MICRO CREDIT PROJECT

SUBMITTED BY
Akshatha Aravind

Acknowledgement

I go through different articles, references and technical sites for the valuable information regarding the project.I would like to mention some here,

Articles:

1)"A machine learning approach to micro credit scoring" by Titus Nyrako, Paresh Date and Corina Constantinesou.

2)"Rural Micro Credit Assessment Using Machine Learning" by Henry Ivan Condori, Guina Sotonayos, Alzamwora.

3)"How do machine learning and non-traditional data affect credit scoring?New evidence from a Chinese finch firm" by Leonardo Gambacorta ,Yiping Huang, Han Qiu, and Wang.

4)"Machine Learning Technologies for Digital Credit Scoring in Rural Finance" by Anil Kumar, Sunil Sharma, Mehregan Mahdavi.

Site References:

1)Science Direct, Dataiku, ADB(Asian Development Bank),Kaggle, Reaearchgate, Risk Journals,IJSR(International journal of Science and Research),Journal of finance management.

INTRODUCTION

Financial services are very common in the society and micro-lending markets and Micro finance is a category of financial services targeting individual and small businesses who lack access to conventional banking and related services, here credit history is a significant impediment to assessing individual borrowers' creditworthiness and therefore deciding fair interest rates. This research compares various machine learning algorithms on real micro-lending data to test their efficacy at classifying borrowers into various credit categories.MFI(Micro Financial Institutions) are usually offer Micro Credit Schemes(MCS) to Micro enterprise activities.The uneducated low income level people are not care about finances, because

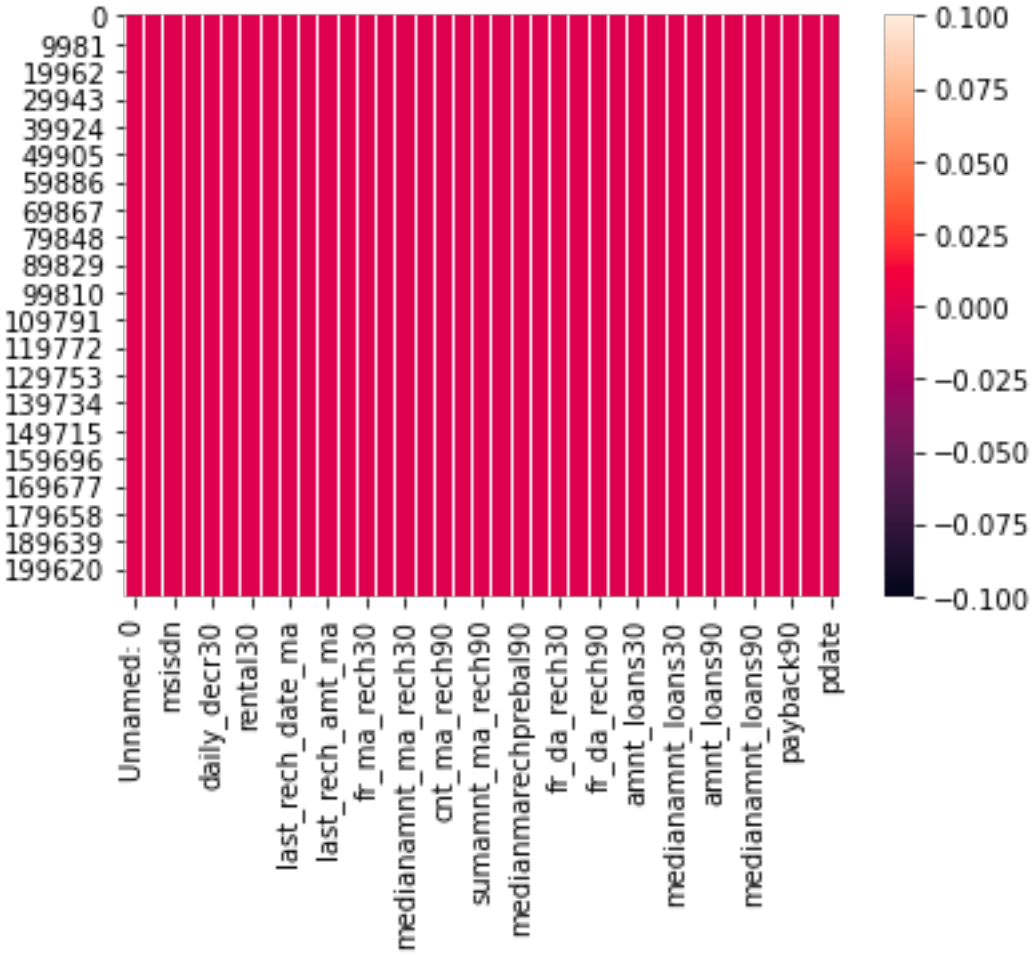
a)Don't have an intention to save money
b)Education level is too low
c)lack of creditworthiness
d)Not interested to repay loans

Thus they are usually not eligible for regular financial services(banking),here MCS are introduce to offer affordable finance assistance without involvement of regular banking system.

