

Micro Credit Defaulter Project

Submitted by:

Ankita Srivastava

**ACKNOWLEDGMENT**

I found all references from google where some author Explain this data structure.my data provided by Flip Robo technologies PVT LTD.

My SME Miss. Gulshana Mam helped me to make this project. I was not Use any other resources.

**INTRODUCTION**

* Business Problem Framing

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.

.

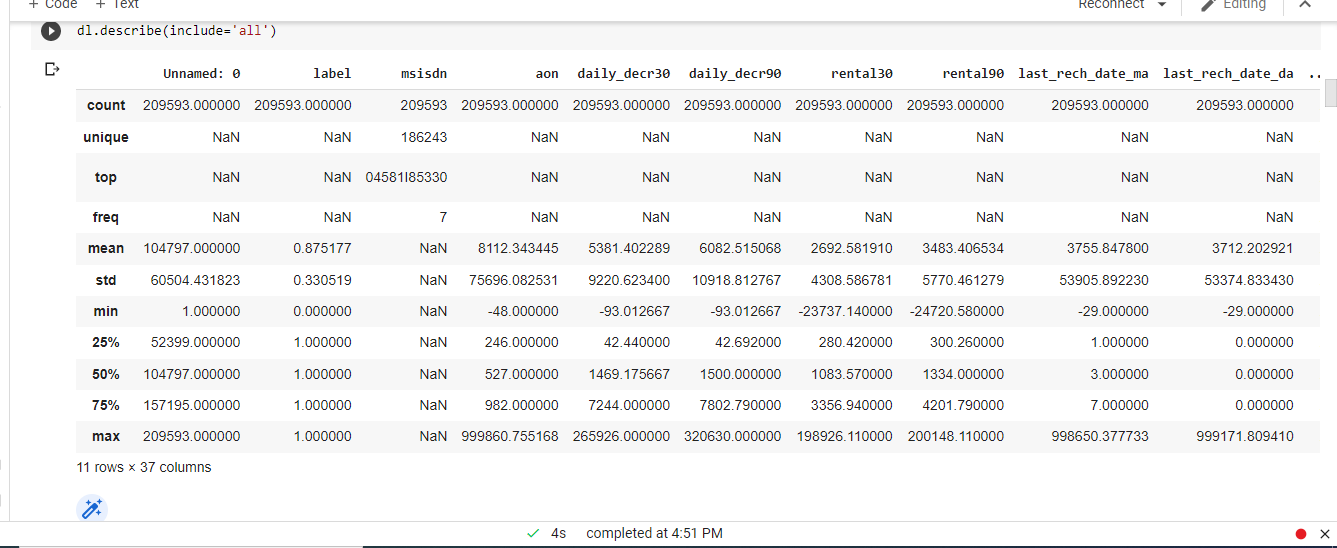
* **Conceptual Background of the Domain Problem**

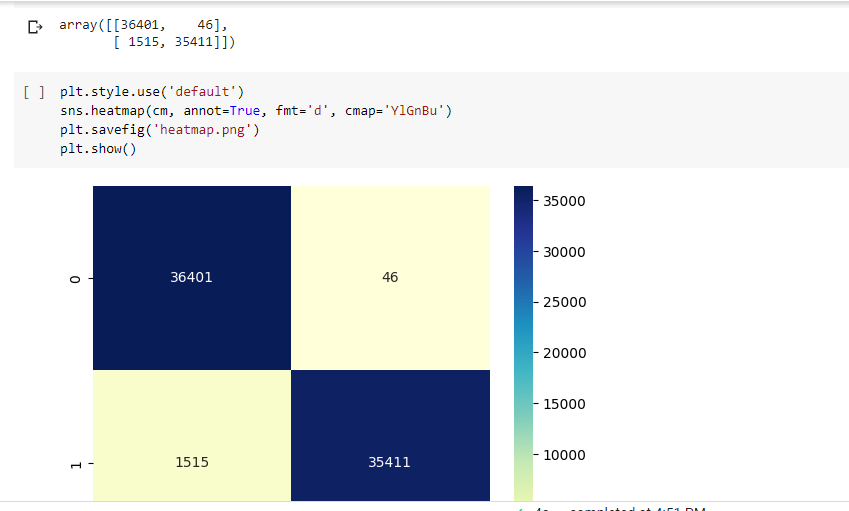
We are working with one such client that is in Telecom Industry. They are a fixed wireless telecommunications network provider.

They are 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).

* Review of Literature

In this research I found that some persons take the loan but they are not pay in the given time, so I treat them as a defaulter and some person did not recharge his account from 30 dayes.





