Enhancing User Engagement with ML and LLM Capstone Project Summary

Introduction

User engagement is crucial for online service platforms. This capstone project aimed to enhance user engagement for an online broker comparison platform through innovative machine learning (ML) and large language model (LLM) applications. The project involved developing a predictive model for recommending relevant forum posts and integrating this with an LLM-driven virtual assistant to create a personalized user experience.

Project Goals and Rationale

The project had two primary goals. First, it sought to improve user engagement on the platform by predicting and recommending forum posts that would interest individual users. Second, it aimed to experiment with LLM technology to create a virtual assistant capable of providing users with datadriven content recommendations and easy access to various platform tools. These goals were driven by the need to address a "low stickiness" problem, where users seldom return to the platform after signing up with a recommended broker.

Analytical Problem and Scope

The central analytical problem was to understand what factors drive user engagement with forum content. The project aimed to develop a robust recommendation model to enhance user interaction on the platform by addressing this question. The scope of the project included:

- Developing a predictive ML model for post recommendations.
- Creating an LLM model integrated with the predictive model and platform tools.

 Ensuring the LLM model could intelligently interact with users and utilize predefined tools when needed.

System Design

The project design consisted of four main components:

- LLM Model: The central processor provides context-aware text generation and tool selection capabilities.
- Predictive ML Model: A histogram-based gradient-boosting classification tree selected after comparisons with a logistic regression model and a random forest model.
- Platform Tools Integration: Simulated tools such as a user questionnaire for investment preferences, broker recommendations, and a broker scam shield.
- Chat Interface: A simple terminal interface for users to interact with the LLM.

Data Collection and Feature Engineering

The project used data from the r/Investing subreddit due to limitations in accessing the platform's actual data. This subreddit was chosen for its relevance to investing topics similar to those discussed on the platform. Data collection involved gathering post and comment data via the Reddit API, supplemented with synthetically generated user profile data. The project also leverages LLM to deduce post categories and asset types, enriching the dataset further.

Key Outcomes and Benefits

The project's primary outcomes included a working prototype of a predictive recommendation model and an LLM virtual assistant. These tools are expected to significantly improve user engagement by providing personalized content and intuitive interaction with the platform. The project highlighted the importance of good data quality and advanced ML techniques in building effective recommendation systems.

Learning Experience and Lessons Learned

The project provided valuable learning experiences in building a good data analytics framework and integrating ML and LLM technologies. Key lessons included:

- The importance of data quality: Synthetic data can be helpful but has limitations compared to real-world data.
- Model selection and tuning: Experimenting with different models and fine-tuning parameters is crucial for achieving optimal performance.
- Practical integration: Combining predictive models with LLMs requires careful design to ensure seamless and intelligent interactions.

Conclusion

This capstone project successfully demonstrated the potential of ML and LLM technologies in analyzing and enhancing user engagement on an online broker comparison platform. By developing a predictive recommendation model and integrating it with an LLM virtual assistant, the project created a more interactive and personalized user experience. While there are limitations, such as the use of synthetic data and a simplified LLM, the project lays a solid foundation for future enhancements. Implementing more advanced models and accessing actual platform data can significantly improve the system's predictive accuracy and functionality, ultimately driving higher user retention and engagement. This project underscores the value of combining advanced technologies to solve real-world business challenges and improve user satisfaction.