantities, lab grown meat isnt even required by law for half of our states to mark on packaging, and this "meat" isn't supported by a body." • "The issue with GM and gene editing of organisms which are capable of reproducing, from a scientific point of view, is that if a mistake is made you cannot put the genie back in the bottle" • "Somebody need to explain how our government allowing genetically modified food and now lab grown meat to be given to us Americans in the United States. This stuff can wipe out all of us"
Political resistance and cultural identity	Social influences (existing)	Political opposition and cultural norms (new)	• "Vote for make America great again even if you don't like Trump you're voting for policies that'll give you security, closed borders low cost food and low cost gas, no wars. You vote for Harris you will be eating Bill Gates lab meat and bugs." • Here comes the lab grown meat from Israel that has been rejected by many country's leaders who truly lobes their people! <…>. South Africa needs to act. URGENTLY!!! • "Hungary and Italy join forces to chop up lab-grown meat. Italy's Meloni and Hungary's Orban are fighting hand-in-hand with some major agricultural interest groups to discredit what they call Frankenstein meat"
Economic disruption and market dynamics	Economic/ market forces (new)	Price disruption and supply-demand dynamics (new)	• "It will end because the meat and dairy industry is being disrupted by precision fermentation and lab grown meat." • "that's the question supply and demand can the demand be met by the supply? do we have enough lab grown meat to meet the demand for it ..." • "Lab grown meat is one of the technologies that will enable abundance" • "If the farm lobby gets the process or source of the lab grown meat outlawed then there will be no industrial lab growers capable of competing. The farms would be able to raise the price as they run out of arable land to compensate until a balanced diet is prohibitively expensive."

Several recurring narratives stood out as influencing public perceptions of CM. These narratives were mapped to CAPA model constructs, providing a systematic examination of consumer attitudes and highlighting the complexity of public discourse surrounding this emerging food technology. *Conspiracy*

theories and misinformation was one of the most prominent narratives which linked cultured meat to global control agendas, involving influential figures like Bill Gates and organisations such as the CIA. These narratives framed CM as an instrument of control, reflecting deep-seated distrust toward powerful entities and technological advancements. This theme was mapped to "Information influence" as it illustrates how misinformation influences public perception by amplifying distrust. Then, *Insect and bug comparison* was another narrative that leveraged cultural food worries to evoke disgust and perceptions of impurity. It positioned CM as unnatural, appealing to deep-seated neophobic responses. This theme was aligned with the "Product attributes" (unnaturalness) construct in CAPA, as it emphasises consumer perceptions of contamination and impurity. *Health risks and safety concerns* frequently associated cultured meat with cancer, tumours, and other long-term health risks, reflecting public distrust in novel food technologies and anxiety about unforeseen health consequences. This theme was mapped to "Product attributes (safety)" as it shows public concerns about food safety and distrust in biotechnology. *GMO and ethical concerns* reflected moral discomfort with technological manipulation of nature, challenging ethical boundaries. This theme was allocated to Ethical issues (artificiality and manipulation of nature) construct, as it highlights moral concerns about unnaturalness. Lastly, *Political resistance and cultural identity* narrative emphasised political resistance linked to ideological opposition and regulatory distrust. It was placed within "Social influences" (political narratives and regulatory authority) construct, reflecting concerns about regulatory overreach and political control.

4.2 Discussion

The analysis revealed a range of cultural, political, ethical, and economic narratives influencing public perceptions of cultured meat (CM). *Conspiracy theories and misinformation* emerged as important themes, reflecting distrust toward influential figures and ideological resistance linked to political identity and anti-establishment opinions. This supports previous findings on consumer scepticism toward novel food technologies (Kouarfaté & Durif, 2023). The intensity of these narratives suggests a need to explicitly incorporate *Political and ideological narratives* into the CAPA framework. *Health risks and safety concerns*, particularly cancer fears, highlighted public distrust in biotechnology and perceptions of unnaturalness. These narratives validate the "Product attributes" construct but reveal a need to expand it to include specific *Health risk narratives* given the prominence and emotional charge of cancer-related concerns (Hanan et al., 2024; To et al., 2024). Ethical discomfort related to *artificiality and manipulation of nature* challenged moral boundaries, suggesting a need to extend "Ethical issues" construct to include *Moral discourses and ethical ambiguity*, reflecting ideological resistance beyond traditional ethical debates (Monaco et al., 2024; Tsvakirai, 2024). *Economic and market dynamics* emphasised concerns about *price disruption and monopolisation*. These narratives indicate the need for a new construct such as, *Economic and market forces*, to systematically capture public concerns about economic disruption and industry transformation (Lewisch & Riefler, 2024).

