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

Problem Identification: The objective of this project is to develop a predictive model that accurately estimates the transaction values of potential customers. By leveraging historical transactional data, this model will provide valuable insights for commercial banks and retailers, enabling them to make informed decisions to enhance customer experiences and optimize business strategies.

Client and Their Interest: The key clients encompass commercial banking institutions, e-commerce platforms, and retailers. These stakeholders are deeply invested in understanding customer behaviors and preferences to tailor their services, marketing campaigns, and product offerings. The ability to predict transaction values will allow them to allocate resources more effectively and enhance customer engagement.

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