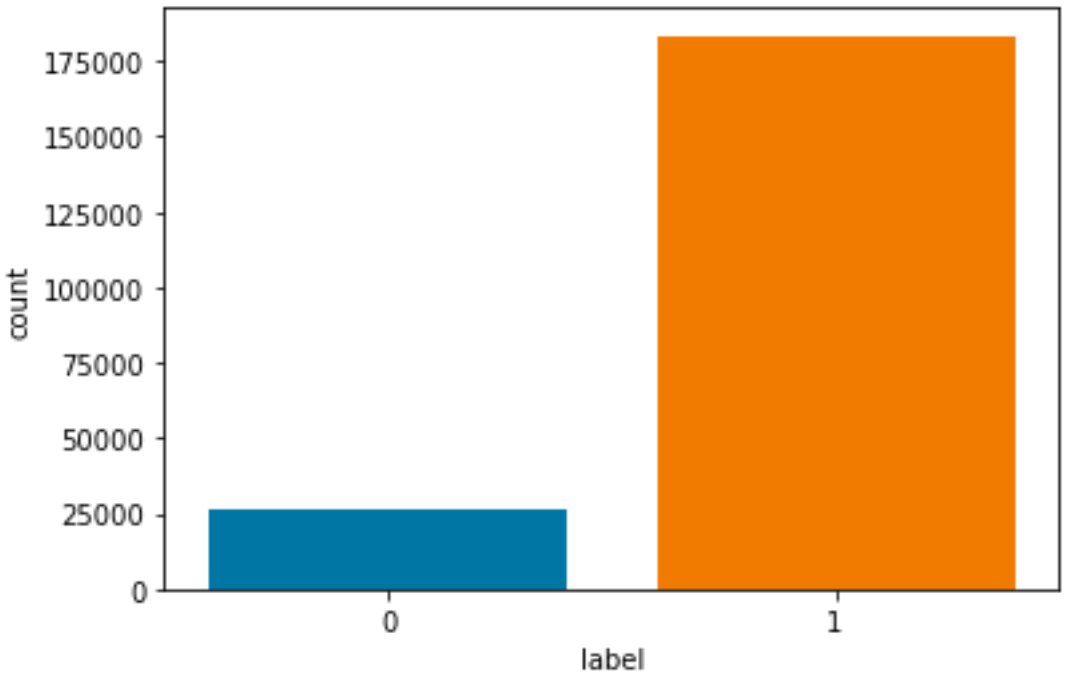
Here this scheme is collaborating with telecom industry, the client is a fixed wireless telecommunication service provider, providing various budget model services in low income level and providing micro credit mobile balance by the help of MFI which have to be paid back in 5 days.Data base consisting various datas relating the balance ,recharge method, loan handling, account maintenance etc.By analysing and using datas we have to credit the chance of pay back in 5 days.