The findings validate the CAPA model's constructs of *Information influence, Product attributes, Ethical issues*, and *Social influences* but also highlight areas for refinement. Specifically, incorporating *Political and ideological narratives, Specific health risk narratives, Moral discourses and Ethical ambiguity,* as well as *Economic and market forces* would improve the model's capacity to systematically capture the cultural, political, and economic complexities influencing consumer acceptance of CM.

These preliminary findings, based on the initial 50 topics, indicate that analysing the complete dataset of 267 topics could further refine or reshape the CAPA framework to capture emerging consumer concerns. These findings suggest that while the CAPA model remains largely valid, public discourse introduces additional dimensions that were not fully represented in the original framework. The emergence of politically charged resistance, conspiracy narratives, and explicit economic concerns points to the need for a more expansive model. As analysis of the remaining 217 topics progresses, it is likely that further refinements, or even reconfigurations, of CAPA may be warranted to fully reflect the complexity of consumer attitudes toward CM. Understanding these refinments is important not only

theoretically but also practically, as they can directly inform communication strategies and policymaking.

From a strategic perspective, clear and transparent messaging is important to counter misinformation and ideological resistance (Specht et al., 2020; Kouarfaté & Durif, 2023). Culturally sensitive engagement, along with communication that aligns with traditional values, can address public apprehension effectively. Addressing ethical discomfort also requires clear ethical positioning. Proactive advocacy focusing on safety standards and ethical labelling can help mitigate regulatory distrust and ease market acceptance challenges.

Given CM's pre-commercial status, public attitudes remain fluid, susceptible to misinformation and media influence (Deliza et al., 2023; Lewisch & Riefler, 2023). Thus, proactive engagement and strategic communication are necessary even prior to commercialisation. Integrating the extended CAPA model into Decision Support Systems (DSS) could offer policymakers strategic foresight through monitoring public narratives. Rather than enabling immediate real-time interventions typical for commercialised products, DSS integration can facilitate early detection of misinformation and resistance, supporting proactive policy formulation and targeted communication strategies. Such anticipatory applications distinguish this proactive DSS use from the crisis management functions relevant post-commercialisation.

5. Conclusion

This preliminary study validated and extended the Cultured Meat Attitude and Perception Assessment (CAPA) model by analysing public narratives on X (Twitter) to understand consumer perceptions of Cultured Meat (CM). The analysis confirmed the relevance of CAPA constructs such as Information Influence, Product Attributes, Ethical Issues, and Social Influences, but also highlighted the need for further refinement due to newly emerged narratives: Conspiracy Theories, Political Resistance, Specific Health Risk Narratives, and Economic and Market Forces. These extensions enrich the model, providing a better understanding of consumer attitudes and decision-making.

For industry stakeholders and policymakers, the findings highlight the need for transparent communication to counter misinformation and ideological resistance. Clear messaging about safety standards and product benefits can address health risk perceptions and ethical concerns. To navigate political resistance and economic anxieties, strategic communication should be tailored to public concerns about economic disruption and regulatory overreach. This would provide actionable insights for strategic communication, product positioning, and policy development, supporting consumer acceptance of sustainable and ethical food innovations.

Given that this paper is based on an initial analysis of 50 out of 267 topics, the findings should be interpreted cautiously. Future research will expand on these preliminary results by analysing the remaining topics and incorporating sentiment analysis to provide a richer understanding of consumer attitudes. Additionally, extending research across other social media platforms would enable cross-validation and strengthen the robustness of insights.

This study contributes to the decision support systems literature by refining the CAPA framework and offering a systematic method for capturing public perceptions through social media analysis. Future research could explore the evolution of consumer narratives, conduct cross-cultural comparisons, and investigate the impact of political and ideological influences on consumer acceptance of cultured meat.