* Motivation for the Problem Undertaken

I want to find which type of person defaulter, who is not paying loan at time or who is some time not paying. That is our objectives.

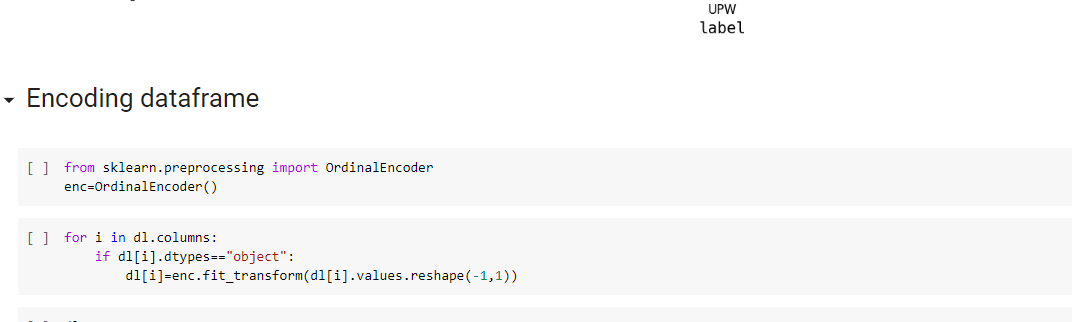
There is some industry who is understand the importance of communication and how it affects a person’s life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour. That can motivate to anyone.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

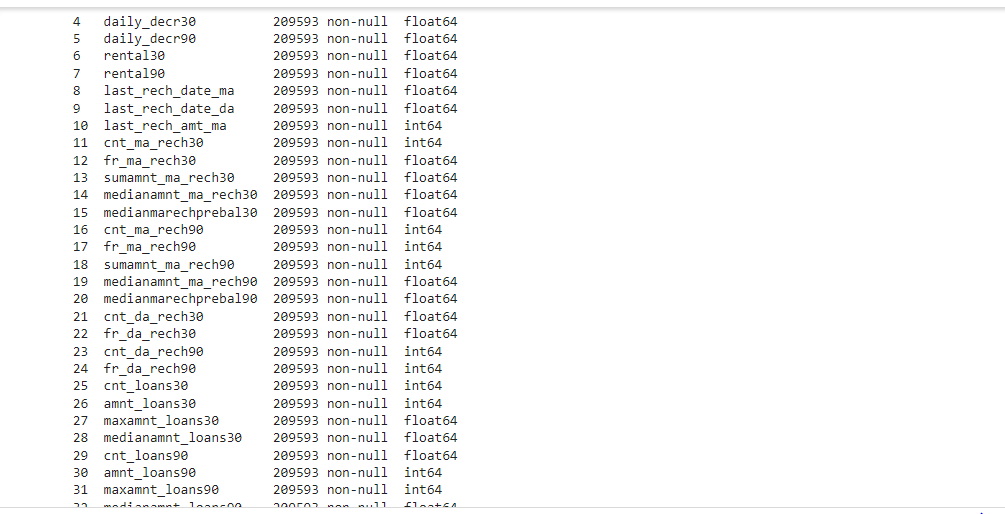
I use some Encoding Technique to spread all data in same formate numerical form.Here some features have skewness when I remove it I calculate how much I lost the data by calculation

[ (past data **-**present data)/past data\*100]



* Data Sources and their formats

The sample data is provided to us from our client database. It is hereby given to you for this exercise. In order to improve the selection of customers for the credit, the client wants some predictions that could help them in further investment and improvement in selection of customers.



