

**Micro Credit Defaulter Prediction**

Submitted by:

ASHISH YADAV

**ACKNOWLEDGMENT**

We would like to express our deep and sincere gratitude to FlipnWork for giving us the opportunity to do this project. As a great bridge between academic and industry, this program educated us how to perform theoretical methodology in real life. We would like to express our sincere thankfulness to my mentor Sapna Verma for the continuous support , for their patience, enthusiasm, motivation and immense knowledge. As our academic mentor, Sapna Verma supported and helped in this project. Additionally, we would also like to thank all our friends who offered us some help, such as Divyanshu who helped me during this project

**INTRODUCTION**

**Problem Statement :**

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.

Many microfinance institutions (MFI), experts and donorsare supporting the idea of using mobile financial services (MFS) which they feel are more convenient and efficient, and cost saving, than the traditional high-touch model used since long for the purpose of delivering microfinance services. Though, the MFI industry is primarily focusing on low income families and are very useful in such areas, the implementation of MFShas been uneven with both significant challenges and successes.

Today, microfinance is widely accepted as a poverty-reduction tool, representing $70 billion in outstanding loans and a global outreach of 200 million clients.

We are working with one such client that is in Telecom Industry. 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.

They understand the importance of communication and how it affects a person’s life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

They arecollaborating with an MFI to provide micro-credit on mobile balances to be paid back in 5 days. The Consumer is believed to be defaulter if he deviates from the path of paying back the loaned amount within the time duration of 5 days. For the loan amount of 5 (in Indonesian Rupiah), payback amount should be6(in Indonesian Rupiah), while, for the loan amount of 10(in Indonesian Rupiah), the payback amount should be 12(in Indonesian Rupiah).

The sample data is provided to us from our client database. It is hereby given to you for this exercise. In order to improve the selection of customers for the credit, the client wants some predictions that could help them in further investment and improvement in selection of customers.

**Exercise:**

Build a model which can be used to predict in terms of a probability for each loan transaction, whether the customer will be paying back the loaned amount within 5 days of insurance ofloan. In this case, Label ‘1’ indicates that the loan has been payed i.e. Non- defaulter, while, Label ‘0’ indicates that the loan has not been payed i.e. defaulter.

**Business Goal :**

Using micro credit as a poverty-reduction tool,by focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

**Domain Understanding :**

The telecom  sector continues to be at the epicenter for growth, innovation, and disruption for virtually any industry. Mobile devices and related broadband connectivity continue to be more and more embedded in the fabric of society today and they are key in driving the momentum around some key trends such as video streaming, Internet of Things (IoT), and mobile payments. Our client is also a telecom player . 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.

They understand the importance of communication and how it affects a person’s life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

**Literature :**

• The main steps in our research were the following.

• **Exploratory Data Analysis (EDA):** By conducting explanatory data analysis, we obtain a better understanding of our data. This yields insights that can be helpful later when building a model, as well as insights that are independently interesting.

•**Balancing Dataset :** In order to balance the imbalance dataset ,we use technique like SMOTE.

• **Modeling:** We apply Decision Tree , Logical Regression models for prediction of the micro credit defaulter prediction

**Analytical Problem Framing**

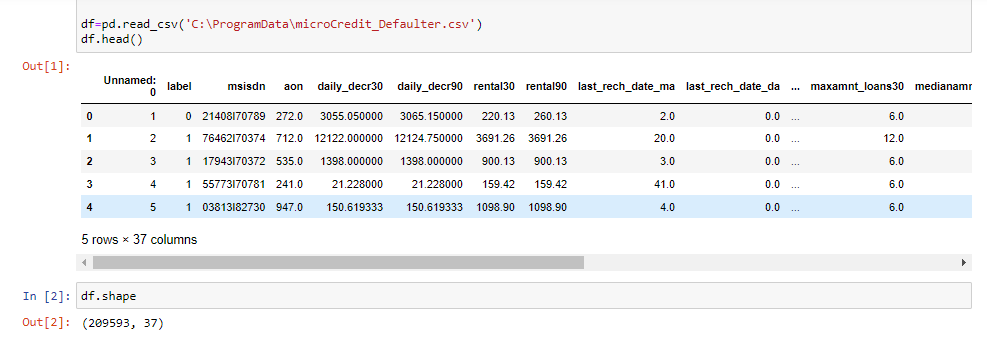
**• Mathematical/ Analytical Modeling of the Problem**

Telecom Industry client iscollaborating with an MFI to provide micro-credit on mobile balances to be paid back in 5 days.

