#### 

#### MICRO CREDIT FINANCE PROJECT ON TELECOM

#### Submitted by:

#### AVINASH SINGH

#### ACKNOWLEDGMENT

#### In preparation of this Project, I had to take the help and guidance of some video tutorials , article on medium and several google pages. As the completion of this Project gave me much pleasure. I would like to expand my gratitude to all those who have directly and indirectly who helped me to use wide thinking to explore here in dataset.

#### In addition, a thank you to my mentor Nitin Mishra, who introduced me to the Methodology of work, and who support always wherever I stuck ever. I also thank FlipRobo Technologies for providing such an opportunity to work on various types of projects which gradually improve my vision to apply for the datasets.

#### INTRODUCTION

#### Business Problem Framing:

#### The low income families in rural and slums area have always the issue to fulfilling the needs they have and the abilities to complete them sometimes happen and sometimes not. And the reason is non other than finance. A Microfinance Institution is an organization that offers financial services to these low income populations. The financial services becomes very useful when targeting especially the unbanked poor families living in remote areas with not much sources of income.

#### Conceptual Background of the Domain Problem:

#### Microfinance is a category of financial services targeting individuals and small businesses who lack access to conventional [banking](https://en.wikipedia.org/wiki/Banking) and related services. Microfinance includes [microcredit](https://en.wikipedia.org/wiki/Microcredit), the provision of small loans to poor clients; [savings](https://en.wikipedia.org/wiki/Savings_account) and [checking accounts](https://en.wikipedia.org/wiki/Checking_account); [micro-insurance](https://en.wikipedia.org/wiki/Microinsurance) ; and [payment systems](https://en.wikipedia.org/wiki/Payment_system), among other services. Microfinance services are designed to reach excluded customers, usually poorer population segments, possibly socially marginalized, or geographically more isolated, and to help them become self-sufficient.

#### Review of Literature:

#### Microfinance was defined initially as the provision of microloans to poor entrepreneurs and small businesses lacking access to credit. The two main mechanisms for the delivery of financial services to such clients were: (1) relationship-based banking for individual entrepreneurs and small businesses and (2) group-based models, where several entrepreneurs come together to apply for loans and other services as a group. Over time, microfinance has emerged as a larger movement whose object is: *"a world in which as everyone, especially the poor and socially marginalized people and households have access to a wide range of affordable, high quality financial products and services, including not just credit but also savings, insurance, payment services, and fund transfers.”*.

#### Motivation for the Problem Undertaken:

#### The Project here is also one of the data collected by Micro finance company with respect to telecom industry. The data presented the various types of customers taking the loans from the company who paus on time and some are not paying it and never even taking an excuse. So using the above data, I have to see the person is defaulter or non defaulter by exploring his previous 30 and 90 days record represented from the data statistics.

#### 

#### Analytical Problem Framing

#### Mathematical/ Analytical Modelling of the Problem:

#### Dataset has been checked for null value, datatypes to see that there will be any object type dataset present or not. Our target column is “label” which indicating whether the user paid back the credit amount within 5 days of issuing the loan{1:success, 0:failure}.The target column which is ‘Label’ here shows the data which is imbalanced. On displaying the statistical summary for the data, it is observed that each and every feature here has the outliers so applying z-score to cut it off

#### Data Sources and their formats:

#### I received the dataset through my mentor and is in the .csv file. Some of the info such as series number, msisdn are seems unusable with respect to the problem.

#### 

#### Data Pre-processing:

#### The first step in the data pre-processing is to obtained the whole information about the dataframe including the index dtype and columns, non-null values and memory usage. Next step is t do EDA which include count plotting, the target column “Label”.

#### Visualization:

#### Average main balance account vs loan payback rate within 5days.

#### Frequency of main account recharged in last 30 days vs loan payback rate within 5 days.

#### Number of loans taken by user in last 30 days vs loan payback within 5 days.

#### Dropping Unnamed:0,msisdn,pcircle,pdate column because it does not give any predictive information.

#### On verifying the correlation among the data, some of the feature seem to be highly correlated through data, so I drop it to remove the ambiguity while operations.

#### Data Inputs- Logic- Output Relationships:

#### The output of the data is completely depends on the input we provide. So, its best the scale the input correctly and remove the clutter to get the data that adds value to the prediction.

#### Hardware and Software Requirements and Tools Used:

#### Hardware And Software required for this project:

#### Laptop with I3 processor and 8 GB Ram

#### Google colab

#### Python Pandas for processing

#### Scikit learn library

#### Model/s Development and Evaluation

#### Identification of possible problem-solving approaches (methods):

#### From very basics step after importing dataset, I checked df.info() i.e it give all the imp information of dataset such as datatype, null value, shape of dataset. On displaying the statistical summary for the data, it shows the huge number of outliers on all the features, so applying z-score to minimise it.

#### Testing of Identified Approaches (Algorithms)

#### Following are the algorithms that I applied on this dataset. Random search is used for hyper parameter tuning the best model.

