

**MICRO CREDIT PROJECT**

Submitted by:

CHAITRA H N

**ACKNOWLEDGMENT**

I’d like to thank FlipRobo Technologies for this opportunity.

Resources Used:

1. <https://stackoverflow.com/>
2. <https://github.com/>

Mentor:

Khushboo Garg

**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

It is important to understand that the consumers are likely to take a loan for 5 (in Indonesian Rupiah) or 10 (in Indonesian Rupiah). There are possibilities that the consumer might be able to return it in 5 days. However, this would depend on the status and ability of the consumer. Without proper digging into the data, we can only consider that consumers are more likely to return the loan. We might want to consider if there are consumers who haven’t returned the loan for 30 or 90 days.

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

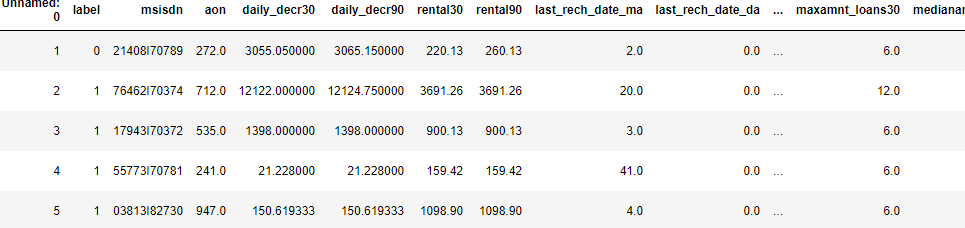
This project helps us to understand the nature of consumers while taking a loan. The amount is small and should be easy to clear unlike a credit card or a home loan. This project gives lot of insights on pre-paid users and the frequency of returning the loan.

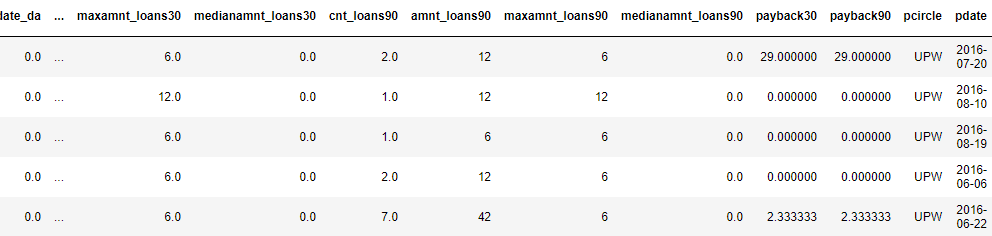
**Analytical Problem Framing**

* Mathematical/ Analytical Modelling of the Problem

1. Understanding outliers
2. Valid data range of each feature
3. Standard deviation for each feature
4. Scaling the data

* Data Sources and their formats
* Data Source: Flip Robo Technologies
* Origin: Client
* Format: Comma Separated Values (CSV)
* Data Types:
  + datetime64[ns] - 1
  + float64 - 21
  + int64 – 12
  + object – 2





* Data Preprocessing Done
* Some of the columns that are not adding any value to the predicted variable are dropped off such as Unnamed:0,msisdn,pcircle,pdate.
* Removed outliers – There are few outliers that are mandatory and wasn’t removed to prevent loss of other data in the row
* Converted the currency to Indonesian Rupiah
* Data Inputs- Logic- Output Relationships

The output of the data is completely depends on the input we provide. So,its best to 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
* numpy – mathematical solution library
* pandas – handle dataframes
* warnings – ignore warnings when as it arises (Optional)
* seaborn - visualization
* matplotlib.pyplot - visualization
* StandardScaler – scale data
* train\_test\_split – split train and test from pre-processed data
* accuracy\_score – evaluation
* roc\_auc\_score – evaluation
* confusion\_matrix – evaluation
* roc\_curve – evaluation
* classification\_report – evaluation
* plot\_confusion\_matrix – visualization
  + - **Machine Learning Algorithms**

