

Micro-Credit Defaulter

Submitted by:

Laxman Singh

**ACKNOWLEDGMENT**

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

**INTRODUCTION**

* Business Problem Framing

The MFI offer offers financial services to low income populations and our client is collaborating with an MFI to offer micro-credit on mobile balances to be paid back in 5 days. On the basis of the sample data we need to 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

* Conceptual Background of the Domain Problem

The loans offered are micro-credit on mobile balances to be paid back in 5 days. The payback of the loan may be dependent on various factors like the credit given, loan amount, payback history so we need to use the sample data and predict the probability of the customer being a loan defaulter

* Review of Literature

Various permutations and combinations have been used to identify the likely hood of the loan being returned.

* Motivation for the Problem Undertaken

MFI are a great source of help for low income families by rendering service in a flexible format and low complexity however it can lead to a very big impact on the economy if these loans turn to NPA hence it is very critical to predict the likelihood of loan being returned

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

Probability of loan being returned has been checked on the basis of various permutations and combination of the nature of the buyers as well as their history of finances from the sample data

* Data Sources and their formats

Sample data was provided by the client in a .CSV format improve the selection of customers for the credit

* Data Preprocessing Done

There were no null values in the data sets however few outliers were present which have been removed

* Data Inputs- Logic- Output Relationships

After passing the various parameter related to the consumer history and pattern of usage, the output is predicted to identify the risk and accordingly select the customer for the services

* Hardware and Software Requirements and Tools Used

The model has been built in Jupyter notebook using python

**Model/s Development and Evaluation**

* Testing of Identified Approaches (Algorithms)

This is a classification problem so following Classification Algorithms have been used to build the model.

KNeighborsClassifier , Support Vector , LogisticRegression , DecisionTreeClassifier, GaussianNB, GradientBoostingClassifier , AdaBoostClassifier , ExtraTreesClassifier

* Visualizations

Density plots have been used to check the data distribution and other plots have been used by grouping various features

* Interpretation of the Results

Post analysing the sample data through various visualisation techniques, it was observed that customer with maximum number of loans, less of very limited recharges can be a potential risk

**CONCLUSION**

* Key Findings and Conclusions of the Study

After conducting trials on various algorithms we can identify that Random Forest Classifier has given the best results in terms of accuracy score and cross Val score hence it is the final model selected