

Micro Credit Defaulter

Submitted by:

Akata Yadav

**ACKNOWLEDGMENT**

I would like to acknowledge Mr. Mohd Kashif (FlipRobo) for his vital cooperation and help in ensuring the successful completion of my assignment. He deserves the utmost credit for the assignment’s outcome.

Finally, I would want to convey my sincere thanks Datatrained Academy and their guidance without them, the task would not have been accomplished.

The website that I referred are:

https://learning.datatrained.com

<https://www.w3schools.com>

<https://www.freecodecamp.org>

<https://github.com>

<https://www.geeksforgeeks.org>

**INTRODUCTION**

* Business Problem Framing

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.

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.

We are working with one such client that is in Telecom Industry. They are 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).

* Conceptual Background of the Domain Problem

This project is all about predicting the Micro Credit Defaulter.

* Review of Literature

## **Features**

* 1. msisdn: mobile number of user
  2. aon: age on cellular network in days
  3. daily\_decr30: Daily amount spent from main account, averaged over last 30 days (in Indonesian Rupiah)
  4. daily\_decr90: Daily amount spent from main account, averaged over last 90 days (in Indonesian Rupiah)
  5. rental30: Average main account balance over last 30 days
  6. rental90: Average main account balance over last 90 days
  7. last\_rech\_date\_ma: Number of days till last recharge of main account
  8. last\_rech\_date\_da: Number of days till last recharge of data account
  9. last\_rech\_amt\_ma: Amount of last recharge of main account (in Indonesian Rupiah)
  10. cnt\_ma\_rech30: Number of times main account got recharged in last 30 days
  11. fr\_ma\_rech30: Frequency of main account recharged in last 30 days
  12. sumamnt\_ma\_rech30: Total amount of recharge in main account over last 30 days (in Indonesian Rupiah)
  13. medianamnt\_ma\_rech30: Median of amount of recharges done in main account over last 30 days at user level (in Indonesian Rupiah)
  14. medianmarechprebal30: Median of main account balance just before recharge in last 30 days at user level (in Indonesian Rupiah)
  15. cnt\_ma\_rech90: Number of times main account got recharged in last 90 days
  16. fr\_ma\_rech90: Frequency of main account recharged in last 90 days
  17. sumamnt\_ma\_rech90: Total amount of recharge in main account over last 90 days (in Indonasian Rupiah)
  18. medianamnt\_ma\_rech90: Median of amount of recharges done in main account over last 90 days at user level (in Indonasian Rupiah)
  19. medianmarechprebal90: Median of main account balance just before recharge in last 90 days at user level (in Indonasian Rupiah)
  20. cnt\_da\_rech30: Number of times data account got recharged in last 30 days
  21. fr\_da\_rech30: Frequency of data account recharged in last 30 days
  22. cnt\_da\_rech90: Number of times data account got recharged in last 90 days
  23. fr\_da\_rech90: Frequency of data account recharged in last 90 days
  24. cnt\_loans30: Number of loans taken by user in last 30 days
  25. amnt\_loans30: Total amount of loans taken by user in last 30 days
  26. maxamnt\_loans30: maximum amount of loan taken by the user in last 30 days
  27. medianamnt\_loans30: Median of amounts of loan taken by the user in last 30 days
  28. cnt\_loans90: Number of loans taken by user in last 90 days
  29. amnt\_loans90: Total amount of loans taken by user in last 90 days
  30. maxamnt\_loans90: maximum amount of loan taken by the user in last 90 days
  31. medianamnt\_loans90: Median of amounts of loan taken by the user in last 90 days
  32. payback30: Average payback time in days over last 30 days
  33. payback90: Average payback time in days over last 90 days
  34. pcircle: telecom circle
  35. pdate: date

## **Target**

label: Flag indicating whether the user paid back the credit amount within 5 days of issuing the loan{1:success, 0:failure}

