

MICRO CREDIT PROJECT

Submitted by:

ALAN TRESA ANANYASE

**ACKNOWLEDGMENT**

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

Resources Used:

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

Mentor:

Nitin Mishra

**INTRODUCTION**

* Business Problem Framing

To identify a consumer is a defaulter or not. Our Client is collaborating with an MFI to 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

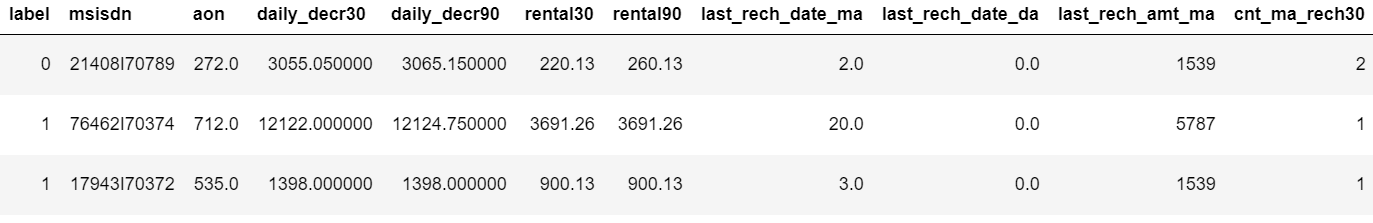
It is important to understand that the consumers are likely to take a loan for 5 (in Indonesian Rupiah) or 10 (in Indonesian Rupiah). There are possibilities that the consumer might be able to return it in 5 days. However, this would depend on the status and ability of the consumer. Without proper digging 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.

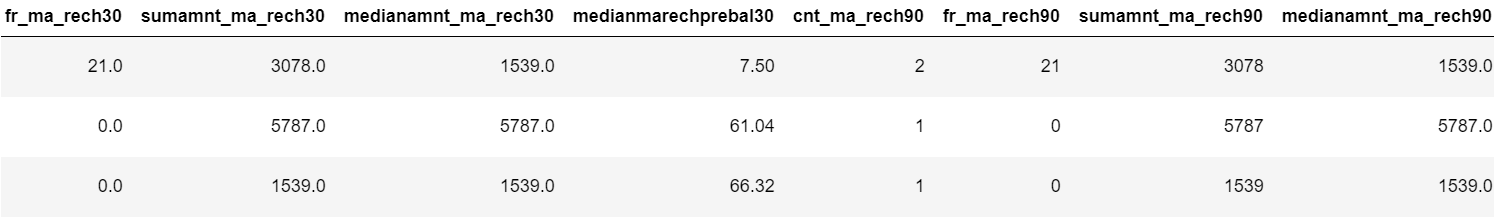
* Review of Literature
  + Label - Flag indicating whether the user paid back the credit amount within 5 days of issuing the loan {1:success, 0:failure}
  + Msisdn - mobile number of user
  + Aon - age on cellular network in days
  + daily\_decr30 - Daily amount spent from main account, averaged over last 30 days (in Indonesian Rupiah)
  + daily\_decr90 - Daily amount spent from main account, averaged over last 90 days (in Indonesian Rupiah)
  + rental30 - Average main account balance over last 30 days
  + rental90 - Average main account balance over last 90 days
  + last\_rech\_date\_ma - Number of days till last recharge of main account
  + last\_rech\_date\_da - Number of days till last recharge of data account
  + last\_rech\_amt\_ma - Amount of last recharge of main account (in Indonesian Rupiah)
  + cnt\_ma\_rech30 - Number of times main account got recharged in last 30 days
  + fr\_ma\_rech30 - Frequency of main account recharged in last 30 days
  + sumamnt\_ma\_rech30 - Total amount of recharge in main account over last 30 days (in Indonesian Rupiah)
  + medianamnt\_ma\_rech30 - Median of number of recharges done in main account over last 30 days at user level (in Indonesian Rupiah)
  + medianmarechprebal30 - Median of main account balance just before recharge in last 30 days at user level (in Indonesian Rupiah)
  + cnt\_ma\_rech90 - Number of times main account got recharged in last 90 days
  + fr\_ma\_rech90 - Frequency of main account recharged in last 90 days
  + sumamnt\_ma\_rech90 - Total amount of recharge in main account over last 90 days (in Indian Rupee)
  + medianamnt\_ma\_rech90 - Median of number of recharges done in main account over last 90 days at user level (in Indian Rupee)
  + medianmarechprebal90 - Median of main account balance just before recharge in last 90 days at user level (in Indian Rupee)
  + cnt\_da\_rech30 - Number of times data account got recharged in last 30 days
  + fr\_da\_rech30 - Frequency of data account recharged in last 30 days
  + cnt\_da\_rech90 - Number of times data account got recharged in last 90 days
  + fr\_da\_rech90 - Frequency of data account recharged in last 90 days
  + cnt\_loans30 - Number of loans taken by user in last 30 days
  + amnt\_loans30 - Total amount of loans taken by user in last 30 days
  + maxamnt\_loans30 - maximum amount of loan taken by the user in last 30 days
  + medianamnt\_loans30 - Median of amounts of loan taken by the user in last 30 days
  + cnt\_loans90 - Number of loans taken by user in last 90 days
  + amnt\_loans90 - Total amount of loans taken by user in last 90 days
  + maxamnt\_loans90 - maximum amount of loan taken by the user in last 90 days
  + medianamnt\_loans90 - Median of amounts of loan taken by the user in last 90 days
  + payback30 - Average payback time in days over last 30 days
  + payback90 - Average payback time in days over last 90 days
  + pcircle - telecom circle
  + pdate - date
* Motivation for the Problem Undertaken