**Model/s Development and Evaluation**

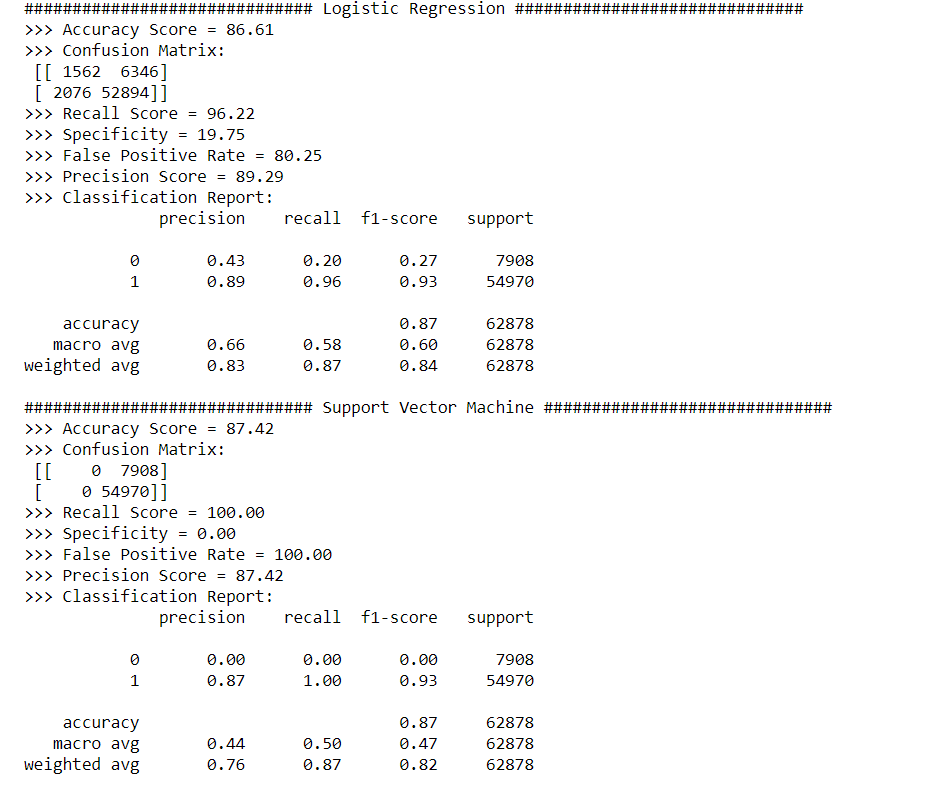
* Identification of possible problem-solving approaches

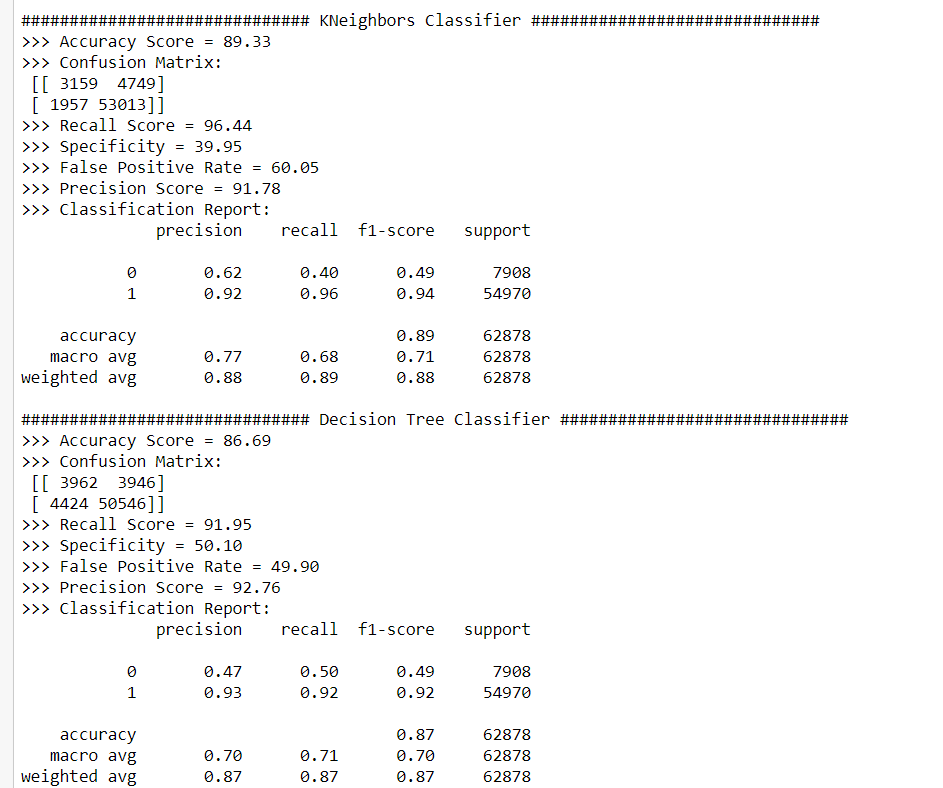
Before taking any further steps, I checked for the datatype first and then dropping some unnecessary columns. On displaying the statistical summary for the data,it shows the huge number of outliers on all the features,so applying z-score to minimise it.Then applying other models to make the prediction better using some hyperparameter tuning.

