Amit Peled, Jake Flynn, Mark Ralston Daniel
12/11/2024
DATA 590

Capstone 2024: Project Proposal

Introduction

Multi-criteria decision-making (MCDM) has emerged as a crucial methodology for addressing complex organizational challenges in today's data-driven business environment. As organizations face increasingly complex decisions involving multiple, often conflicting criteria, the need for sophisticated decision support tools has become paramount. These decisions span across various domains, from supplier selection and risk assessment to resource allocation and sustainability planning, where decision-makers must balance numerous factors while considering both quantitative and qualitative inputs.

Recent advances in data science and machine learning have opened new possibilities for enhancing MCDM methodologies, particularly in handling dynamic, time-series data and incorporating predictive analytics. By combining traditional MCDM approaches with modern computational techniques, organizations can now develop more robust and adaptable decision support systems. Our project seeks to contribute to this evolution by developing a production-ready system that can process multiple decision criteria over time, predict future rankings, and provide actionable insights for stakeholders. Through the integration of Bayesian methods, operations research, and advanced analytics, we aim to create a framework that not only addresses current decision-making challenges but also adapts to emerging needs across different industries.

Problem Statement

The objective of our project is to enhance and refine an existing risk ranking model initially developed in a 2022 project, adapting it to a production environment capable of providing accurate and actionable risk assessments. Leveraging advanced methodologies from decision analysis, Bayesian methods, operations research, and multi-criteria decision analysis (MCDA), our aim is to create a robust framework that predicts and analyzes risk rankings over time while addressing real-world constraints. The completed model will consist of an adaptable, production-ready MCDA system designed to handle multivariate rank forecasting problems. Inputs to the model include 17 ordinal decision factors, such as cost and location, associated with 1,222 subjects (e.g., suppliers), captured at irregular timestamps. These variables influence risk rankings, which range from 1 (high risk) to 265 (low risk), and are modeled using a dataset of 31,933 synthetic records generated through an R script. The primary type of risk considered relates to operational and logistical factors impacting supplier evaluations.

Our project involves two key components: a comprehensive literature review and a rank data analysis phase. The literature review will focus on domain knowledge acquisition by exploring cutting-edge research in risk analysis, Bayesian methods, and operations research. We will also investigate the role of Large Language Models (LLMs) in improving model interpretability and accuracy, assessing their relevance to our objectives. This step ensures a strong theoretical foundation for subsequent analysis.

In the rank data analysis phase, we aim to address two central questions: can historical data be used to predict future risk rankings accurately, and what are the key factors driving shifts in rankings? To answer these, we will utilize our synthetic dataset to simulate real-world scenarios while maintaining data security and privacy. By applying iterative validation checks and regular stakeholder feedback, we aim to develop a model that is not only practical but also generalizable across industries.

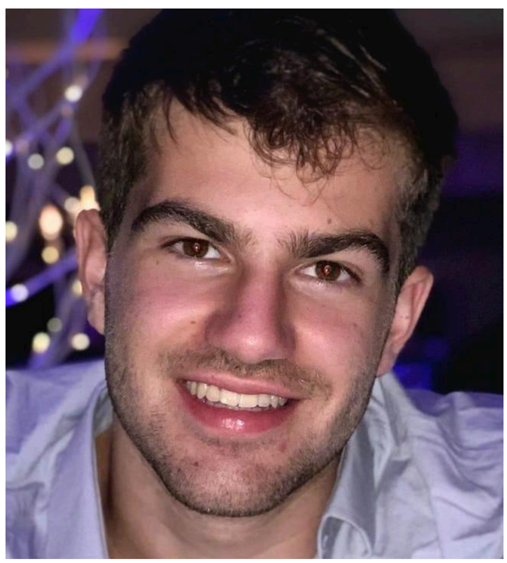
However, several risks must be managed carefully to ensure project success. Our reliance on synthetic data introduces a technical challenge, as it may not fully capture the complexity of real-world scenarios. To mitigate this, we will validate the data generation process with Dr. Song and explore potential alternative datasets. Another significant challenge lies in modeling relationships between 17 ordinal variables, requiring sophisticated approaches to effectively represent complex dependencies. Regular validation and iterative refinement of the MCDA model will address this risk. Additionally, our ambitious timeline and the limited availability of key stakeholders, such as Dr. Song, pose project management challenges. To mitigate these, we have incorporated buffer periods, scheduled key meetings in advance, and prioritized features to ensure efficient progress.

