

Micro Credit Loan

Submitted by:

Boobalan Murugan.

**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

MFI want to know about customer credit worth before offering the finance to the customer as a capable to repay the loan back to MFI.

* Conceptual Background of the Domain Problem

Telecom Industry want to improve their service through their customer. Hence, they are checking with MFI.

This credit will allow to eligible customer who will pay back the amount within the limit.

* Review of Literature

This is a comprehensive summary of the research done on the topic. The review should enumerate, describe, summarize, evaluate and clarify the research done.

* Motivation for the Problem Undertaken

I need to build the model for this project to find the right customer for this credit limit.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

This Data contain 2L + rows in this dataset. Here, I can see the lot of outlier and skewness here. Model will not perform with this error. Hence, I used to remove outlier through “ZSCORE” and removed to skewness. Then I choose classification model.

* Data Sources and their formats

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.

The Dataset in the excel type under CSV.

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.

* Data Preprocessing Done

I can see the lot of outlier and skewness here. Model will not perform with this error. Hence, I used to remove outlier through “ZSCORE” and removed to skewness. Then I choose classification model.

* Data Inputs- Logic- Output Relationships

Label is an Output column and except columns are input. Here, It will show the customer credit worth based on that we can come to the conclusion. Also, we can remove ID & Date columns from the dataset which will not help to built the model.

* Hardware and Software Requirements and Tools Used

I have used the below listed tool to build the model listed below:

* Jupiter Notebook
* Pandas
* Numpy
* Seaborn
* ZSCORE
* Standard Scaling
* Logistic Regression.

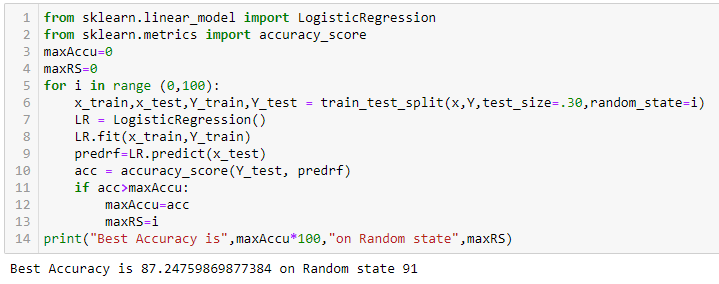
**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

ZSCORE

Standard Scaling.

* Testing of Identified Approaches (Algorithms)
* Logistic Regression
* Decision Tree
* Random Forest.
* Run and Evaluate selected models



* Key Metrics for success in solving problem under consideration

I have used the accuracy score metric to evaluate the model performance for all the algorithm.

* Interpretation of the Results

Model final accuracy score is 91% - Visualization help me to understand the data and type. I performed the EDA which is outlier removal, skewness removal and standard scaling because it will help me to build model with high level accuracy.

Then I choose few algorithm to train my model which are :

Logistic Regression, Decision Tree & Random Forest. Hence, I am getting a final score is 91%.

**CONCLUSION**

* Key Findings and Conclusions of the Study

From the Dataset Observation which are listed below:

* 2L + rows and 36 columns
* Taking long time to process the data
* Not able to visualize the data for outlier
* Heatmap will contain lot of data.
* Learning Outcomes of the Study in respect of Data Science
* Model final accuracy score is 91% - Visualization help me to understand the data and type. I performed the EDA which is outlier removal, skewness removal and standard scaling because it will help me to build model with high level accuracy.
* Then I choose few algorithm to train my model which are :

Logistic Regression, Decision Tree & Random Forest. Hence, I am getting a final score is 91%.

* Limitations of this work and Scope for Future Work

Need to apply the another metric and one more algorithm which is SVC to improve the performance and accuracy.