Dataset:Analysis

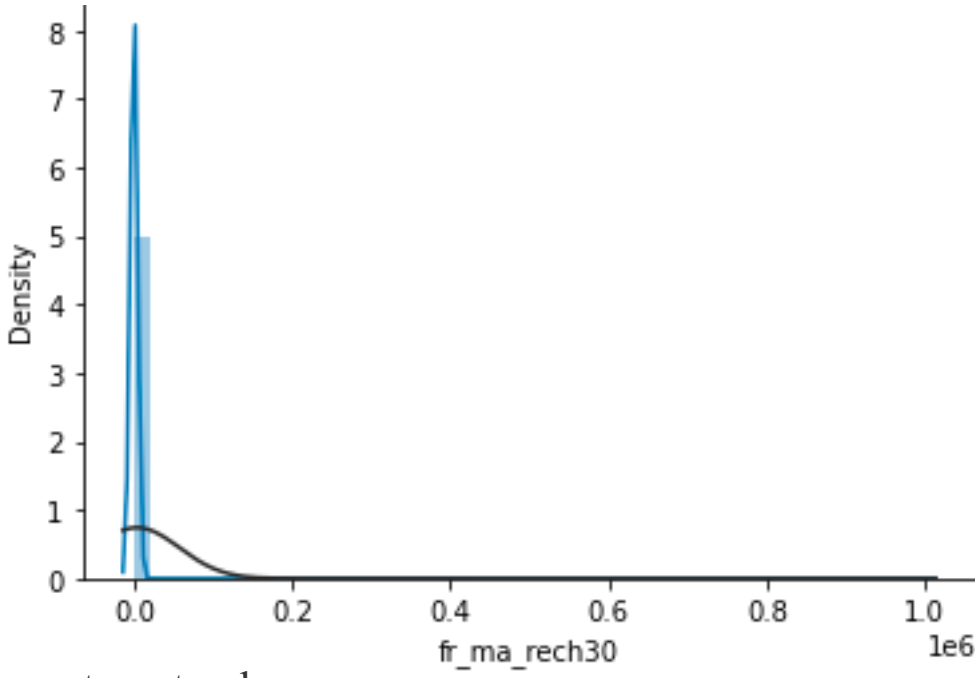
Dataset consist of 209593 rows and 37 different columns, here Unnamed: 0', 'label', 'msisdn', 'aon', 'daily_decr30', 'daily_decr90' 'rental30', 'rental90', 'last_rech_date_ma', 'last_rech_date_da', 'last_rech_amt_ma',cnt_ma_rech30', 'fr_ma_rech30','sumamnt_ma_rech30', 'medianamnt_ma_rech30', 'medianmarechprebal30', 'cnt_ma_rech90','fr_ma_rech90', 'sumamnt_ma_rech90','medianamnt_ma_rech90', 'medianmarechprebal90', ‘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', ‘pdate' are the different columns.Most of the columns are in int ,float datatypes except tree object data type columns, moreover there is no null values in the data set



Our target column name is ‘label’ it is a double valued column with values 0 and 1.



From the figure it is clear that the target is not balanced so we have to balance it by sampling process.



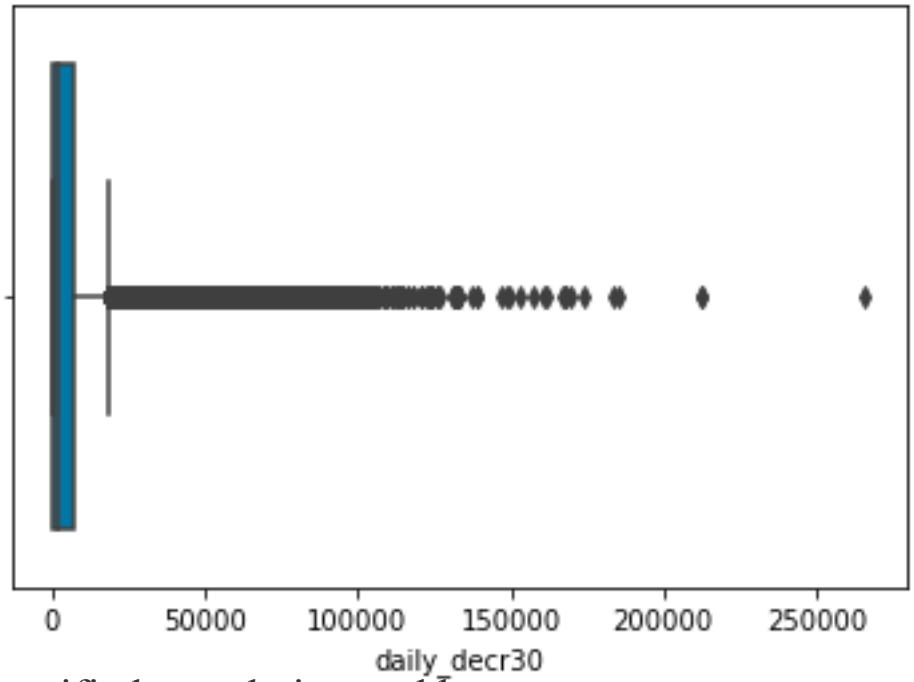
Analysis of feature columns gives more about the datasets and we can also find the value counts and data distributions of different columns and also we can find the dependency to the target column.

