

Loan Predictor(Defaulter or non Defaulter)

Submitted by:

Aamir Rahman

**ACKNOWLEDGMENT**

I would like to express my special thanks of gratitude to FlipRobo, who gave me the golden opportunity to do this wonderful project of Loan predictor.

Secondly, I would also like to thank DataTrained teachers and mentors who teach me all the basic and professional concept for building the project.

**INTRODUCTION**

* Business Problem Framing

Many banks had the issue regarding loan facility. There was always a doubt whether their client is eligible for the loan or not.

So now it’s possible to know to whom they should give the loan by looking at their client history like how many transaction are made and what is their client’s annual income, etc.

* Conceptual Background of the Domain Problem

The person should be aware of the basic statistics, machine learning algorithms and some loan related concepts.

* Motivation for the Problem Undertaken

The main objective behind taking this project is to improve my skill in statistical as well analytical knowledge in machine learning and artificial intelligence.

**Analytical Problem Framing**

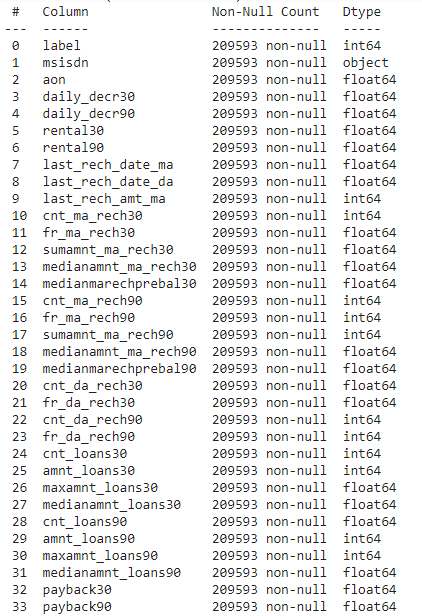
* Mathematical/ Analytical Modeling of the Problem

There are like

Describe the mathematical, statistical and analytics modelling done during this project along with the proper justification.

* Data Sources and their formats

The data is collected from a bank with various attributes. Below is the data summary of all the attributes present.



* Data Preprocessing Done

For data preprocessing various steps are taken into consideration like:

1. Removing redundant columns.
2. Removing any duplicacy present in the data.
3. Checking outliers using box plot and treating them suitable approach.
4. Removing skewness from the data by using power transformation.
5. Balancing the target variable using over sampling algorithm.

* Data Inputs- Logic- Output Relationships

The main relationship between the input variable and the output variable is their correlation and covariance value. The value must lie between -1 to 1 for correlation and 0 to 1 for covariance for a strong relationship between input and the output.

For eg. cnt\_loans90(number of loans taken by user in last 90 days)

By examining this column we can establish a relation between input and output, whether the user had taken the loan or not if he had taken then is he able to pay it or not.

* 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

List of tools used in the project:

1. Scientific computing libraries:- Pandas, Numpy, Scipy
2. Visualization libraries:- Matplotlib, Seaborn
3. Algorithmic libraries:- Scikit learn, Stats model

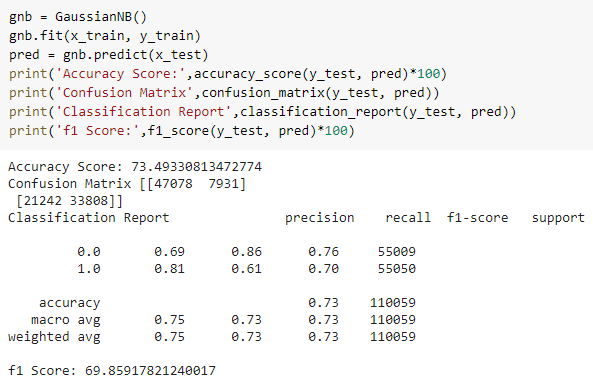
**Model/s Development and Evaluation**

* Testing of Identified Approaches (Algorithms)

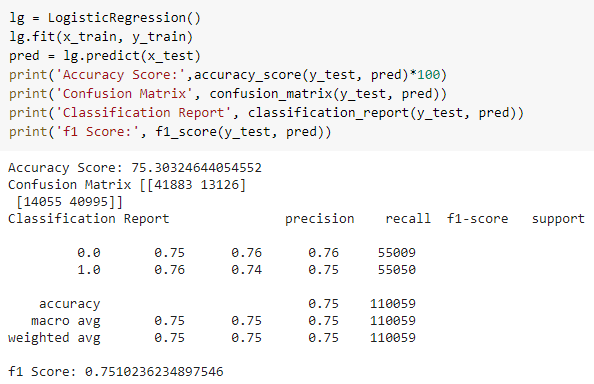
1. GaussianNB
2. Logistic Regression
3. Decision Tree
4. Random Forest
5. Cross Validation