Acknowledgments. This research was conducted as part of a Doctoral Training Account (DTA) funded by the University of Plymouth.

References

Ajzen, I. (1991) The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

An, Y., Oh, H., & Lee, J. (2023) Marketing insights from reviews using topic modelling with BERTopic and deep clustering network. *Applied Sciences*, 13(16), 9443. https://doi.org/10.3390/app13169443

Boot, A. B., Tjong Kim Sang, E., Dijkstra, K., & Zwaan, R. A. (2019) How character limit affects language usage in tweets. *Palgrave Communications*, 5, 76. https://doi.org/10.1057/s41599-019-0280-3

Braun, V., & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Bryant, C., & Barnett, J. (2020) Consumer acceptance of cultured meat: An updated review (2018–2020). *Applied Sciences*, 10(15), 5201. https://doi.org/10.3390/app10155201

Deliza, R., Rodríguez, B., Reinoso-Carvalho, F., & Lucchese-Cheung, T. (2023) Cultured meat: A review on accepting challenges and upcoming possibilities. *Current Opinion in Food Science*, 52, 101050. https://doi.org/10.1016/j.cofs.2023.101050

Egger, R., & Yu, J. (2022) A topic modelling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in Sociology*, 7. https://doi.org/10.3389/fsoc.2022.886498

GFI (2025) Consumer snapshot: Cultivated meat in the U.S. The Good Food Institute. https://gfi.org/wp-content/uploads/2025/01/Consumer-snapshot-cultivated-meat-in-the-US.pdf

González-Rostani, V., Incio, J., & Lezama, G. (2024) Social media versus surveys: A new scalable approach to understanding legislators' discourse. *Legislative Studies Quarterly*. https://doi.org/10.1111/lsq.12481

Ju, X. (2024) A social media competitive intelligence framework for brand topic identification and customer engagement prediction. *PLOS ONE*, 19(11), e0313191. https://doi.org/10.1371/journal.pone.0313191

Hanan, F. A., Karim, S. A., Aziz, Y. A., Ishak, F. A. C., & Sumarjan, N. (2024) Consumer's cultured meat perception and acceptance determinants: A systematic review and future research agenda. *International Journal of Consumer Studies*, 48, e13088. https://doi.org/10.1111/ijcs.13088

Kouarfaté, B. B., & Durif, F. (2023) Understanding consumer attitudes toward cultured meat: The role of online media framing. *Sustainability*, 15(24), 16879. https://doi.org/10.3390/su152416879

Lewisch, L., & Riefler, P. (2023) Cultured meat acceptance for global food security: A systematic literature review and future research directions. *Agricultural and Food Economics*, 11, 48. https://doi.org/10.1186/s40100-023-00287-2

Mancini, M. C., & Antonioli, F. (2019) Exploring consumers' attitude towards cultured meat in Italy. *Meat Science*, 150, 101–110. https://doi.org/10.1016/j.meatsci.2018.12.014

Mendonça, M., & Figueira, Á. (2024) Topic Extraction: BERTopic's Insight into the 117th Congress's Twitterverse. Informatics, 11(8). https://doi.org/10.3390/informatics11010008

Monaco, A., Kotz, J., Al Masri, M., Allmeta, A., Purnhagen, K. P., & König, L. M. (2024) Consumers' perception of novel foods and the impact of heuristics and biases: A systematic review. *Appetite*, 196, 107285. https://doi.org/10.1016/j.appet.2024.107285

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017) Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1-13. https://doi.org/10.1177/1609406917733847

Onwezen, M. C., & Dagevos, H. (2024) A meta-review of consumer behaviour studies on meat reduction and alternative protein acceptance. *Food Quality and Preference*, 114, 105067. https://doi.org/10.1016/j.foodqual.2023.105067

Pilařová, L., Kvasničková Stanislavská, L., Pilař, L., Balcarová, T., & Pitrová, J. (2022) Cultured meat on the social network Twitter: Clean, future, and sustainable meats. *Foods*, 11(17), 2695. https://doi.org/10.3390/foods11172695

