

Submitted by:

D.Mary Shereesha

**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

In the developing countries, the percentage of population who are below poverty line is more and are uneducated to earn their living.

And these are situation which lead to drop of economy in the country. If we can address this problem in the country and give these people a means of earnings, it would address this problem.

* Conceptual Background of the Domain Problem

These financial schemes are offered chiefly for micro enterprise activities such as agricultural activities, artisan activities, etc. These schemes assist low-income individuals to engage themselves in income-generating activities for a sustainable livelihood. Micro credit schemes are provided to self-help groups as well in order to enable them to set up ventures and earn incomes. These schemes are helpful to tribal people as well.

* Review of Literature

## The Internet review of MCS helped to learn Common Activities that are funded by MCS

## Evolution and Development of Micro Credit Scheme

## How are Micro Credit Schemes distributed to beneficiaries?

## Characteristics of Micro Credit Schemes

* Motivation for the Problem Undertaken

## The great benefits of Microcredit schemes that assist low-income individuals to be financially secure and independent is the motivation to take it up and address it

**Analytical Problem Framing**

* Mathematical/ Analytical Modelling of the Problem

A classification Machine Learning model could help the banks to check whether the customer would repay the loan or no.

This Machine Learning model would be trained with the Loan history and details of the customers.

* Data Sources and their formats

The Data in this Data source is mostly the history of the previous loan and there mode of repayments done for there previous loan in last 30 days and in last 90 days.

* Data Pre-processing Done

The data is all in numerical data type and did not need encoding

The data have extreme outliers which need addressing

Most of the features were collinear and need some of the columns need to be dropped to feed into the model.

* Data Inputs- Logic- Output Relationships

The Logic behind chosen the previous loan details taken and replayed in last 30 days and last 90 days would give an estimate or status of the customer and their awareness of **proper knowledge to take care of the finances carefully**

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

There is a strong ability to handle finances proficiently even by low-income induvial

* .Hardware and Software Requirements and Tools Used

The Model was developed in Python

Pandas, Numpy, StatsModel, SKlearn Libraries are used to develop the model

Matpltlib and seaborn are the libraries used for visualization

**Model/s Development and Evaluation**

* Identification of possible problem-solving approaches (methods)

Defined the Year as Time data

Identified the multicollinear features and drop no required features

Identified the skewness and tried to rectify it to some extent without losing lot of data

Trained-tested the classification Machine learning model

Metrics to estimate its robustness

Checked for its overfitting-underfitting problems

Hyper tunned the selected model

* Testing of Identified Approaches (Algorithms)/ Run and Evaluate selected models

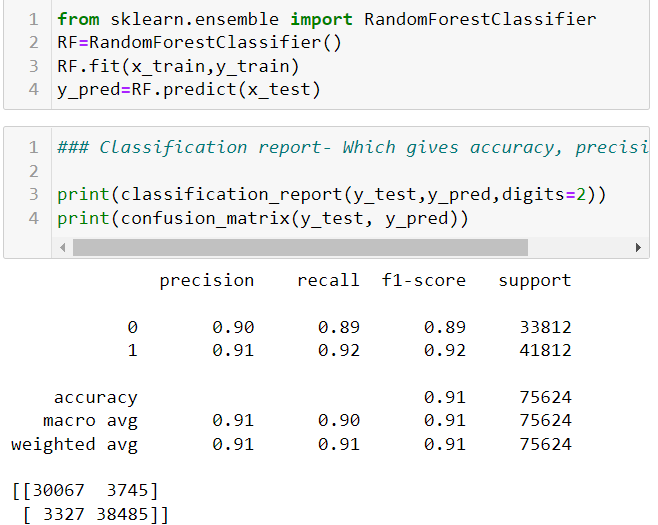
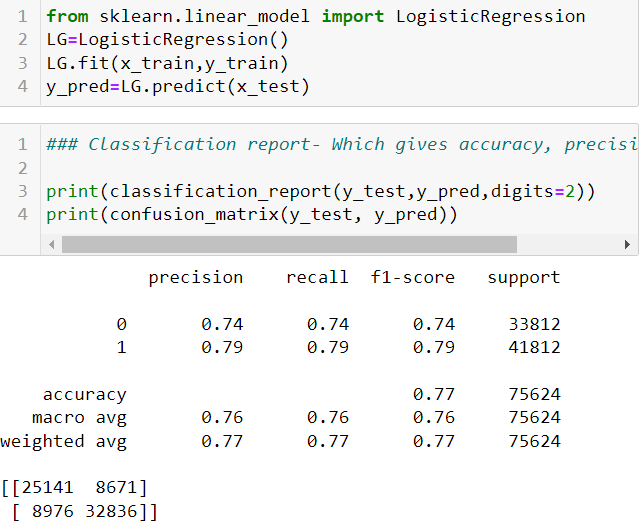
The identified models are