Despite these challenges, our project offers substantial benefits. Academically, it contributes to the understanding of MCDA methodologies through novel implementations. Practically, it aims to deliver a user-friendly tool for complex decision-making that can be adapted across domains. This project provides team members with valuable experience in data science, software engineering, and project management. By managing risks through weekly check-ins, contingency plans, and clear communication channels, we aim to ensure the successful delivery of a flexible, scalable, and impactful risk ranking model.

Team Bios

Jake Flynn gained hands-on experience in manufacturing data analytics during internships at Illumina and Bristol Myers Squibb, where he applied predictive modeling techniques to enhance manufacturing processes. His projects included creating a model and a web application to optimize manufacturing scheduling, aimed at reducing turnaround time. At Starbucks, Jake developed robust data pipelines to process HR data and contributed to store segmentation analysis. Jake is proficient in dashboard creation and data visualization using tools such as Tableau, Plotly Dash, and R Shiny. In addition, Jake has data engineering expertise in designing and automating data pipelines using dbt. Throughout his time as a teaching assistant during both undergraduate and graduate school, Jake has developed the ability to simplify technical concepts and clearly communicate to diverse audiences. After graduating from the MSDS program, Jake is interested in applying data science techniques to solve problems in healthcare, manufacturing, and retail.

Mark Ralston Daniel brings a unique combination of healthcare research expertise and industry experience in software engineering and data science. At Seattle Children's Hospital, he has demonstrated strong data integration skills by successfully merging hospital datasets and engineering efficient database solutions that improve research workflow. His role as a Data Science Intern at Nordstrom focused on developing machine learning pipelines for retail predictions, showcasing his ability to work with large-scale time-series data and modern ML frameworks. During his time at the Royal Bank of Canada, Mark developed full-stack applications that improved testing efficiency, displaying his versatility across the technical stack. His academic projects, including the development of TLDhubeR, and a RAG-powered podcast search engine, demonstrate his ability to apply cutting-edge technologies to solve complex problems. Mark's diverse skill set spans Python, R, and various ML libraries, complemented by his strong academic performance in both his current Master's in Data Science at UW and his BSc in Computational Neuroscience from McGill. After completing his MSDS program, Mark is interested in applying his combined expertise in healthcare analytics, retail prediction systems, and advanced machine learning techniques to solve complex data challenges in healthcare and industry settings.



Amit Peled combines extensive experience in data science with a background in statistical modeling, data integration, and signal processing. During his time with the Oklahoma City Thunder, Amit developed regression and hierarchical Bayesian models to enhance player performance evaluations, demonstrating his ability to apply complex statistical techniques to real-world sports analytics. Additionally, Amit’s expertise in database engineering and data integration was evident in his work implementing recursion-based mapping functions across multiple datasets, streamlining data flow and improving the quality of analytics. He led a signal processing project that developed a heart rate monitoring app using Python and OpenCV, which could accurately measure heart rate through smartphone cameras, showcasing his innovative approach to health technology. Amit’s skill set spans Python, R, and SQL, with a proven track record of communicating findings effectively to a wide range of stakeholders. Passionate about AI-based solutions, he is interested in using predictive modeling to analyze data and drive efficiency in various sectors. After completing his MSDS program, Amit is excited to apply his data science expertise to tackle challenging data problems.

Data Pipeline & Exploration

Data Streams

- The data was synthetically generated from an R script.
- 17 ordinal variables (decision factors like cost and location) associated with 1,222 subjects (e.g., suppliers). These factors influence risk rankings, which are ordinal values ranging from 1 (high risk) to 265 (low risk). Data points are collected at irregular timestamps.

Volume of Data

- Observations: 31,933 records.
- Attributes: 17 attributes (ordinal variables).

Data Location

- Google Drive on a Small CSV file.