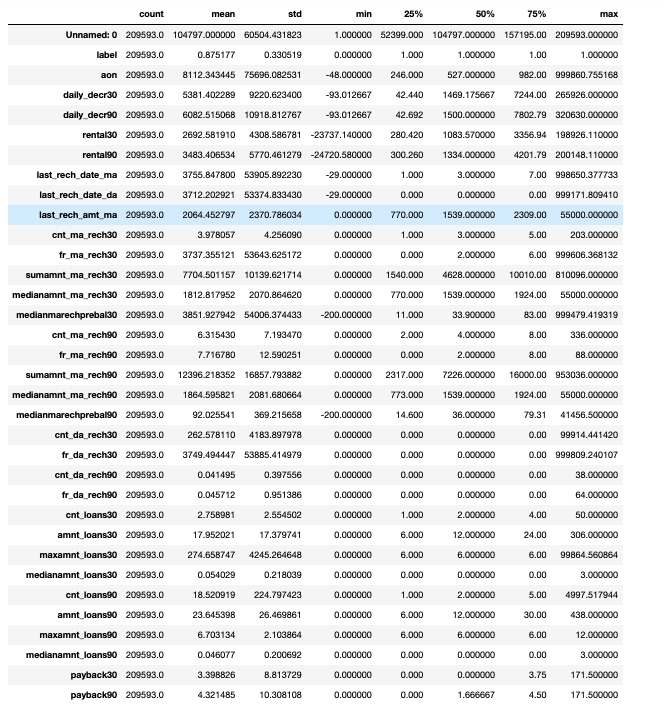
**Information of the Dataset.**

* + RangeIndex: 0 to 209592
  + Data columns: 37
  + dtypes: float64(21), int64(13), object(3)
* Motivation for the Problem Undertaken

This project is on the data science and machine learning model, build the model to predict the Micro Credit Defaulter based on some features.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

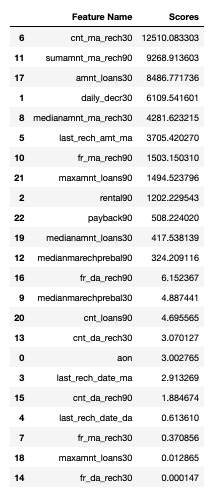


**Description of the dataset.**

* Data Sources and their formats
  1. Information of the dataset.
  2. Description of the dataset.
  3. Visualization.
  4. Correlation present in the dataset.
  5. Skewness is present.
  6. Outliers are present in the dataset.
  7. Target column is Imbalanced.
* Data Preprocessing Done
  1. Shape of our dataset is 209593, 37
  2. Dropped unwanted columns.
  3. Dropped correlated columns.
  4. Dropped columns whose score with the target column is less than 1.
  5. Apply Transformation - PowerTransformer (method = “yoe-Johnson”)
  6. Apply Oversampling on target column.
* Data Inputs- Logic- Output Relationships
  1. Correlation:

|  |  |  |
| --- | --- | --- |
| Column 1 | Correlation | Column 2 |
| daily\_decr90 | 98% | daily\_decr30 |
| rental30 | 96% | rental90 |
| cnt\_ma\_rech90 | 89% | cnt\_ma\_rech30 |
| sumamnt\_ma\_rech30 | 89% | sumamnt\_ma\_rech90 |
| medianamnt\_ma\_rech90 | 86% | medianamnt\_ma\_rech30 |
| cnt\_loans30 | 96% | amnt\_loans30 |
| amnt\_loans90 | 90% | amnt\_loans30 |
| medianamnt\_loans90 | 91% | medianamnt\_loans30 |
| payback30 | 83% | payback90 |

2. Feature Selection



Selected all the Features with score more than 1.