Pivoraite, G., Liu, S., Roh, S., Zhao, G. (2024) Framework for Understanding Consumer Perceptions and Attitudes to Support Decisions on Cultured Meat: A Theoretical Approach and Future Directions. In: Duarte, S.P., Lobo, A., Delibašić, B., Kamissoko, D. (eds) Decision Support Systems XIV. Human-Centric Group Decision, Negotiation and Decision Support Systems for Societal Transitions. ICDSST 2024. Lecture Notes in Business Information Processing, vol 506. Springer, Cham. https://doi.org/10.1007/978-3-031-59376-5_9

Statista (2024) Global willingness to try cultured meat. Statista. https://www.statista.com/statistics/1455302/global-willingness-to-try-cultured-meat/

Siddiqui, S. A., Khan, S., Farooqi, M. Q. U., Singh, P., Fernando, I., & Nagdalian, A. (2022) Consumer

behaviour towards cultured meat: A review since 2014. Appetite, 179, 106314. https://doi.org/10.1016/j.appet.2022.106314

Specht, A., Rumble, J., & Rhoades, E. (2020) “You call that meat?” Investigating social media conversations and influencers surrounding cultured meat. *Journal of Applied Communications*, 104(1). https://doi.org/10.4148/1051-0834.2303

Takats, C., Kwan, A., Wormer, R., Goldman, D., Jones, H. E., & Romero, D. (2022) Ethical and methodological considerations of Twitter data for public health research: Systematic review. *Journal of Medical Internet Research*, 24(11), e40380. https://doi.org/10.2196/40380

Tommasel, A., & Godoy, D. (2019) Short-text learning in social media: A review. The Knowledge Engineering Review, 34, e0. Cambridge University Press. https://doi.org/10.1017/S0269888919000018

Tsvakirai, C. Z., Nalley, L. L., & Makgopa, T. (2024) The valency of consumers’ perceptions toward cultured meat: A review. *Heliyon*, 10, e27649. DOI: 10.1016/j.heliyon.2024.e27649

Wunsch, N.-G. (2024) Cultured meat: Statistics and facts. Statista. https://www.statista.com/topics/8043/cultured-meat/#topicOverview

X (2024) Terms of Service. https://x.com/en/tos

Xanthe Lin, J. W., Maran, N., Lim, A. J., Ng, S. B., & Teo, P. S. (2025) Current challenges and potential solutions to increase acceptance and long-term consumption of cultured meat and edible insects: A review. *Future Foods*, 11, 100544. https://doi.org/10.1016/j.fufo.2025.100544

Posters

Incorporating Sentiment Analysis into Doctor-Patient Co-Decision in Primary Care

Rodrigo Aznar, Ayrton Sarango, Juan Aguarón, José María Moreno-Jiménez, Jorge Navarro*, and Alberto Turón

Grupo Decisión Multicriterio Zaragoza (GDMZ), Departamento de Economía Aplicada, Facultad de Economía y Empresa, Universidad de Zaragoza, 50005, Zaragoza, Spain
* jnavarrol@unizar.es

Abstract: The cognitive approach in decision making, particularly in the multi-criteria domain, pursues the extraction and dissemination of the arguments that support different positions and decisions. In the context of the Knowledge and Artificial Intelligences Society, scientific decision-making must integrate the objective and rational aspects of the scientific methods with the subjective and emotional aspects associated with the human factor. Traditionally, the incorporation of the emotional in decision support systems (DSS) requires the analysis of texts that reflect the positions of the actors involved in the decisional process. This study analyses the texts associated with 42 audios of doctor-patient conversation in Primary Care consultations looking at the evolution of identified affective aspects, specifically, mood and emotions. The sentiments and emotions in the texts were analysed using the SSA-DSS (Spanish Sentiment Analysis-Decision Support System) software, developed by the GDMZ, in particular, using its graphic visualisation tools. The analysis focused on two of the six diseases covered by the 42 audios: Hypertension and Dyslipidemia disease. In order to identify the differences between consultations for both diseases, a comparative analysis of the evolution of sentiment and emotional factors in doctor-patient interactions has been carried out. In addition to the characterisation of the emotions of these two diseases, the work proposes an appropriate structure (protocol) for the general analysis of the sentiment and emotional factors in the interactions between doctors and patients.

