FLIP ROBO

MICRO CREDIT DEFAULTER PROJECT

A project report submitted in

partial fulfilment of the requirement for the award of internship provided by

FlipRobo Technologies, Bangalore

Submitted by:

ELTON GRIVITH D SOUZA

engardaelton.work@gmail.com

OCTOBER 2021

ACKNOWLEDGMENT

INTRODUCTION

Problem Statement

A **Microfinance Institution (MFI)** is an organisation that offers financial services to low income populations. MFS becomes very useful when targeting especially the unbanked poor families living in remote areas with not much sources of income. The Microfinance services (MFS) provided by MFI are Group Loans, Agricultural Loans, Individual Business Loans and so on.

Many MFIs' experts and donors are supporting the idea of using mobile financial services (MFS) which they feel are more convenient and efficient, and cost saving, than the traditional high-touch model used since long for the purpose of delivering microfinance services. Though, the MFI industry is primarily focusing on low income families and are very useful in such areas, the implementation of MFS has been uneven with both significant challenges and successes.

Today, microfinance is widely accepted as a poverty-reduction tool, representing $70 billion in outstanding loans and a global outreach of 200 million clients.

The client is a fixed wireless telecommunications network provider. They have launched various products and have developed its business and organisation based on the budget operator model, offering better products at lower prices to all value conscious customers through a strategy of disruptive innovation that focuses on the subscriber.

They understand the importance of communication and how it affects a person's life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

Business Goal

The requirement of this project is to collaborate with an MFI to provide micro-credit on mobile balances to be paid back in 5 days.

The Consumer is believed to be defaulter if he deviates from the path of paying back the loaned amount within the time duration of 5 days.

For the loan amount of 5 (in Indonesian Rupiah), payback amount should be 6 (in Indonesian Rupiah), while, for the loan amount of 10 (in Indonesian Rupiah), the payback amount should be 12 (in Indonesian Rupiah).

In order to improve the selection of customers for the credit, the client wants some predictions that could help them in further investment and improvement in selection of customers.

Data Sources and their formats

The data source was provided by **FlipRobo Technologies** stating the source of the data is from a client who is a fixed wireless telecommunications network provider. The format as provided by the company was a CSV file containing 209593 tuples with 37 features.

Data Pre-processing

Since the data had a lot of features we first conducted VIF to reduce the number of features. The threshold was taken to be 5 which reduced our data size by five features

The provided data was heavily imbalanced. This was treated by upsampling using imblearn's' ADASYN algorithm. This increased the data size to 367057 features.

The data also consisted of some unrealistic values. But this was ignored to avoid overfitting our model. But we reduced its influence by fixing the skew of our data.

Skew was treated by using scikits' power transform method. yeo-johnson algorithm used for the treatment. This process was followed by normalisation to make the distribution more Gaussian.

The data was also numerically encoded and scaled using standard scaler to make the data easier to work with and also to accelerate the computations.

Data Relationships

The relationships analysed from the data was done through exploring the correlation, distribution of feature values, distribution of feature sets, distribution of feature values in regard to the target feature etc.

The resulting relationships explored the features that have high correlation with the target label. This was done by using univariate and multivariate feature exploration.

Hardware and Software Requirements and Tools Used

The development of this project required the following tools:

- Python 3.8

- CUDA 10

- Scikit-learn 1.0

Model Development and Evaluation

Model Development

The model was trained on a 25% split of the training data which was fed to a pipeline of different classification algorithms. These models are listed below:

- Decision Tree Classifier

- Random Forest Classifier

The model was evaluated based on the accuracy, precision, recall and CV using AUC/ROC plots.

Hyper Parameter Tuning *(HPT)*

Parameter tuning was done using GridSearchCV on Decision Tree Classifier. This model was chosen based on its CV scores. The grid parameters that were used are given below:

- criterion: ['gini','entropy']

- max_depth: range(8,26,4)

- min_samples_leaf: range(22,40,4)

- min_samples_split: range(10,50,4)

The best parameters obtained from this process was then used to provide the submission model.

CONCLUSION

The submitted model provides consistent results that can be validated under other circumstances. The result obtained from the provided data can be assumed to be competitive and can provide better insights to the client. The provided solution satisfies all the requirements stated by the client and hence this project can be considered successful.

Future Work could include cleaner, balanced and localised data with more independent feature sets to provide better estimations and analytical relationships.