Client is 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.

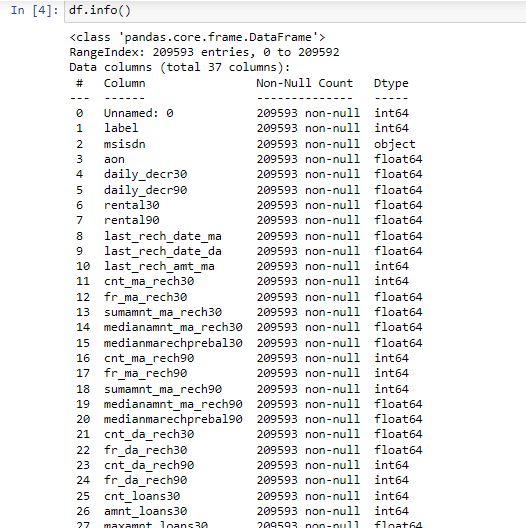
They understand the importance of communication and how it affects a person’s life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

**Dataset :**

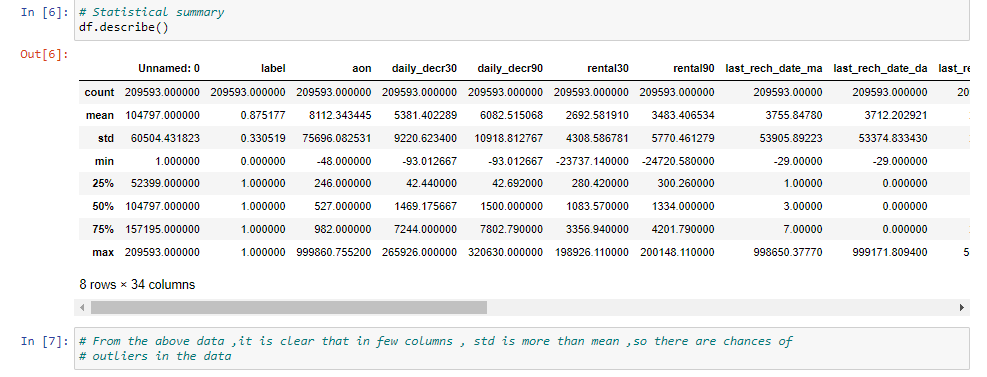
****

**EXPLORATORY DATA ANALYSIS**

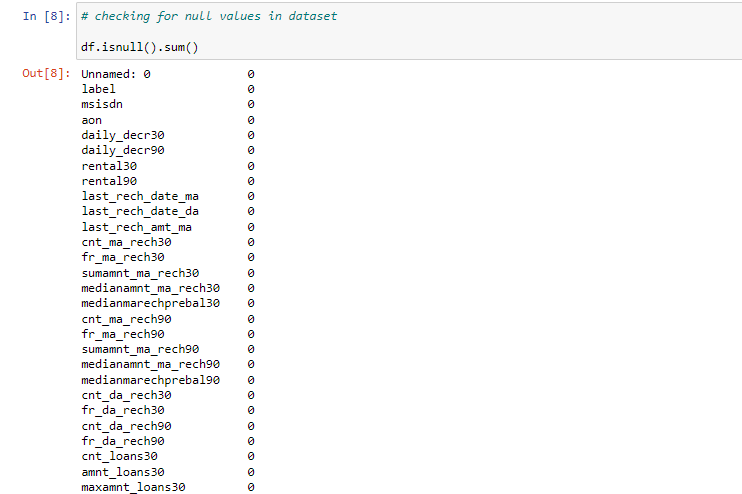
* Data exploration is the first step in data analysis and typically involves summarizing the main characteristics of a data set, including its size, accuracy, initial patterns in the data and other attributes. It is commonly conducted by data analysts using visual analytics tools, but it can also be done in more advanced statistical software, Python. Before it can conduct analysis on data collected by multiple data sources and stored in data warehouses, an organization must know how many cases are in a data set, what variables are included, how many missing values there are and what general hypotheses the data is likely to support. An initial exploration of the data set can help answer these questions by familiarizing analysts with the data with which they are working.