Keywords: Doctor-patient relationship; Decision support system; Primary care; Artificial intelligence; Sentiment analysis

Improvement Strategies in Public Administration. Technological Trees for Intellectual Capital

Victoria Muerza[1], María Teresa Escobar[1,*], Emilio Larrodé[2] and José María Moreno-Jiménez[1]

[1] Zaragoza Multicriteria Decision Making Group (GDMZ), University of Zaragoza, Zaragoza, Spain
[2] Department of Mechanical Engineering, University of Zaragoza, Zaragoza, Spain
[*] mescobar@unizar.es

Abstract: Public Administration is defined as the set of institutions and organizations that manage resources and services and implement policies aimed at ensuring social welfare, promoting transparency and the efficiency of their services. The fundamental difference between the public and private sectors lies in how their effectiveness is understood. The private sector seeks economic profitability, whereas the public sector aims to improve social profitability – that is, the quality of life of citizens by guaranteeing their rights and freedoms. Leveraging the synergies between the public and private sectors, this research transfers to the public sector the methodology developed by Muerza et al. (2014) in the business area to enhance its competitiveness and ensure the long-term viability of companies. The difference lies in that, in the public sector, the human factor and intellectual capital are essential elements, as the knowledge, experience, and adaptability of its civil servants drive the development and continuous improvement of public services. This intellectual capital becomes a strategic asset, fostering innovation and institutional learning. To this end, we have designed a methodology in the following phases: (i) Taxonomy for Technologies in Public Administration; (ii) Identification of the relevant technologies for the institution, e. g. research, cooperation, training for research, dissemination; (iii) Construction of technological trees based on the relevant technologies, as structured visual tools to support innovation planning; and (iv) Multicriteria selection of the strategy that provides the greatest future social value, through the development of a Decision Support System. We focus the analysis on the "*Instituto Nacional de la Administración Pública*" (INAP) to illustrate our proposal.

Keywords: Decision-Making Process; Public Administration; Knowledge Society; Taxonomy of technologies; Technological Trees

Acknowledgments

This study was supported by the Regional Government of Aragón project "S35_23R: Grupo Decisión Multicriterio Zaragoza (GDMZ)"; the Spanish Ministry of Science and Innovation Projects "Codecisión Cognitiva y Colaborativa. Aplicaciones en Salud" (ref. PID2022-139863OB-I00), and "Red Temática de Decisión Multicriterio" (ref. RED2022-134540-T).

Optimizing Forest Management Strategies for Balancing Biodiversity and Production

Milena Lakićević[1]*

[1] University of Novi Sad, Faculty of Agriculture, Novi Sad, Serbia

* milena.lakicevic@polj.edu.rs

Abstract: Forest management can be enhanced through the integration of simulation models and decision support systems. This study explores the combined application of Samsara2, an empirical forest model, and PROMETHEE, a multi-criteria decision support method, to optimize forest management strategies. Through a case study of forest stands, the research examines how different balances between biodiversity and production influence optimal management decisions. Unlike traditional approaches that rely on fixed criteria importance, this study assesses outcomes based on varying criteria weights. Here, biodiversity and production are two competing criteria, each consisting of a set of related indicators. The research tests the results as the importance of biodiversity varies from 0 to 1, with the corresponding decline in production importance varying inversely on the same scale. The proposed approach is particularly valuable in cases where decision-makers can estimate approximate criteria importance, as well as when criteria importance remains undefined. By offering a flexible and robust decision-making framework, this research contributes to improved forest management planning, ensuring that biodiversity conservation and production goals are effectively balanced based on context-specific priorities.

Keywords: Forest Management; Decision Support Systems; Multi-Criteria Analysis; PROMETHEE

A DSS to evaluate intervention programs based on physical activity and sport for the prevention of drug use

Antonio Jiménez-Martín[1*], Marcos Asensio-Hernández[1,2], Pedro J. Jiménez[2] and Alberto Dorado[3]

[1] Decision Analysis and Statistics Group, Universidad Politécnica de Madrid, Boadilla del Monte, Spain
[2] Facultad de Ciencias de la Actividad Física y del Deporte, Universidad Politécnica de Madrid, Madrid, Spain
[3] Facultad de Ciencias del Deporte, Universidad de Castilla la Mancha, Toledo, Spain
[*] antonio.jimenez@upm.es

Abstract: Drug consumption is an ongoing social problem today. Intervention programs for the prevention of drug use by adolescents based on physical activity and sport have recently attracted a lot of attention, and the evaluation of such programs constitutes a multi-criteria decision-making problem. In this work we introduce a web-based decision support system based on an additive multi-attribute utility function that accounts for a decision-making context with partial/incomplete information. The evaluation of intervention programs is structured in an objective hierarchy that includes four main dimensions: i) the facilities where the program is to be developed, their versatility and management formulas; ii) the experience of the applicant organization; iii) the technical proposal, which includes the work team, program and results dissemination and the methodology underlying the program, intervention sessions and evaluation, and iv) the organization's financial solvency and budget soundness.