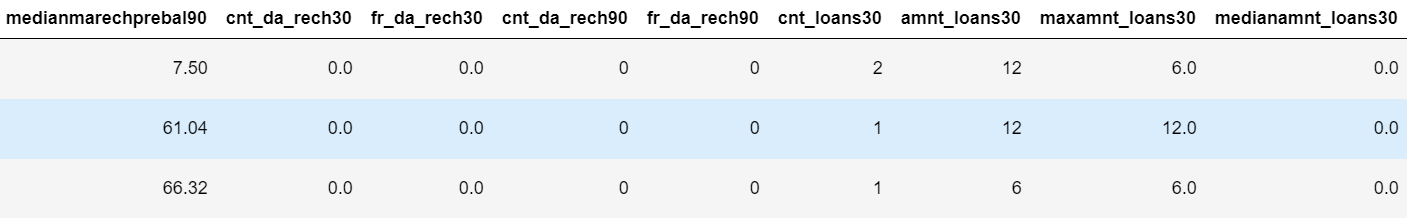
This project helps us to understand the nature of consumers while taking a loan. The amount is small and should be easy to clear unlike a credit card or a home loan. This project gives lot of insights on pre-paid users and the frequency of returning the loan.

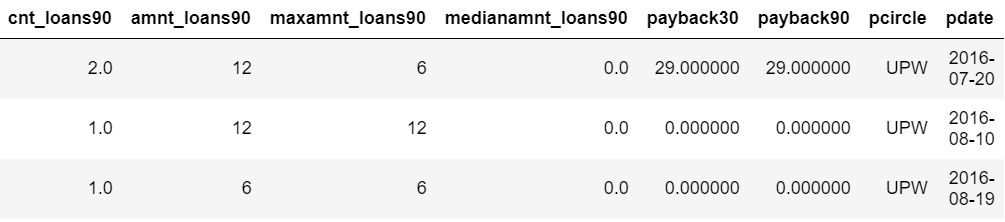
**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem
* Understanding outliers
* Valid data range of each feature
* Standard deviation for each feature
* 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
* Snapshot of the data:









* Data Pre-processing Done
* 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
* Reviewed the mean and standard deviation of the data
* Checked the count and used SMOTE to over-sample the data
* Data Inputs- Logic- Output Relationships

The data has been scaled using Standard Scaler before training.

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

No pre-assumptions

* 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
* matplotlib – visualization of dates
* datetime - handle date
* StandardScaler – scale data
* train\_test\_split – split train and test from pre-processed data
* GridSearchCV – hyperparameter tuning
* cross\_val\_score – cross validation
* accuracy\_score – evaluation
* roc\_auc\_score – evaluation
* confusion\_matrix – evaluation
* roc\_curve – evaluation
* classification\_report – evaluation
* plot\_confusion\_matrix – visualization
* Joblib – save pkl file

**Machine Learning Algorithms**

* Logistic Regression
* Gaussian Naive Bayes
* Support Vector Classifier
* K-Neighbors Classifier
* Decision Tree Classifier
* XGB Classifier
* Random Forest Classifier
* Bagging Classifier
* Extra Trees Classifier
* Gradient Boosting Classifier

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Describe the approaches you followed, both statistical and analytical, for solving of this problem.

* Testing of Identified Approaches (Algorithms)

Listing down all the algorithms used for the training and testing.

* Run and Evaluate selected models

Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

* Key Metrics for success in solving problem under consideration

What were the key metrics used along with justification for using it? You may also include statistical metrics used if any.

* Visualizations

Mention all the plots made along with their pictures and what were the inferences and observations obtained from those. Describe them in detail.

If different platforms were used, mention that as well.

* Interpretation of the Results

Give a summary of what results were interpreted from the visualizations, preprocessing and modelling.

**CONCLUSION**

* Key Findings and Conclusions of the Study

Describe the key findings, inferences, observations from the whole problem.

* Learning Outcomes of the Study in respect of Data Science

List down your learnings obtained about the power of visualization, data cleaning and various algorithms used. You can describe which algorithm works best in which situation and what challenges you faced while working on this project and how did you overcome that.

* Limitations of this work and Scope for Future Work

What are the limitations of this solution provided, the future scope? What all steps/techniques can be followed to further extend this study and improve the results.