Micro-Credit Defaulter Model

Submitted by:

Charchit Pathak

ACKNOWLEDGMENT

This dataset , data sources are provide by “FLIPROBO” and I also take references from some other loan prediction datsets, which done it by me. The guidance and helped provide by our “SME” to completion of the project.

INTRODUCTION

Business Problem Framing

Microfinance Institution (MFI) is an organization 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 microfinance institutions (MFI), 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.

We are working with one such client that is in Telecom Industry. They are a fixed wireless telecommunications network provider. They have launched various products and have developed its business and organization 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.

They are collaborating 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). The sample data is provided to us from our client database. It is hereby given to you for this exercise. 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.

Conceptual Background of the Domain Problem

In this model we understand the loan defaulter and the credit process through a telecom system model prediction and 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.
The sample data is provided to us from our client database. It is hereby given to you for this exercise. 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.
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).

Review of Literature & Motivation for the Problem Undertaken

The sample data is provided to us from our client database. It is hereby given to you for this exercise. 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. 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. A Microfinance Institution (MFI) is an organization 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.

Our objective behind to make this project, this domain and what is the motivation behind.Build a model which can be used to predict in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount within 5 days of insurance of loan. In this case, Label '1' indicates that the loan has been payed i.e. Non-defaulter, while, Label '0' indicates that the loan has not been payed i.e. defaulter.

The sample data is provided to us from our client database. It is hereby given to you for this exercise. 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.

Analytical Problem Framing

Mathematical/ Analytical Modeling of the Problem

Model which can predicts in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount within 5 days of insurance of loan. For that problem we are using the pandas , matplotlib, python and compare the models to understand the dataset also we describe the dataset model to know the mathematical values of the datset and also use for identify correleation of the dataset variable because it is useful to understand and comparision purpose use and also identifying the values the datasets.

Data Sources and their formats

We will do the prediction of the Micro - Credit Defaulter Model that dataset provide by the FLIP-ROBO and that in that dataset we are trying to predict the in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount within 5 days of insurance of loan. In this case, Label ‘1’ indicates that the loan has been payed i.e. Non- defaulter, while, Label ‘0’ indicates that the loan has not been payed i.e. defaulter.

Data Inputs- Logic- Output Relationships

The inputs of data models are pandas, models, matplotlib, seaborn that are the models are used to prediction the dataset and that dataset are used to analysis with the help of models and we do comparision , description , correlation , visualization. Process for getting output

Model/s Development and Evaluation

Identification of possible problem-solving approaches (methods)

The analytical approaches followed, for solving of this problem using the models comparision between the variable using the different models and find out the which one is suitable for dataset and fullfill the dataset approached and also used to solve different classifier, regression, roc_auc curve , confusion matrix.

Testing of Identified Approaches (Algorithms)

For training and testing approach used the train- test split model for the dataset and the used metrics model for predict test and the fit model for find the accuracy, prediction test , auc_roc_curve. Build a model which can be used to predict in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount within 5 days of insurance of loan.

Key Metrics for success in solving problem under consideration

the key metrics used for the predicting the dataset their model and with their variable and the matrices are generate the output result just like comparision Effect of Daily amount spent over last 30 days on Delinquency and Daily amount spent over last 30 days(in Indonesian Rupiah) for Daily amount spent from main account, averaged over last 90 days (in Indonesian Rupiah), evaluating the result roc , there are methods are used and we will see more to see solutions.

Interpretation of the Results

A summary of results were interpreted from the visualizations, processing and modelling. The dataset for prediction the output are generating by use of various model but the satisfied result is produced by randomforestclassifier . First we generate data and know about the information and the types of data then datacleaning and EDA process for correlating ,missing value,and the dropout the unused columns.then use various classfier and visualization , auc_roc curve then conclusion .

CONCLUSION

Conclusions of the Study

The conclusion of the study predict in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount days of insurance of loan. Label ‘1’ indicates that the loan has been payed i.e. Non- defaulter. The sample data is provided to us from our client database. It is hereby given to me for this exercise. 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.

Learning Outcomes of the Study in respect of Data Science

The outcome in respect of datascience, data is provided to us from our client database. It is hereby given to for this exercise. 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. A model which can be used to predict in terms of a probability for each loan transaction.

Limitations of this work and Scope for Future Work

- There are no null values in the dataset.
- There may be some customers with no loan history.
- The dataset is imbalanced. Label ‘1’ has approximately 87.5% records, while, label ‘0’ has approximately 12.5% records.
- For some features, there may be values which might not be realistic. You may have to observe them and treat them with a suitable explanation.
- You might come across outliers in some features which you need to handle as per your understanding.

The scope for datascience in that type of model is not only useful for loan defaulter prediction also for various fields like agriculture, bank .that type of dataset prediction is useful for 5 days but also yearly approached. the working with one such client that is in Telecom Industry. They are a fixed wireless telecommunications network provider. They have launched various products and have developed its business and organization 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. it 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.