Default attributes, component utilities and weights are proposed by the system based on expert knowledge, which can be updated by the user if deemed appropriate. The system accounts for imprecision concerning the quantification of the decision-makers preferences, both for component utilities and weights, leading to classes of utility functions and weight intervals, respectively. Weights are hierarchically elicited providing the possibility of using different weighting methods at the different levels and branches of the objective hierarchy.

The additive multi-attribute utility function is used to derive an overall average utility for each intervention program, on which the ranking of intervention programs is based, and overall minimum and maximum utilities, which give us further insight into the robustness of such ranking. Different sensitivity analysis tools are also provided by the system that take advantage of the imprecise available information to provide more meaningful information and reduce the number of intervention program of interest.

Keywords: decision support systems; multi-criteria decision making; drug use prevention; physical activity and sport; intervention programs

Acknowledgments: This work was supported by the grants PID2021-122209OB-C31 and RED2022-134540-T funded by MICIU/AEI/10.13039/501100011033.

The EnCycLEd Project Promotes "The Importance of Cybersecurity Education for Leveraging the Development of AI-based Decision Systems"

Matthias Kettemann[1*], Jason Papathanasiou[2], Sebastian Harnisch[3], Iryna Volnytska[4], Andrii Paziuk[5], Sergii Melnyk[6], Dimitris Slavoudis[7], Anamaria Magri Pantea[8], Fátima Dargam[9]

[1] Universität Innsbruck - Department of Theory and Future of Law, Innsbruck, Austria
[2] University of Macedonia, Thessaloniki, Greece
[3]Universität Heidelberg, Heidelberg, Germany
[4]Science, Entrepreneurship and Technologies University, Kyiv, Ukraine
[5]Ukrainian Academy of Cybersecurity, Kyiv, Ukraine
[6]Kyiv Vocational College with Enhanced Military and Physical Training, Kyiv, Ukraine
[7]American Farm School, Thessaloniki, Greece
[8]AscendConsulting.eu, Malta
[9]REACH Innovation, Graz, Austria
[*] e-mail: coordination@encycled.eu

Abstract: The EnCycLEd project[30] (*Enhancing Cybersecurity Literacy in Education*) aims at improving awareness of cybersecurity threats and opportunities among students, individuals and organizations. This way, the project fosters digital security readiness and resilience by providing cybersecurity educational programs and resources, encouraging also the interest in cybersecurity studies and career pathways. Hence, the project leverages digital skills among students and inspires the next generation of cybersecurity professionals to be better prepared to develop safer systems, including AI-based decision systems.

The project boosts digital readiness and cybersecurity resilience for educational. It focuses on equipping students, teachers, and educational professionals with key digital skills and competences. As key elements, the project develops educational curricula and a platform, which include: cybersecurity fundamentals; practical defence skills and critical thinking; interactive games simulating real-world challenges; interviews with insights from cybersecurity experts; simulation exercises in real world scenarios; as well as pathways and study options to assist teachers in guiding students' career choices. The EnCycLEd web-platform counts with a curated resource collection of the produced material for the curriculum and provides a trusted space for the development and sharing of educational materials, prioritizing safety and security.

The project will be validated at various educational institutions, focusing on collecting feedback and on evaluating the project's impact. For this, the EnCycLEd partners will conduct training sessions, workshops, and presentations for both teachers and students. This includes developing didactic workshops and pilot programs. Additionally, the developed toolkit will be tested within the EnCycLEd platform at the pilot-schools.

In conclusion, cybersecurity awareness is essential in education to protect data, promote responsible AI use, and prepare students for digital threats. The EnCycLEd project fosters a culture of safety and ethics in technology, raising awareness among students and teachers to avoid cyber threats and ensure data privacy, especially in the development of AI decision systems.