**Statistical Summary :**

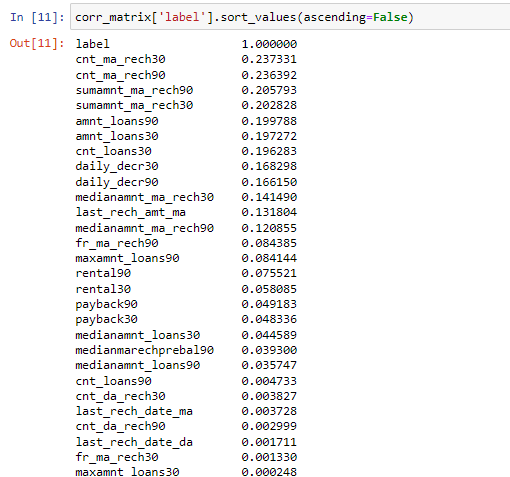
****

**Checking null values in dataset**

****

**CORRELATION :**

****

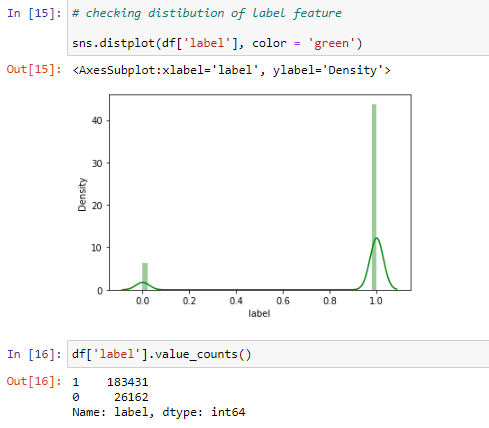
****

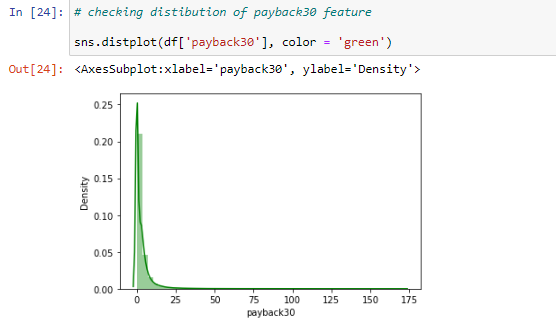
**DATA VISUALIZATION :**

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. In the world of Big Data, data visualization tools and technologies are essential to analyse massive amounts of information and make data-driven decisions.

