

MICRO-CREDIT- PROJECT

Submitted by:

POORNIMA SINGH

**ACKNOWLEDGMENT**

I would like to express my special thanks of gratitude to my teacher (Ms. Deepika) as well as our mentors from Fliprobo who gave me the golden opportunity to do this wonderful project on the topic of Micro Credit , which also helped me in doing a lot of Research and I come to know about so many new things I am really thankful to them.

Secondly I would also like to thank my parents and friends who helped me a lot in finalizing this project with in the limited time frame.

**INTRODUCTION**

* Business Problem Framing

1. A Microfinance Institution (MFI) is an organization that offers financial services to low income populations. Microfinance services (MFS) becomes very useful when targeting especially the unbanked poor families living in remote areas with not much sources of income. The MFS provided by MFI are Group Loans, Agricultural Loans, Individual Business Loans and so on.
2. Today, microfinance is widely accepted as a poverty-reduction tool, representing $70 billion in outstanding loans and a global outreach of 200 million clients.
3. We are working with one such client that is in Telecom Industry from Indonesia. 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

1. In this case, Label ‘1’ indicates that the loan has been payed i.e. Non- defaulter, while, Label ‘0’ indicates that the loan has not been payed i.e. defaulter.
2. 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.
3. 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.

* Review of Literature

1. As we know that the given dataset is imbalance dataset. According to the problem statement we have to find out fraud and non-fraud customers which has taken loan from the Telecom Industry from Indonesia as per their need.
2. 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. For this we have to analysis given data.
3. In Data Analysis process we are inspecting, cleansing, transforming and modelling data with the goal of discovering useful information, informing conclusions and supporting decision-making.

* Motivation for the Problem Undertaken

Interest for increasing knowledge motivates to do research in given dataset. Research adds to existing knowledge in a systematic way.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

1. In statistics, exploratory data analysis (EDA) is an approach to analysing data sets to summarize their main characteristics, often with visual methods.
2. We shall look at various EDA methods like- Descriptive Statistics, of the dataset we are dealing with, including same measures and features of sample.
3. Multiple libraries are used to perform basic EDA as pandas, numpy, seaborn, matplotlib, sklearn etc.
4. We do Feature Engineering like- dealing with missing values, plotting and removing outliers, checking and removing skewness upto some extent. After that scaling the data.
5. Modelling and doing evaluation of Data using various matrixes.
6. At last with appropriate model result we have to save the data.

* Data Sources and their formats

1. Data has provided by the company. Data is in csv format.
2. This dataset has imbalance data.

* Data Preprocessing Done

steps followed for the cleaning of the data

1. EDA
2. Feature Engineering
3. Feature Scaling

next actions steps over that

1. Best Fit the Model
2. Testing or predict the Model
3. Evaluate
4. Conclusion
5. Saving the Model

* Data Inputs- Logic- Output Relationships
* In this dataset we have a client in Telecom Industry from Indonesia . 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.
* In this dataset we have provided various prospects of that company’s customers like amount recharge last 30 day and 90 days, last recharge in customers main account and data account etc. this are data inputs or independent data given by the company of their customers.
* We have to correlated output on the basis of these inputs.
* if the customers payback its taken loan amount with in days provided by the company then the customer is reliable but if the customer doesn’t payback its taken loan amount within the given time then he is fraud.
* State the set of assumptions (if any) related to the problem under consideration

Here, you can describe any presumptions taken by you.

* Hardware and Software Requirements and Tools Used
* Tools Used

1. Anaconda
2. Jupyter

* Library Used

1. Analysing : Numpy, Pandas, Sci-kit learn
2. Visualization: Matplotlib, Seaborn

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

1. df.describe()

* The mean is more than median (50th percentile) in all columns
* There is a large differences in 75th percentile and max in all columns.
* Most of the columns have zero as their min value.
* The first two observations suggest that there are extreme outliers present in all the columns.

1. **Exploring Data Variables**

* df.label.value\_counts()

Target /Dependent variables is discrete and categorical. In this dataset target variables have 2 values fraud and normal.

1. **Data Visualization**

* To check missing values.
* Dataset has no missing values.
* Testing of Identified Approaches (Algorithms)

Listing down all the algorithms

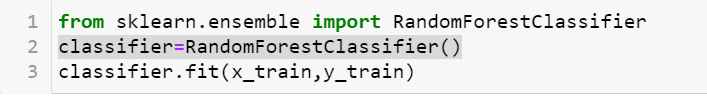
* KNN=KNeighborsClassifier(n\_neighbors=6)
* LR=LogisticRegression()
* DT=DecisionTreeClassifier(random\_state=6)
* GB=GaussianNB()

Ensemble Model

* classifier=RandomForestClassifier()
* Run and Evaluate selected models

1.