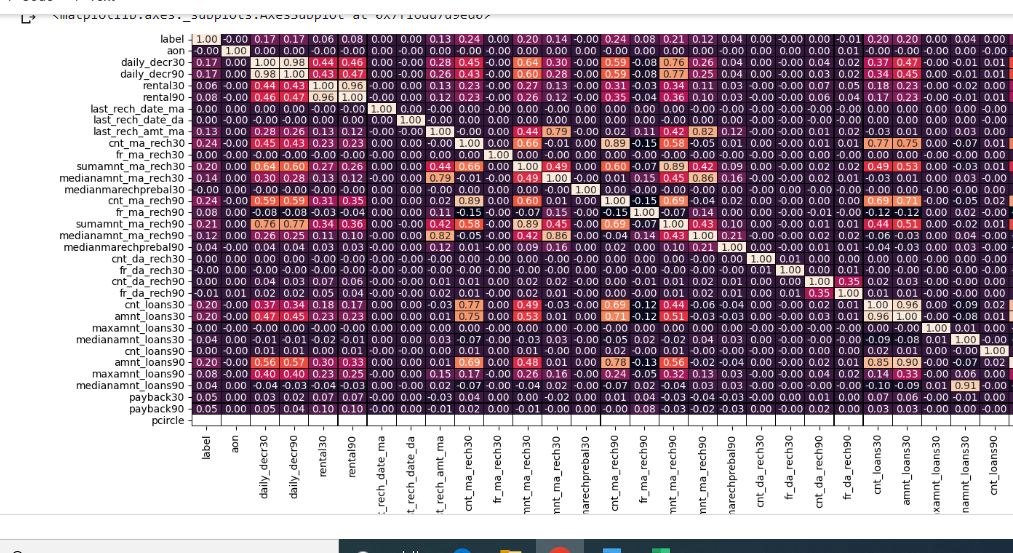
* Data Preprocessing Done

Firstly I import all libraries which is necessary for our data.i find there is no null values presenta.after that I check the target type which shows categorical features and it is not proper balanced.then I find the unique value, by which I can clear to me which column I want to drop.

I Assume that there is some NAN vlues in last rowes but I was wrong.



* Data Inputs- Logic- Output Relationships

****

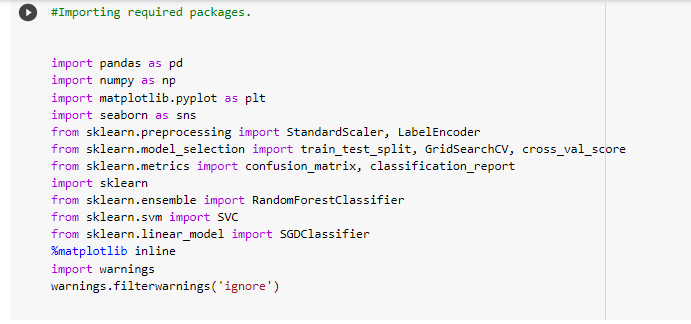
* State the set of assumptions (if any) related to the problem under consideration

No any pre assumtion by me.

* Hardware and Software Requirements and Tools Used

Hardware-this is required 8 GB Ram.

Software-I use google colab to execute this programme.



**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

**I** use Standard scaler for scaling the data after removing outliers.

Random over sampler use for overcome the oversampling.

* Testing of Identified Approaches (Algorithms)

I use Logistic regression, KNN, Random Forest classifier to build the model.

* Run and Evaluate selected models



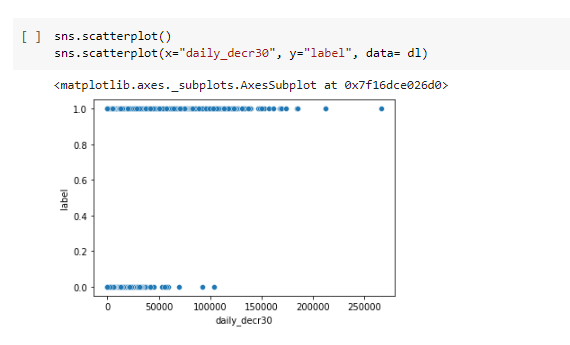
we can see that all model are well trained..but Random Forest Classifier gives higest Accuracy= 97.83 and F1 score is also good=97.8

* Visualizations

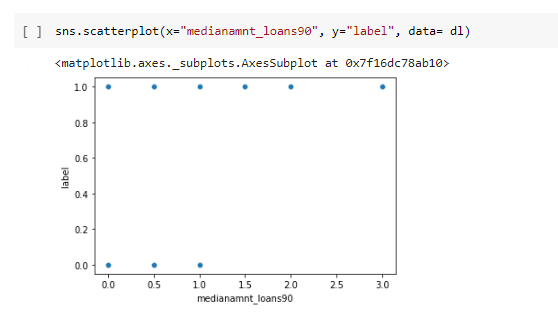
Here I mention Scatter plot for showing dependency of target and features.

Line plot show the correlation b/w Target and features.