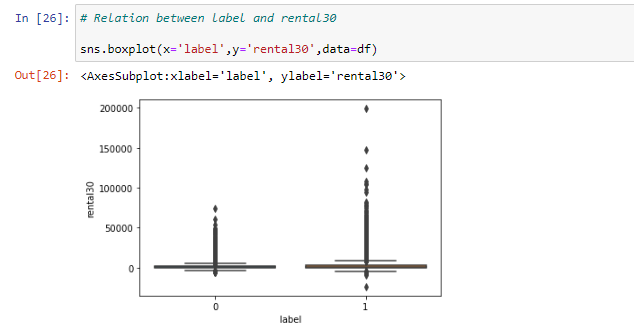
**Distribution Plots :**

****

****

****

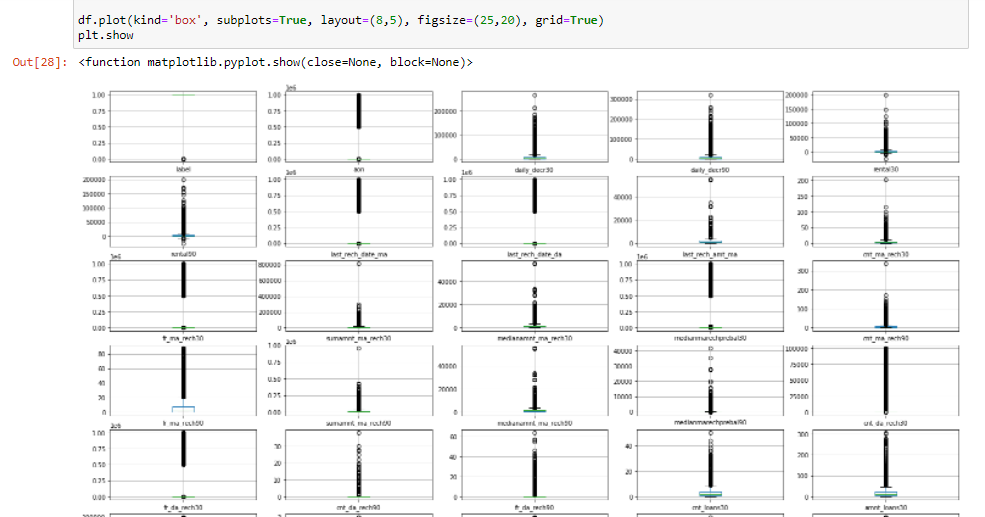
**BOX PLOTS :**

****

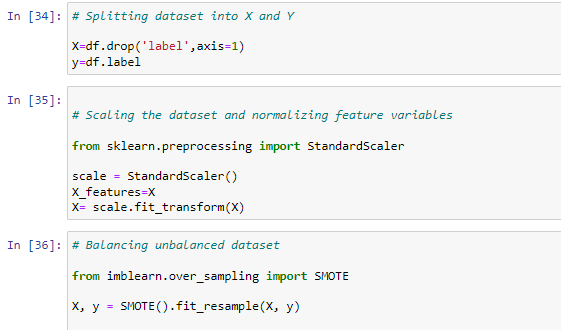
****

**OUTLIERS :**

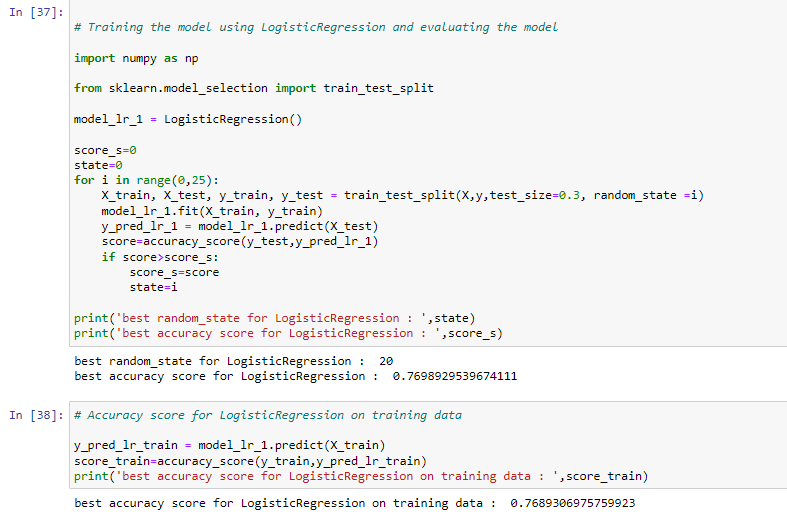
* An outlier is an object that deviates significantly from the rest of the objects. They can be caused by measurement or execution error. The analysis of outlier data is referred to as outlier analysis or outlier mining.
* It is a data point that is noticeably different from the rest. They represent errors in measurement, bad data collection, or simply show variables not considered when collecting the data.

