ie University Microsoft

SOKONEWS MEDIA

Executive Summary

Completed by:

***Hugo** A. Borjórquez Gallardo, **Sofia** Depoortere, **Filippo** Lisanti, **Fernando** Moreno Borrego, **Christopher** Stephan*

Master in Business Analytics & Big Data **March 2025**

ETHICS STATEMENT

This capstone project adheres to strict ethical guidelines to ensure confidentiality, transparency, integrity, and fairness throughout all phases of our research and collaboration. We recognize the importance of ethical responsibility in developing AI-driven personalization systems, particularly in the context of news recommendations, where trust, neutrality, and user privacy are critical. Our ethical commitments include:

- Confidentiality: We will safeguard any proprietary, sensitive, or personally identifiable information provided by the company or collected during the project. No confidential data will be shared, stored, or used beyond the agreed-upon scope.
- Compliance: We will strictly adhere to all relevant laws, regulations, and institutional policies, as well as best practices in data protection. We will ensure that all data handling, model training, and deployment practices align with legal and ethical requirements.
- Integrity: All research, analysis, and reporting will be conducted honestly, transparently, and without bias. We are committed to ensuring that our recommendations and insights are evidence-based, free from manipulation, and ethically sound.
- Respect for Stakeholders: We will treat all stakeholders—including users, business partners, and regulatory bodies—with fairness, professionalism, and respect. Our goal is to produce outcomes that benefit both the company and the broader community, while minimizing potential harms such as algorithmic bias, misinformation, or filter bubbles.

By embedding these ethical principles into the core of our research and system design, we establish a clear framework for responsible AI development. This explicit commitment not only guides our decision-making at every stage but also assures stakeholders of the integrity, fairness, and accountability of our project.

LINK TO GITHUB REPOSITORY

The following is the link to the github repository where you can find supplementary information, including:

- Technical notebook implementations
- Report Documents, such as the annex, supporting documents, and diagrams.

https://github.com/Christopher-Stephan/Microsoft_Capstone_Project

The datasets used for model implementations can be found in the following link:
https://msnews.github.io/

Processed datasets stored in .parquet can be found in the following google drive link:
https://drive.google.com/drive/folders/1JU3C_SzxmmjvE2jJ2Hw_Ab0-nR-HaE6k?usp=sharing

If you have any questions or inquiries regarding this, do not hesitate to contact us at:

- christopher.stephan@student.ie.edu

EXECUTIVE SUMMARY

As the digital news landscape undergoes rapid transformation, SokoNews is faced with the critical challenge of increasing user engagement through personalized content delivery. In an era where readers are inundated with vast amounts of information, a well-implemented recommendation system is crucial in capturing audience interest, improving content consumption, and ultimately driving business growth. This report explores the development of a data-driven recommendation system aimed at optimizing content suggestions based on user preferences. Given the absence of an in-house dataset, our study leverages external datasets, notably the Microsoft MIND dataset, to evaluate various recommendation methodologies. The goal is to create a proof-of-concept system that enhances engagement while maintaining scalability, ease of deployment, and operational sustainability, ensuring that SokoNews remains competitive in an increasingly AI-driven media landscape.

Unlike traditional recommendation domains such as movies or e-commerce, personalized news recommendation presents unique challenges. The cold-start problem, which applies to both new users and newly published articles, complicates the process of generating relevant recommendations without sufficient historical data. Additionally, news content has an exceptionally short lifecycle, with most articles seeing peak engagement within the first 24 hours of publication. These factors necessitate advanced modeling techniques that move beyond conventional collaborative filtering methods. Our approach integrates Natural Language Processing (NLP), entity recognition, and click-through rate (CTR) prediction models to address these challenges effectively and make real-time recommendations that adapt to dynamic user preferences and evolving news trends.

The study formulates several hypotheses that guide the development of the recommendation system. These include the significance of article titles in driving engagement, entity-based similarities between articles and user interests, historical reading preferences, recency-based preferences for fresher content, and category-based filtering as a baseline approach. To validate these hypotheses, we implement two key machine learning models: a BERT-based content similarity model, which processes textual data to generate personalized recommendations, and a CTR prediction model powered by LightGBM to enhance engagement through entity embeddings and behavioral data, as wel as an LLM-powered conversational recommender. By combining content-based and interaction-based methodologies, our models ensure a comprehensive personalization framework that caters to diverse user behaviors.

The implementation of the recommender system is expected to yield significant financial gains. Industry benchmarks indicate that personalized recommendations can lead to a 50% increase in total impressions, significantly enhancing user retention and engagement. In terms of advertising revenue, the introduction of personalized recommendations is projected to generate an additional €1,661,080 in CPM and CPC-based earnings within the first year (Mediahaus, 2021). Meanwhile, subscription revenue is expected to experience a substantial boost, with conversion rates for casual readers doubling from 2% to 4%, translating into an estimated additional €38.1 million in revenue. This results in a total first-year revenue impact of €39.8 million and a compelling ROI of 8.2, firmly establishing the financial viability of the recommendation system.

The total investment required for implementation is estimated at €3.46 million, encompassing cloud infrastructure, AI model training, personnel, and compliance costs. Given the need for a scalable and reliable system, our design transitions from a local prototype to an Azure-based cloud infrastructure, featuring automated pipelines for real-time data ingestion, processing, and recommendation delivery.

Ensuring compliance with local and emerging AI regulations remains a core focus, safeguarding user data privacy and ethical AI deployment. This will reinforce trust among SokoNews users and ensure that personalization efforts remain transparent and responsible.

Looking ahead, a future-proof, enterprise-grade recommendation system will require the integration of multi-modal data sources, including images, videos, and sentiment analysis. Additionally, real-time clickstream processing and adaptive learning mechanisms will be essential to refining recommendations as user behavior evolves. Our envisioned architecture leverages Azure Data Lake, Synapse, and Machine Learning services to enable scalable AI deployment while continuously improving the quality of recommendations. Furthermore, SokoNews plans to transition towards building a proprietary, in-house dataset, allowing for more refined personalization strategies tailored specifically to its audience. This shift will not only improve recommendation accuracy but also enhance data ownership, ensuring long-term sustainability and independence from third-party datasets.

The proposed personalized news recommender system represents a transformative opportunity for SokoNews, addressing both business and technological objectives. By integrating CTR prediction, BERT-based content modeling, and LLM-powered conversational recommendations, we have developed a robust, scalable, and modular system capable of tackling the complexities of news personalization. The projected first-year revenue increase of €39.8 million and market-leading ROI validate the investment and highlight the system’s potential to drive long-term engagement and profitability. Beyond financial gains, this initiative fosters user trust through transparent recommendation mechanisms, enhances content discovery, and ensures that even new users receive meaningful, context-aware article suggestions from their very first interaction.

SokoNews is well-positioned to take advantage of the AI-driven transformation in the media industry. Future iterations of the recommendation system will focus on real-time content updates, hybrid collaborative-content filtering approaches, and AI-driven summarization tools that allow users to consume news in a more efficient and engaging manner. Ethical considerations—including bias mitigation, privacy protection, and regulatory compliance—remain integral to our approach, ensuring fair and transparent personalization. By continuously refining its AI-driven strategy, SokoNews will not only enhance user engagement and retention but also maintain a competitive edge, establishing itself as a leader in personalized news delivery for the digital era.