Capstone Project Submission

Credit Card Default Prediction
Project summary

Contribution - Individual Project

Member - Onkar Anil Pawar

E-mail – pawaronkar177@gmail.com

Problem Statement :

This project is aimed at predicting the case of customers' default payments in Taiwan. From the perspective of risk management, the result of predictive accuracy of the estimated probability of default will be more valuable than the binary result of classification - credible or not credible clients. We can use the K-S chart to evaluate which customers will default on their credit card payments.

Data Exploration:
This dataset contains information on default payments, demographic factors, credit limit, history of payments, and bill statements of credit card clients in Taiwan from April 2005 to September 2005. It includes 30,000 rows and 25 columns, and there is no credit score or credit history information. The data dictionary is available in Appendix A.

Modeling Preparation:
 Since there are labeled data and the expected outcome is the probability of customer default, we define this as supervised machine learning and it is a binary classification problem. For better model performance, we first take a few preprocessing steps to prepare for modeling.
● Feature Selection
● Check Class Imbalance
● Transform Categorical Column
● Split Training and Test Data

Predective Modeling:
This analysis uses 3 classification models - Logistic Regression, Random Forest, and XGBoost. Since Random Forest and XGBoost are tree based algorithms, rescaling is only performed on Logistic Regression, not on these 2 models. For each model, we first try the model's default parameters and train each model without SMOTE and with SMOTE samplings.
●SMOTE Oversampling
● Hyperparameters Tuning
● Performance Metrics
● Feature Importance

Conclusion: More credit card default for a limit balance of about 10000. It might mean that credit cards might be too easy to be issued for people who have low credit scores. The variance of the default rate for a limit balance over 500,000 NTD is higher than other ranges of limit balance. It is a lower default rate for cardholders who have a higher education level. Moreover, the default rate for clients whose age over 60 was higher than mid-age and young people. The best-fit algorithm for predicting limit balance is the bagging approach. The best-fit algorithm for predicting whether a client default next month is a classification tree.

GitHub Link:- https://github.com/Onkar-TAE/Credit-Card-Default-Prediction.git