Access Software

- Preprocessing: Python
- Modeling and Prediction: Python
- Environment: Jupyter Notebooks

Summary of Previous Work

We reviewed the paper "Identifying Best Practice Melting Patterns in Induction Furnaces," which introduces a methodology combining time-series K-means clustering with Multi-Criteria Decision Making (MCDM) to optimize energy efficiency in industrial melting processes. The authors use time-series clustering to group melting profiles based on temperature data, followed by MCDM techniques such as TOPSIS and VIKOR to evaluate cluster performance based on energy consumption and carbon emissions. The findings suggest that implementing the identified best practices could lead to an 8.6% reduction in electricity costs and an 8.44% decrease in CO_2 emissions, offering significant environmental and economic benefits for the foundry industry. However, the study is limited by its use of data from a single furnace over a short period, which raises questions about the generalizability of the study.

The second paper we reviewed, "A Comparative Analysis of Multi-Criteria Decision-Making (MCDM) Methods," focuses on analyzing three algorithms: multi-MOORA, TOPSIS, and VIKOR. Using 1,600 randomly generated decision matrices, the researchers examined how the different methods ranked alternatives and assessed the consistency of these rankings using Spearman's correlation coefficient. The results reveal significant variability across methods, particularly with the VIKOR algorithm, which showed sensitivity to parameter changes. These findings highlight the need for careful selection of MCDM methods based on the specific context of the decision problem. The paper also underscores the importance of robust parameter tuning and emphasizes that the reliability of results may vary depending on the algorithm and decision matrix design.

Finally, we examined the paper "How Do the Criteria Affect Sustainable Supplier Evaluation?" which leverages fuzzy MCDM methods to address the challenge of ranking suppliers under uncertainty. The study introduces Triangular Fuzzy Numbers (TFNs) to account for incomplete or uncertain data and uses fuzzy extensions of methods such as TOPSIS and VIKOR to create more reliable rankings. Sensitivity analysis is applied to test the robustness of these rankings and identify the most influential criteria. This approach demonstrates significant potential for industries like energy and supply chain management, where sustainability and supplier reliability are critical. The study offers valuable insights for improving decision-making frameworks but notes challenges related to fixed criteria and scalability to larger datasets.

Proposed Solutions with Deliverables

Phase 1: Project Setup

- Conduct an in-depth literature review of Multi-Criteria Decision Analysis (MCDA) techniques to identify methods for the project.
- Select and review research papers that outline methodologies relevant to our project. Choose one and finalize what method we will use for the Boeing data.
- Engage with stakeholders, including Dr. Song, to discuss project objectives, requirements, and initial plans.

Phase 2: Data Collection

- Choose criteria that will drive the risk ranking model. KPIs, guardrail metrics, etc.
- Dr. Song will give us access to Boeing synthetic data.

Phase 3: Model Development

- Create a prototype for the MCDA-based model, including algorithms and decision-making frameworks.
- Build and test the MCDA model using synthetic data to ensure accuracy and robustness in rank predictions.
- Conduct a formal review of project progress, including feedback from stakeholders, to assess alignment with objectives and adjust plans as necessary.

Phase 4: Web Application Development

- Develop an initial prototype of the web application using Dash.
- Refine and finalize the web application, incorporating feedback from Dr. Song.
- Create comprehensive documentation for the model and application.
- Design and prepare a professional poster summarizing the project for presentation.
- Practice delivering the project presentation.
- Present the completed project.

Analysis of Risks and Benefits

Our project faces several key technical and operational risks that require careful management. The primary technical challenge stems from our reliance on synthetic data, which may not fully capture real-world complexities. To address this, we will work closely with Dr. Song to validate our data generation process and potentially source alternative datasets. The MCDA model development presents another significant risk, as effectively modeling relationships between 17 ordinal variables requires sophisticated approaches. We will mitigate this through regular validation checks and iterative refinement based on stakeholder feedback.

From a project management perspective, our ambitious timeline poses challenges, particularly during the web application development phase. We have implemented buffer periods for critical path activities and established clear feature prioritization to manage this risk. Additionally, we recognize that stakeholder availability, including Dr. Song's time for consultation, may be limited. To address this, we have scheduled key meetings in advance and established clear communication channels.

Despite these risks, our project offers substantial benefits. Academically, it will advance understanding of MCDA methodologies and contribute to the field through novel implementations. The practical value includes developing a user-friendly tool for complex decision-making that can be adapted across different domains. Furthermore, the team will gain valuable experience across data science, software engineering, and project management.