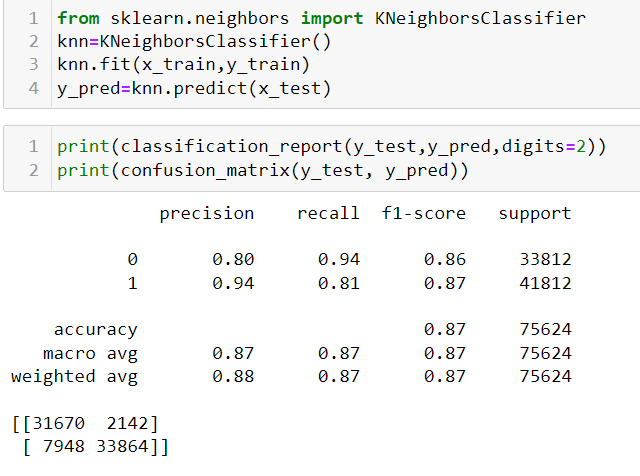
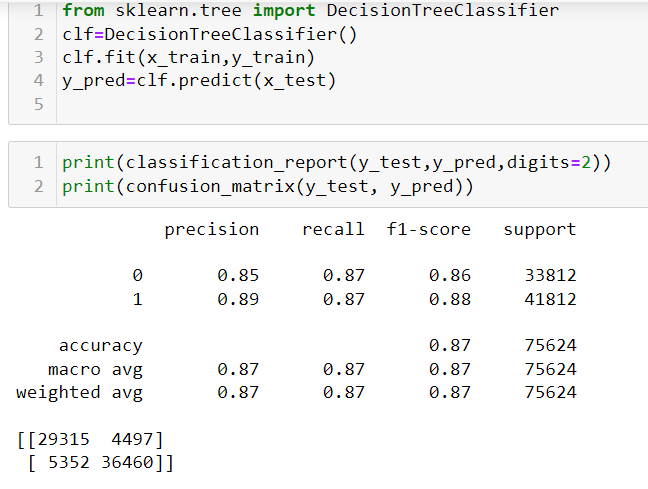
Random Forest Classifer

Logistic Regresson

Kneigbour classifer

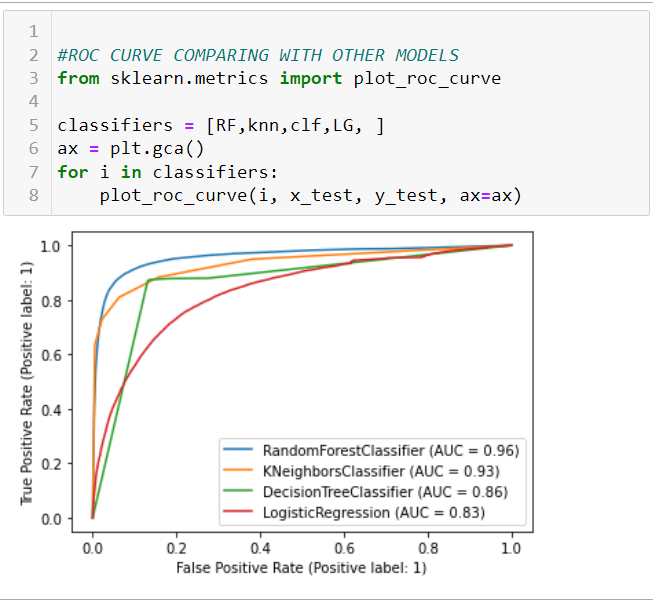
Decision Tree Classifier

* Key Metrics for success in solving problem under consideration/ Visualizations

AUC and ROC curve help me to choose Random forest classifier as the best model



* Interpretation of the Results

The accuracy at which we predict that the customers can repay the borrowed loan or no can be predicted to 90% accuracy by carefully examination of the previous Loan history

**CONCLUSION**

* Key Findings and Conclusions of the Study/Learning Outcomes of the Study in respect of Data Science

The accuracy at which we predict that the customers can repay the borrowed loan or no can be predicted to 90% accuracy by carefully examination of the previous Loan history

* Limitations of this work and Scope for Future Work

The limitation is extreme abnormalities in data entry of the data, collecting the data correctly is a important to further improve the process.