3. Applied PowerTransformer to remove skewness.

4. Used Oversampling method to balance the target column.

* Hardware and Software Requirements and Tools Used

Anaconda-navigator

jupyter notebook

matplotlib-inline==0.1.6

numpy==1.23.2

packaging==21.3

pickleshare==0.7.5

platformdirs==2.5.2

prompt-toolkit==3.0.30

pyparsing==3.0.9

python-dateutil==2.8.2

scikit-learn==1.1.2

scipy==1.9.0

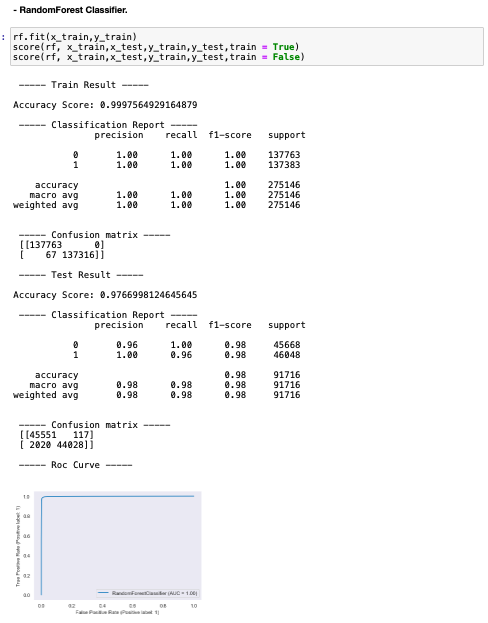
sklearn==0.05

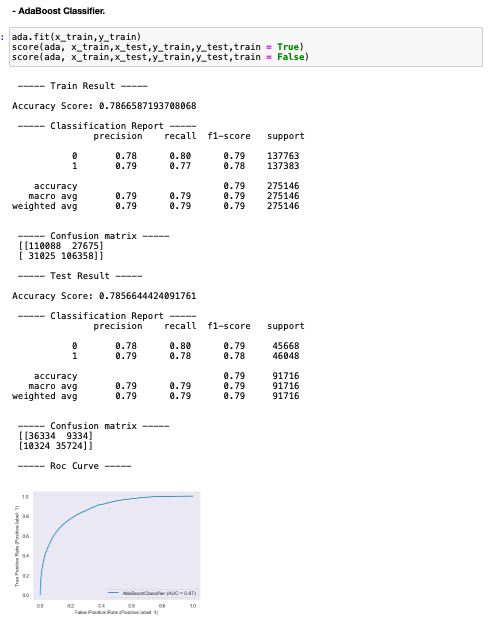
**Model/s Development and Evaluation**

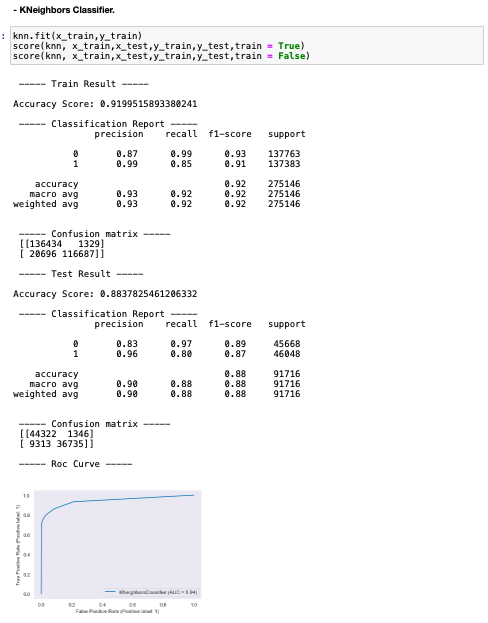
* Identification of possible problem-solving approaches (methods)
  + EDA
  + Visualization
  + Correlation
  + Features Selection
  + Normal Distribution
  + Outliers
  + Imbalanced Target Column
  + Final Dataset
  + Model Building
  + Cross-Validation
  + Saving the model.
* Testing of Identified Approaches (Algorithms)

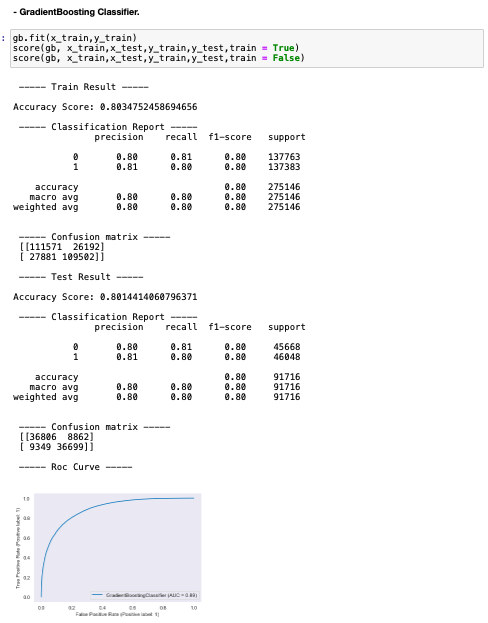
Algorithms used for the training and testing:

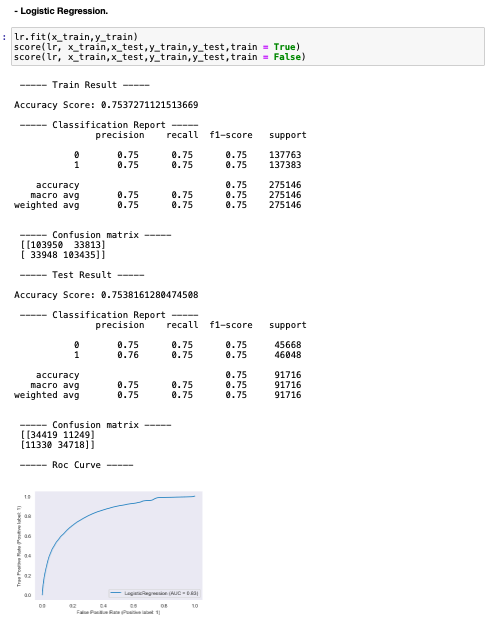
* + RandomForest Classifier.
  + AdaBoost Classifier.
  + KNeighbors Classifier.
  + GradientBoosting Classifier.
  + Logistic Regression.
* Run and Evaluate selected models











* Key Metrics for success in solving problem under consideration
  + Cross-Validation

Score is approx. 97%

* Interpretation of the Results

RandomForest Classifier, is giving the best score among all other models.

**CONCLUSION**

* Key Findings and Conclusions of the Study

Cross Validation Score and model Accuracy score is very close, so we can say that our model is working well and doesn't having the overfitting/under fitting problem present.