* Run and Evaluate selected models

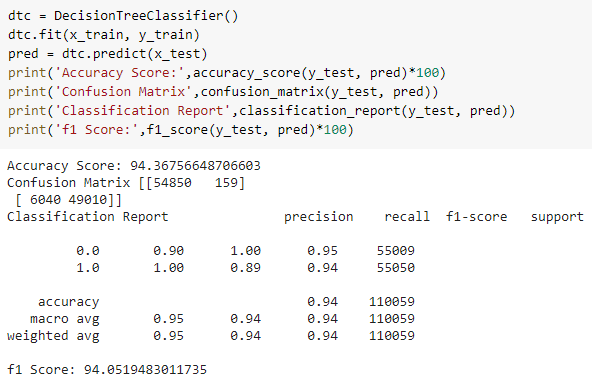
1. GaussianNB



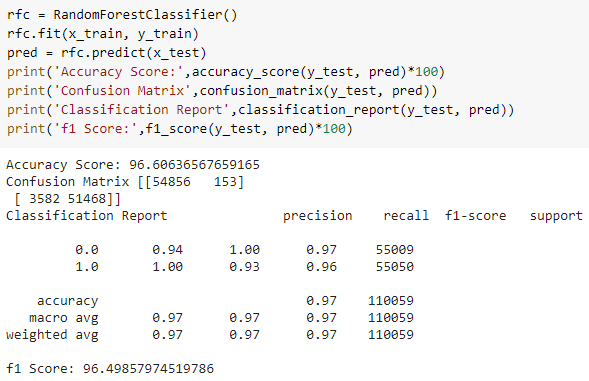
1. Logistic Regression



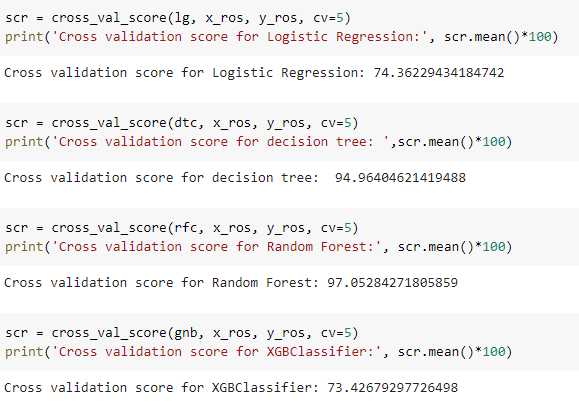
1. Decision Tree



1. Random Forest



1. Cross Validation

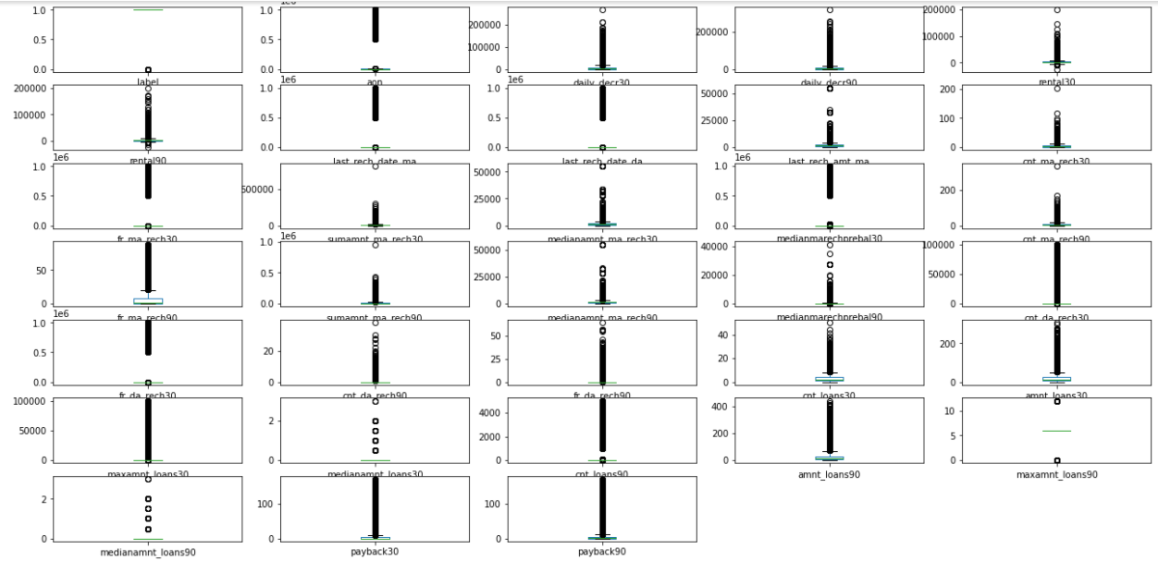


* Key Metrics for success in solving problem under consideration

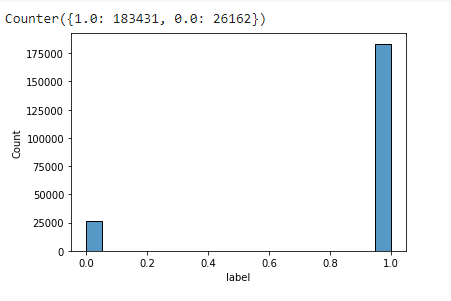
This project is classification problem. These are the metrics used in project:-

1. Accuracy Score
2. Confusion matrix
3. Classification report
4. F1 score

* Visualizations
* Boxplot



* Histplot of unbalanced data



* Roc Curve