Preprocessing steps

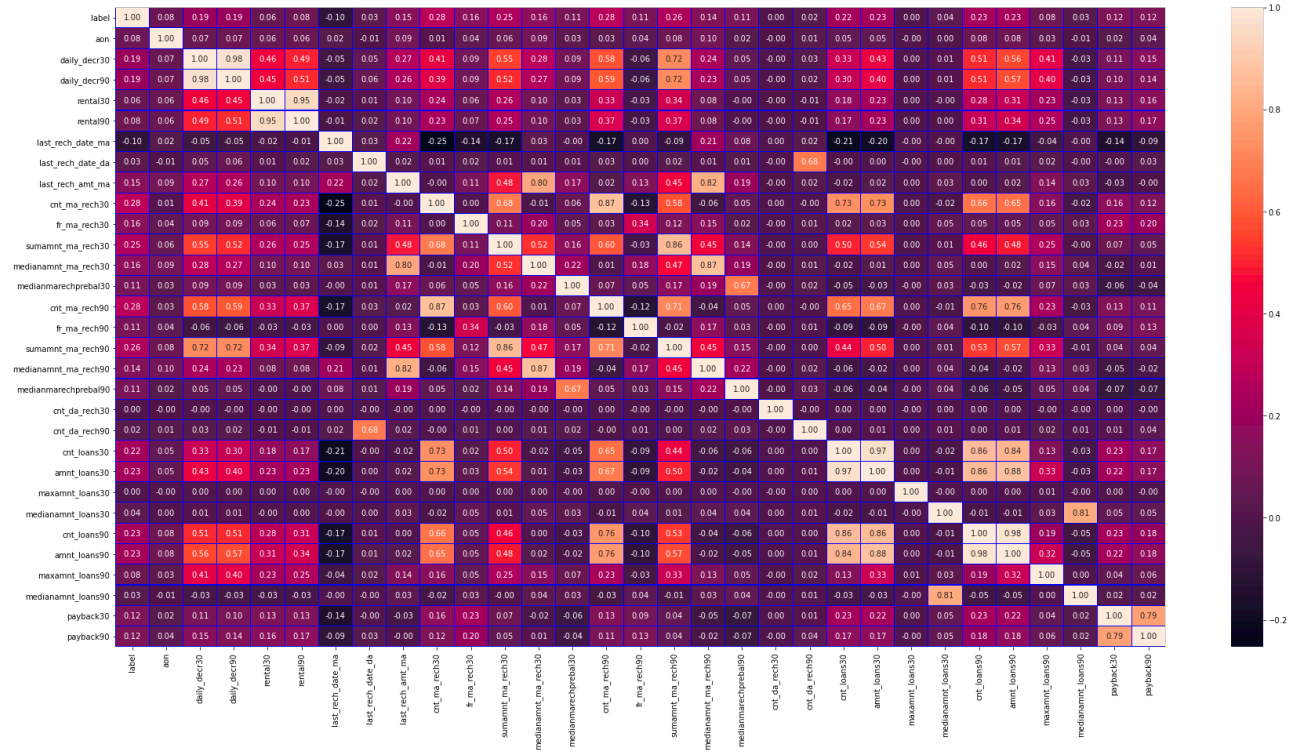
After analysing the feature columns we undergone different preprocessing steps

a)Dropped some unwanted columns: some univariate columns and no value columns are there

b)Using box plot checked the outliers:box plot conveys the presence of outlier



c)verified correlation and heat map



d)removed outliers and skewness

aon	0.307788
daily_decr30	-0.574458
daily_decr90	-0.537153
rental30	0.278805
rental90	0.279948
last_rech_amt_ma	-0.127823
cnt_ma_rech30	-0.016019
fr_ma_rech30	0.132234
sumamnt_ma_rech30	-0.409089
medianamnt_ma_rech30	-0.252729
cnt_ma_rech90	-0.019191
fr_ma_rech90	0.140178
sumamnt_ma_rech90	-0.312352
medianamnt_ma_rech90	-0.113306
cnt_da_rech90	0.000000
cnt_loans30	0.035974
amnt_loans30	0.009452
medianamnt_loans30	0.000000
cnt_loans90	0.087113
amnt_loans90	-0.001444
maxamnt_loans90	0.424586
medianamnt_loans90	0.000000
payback30	0.292928
payback90	0.200700

e)Checked multicollinearity with VIF: Using VIF value checked the multicollinearity and it is made in range.

f)Sampled the Target column:Using SMOTE Balanced the target

```
0     153614
1     153614
```

g)Scaled the feature columns:Using standard scaler scaled the features.

Model development

NAME OF ALGORITHM	ACCURACY SCORE	CROSS VALIDATION MEAN	DIFFERENCE
LOGISTIC REGRESSION	79	78	1
DECISION TREE classifier	89	89	0
KNN classifier	85	86	-1
Random forest classifier	93	94	-1

Thus here preferred DTC for further processing.Select best parameter from hyper parameter tuning and passed the best parameter values to the model.Finally saved the model in jib lib in the name” Best_micro_credit_model.pk1”.

CONCLUSION

- By analysing target variable we come to conclusion that ,this is a classification type model.
- By analysing features we dropped some unwanted feature columns.
- Verified the correlation.
- To check multicollinearity applied VIF.
- To balancing target applied SMOTE.
- Splitted x and y and applied Algorithms
- Preferred DTC for Hyper parameter tuning.
- Got 81.4% Accuracy and saved the model