#### Logistic Regression

#### Gaussian NB

#### Decision Tree Classifier

#### K Neighbors Classifier

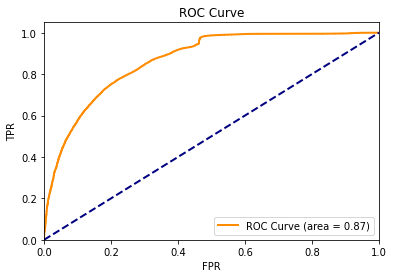
#### Random Forest Classifier

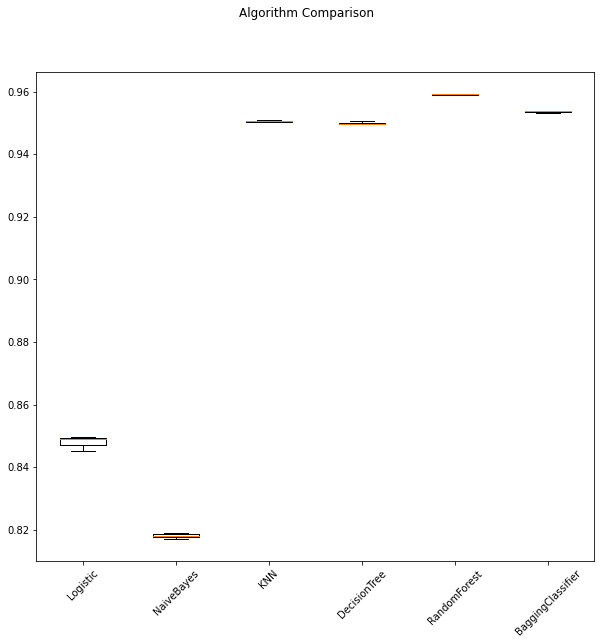
#### Bagging Classifier

#### Run and Evaluate selected models

#### Logistic regression:

#### 





* **Final score**
* Logistic: 0.847968
* Naïve Bayes: 0.818116
* KNN: 0.950553
* Decision Tree: 0.949900
* Random Forest: 0.959107
* Bagging Classifier: 0.953642
* From the above results it is observed that Random Forest is the best performing model. By comparing all algorithms bias error and variance error, random forest is observed to be the best so it would be used to predict loan defaulters. The test of random forest with base estimator (Decision Tree (which is default for random forest), n\_estimators=7) model successfully achieved a weighted F1\_score of 95%, suggesting high level of strength of this model to classify loan defaulter’s.

#### Metrics for success in solving problem under consideration

#### Key metrics are F1-score, Precision, Recall and Roc-Auc Score.

#### Visualizations

#### 

#### From above bar plot we can observe that how customers with different main balance levels are paying back the loan with in five days. The high balance level people are with 100% rate i.e they are paying loan within 5 days. Coming to the average and low balance people it is observed that around 10%-12% of people are not paying the loan within 5 days. Coming to low balance level people, it is observed that around 30% of people are not paying back the loan with in 5 days of time. The 30% of people with no balance or negative balance people are creating a major loss to the company without paying back the loan within five days of time.

#### 

#### From above bar plot we can observe that how customers with different frequency levels (main account recharge) are paying back the loan within five days. The is no 100% rate in any of the frequency levels to pay back the loan within 5 days. Coming to the average and low & medium frequency people it is observed that around 5%-6% of people are not paying the loan within 5 days.Coming to low frequency level people, it is observed that around 25% of people are not paying back the loan with in stipulated 5 days of time. The 25% people who are not getting their main account recharge for 30 days creating a major loss to the company without paying back the loan within five days of time.

#### 

#### From above bar plot we can observe that how customers with different loans levels taken are paying back the loan within five days. In the data set people not taken loans are labelled as ‘1’. So we should not consider the people with no loans labelled in the above graph. Considering the remaining levels, there is no 100% rate in any of the loan levels to pay back the loan within 5 days. Coming to the high number of loan level people it is observed that around 25% of people are not paying the loan within 5 days. Only 2% of the people from low number of loans category are not paying the loan within 5 days. This is followed by the people with medium number of loans having defaulters of 7% approximately.

#### 

#### 

#### CONCLUSION

#### Key Findings and Conclusions of the Study

#### Some of the observations I infer from the whole project are as follows:

#### We observe that how customers with different main balance levels are paying back the loan with in five days. The high balance level people are with 100% rate i.e they are paying loan within 5 days. Coming to the average and low balance people it is observed that around 10%-12% of people are not paying the loan within 5 days. Coming to low balance level people, it is observed that around 30% of people are not paying back the loan with in 5 days of time. The 30% of people with no balance or negative balance people are creating a major loss to the company without paying back the loan within five days of time.

#### Next observation that how customers with different frequency levels (main account recharge) are paying back the loan within five days. The is no 100% rate in any of the frequency levels to pay back the loan within 5 days. Coming to the average and low & medium frequency people it is observed that around 5%-6% of people are not paying the loan within 5 days. Coming to low frequency level people, it is observed that around 25% of people are not paying back the loan with in stipulated 5 days of time. The 25% people who are not getting their main account recharge for 30 days creating a major loss to the company without paying back the loan within five days of time.

#### How customers with different loans levels taken are paying back the loan within five days. In the data set people not taken loans are labelled as ‘1’. So we should not consider the people with no loans labelled in the above graph. Considering the remaining levels, there is no 100% rate in any of the loan levels to pay back the loan within 5 days. Coming to the high number of loan level people it is observed that around 25% of people are not paying the loan within 5 days. Only 2% of the people from low number of loans category are not paying the loan within 5 days. This is followed by the people with medium number of loans having defaulters of 7% approximately.

#### In order to decrease loss to the company, the company should start some marketing strategies like sms alerting and notifications and others on the people with all loan levels and especially on low & high level people notifying them to pay the loan back within five days of time.