Executive Summary: QNB FinansBank Final Report

Problem Definition

The problem revolves around providing customers with financial assistance through instant loans, a method utilizing the unused credit card limit. The traditional loan application process involves time-consuming controls, and cash advances from ATMs incur high-interest rates. The instant loan option, however, offers a quicker process with lower interest rates, leveraging the unused credit card limit. The unique challenge is managing the opportunity cost for banks, as they must keep a portion of each credit card limit in cash. Predicting which customers are likely to accept instant loans is crucial for optimizing credit utilization and improving overall efficiency. The project is framed as a binary classification problem, where the outcome is whether a customer is likely to accept the offered instant loan.

Initial Data

The provided dataset, comprising training and test data, was received in Excel format. The training data consists of 51 columns of customer information, with undisclosed meanings. There are 30,000 rows. Similarly, the test data includes 50 columns of customer information, with 15,000 rows. Initial data analysis revealed missing values in 32 columns, with some exceeding 50%. Addressing these missing values becomes a crucial step in preparing the data for subsequent model development.

Approach:

1. Preprocessing:

The initial phase focused on preparing the dataset for modeling.

- Data Analysis: Through exploratory data analysis, the team visualized the dataset, revealing imbalances and missing values. Techniques such as boxplots, count graphics, and heatmaps were employed.

- Handling Missing Values: The team experimented with multiple imputation methods, assessing their efficacy through RMSE scores. A highlight was using XGBoost for prediction, a unique approach for filling missing values.

- Feature Selection: Different methods, including SelectKBest and Recursive Feature Elimination, were employed to identify the most influential features for modeling.

2. Training:

This stage involved preparing the data for various models and refining them for optimal performance.

- **Preparation to Train:** Categorical variables underwent OneHotEncoding to facilitate training diverse models, including GaussianNB, Decision Tree, KNN, Logistic Regression, RandomForest, XGBoost, and LightGBM.

- **Hyperparameter Tuning:** Leveraging RandomizedSearch, the team fine-tuned hyperparameters, a crucial step in enhancing model accuracy.

3. Prediction:

With models trained and optimized, the focus shifted to evaluating their predictive power.

- **Prediction Stage:** The team assessed model performance on the test set, comparing results across classifiers. Metrics like AUC, Gini Index, and confusion matrices were pivotal in gauging effectiveness.

- **Performance Comparison:** Detailed analysis of metrics facilitated a comprehensive evaluation, offering insights into each model's strengths and weaknesses. Subsequently, a consolidated model was formulated.

4. Consideration of Alternatives:

Addressing challenges like class imbalance and potential model bias demanded thoughtful consideration of alternatives.

- **Handling Class Imbalance:** The team explored methods such as oversampling, undersampling, and SMOTE to tackle class imbalances. Ultimately, adjusting class weight parameters emerged as a practical solution.

Results:

The culmination of the group's efforts is a set of custom models tailored for QNB FinansBank's binary classification problem.

- **Model Customization:** The team created custom models capable of making moderate predictions regarding customer acceptance of instant loan offers.

- **Tools for Adaptation:** Providing a wrapper class and custom_ml class ensures flexibility, allowing easy adaptation and modification of models for future improvements.

- **Promising Outcomes:** The results, while promising, leave room for refinement. Further hyperparameter tuning and adjustments to the initial search space could yield enhanced predictive capabilities.

- **Operational Impact**: QNB FinansBank now possesses a toolset to strategically predict customer responses, potentially reducing opportunity costs associated with maintaining unused credit limits. The consolidated model contributes to a 0.015 increase in QNB FinansBank's prediction rate.