



## MICRO CREDIT DEFAULTER PROJECT

Submitted by:

Sreekari I

### **ACKNOWLEDGMENT**

This project is on the micro credit defaulter. It includes the data provided by a telecom industry. The sample data is provided to us from our client database.

# INTRODUCTION

- **Business Problem Framing**

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.

- **Conceptual Background of the Domain Problem**

To identify the low income families to provide micro credit for mobile recharges.

- **Motivation for the Problem Undertaken**

The Micro Finance Institutions are focusing on providing their services and products to low income families and poor customers that can help them in the need of hour. However, there are some customers behave negatively after their application are approved. To prevent this situation, we have to find some methods to predict customers' behaviours. Machine learning algorithms have a pretty good performance on this purpose, which are widely-used by the banking. Here, I will work on loan behaviours prediction using machine learning models.

## **Analytical Problem Framing**

- **Mathematical/ Analytical Modelling of the Problem**

In this project, I have used Logistic Regression, Decision Tree and Random Forest Classifier, Naive Bayes Classification

- **Data Sources and their formats**

The sample data is provided to us from our client database. It is in the csv format and excel sheet.

- **Data Pre-processing Done**

I have done exploratory data analysis on the data

- **State the set of assumptions (if any) related to the problem under consideration**

1. 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.
2. The Consumer is believed to be a non defaulter if the loaned amount is paid within the time duration of 5 days.
3. Defaulters are labelled as '0'
4. Non-defaulters are labelled as '1'

- **Hardware and Software Requirements and Tools Used**

I have used jupyter notebook for the entire project.

## **Model/s Development and Evaluation**

- **Identification of possible problem-solving approaches (methods)**

The classification models used for this analysis are: Logistic Regression, Decision Tree and Random Forest Classifier, Naive Bayes Classification

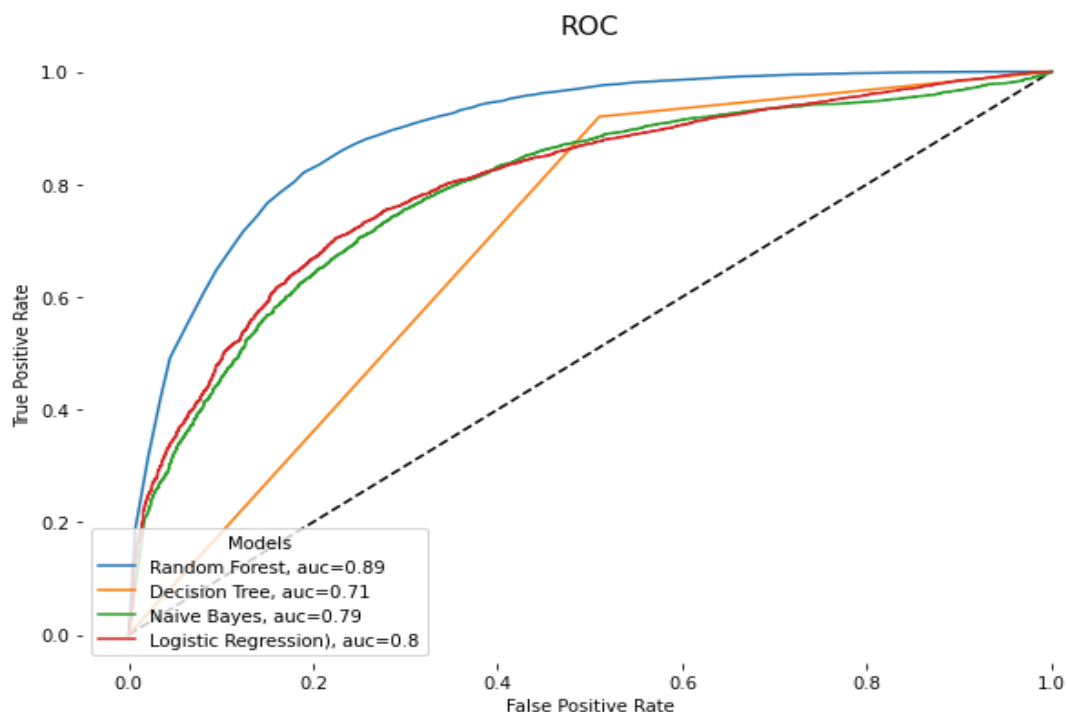
- Testing of Identified Approaches (Algorithms)

To build machine learning models the original data was divided into features (X) and target (y) and then split into train (80%) and test (20%) sets. Thus, the algorithms would be trained on one set of data and tested out on a completely different set of data (not seen before by the algorithm).

- Visualizations

In this project, I have used seaborn library and matplotlib. Heatmap, barplot, boxplot were used for the visualization

- Interpretation of the Results



From the above ROC curve, we can see that the random forest is the best model for this classification.

## CONCLUSION

Here, you can see the random forest prediction works very well, which means it can confidently tell you who is the good customer. By contrast, it has a pretty low recall when predicting the loan default behaviours. In laymen's terms, recall means how many cases are predicted correctly among all the true conditions. They only cover a small part of the total amount of customers with non default behaviours.

To improve the recall of the model, we can use the the probabilities predicted by the model and set threshold by ourselves. The threshold is set based on several factors such as business objectives. It is different case by case. In credit behaviour prediction, for example, micro finance institutions want to control the loss to a acceptable level, so they may use a relatively low threshold. This means more customers will be grouped as "defaulters" and their profiles will be checked carefully later by the credit risk management team. In this way, the telecom industry can detect the default behaviours in the earlier stage and conduct the corresponding actions to reduce the possible loss.