Our risk management approach focuses on weekly team check-ins to assess risk status and validate technical components. We have prepared contingency plans including alternative data sources, simplified model versions, and flexible feature prioritization for the web application. This balanced approach to risk management will help ensure project success while maintaining our ambitious goals.

Proposed Schedule

Task Description	Assigned To	Start Date	Due Date	Status	Dependencies	Deliverables	Priority	Est. Hours	Progress %
Phase 1: Project Setup									
Finalize Paper to Recreate	Mark, Jake, Amit	12/15/24	12/20/24	To-Do	None	Meeting with Dr. Song	High	10	12
MCDA Methodology Research	Mark, Jake, Amit	12/20/24	12/31/24	To-Do	Team Organization	Research Summary Report	High	20	7
Stakeholder Meeting	Mark, Jake, Amit, Dr. Song	12/22/24	12/22/24	To-Do	Team Organization	Meeting Minutes	Medium	2	0
Phase 2: Data Collection									
Define Decision Criteria	Amit	1/1/25	1/15/25	To-Do	Research Complete	Criteria Framework	High	15	0
Data Collection Planning	Jake	1/15/25	1/31/25	To-Do	Decision Criteria	Data Collection Protocol	High	25	0
Stakeholder Planning	Mark, Jake, Amit, Dr. Song	1/20/25	1/20/25	To-Do	Decision Criteria	Requirements Document	Medium	4	0
Phase 3: Model Development									
MCDA Model Design	Mark, Jake, Amit, Dr.	2/1/25	2/15/25	To-Do	Data Collection Plan	Spec Sheet	High	30	0

	Song								
Model Implementation	Mark, Jake, Amit, Dr. Song	2/15/25	2/28/25	To-Do	Model Design	Working MCDA Model	High	40	0
Mid-Project Review	Mark, Jake, Amit, Dr. Song	2/15/25	2/15/25	To-Do	Model Design	Progress Report	Medium	4	0
Phase 4: Web Application Development	TBD								
Dash App Prototype	Mark, Jake, Amit, Dr. Song	2/17/2025	3/10/2025	To-Do	Model Implementation	Dash App Prototype	High	100	0
Dash App Finalized	Mark, Jake, Amit, Dr. Song	2/25/2025	3/14/2025	To-Do	Model Implementation	Dash App Final	High	60	0
Complete Documentation	Mark, Jake, Amit, Dr. Song	3/5/2025	3/15/2025	To-Do	Model Validation	User Manual	Medium	15	0
Get Poster Ready For Print	Mark, Jake, Amit, Dr. Song	3/10/2025	3/16/2025	To-Do	All Previous Tasks	Presentation Deck			
Rehearse	Mark, Jake, Amit, Dr. Song	3/10/2025	3/16/2025	To-Do	All Previous Tasks	Presentation Deck	High	10	0
Poster Presentation	Mark, Jake, Amit, Dr. Song	3/17/25	3/17/25	To-Do	All Previous Tasks	Final Deliverable Package	High	5	0

References

Ceballos, B., Lamata, M. T., & Pelta, D. A. (2016). A comparative analysis of multi-criteria decision-making methods. *Progress in Artificial Intelligence, 5*(2), 101–116. https://doi.org/10.1007/s13748-016-0093-1

Habib, M. S., & Alhaj, A. (2022). *How Do the Criteria Affect Sustainable Supplier Evaluation? A Case Study Using Multi-Criteria Decision Analysis Methods in a Fuzzy Environment.* International Journal of Sustainable Development and Planning, 17(1), 1-13.

Howard, D. A., Jørgensen, B. N., & Ma, Z. G. (2023). Identifying Best Practice Melting Patterns in Induction Furnaces: A Data-Driven Approach Using Time Series K-Means Clustering and Multi-Criteria Decision Making. In B. N. Jørgensen, L. C. Pereira da Silva, & Z. Ma (Eds.), *Energy Informatics: Third Energy Informatics Academy Conference, EI.A 2023, Campinas, Brazil, December 6–8, 2023, Proceedings, Part I (pp. 271–288).* Springer. https://doi.org/10.1007/978-3-031-48649-416