Keywords: Cybersecurity Awareness; Erasmus+ Project; Education; Digital Readiness; Cybersecurity Training

[30] The EnCycLEd Project (https://www.encycled.eu/) is an Erasmus+ Project (Nr.2023-1-AT01-KA220-SCH-000166888), co-funded by the European Union, coordinated by the Universität Innsbruck - Department of Theory and Future of Law, including the following project partners: University of Macedonia, Universität Heidelberg, Heidelberg, Science, Entrepreneurship and Technologies (SET) University, Ukrainian Academy of Cybersecurity, Kyiv Vocational College with Enhanced Military and Physical Training, American Farm School, Ascend Consulting, and REACH Innovation.

An integrated tool for Road Safety Management

Sérgio Pedro Duarte[1,2*], Pedro Rodrigues[3], Telma Silva[4], João Neves[4], Sara Ferreira[1], António Lobo[1] and, Thiago Sobral[3, 5]

[1] CITTA – Centro de Investigação do Território, Transportes e Ambiente, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
[2] ICS – Instituto para a Construção Sustentável, Porto, Portugal
[3] Departamento de Engenharia e Gestão Industrial, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
[4] ASCENDI IGI – Inovação e Gestão de Infraestruturas, S.A., Porto, Portugal
[5] OPT – Optimização e Planeamento de Transportes, S.A, Porto, Portugal, Portugal

[*] spduarte@fe.up.pt

Abstract: Road Safety Action Plans (RSAP) are strategic instruments that contain the goals and action plans for a specific road network. Ascendi, a Portuguese road concessionaire, has been developing data driven approaches to its Road Safety Plans for periods of four years based on data collected along its motorway network, considering the number of road crashes registered. Previous studies have identified locations where the frequency of road crashes is higher, but this information alone does not provide insights regarding the causes of the crash nor possible countermeasures to be implemented. Recently, crash prediction models were developed for a network with over 600 km in Portuguese motorways. These models have aided the characterization of the network by relating road crashes with road geometry and traffic volume. The main objective of this work is to develop a tool that is integrated into Ascendi's network management system in order to support the design of future RSAP. One of the main challenges is to integrate data from other data sources, such as speed, traffic volume, pavement, and others. We have mapped Ascendi's application architecture and data structure to identify how the proposed tool can be integrated with the existing system. Moreover, we selected data-driven KPI to help characterization of the network. Those KPI are presented in an integrated dashboard fed by the different data sources to support the design of RSAP. By including KPI that are derived from different characteristics of the network, the tool can support be used to validate the impact of countermeasures. Future applications of the tool include the simulation of different traffic volumes and changes to the geometry.

Keywords: road safety; KPI; data-driven tools; crash prediction models

Acknowledgements: The authors acknowledge Ascendi's support in the development of the prediction models and the tool.

Surjective Independence of Causal Influences for Local BN Structures

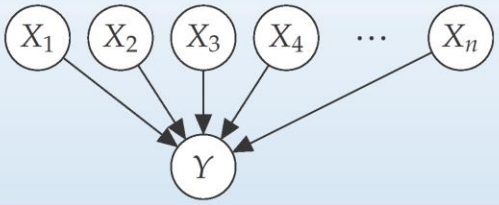
Kieran Drury[1], Martine J. Barons[1] and Jim Q. Smith[1]

Correspondence Email: kieran.drury@warwick.ac.uk

[1]Department of Statistics, University of Warwick, Coventry, United Kingdom

WARWICK
THE UNIVERSITY OF WARWICK

1. Bayesian Network Local Structure [KN11]



2. Bayesian Networks for Decision Support

- Bayesian networks have an **intuitive graphical structure** that can be explained to non-statistician clients
- **Causal Bayesian networks** very naturally **model policy interventions**, allowing us to calculate **expected utility scores** over a set of policies
- **Expert judgement** can easily be embedded into such Bayesian network models, giving clients a sense of **ownership and explainability**
- Bayesian network models can be **updated locally** when new data or judgements are available, providing **timely decision support**
- The process of building such a model through expert judgement typically involves **qualitative and quantitative elicitation stages**

