Capstone Project Proposal: Predicting Transaction Values for Customer-Centric Insights

Problem Identification: In today's data-driven landscape, businesses are presented with a unique challenge: how to extract actionable insights from vast amounts of transactional data. The objective of this project is to tackle this challenge head-on by developing a sophisticated predictive model that goes beyond simple analytics. By accurately estimating transaction values, we aim to provide a comprehensive understanding of customer behaviors and preferences.

The essence of the problem lies in the variability and complexity of customer transactions. Traditional approaches often fall short in capturing the nuances inherent in transaction data, which encompasses a wide range of factors, including timing, products, customer demographics, and macroeconomic trends. Our project seeks to bridge this gap by harnessing the power of machine learning to decipher intricate patterns and correlations within these transactions.

Through the lens of predictive modeling, we endeavor to untangle the hidden relationships that govern transaction values. By doing so, we can unlock a deeper understanding of customer spending behaviors and uncover insights that have the potential to revolutionize business strategies.

Client and Their Interest: Our project's key clients—commercial banking institutions, e-commerce platforms, and retailers—are at the forefront of an industry-wide shift towards data-centric decision-making. These stakeholders understand that customer behavior lies at the core of business success. However, the challenge lies in extracting meaningful insights from the avalanche of transactional data generated every day.

Commercial banks, for instance, are eager to harness the potential of predictive analytics to refine credit offerings and enhance customer experiences. By accurately estimating transaction values, banks can tailor credit limits and rewards programs to align with individual customer preferences and behaviors. This, in turn, leads to greater customer satisfaction, reduced credit risk, and increased customer loyalty (Bhattacharya et al., 2020).

E-commerce platforms and retailers are also acutely aware of the need to personalize offerings and marketing campaigns. Accurate transaction value predictions enable them to optimize inventory management, adjust pricing strategies, and tailor promotional activities based on individual customer profiles. This level of personalization not only leads to higher conversion rates but also fosters brand loyalty and customer retention (Lee & Hosanagar, 2017).

In essence, our clients' interests are not confined to short-term gains; they are committed to building lasting customer relationships and fostering sustainable growth. Our predictive model serves as the bridge between raw transactional data and strategic insights, enabling our clients to navigate the intricate landscape of customer preferences with confidence and foresight.

Context: The project operates at the intersection of data analysis, machine learning, and business strategy. By analyzing customer transaction data, we aim to unearth patterns and trends that empower businesses to tailor their operations for maximum efficiency and customer satisfaction.

Criteria for Success: The success of this project hinges on two main pillars: the accuracy of transaction value predictions and the practicality of implementing the model in real-world scenarios. The predictive model should achieve a low prediction error, as evidenced by a low mean squared error (MSE) or other relevant evaluation metrics. Additionally, the model should be designed with scalability and ease of integration in mind, ensuring seamless adoption by client businesses.

Scope of Solution Space: This project encompasses a comprehensive approach, spanning data preprocessing, exploratory data analysis (EDA), feature engineering, model selection, model training, and evaluation. While the primary focus is on predicting transaction values, exploratory analyses may also uncover valuable insights into customer segmentation and spending behaviors.

Constraints: A few critical constraints to consider include data privacy and security when dealing with transactional data. Computational resources and time availability might impact the complexity of the chosen modeling techniques. Furthermore, the model's interpretability should be balanced with its predictive power to ensure actionable insights.

Stakeholders: The primary stakeholders are commercial banking institutions, e-commerce platforms, and retailers. The insights gained from accurate transaction value predictions will aid in tailoring services and offerings to meet customer demands effectively, resulting in improved customer satisfaction and business growth.

Data Sources: The project will rely on the Santander Value Prediction Challenge dataset available on Kaggle. This dataset encompasses anonymized customer transactional data, allowing us to extract valuable patterns while maintaining data privacy and security.

Project Workflow:

1. Data Preprocessing and Exploration: Clean and prepare the dataset for analysis. Conduct exploratory data analysis to understand the distribution of variables, identify missing values, and uncover preliminary insights.
2. Feature Engineering: Create relevant features that capture customer behaviors, temporal trends, and transaction patterns. This step is essential for enhancing the model's predictive capabilities.
3. Model Selection and Training: Experiment with various machine learning algorithms, such as regression, gradient boosting, and neural networks. Train and fine-tune models using appropriate techniques to achieve optimal performance.
4. Model Evaluation: Evaluate the models using robust validation strategies and relevant evaluation metrics. Select the model that demonstrates the highest predictive accuracy and generalizability.
5. Implementation and Reporting: Implement the chosen model in a practical business context. Develop a comprehensive report detailing the project's methodology, findings, and actionable insights for stakeholders.

Conclusion: By accurately predicting transaction values, this project aims to empower commercial banks and retailers with actionable insights that can guide their business strategies. The fusion of data science techniques and domain knowledge will drive decision-making processes, leading to improved customer satisfaction, optimized marketing efforts, and enhanced revenue generation.

Work by

- Sang Thai – Student
- Bernard Chan – Data Scientist & Machine Learning Specialist