FLIP ROBO

Micro-Credit Defaulter

By

NAGAPPAN KANNAPPAN

ACKNOWLEDGMENT

I would like to express my special thanks to Datatarained & FlipRobo who gave me this golden opportunity for this internship on the topic of Micro-Credit Defaulter.
The sample data is provided to us from FlipRobo's client database. Kaggle & Github are the websites which helped me in completing the project.

Business Problem:

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. Today, microfinance is widely accepted as a poverty-reduction tool, representing $70 billion in outstanding loans and a global outreach of 200 million clients.

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. Here we need to build model which can be used to predict in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount within 5 days of insurance of loan.

Conceptual Background of the Domain Problem:

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. Here the client that is in Telecom Industry is a fixed wireless telecommunications network provider. They have launched various products and have developed its business and organization based on the budget operator model, offering better products at Lower Prices to all value conscious customers through a strategy of disruptive innovation that focuses on the subscriber. 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.

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:

Microfinance is a banking service provided to unemployed or low-income individuals or groups who otherwise would have no other access to financial services.

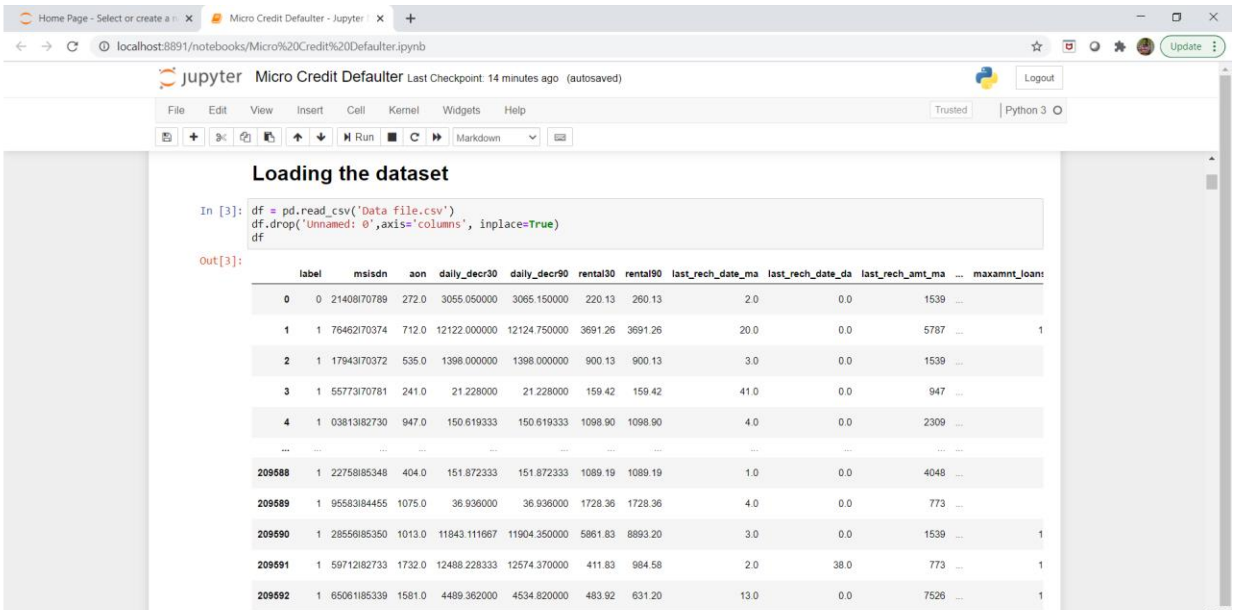
Microfinance allows people to take on reasonable small business loans safely, and in a manner that is consistent with ethical lending practices.

Motivation for the Problem Undertaken:

With the help of this project deserved people will get loan more easily & quickly. Being a part of this project and reducing poverty is a proud feeling & motivation.

Data Sources and their formats:

The sample data is provided to us from FlipRobo's client database.



Mathematical/ Analytical Modelling of the Problem:

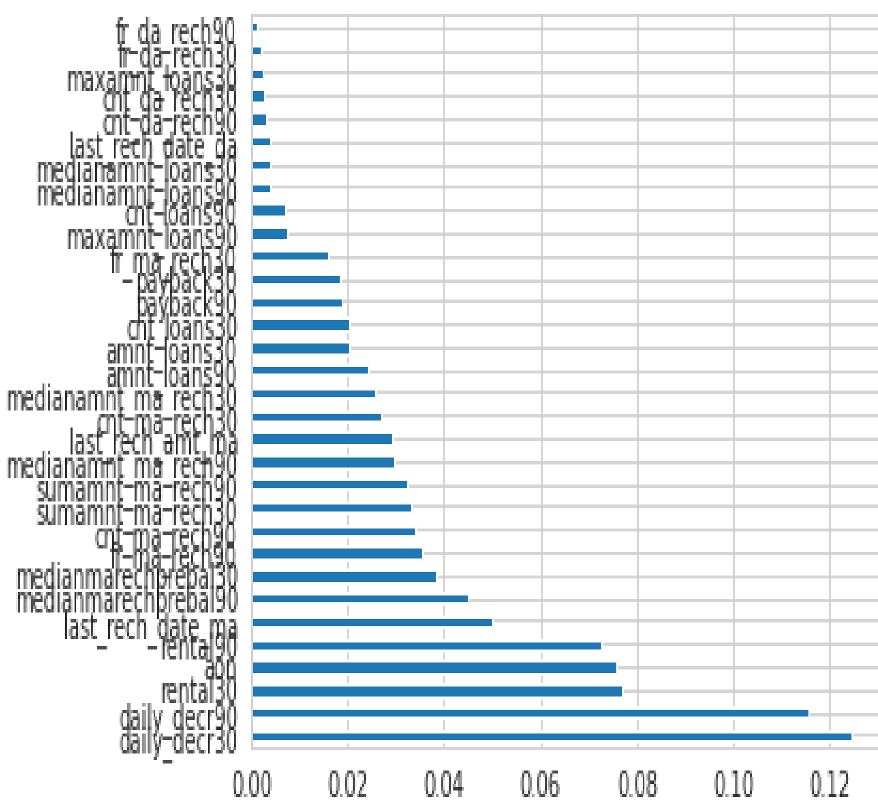
In this case, Label '1' indicates that the loan has been paid i.e. Non- defaulter, while, Label '0' indicates that the loan has not been paid i.e. defaulter. In the provided *dataset*, our target variable "label" is a *categorical* with two categories: " defaulter " and " Non- defaulter ". Therefore we will be handling this modelling problem as classification.

