Information Technology Department

Big Data Analytics (2171607)
(Departmental Elective -2)

Mini Project
Credit Card Approval Prediction

Team Members
Harshil Patel (170170116031)
Umang Patel (170170116039)

Faculty Guide- Prof. Naimisha S Trivedi
Year: 2020-21
Semester VII

Vishwakarma Government Engineering College, Chandkheda

Objective of the Project:

Credit score cards are a common risk control method in the financial industry. It uses personal information and data submitted by credit card applicants to predict the probability of future defaults and credit card borrowings. The bank is able to decide whether to issue a credit card to the applicant. Credit scores can objectively quantify the magnitude of risk.

We are assuming that if the customer paid very late payment (More than 60 days) We consider them as risky customer(Might not pay the bill in the future)

First, We used **Logistic Regression** to train our model but in that case accuracy not as much good. so, we used **Decision Tree Classifier** to improve accuracy of model.

Our objective is to determine an applicant is 'good' or 'bad' client, on the bases of given values of attributes.

List of Dependent Variables:

target_Value

List of Independent Variables:

ID, Gender, Car, Reality, ChldNo, income, incometp,edutp, fam_status, housingtp, DAYS_BIRTH, DAYS_EMPLOYED, FLAG_MOBIL, workphone, phone, email, occytp, familysize, starting_month

Method of Approach used :

1) **Data Collection**
2) **Data Exploration**
3) **Data Pre-processing**
4) **Data Transformation**
5) **Feature Extraction (Information Value table)**
6) **Model Training**
7) **Data Visualization**

Algorithms used for Model Training

Logistic Regression :

Logistic regression models the probabilities for classification problems with two possible outcomes. Logistic model is a common method for credit scoring. Because Logistic is suitable for binary classification tasks and can calculate the coefficients of each feature.

Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extensions exist.

$$Logistic(z) = 1/(1+e^{(-z)})$$

Accuracy of the Model: 61.90%

Decision Tree Classifier :

Decision trees are widely used since they are easy to interpret, handle categorical features. The goal of using a **Decision Tree** is to create a training model that can use to predict the class or value of the target variable by learning simple **decision** rules inferred from prior data(training data).

It is a **tree**-structured classifier, where internal nodes represent the features of a dataset, branches represent the **decision** rules and each leaf node represents the outcome. In a **Decision tree**, there are two nodes, which are the **Decision** Node and Leaf Node.

Accuracy of the Model: 98.45%