

Micro Credit Loan

Submitted by:

Dhara Patel

**ACKNOWLEDGMENT**

I would like to express my special thanks to all those who spent their valuable time for guiding and helping me to complete this project.

I express my gratitude towards my guide **Data Trained Pvt. LTD mentor’s and Flip Robo’s mentor’s**.

I also appreciate my family and friends for their constant inspiration.

**Patel Dhara M.**

**INTRODUCTION**

* Business Problem Framing
* 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.
* 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.
* 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.
* Here in this dataset it 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).
* Conceptual Background of the Domain Problem

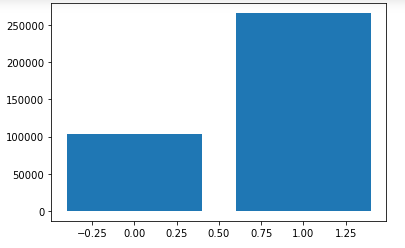
Here the basic concept is related to label mean we have to predict about loan defaulter and non- defaulter.

Lots of details are available in this data set like data pack for 30 days, 90 days, loan amount, etc.

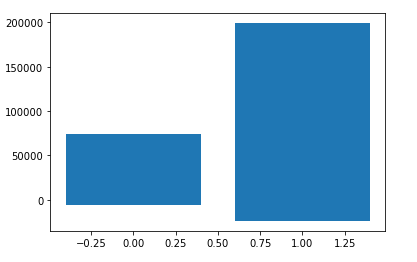
In this graph 0 indicate defaulter and 1 indicate non- defaulter

This graph is about Daily amount spent from main account, averaged over last 30 days (in Indonesian Rupiah)

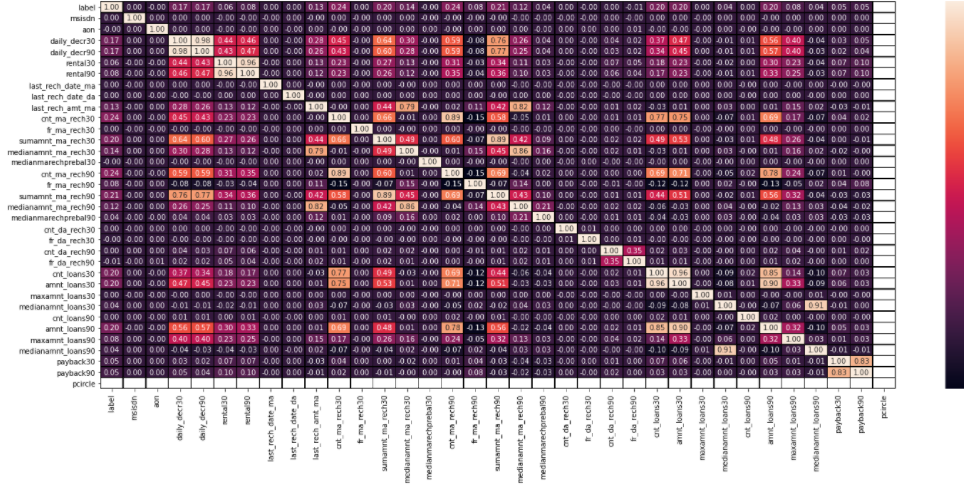
Daily amount spent from main account, averaged over last 90 days (in Indonesian Rupiah)



Given below graph show that Average main account balance over last 30 days and Average main account balance over last 90 days



In all most data there are less defaulters.

In below figure it show the correlation we can see that some columns data has 0 correlation with all most columns. so we will ignore this columns in feature selection. Because of this unnecessary columns may be we decrease our accuracy score so when we have to select features this is the most important task about best features those are useful and important for model developing.

* Review of Literature

Here using given dataset of micro credit loan apply some visualization methods we can see that there are less defaulter in loan.

After all this analysis and EDA phase we select most appropriate feature for model building. msisdn ,aon ,daily\_decr30 , daily\_decr90 , rental30 , rental90

last\_rech\_amt\_ma, cnt\_ma\_rech30, sumamnt\_ma\_rech30, medianamnt\_ma\_rech30, medianmarechprebal30 , cnt\_ma\_rech90, fr\_ma\_rech90, sumamnt\_ma\_rech90, medianamnt\_ma\_rech90, cnt\_da\_rech30

fr\_da\_rech30, cnt\_da\_rech90, fr\_da\_rech90, cnt\_loans30, amnt\_loans30

maxamnt\_loans30, medianamnt\_loans30, cnt\_loans90, amnt\_loans90, maxamnt\_loans90, medianamnt\_loans90, payback30, payback90, pcircle

Model Test Accuracy Score Train Accuracy Score

LogisticRegression() 88% 87%

DecisionTreeClassifier() 86% 99%

KNeighborsClassifier() 87% 90%

GaussianNB() 57% 57%

SVC() 88% 88%

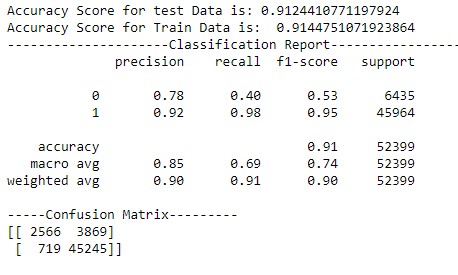
RandomForestClassifier() 91% 100%

AdaBoostClassifier() 90% 90%

GradientBoostingClassifier() 90.95% 90.90%

GradientBoost using Hyper Parameters we got 91.24% accuracy in given dataset.

This below figure gives information about recall , precision,f1 score etc.



So build our model using GradientBoosting method with below hyper parameters

'learning\_rate': 0.1,

'max\_depth': 5,

'min\_samples\_leaf': 2,

'min\_samples\_split': 3,

'n\_estimators': 100

and get 91.24% accuracy.

* Motivation for the Problem Undertaken

The main objective of this project is to find loan defaulter and after done this we can check how to reduce loan defaulter.

**Analytical Problem Framing**

* Data Preprocessing Done

Data cleaning can be done using label encoder for numeric data. Then check data skewness to check outliers and remove outliers using square root method.

* Data Inputs- Logic- Output Relationships

According to the analysis data inputs are . msisdn ,aon ,daily\_decr30 , daily\_decr90 , rental30 , rental90

last\_rech\_amt\_ma, cnt\_ma\_rech30, sumamnt\_ma\_rech30, medianamnt\_ma\_rech30, medianmarechprebal30 , cnt\_ma\_rech90, fr\_ma\_rech90, sumamnt\_ma\_rech90, medianamnt\_ma\_rech90, cnt\_da\_rech30

fr\_da\_rech30, cnt\_da\_rech90, fr\_da\_rech90, cnt\_loans30, amnt\_loans30

maxamnt\_loans30, medianamnt\_loans30, cnt\_loans90, amnt\_loans90, maxamnt\_loans90, medianamnt\_loans90, payback30, payback90, pcircle

and Label is data output.