Data Pre-processing:

Dropping unimportant columns.

```
In [ ]:
#we can drop some features for further processing
df.drop(['pdate','pcircle','msisdn'],axis='columns', inplace=True)
df
```

```
In [ ]:
#we can drop less important featues for further process
df.drop(['last_rech_date_da','cnt_da_rech30','fr_da_rech30','cnt_da_rech90','fr_da_rech90','maxamnt_loans30' ],axis='
df
```



Imbalance dataset is normalized for final modeling.

After splitting the data for input and output standard scalar is applied to standardize the input data.

In [ ]:
```
#split train and test dataset
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 0)
```

In [ ]:
```
#check shape of train dataset
X_train.shape
```

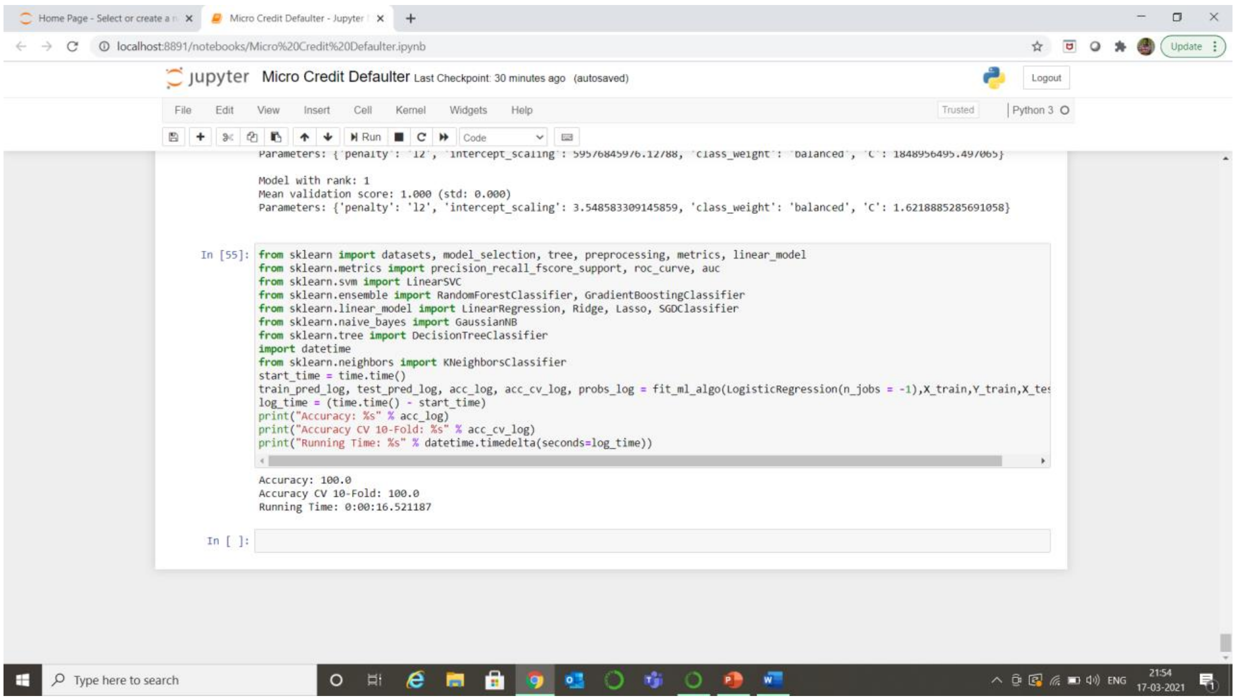
Out[11]: (146715, 26)

After train test split we will apply all the classification algorithms with hyper tuning to find the best scoring one.

Model
Random Forest
Gradient Boosting Trees
Logistic Regression
Linear SVC
Stochastic Gradient Decent
KNN
Decision Tree
Naive Bayes

After applying all the above classification algorithms on the dataset we see that Gradient Boosting trees & Random Forest both fits the best for our objective

We will use Logistic Regression as our final model.



The final model is saved.

Conclusion:

• Conclusions of the Study

The last four days I spend quite a lot of my free time on a current data-science project. A Micro-Credit Defaulter prediction problem at FlipRobo. And yes, it was less sleep than usual but the learning’s were worth it.

- Learning Outcomes of the Study in respect of Data Science

Here I learned about the micro credit industry, visualization, data cleaning, handling outliers and using various algorithms on huge dataset. This was the first time I worked on such huge dataset. It took a lot of time to hyper tune all the algorithms to find out the best one to work with. Working with such huge dataset that took a lot of time to train the algorithms and tuning it for the best prams was worth knowing in this project.

- Limitations of this work and Scope for Future Work

Training the huge dataset was a challenge for me. Balancing the imbalance dataset. Overcoming the outliers. Hyper tuning the algorithms can bring out more satisfactory result