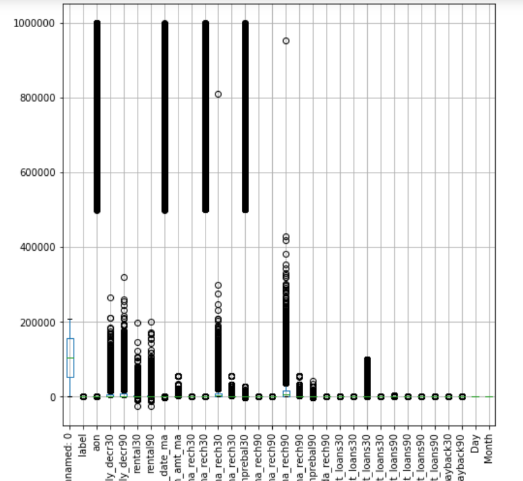


* Key Metrics for success in solving problem under

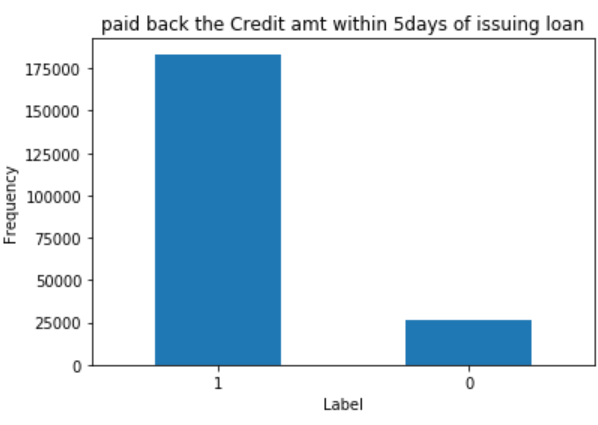
Consideration

* accuracy\_score
* confusion\_matrix
* classification\_report
* roc\_curve, auc
* Visualizations

1. Plotting box plot for understanding outliers:

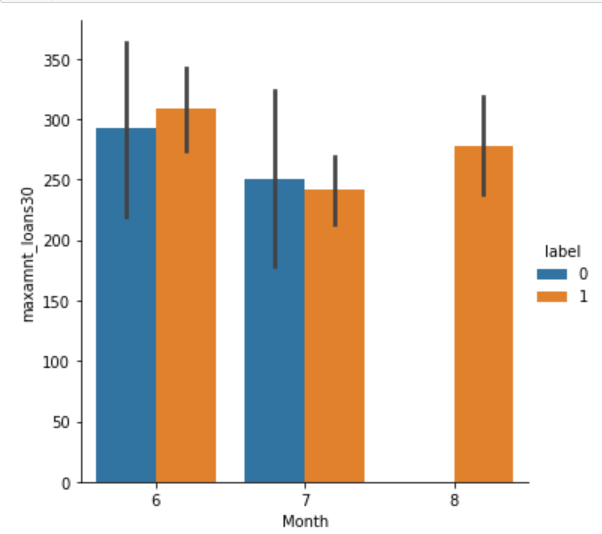


1. Define target variables:



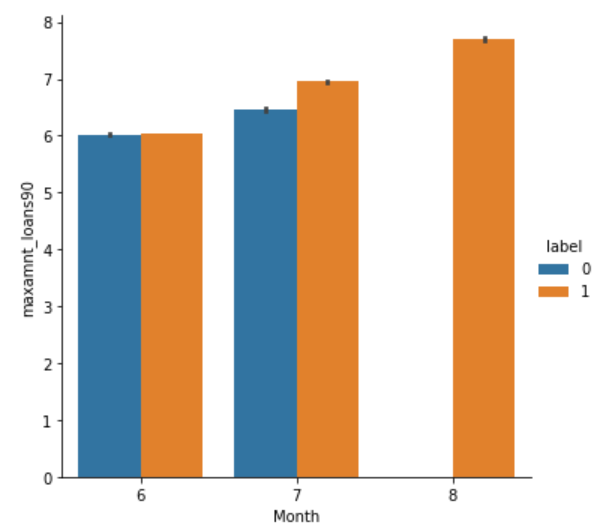
* 1 as non -fraud customers
* 0 as fraud customers

1. Maximum amt of loan taken by the user in last 30 days Vs Month



* Here, we see that in first two month fraud and non-fraud users have nearly equal but in last month Non-fraud users are more than fraud users.

1. Maximum amt of loan taken by the user in last 90 days Vs Month



* In first month fraud users nearly equal to the non-fraud users, in second month non-fraud users are more than fraud users and in third month there is no fraud user.
* Interpretation of the Results

**CONCLUSION**

* Key Findings and Conclusions of the Study
* Those users who has do recharge their no before last date, are Non-Fraud.
* Paid back credit amount with in 5 days have more users who are Non-Fraud.
* Learning Outcomes of the Study in respect of Data Science
* Develop relevant programming abilities.
* Develop the ability to build and assess data based models.
* Execute statistical analysis with professional statistical software.
* Demonstrate skill in data management.
* Demonstrate proficiency with statistical analysis of data.
* Limitations of this work and Scope for Future Work