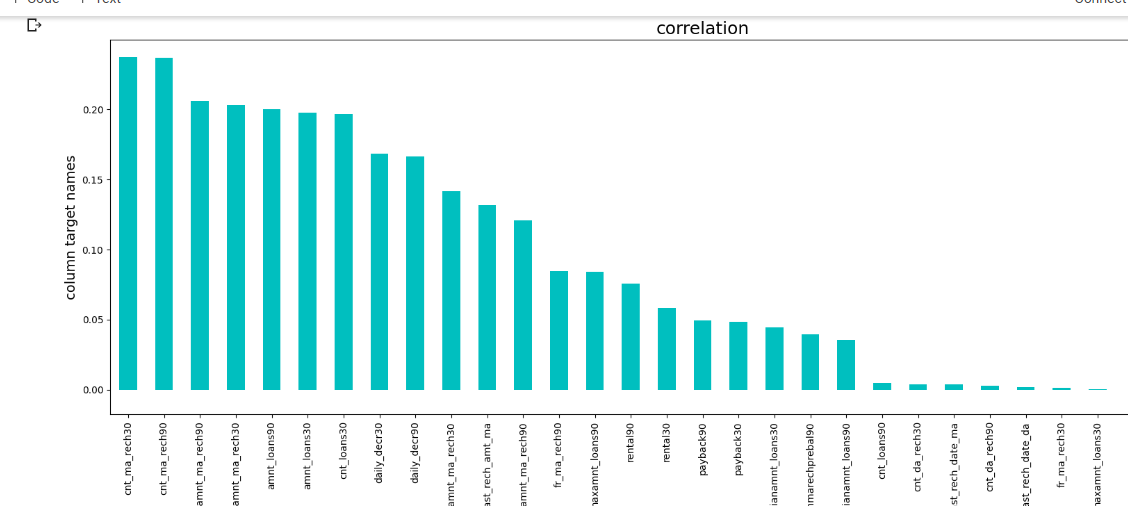
Count plot show the object type data values.



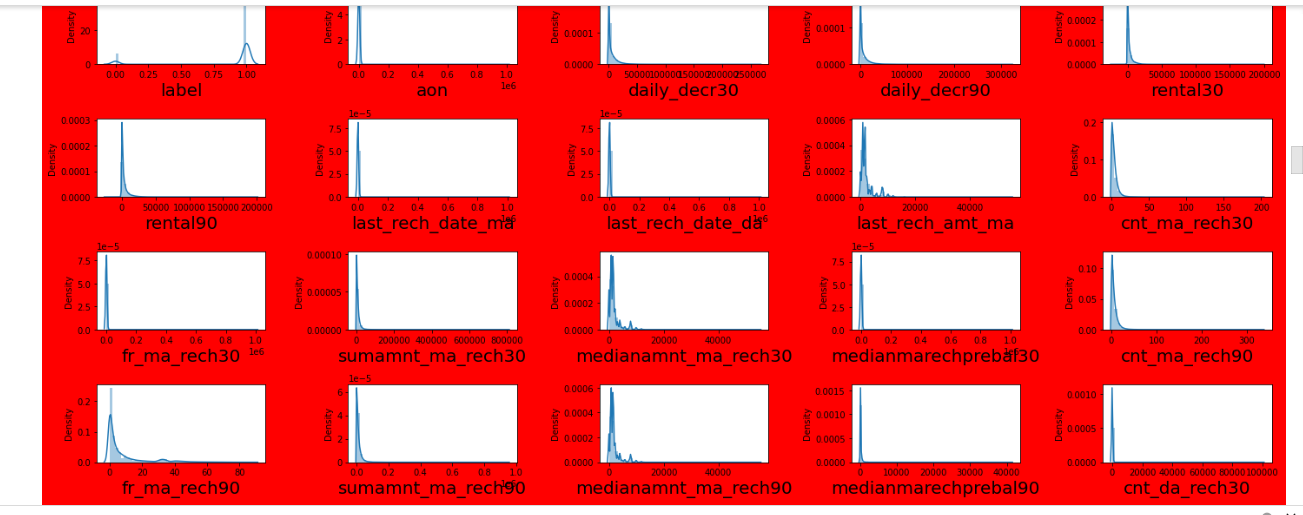
This plot I can see that failure chances is minimum to daily amount spent in main account.



Median of amounts of loan taken by the user in last 90 days is minimum failure chances.



Here I can see that how Target is correlated to features to + and – formate.



Dist plot show the skewness of data. Maximum features are skeweed in right side.

.

