

MICRO CREDIT PROJECT

Submitted by:

ARSH EQBAL

**ACKNOWLEDGMENT**

I would like to thank and express my feelings of gratitude to FlipRobo, who gave me the golden opportunity to do this project on Micro Credit.

Secondly, I would like to thank Data Trained teachers and mentors who have taught me all the basics and professional concept for making this project.

And I have also gathered some insights from the below mentioned websites:-

<https://www.kaggle.com>

<https://www.analyticsvidya.com>

<https://www.learning.datatrained.com>

**INTRODUCTION**

* Business Problem Framing

This project helps us reduce the business problem that are loan defaulters who takes loan and do not pay at the correct time. The real world scenario of this project is Bank Loans deafulters.

* Conceptual Background of the Domain Problem

We analyse the various aspects of datasets relation with the help Exploratory Data Analysis i.e, Drawing Histograms , Heatmap and Boxplots to check for the relationship between the datasets and also for checking the outliers (Extreme Values).

We also check the correlation between the datasets between various attributes in the dataset.

* Review of Literature

We will analyse the dataset using Exploratory Data Analysis. Following are the few methods used:-

1. Histograms
2. Countplot
3. ScatterPlot
4. BoxPlot

Histograms are used to check the dataset is in balance or not.

Countplot are used to check the number of values present in the dataset columns.

Scatter plots are used to see that the datatype is continuous or not.

Boxplot are used to check for Outliers in the dataset.

After finding all this we use Classifiers as it is Classification Problem and using different Classifiers we train our dataset and further we find accuracy of the model .

* Motivation for the Problem Undertaken

Our objective behind making this project is to select the Loan applicants accurately that will return the loan on time and reduce loan defaulting. Motivation behind is the bank models in which thousands of loan defaulter cases are their and to reduce that problem this project has been done.

**Analytical Problem Framing**

* Mathematical/ Analytical Modeling of the Problem

There are mathematical , statistical and analytical modelling done during the project and it is mentioned inside the project with proper justification.

* Data Sources and their formats

The data is collected from a micro service company with various attributes and their data types are mainly float64, int64 and object type.

* Data Preprocessing Done

For data preprocessing following steps have been taken into consideration and are as follows:-

1. Removing Redundant columns.
2. Removing any duplicate data present.
3. Removing Skewness from the data using power transform.
4. Checking for Outliers using BoxPlot and Treating them using suitable approach.
5. Balancing the target variable using oversampling.

* 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 strong relationship between the input and the output.

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

The mobile number given in the dataset was of no use for finding the accuracy and is also having very low correlation with the target variable so we have dropped that column.

* 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**

* Identification of possible problem-solving approaches (methods)

By analysing the dataset we have found that this problem is a Classification Problem so we have used the approaches used for solving the Classification Problem.

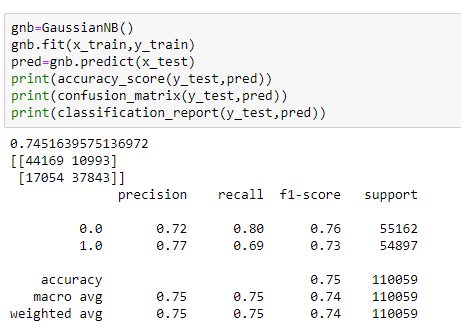
* Testing of Identified Approaches (Algorithms)

List of all the algorithms used:-

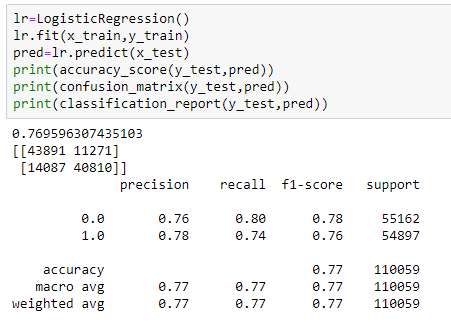
1. GaussianNB
2. LogisticRegression
3. Decision Tree
4. RandomForest
5. AdaBoostClassifer