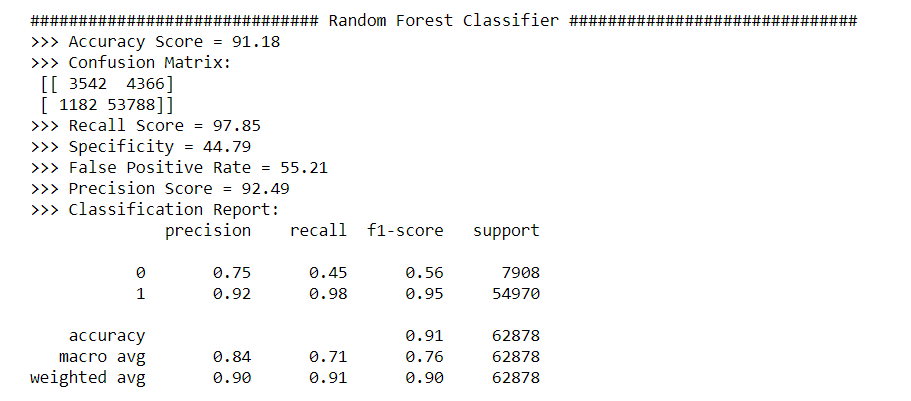
* Testing of Identified Approaches (Algorithms)

Following are the algorithms that I applied on this dataset

* Logistic Regression
* Support Vector Classifier
* K-Neighbors Classifier
* Decision Tree Classifier
* Random Forest Classifier
* Snipped Screenshot of Applied algorithms







* Key Metrics for success in solving problem under consideration

As it is clear from above observations that we have bias in the data labels or imbalanced data. So the performance metrics I consider for this data on all the applied algorithms are F1-score, Precision, Recall and Roc-Auc Score.

* Interpretation of the Results

Random Forest without any hyperparameter tuning gives a good result also shows it is also above the acceptance margin.

**CONCLUSION**

* Key Findings and Conclusions of the Study

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

* The Same data is presented on these dataset using various different features but shows high correlation and can be processed using only one features. Some of these features are “Date”,”rental30”,”rental90”.
* Heatmap of correlation of all the features comprising together shows some are not proper features to used for predictions.
* The value counts of target variable clearly shows the given dataset is imbalanced and in the improper ratio.
* Learning Outcomes of the Study in respect of Data Science

It is clear from data processing and after statistical summary that the whole dataset is imbalanced in nature. And it need a little regularisation or operations to handle it all. We first check it using Linear model any tuning. So it better to process the random forest with some tuning with oversampling and Under-sampling.