* Interpretation of the Results

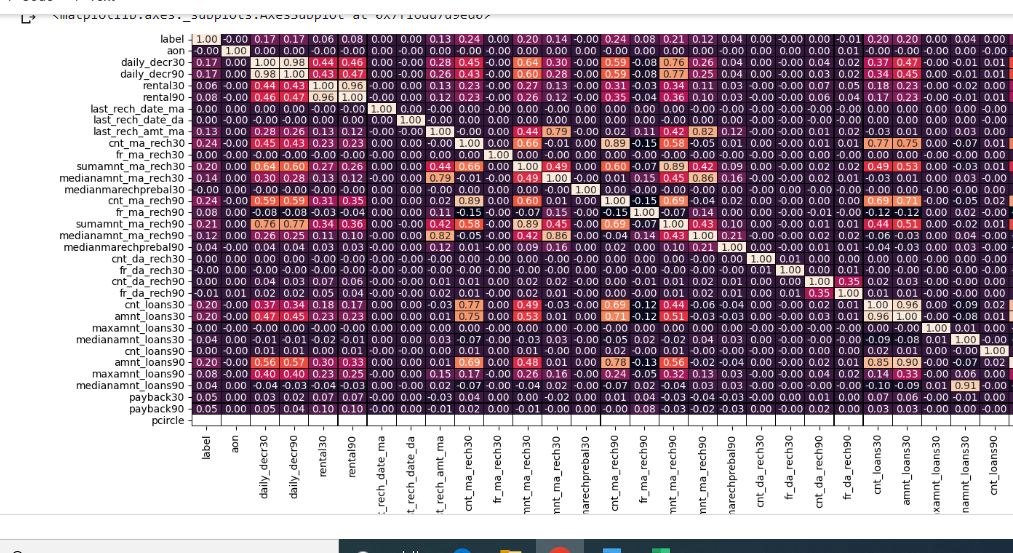
I observe that there is maximum outliers are present. When it was remove from log transformation method , I was loss maximum data. So I use Z transform method to remove outliers.here some fetures are not related to target.

Let’s talk about model. Here I use 3 type model Logistic regression ,KNN, random forest classifier.i found Random forest gives best result with accuracy 100%.

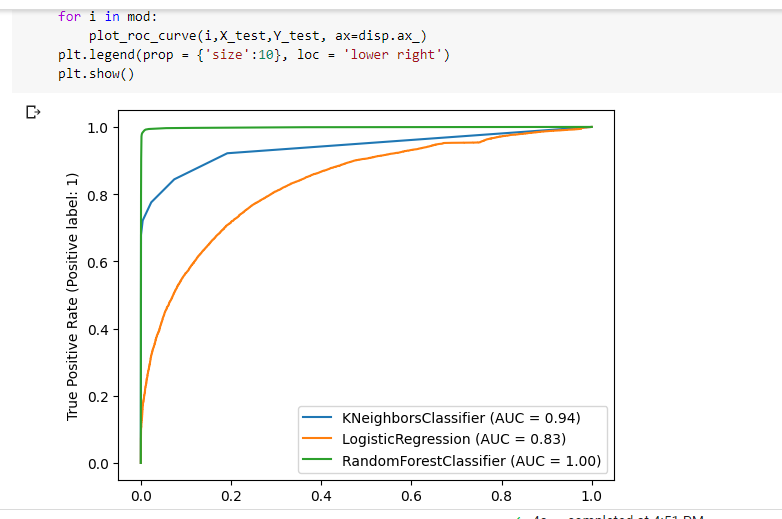
**CONCLUSION**

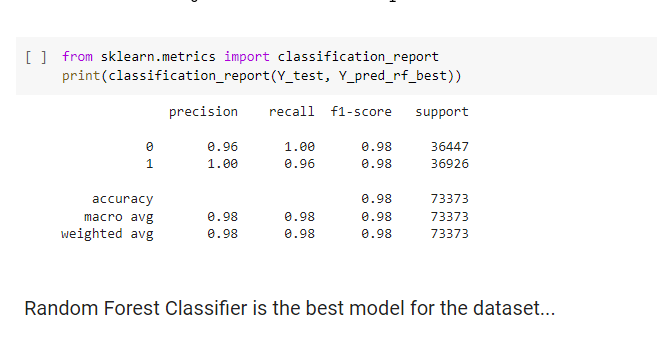
* Key Findings and Conclusions of the Study

I find the some features not affect the defaulter data.



* Learning Outcomes of the Study in respect of Data Science

Some time model exacutaion takes a lot of time, I was it to run in many times but all times hyperparameter Tuning takes a lot of time.When I use google colab it gives a final result.



* Limitations of this work and Scope for Future Work

There is Maximum scope to find the defaulter by this Random Forest Classifier model.Firstly I have to find when was last time for loan submition and who is not pay this money.

Perhaps I build different type many model to predict data I can find Maximum Accuracy.