* Run and Evaluate selected models

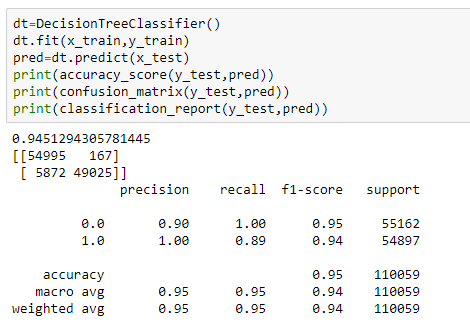
1. GaussianNB



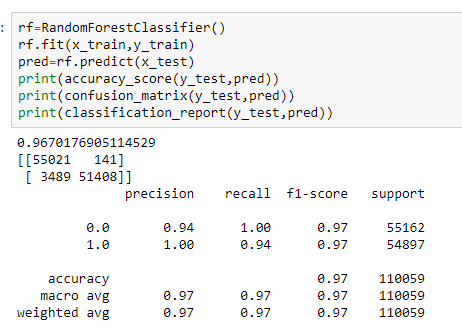
1. Logistic Regression



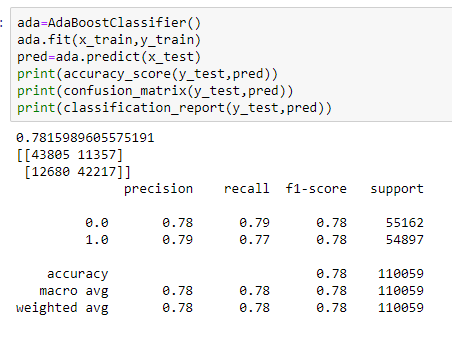
1. Decision Tree



1. Random Forest



1. AdaBoost



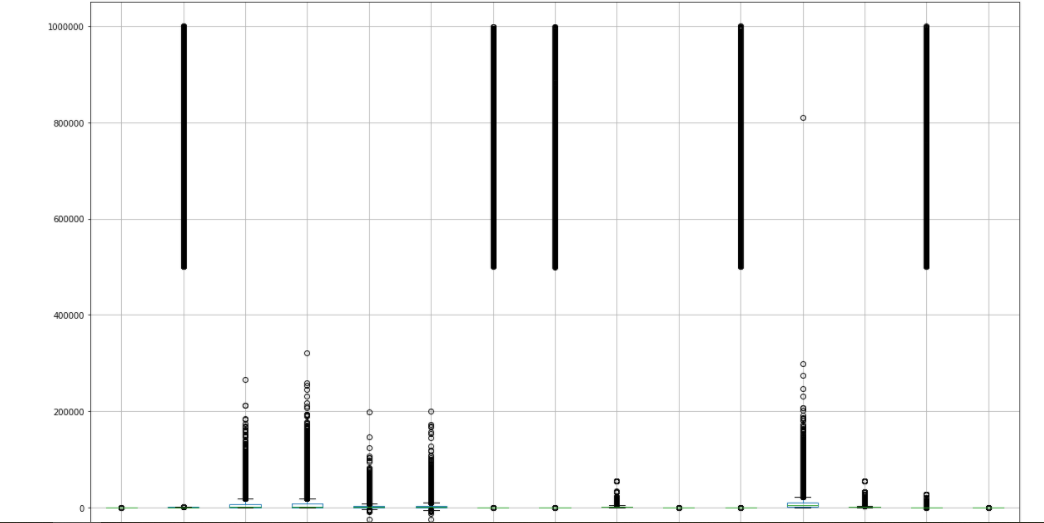
* Key Metrics for success in solving problem under consideration

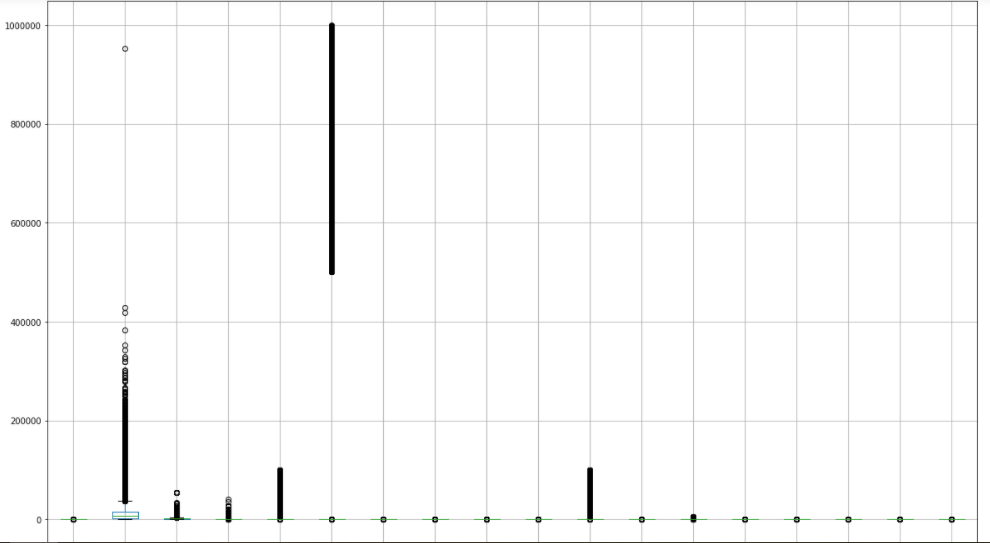
This project is a Classification Problem. These are the metrics used in the project:-

1. Accuracy Score
2. Classification Report
3. Confusion Matrix

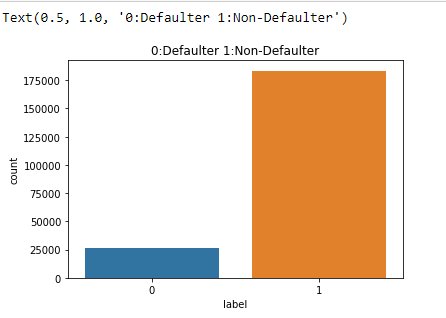
* Visualizations

1. BoxPlot





1. CountPlot



* Interpretation of the Results

Countplot was used to figure out the imbalance in the data.

Boxplot was used to figure out the Outliers in the dataset.

Classification Report gave us the F1 Score , precision values.

Accuracy Score gave us the Accuracy of the classifier used .