3. Qualitative Elicitation

- This refers to drawing the **network structure** through **expert judgement**
- This includes **node definitions**, their **potential values** and which nodes to connect with **directed edges** [KN11]
- The network structure should represent **causal relationships** such that **policy interventions** on the model are justifiable

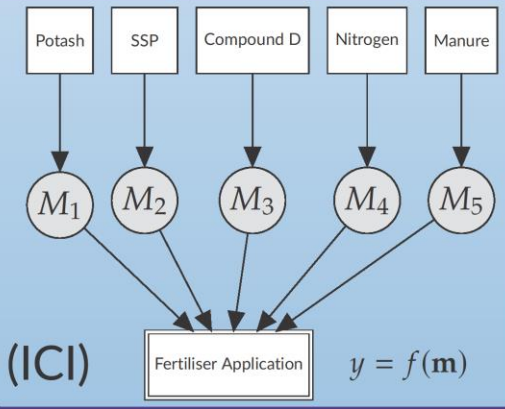
4. Quantitative Elicitation

- Given the network structure, each non-root node has an associated **conditional probability table (CPT)** that needs populating, e.g.:

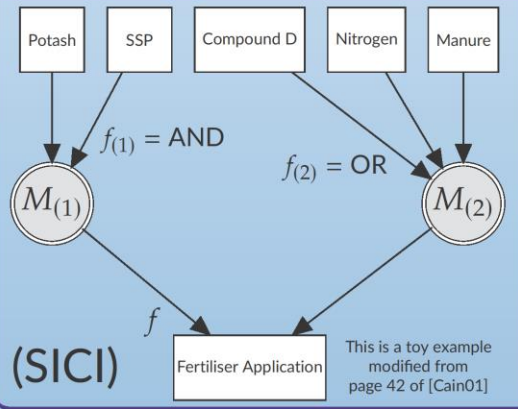
X_1	X_2	$\mathbb{P}(Y = 0 \mid \text{Pa}(Y))$	$\mathbb{P}(Y = 1 \mid \text{Pa}(Y))$
0	0	$p_{0(1)}$	$p_{1(1)}$
0	1	$p_{0(2)}$	$p_{1(2)}$
1	0	$p_{0(3)}$	$p_{1(3)}$
1	1	$p_{0(4)}$	$p_{1(4)}$

- Often, we do not have sufficient, reliable data for this and we must instead **rely on expert judgement** to estimate the unknown entries
- The **number of parameters in a CPT grows exponentially** in the number of parents and the number of potential states across all nodes
- Modifying the local structure reduces this growth to **make the CPT elicitation tractable** for the domain experts

5. Independence of Causal Influences [Heck93]



7. Surjective Independence of Causal Influences



This is a toy example modified from page 42 of [Cain01]

6. Benefits and Limitations of ICI

- ICI significantly **reduces the number of free parameters to be determined in the CPT** of Y
- It also **reduces the complexity of the judgements** required from experts, leading to **reduced cognitive burden** and **more accurate assessments**
- However, **the ICI assumption is very strong** in practice. We therefore **instead partition the parents into blocks that exhibit ICI themselves.**

8. Evaluation of SICI and Future Work

- SICI **allows ICI to be embedded through latent mechanisms** when ICI cannot be an assumed property of the original parent set
- The **number of parameters** required to populate the CPT of Y may now only **grow linearly in the number of elicited mechanisms**
- Our next step is to **modify and analyse methods for populating CPTs within the SICI local structure framework** (see e.g. [MPD16])

References

[Heck93] David Heckerman. "Causal Independence for Knowledge Acquisition and Inference". In: *Uncertainty in Artificial Intelligence*. Elsevier, 1993, pp. 122–127.

[Cain01] Jeremy Cain. *Planning improvements in natural resource management*. UK Centre for Ecology and Hydrology, 2001.

[KN11] Kevin B. Korb and Ann E. Nicholson. *Bayesian Artificial Intelligence*. CRC Press, 2011.

[MPD16] L. Mkrtchyan, L. Podofillini, and V.N. Dang. "Methods for building Conditional Probability Tables of Bayesian Belief Networks from limited judgment: An evaluation for Human Reliability Application". In: *Reliability Engineering & System Safety* 151 (2016), pp. 93–112.

Webpage Link

Index