Micro Credit defaulter model

Submitted by:

Lokesh Gangwani

ACKNOWLEDGMENT

This includes mentioning of all the references, research papers, data sources, professionals and other resources that helped you and guided you in completion of the project.

INTRODUCTION

- Business Problem Framing

Whether the customer will be paying back the loaned amount within 5 days of insurance of loan.

- Conceptual Background of the Domain Problem

Describe the domain-related concepts that you think will be useful for a better understanding of the project.

- Review of Literature

The data is huge and it takes a lot of time to read and find whether to start. Data consist of 209593 rows × 37 columns. Starting with information of columns, read the purpose of the column why these many columns exist and what if it doesn't. With the help of visualization figured out how each column is connected. Among 37 columns just drop one column which has S.no info which is not necessary.

- The motivation for the Problem Undertaken

How data can play a major part in your life from one thing to another. This model reminds me of how organizations are helping people to use the latest technology at a low price.

Analytical Problem Framing

- Mathematical/ Analytical Modelling of the Problem

After EDA used the model for analytical modeling.

Used different statistical models to look simpler.

- Data Pre-processing Done

Used a simple technique with the help of info, Figured out the data type of column. After that divided data into two portions, one is int and another one is float. With the help of astype(int) change the float into int and concat both types of data into one. Now we have the same data type for all the columns and merge them into one.

- Data Inputs- Logic- Output Relationships

Output has only two outcomes 0 & 1 but input consists of a long procedure to follow and to fig. out relationship b/w these columns was a bit difficult.

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

Before the start of coding read the model as much you can and fig. Out what kind of problem this is.

Model/s Development and Evaluation

- Identification of possible problem-solving approaches (methods)

Describe the approaches you followed, both statistical and analytical, for solving this problem.

- Testing of Identified Approaches (Algorithms)

Used RandomForestClassifier, confusion_matrix,classification_report, train_test_split, accuracy_score, roc_curve, roc_auc_score & DecisionTreeClassifier

- Run and Evaluate selected models

Describe all the algorithms used along with the snapshot of their code and what were the results observed over different evaluation metrics.

- Key Metrics for success in solving the problem under consideration

Looking at the dataset and finding the reason what is the purpose of this column in this model would be the key metrics to success.

- Visualizations

Firstly, used a heat map to know the relationship between columns present in the model. Started with bar plot by using multiple columns and comparing each other with the help of visualization. Pair plot also plays a major role to know the relation b/w data. Comparing different columns in the dataset with each other gives clarity on how this dataset going to work.

- Interpretation of the Results

Give a summary of what results were interpreted from the visualizations, preprocessing, and modeling.

CONCLUSION

- Key Findings and Conclusions of the Study

As I'm in the learning stage. This model helped me to clear my concepts and still learning from this model. The only key finding is to know the customer is going to re-pay or not.

- Learning Outcomes of the Study in respect of Data Science

Data is big so; it takes time to read the column's name and what is the objective of that particular column. Once you know how these columns are related to each other afterward used visualization to know more relation between these columns. Used multiple techniques to figure out which model is giving the best outcomes.

- Limitations of this work and Scope for Future Work

The limitation is company can’t provide a loan of the big amount to the consumer because it’ll be difficult for him to pay back or may switch to another network. The future scope is limited because technology is booming. If a company only captures an audience of people using only balance instead of data then this model can be executed well.