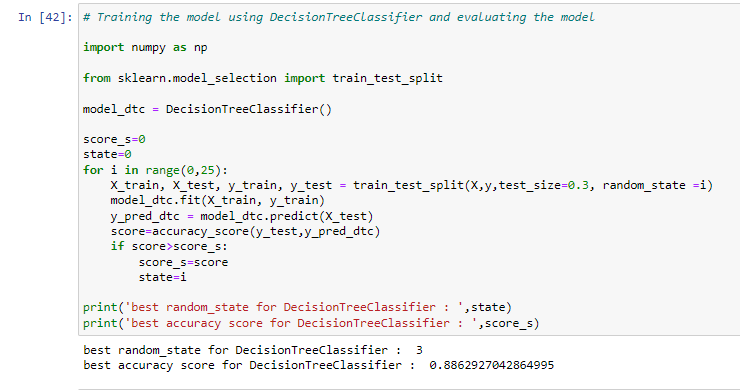
****

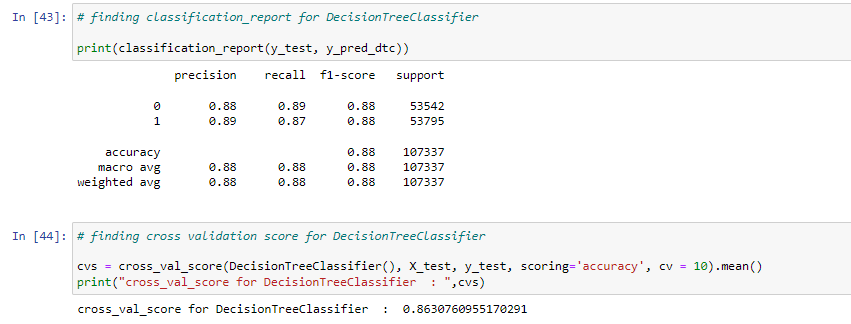
**Splitting,scaling,balancing Dataset :**

****

**EVALUATION OF MODELS**

****

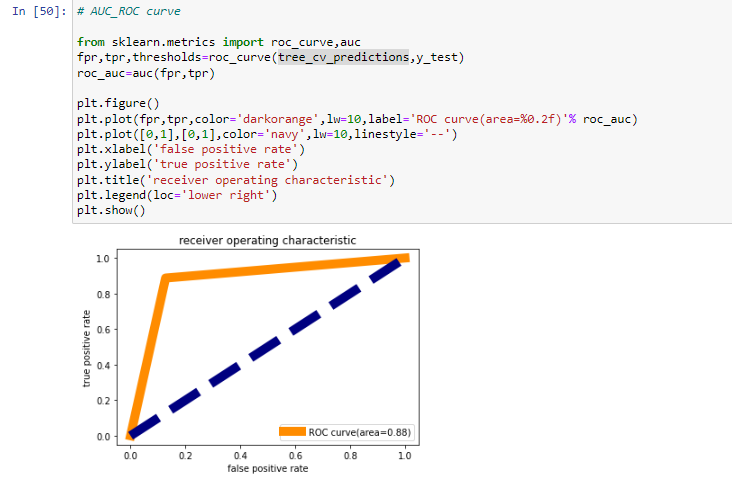
****

****

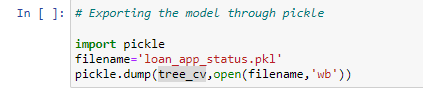
**HYPERPARAMETER TUNING OF DecisionTreeClassifier :**

****

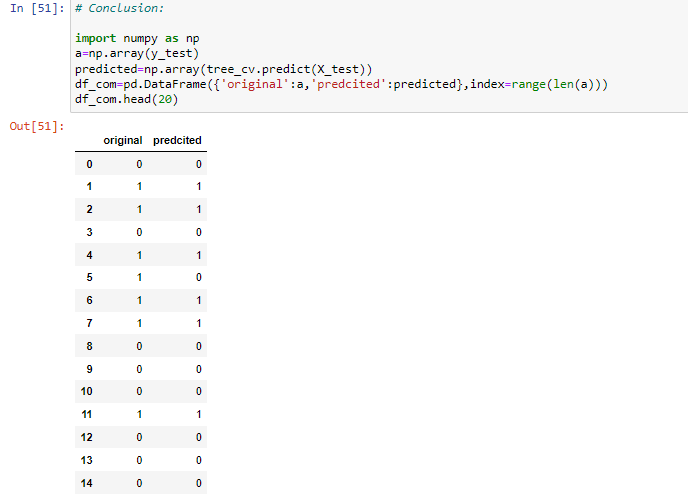
**AUC\_ROC Curve :**

****

**IMPORTING OF MODEL :**

****

**Conclusion :**

****