CREDIT CARD ELIGIBILITY ANALYSIS

A project work done in partial fulfilment of the "**Certificate course on Data Analytics & Business Intelligence**"

Submitted by:

Kirti

Ujjwal Singh

Certificate Course on Data Analytics & Business Intelligence Batch-10

Shaheed Sukhdev College of Business Studies

May, 2024

ACKNOWLEDGEMENT

We would like to express my special thanks of gratitude to our teacher **Dr. Rishi Ranjan Sahay**, who gave us the golden opportunity to do this wonderful project. This project helped me in doing a lot of research and discovering many new things.

We overwhelmed in all humbleness and gratefulness to acknowledge my learning to all those who have helped me to put these ideas, well above the level of simplicity and into something concrete.

Thanking You
Kirti
Ujjwal Singh

DECLARATION

We, **Kirti and Ujjwal Singh**, hereby declare that this project entitled "**Credit Card Eligibility Analysis**" is the original work done by me under the guidance of **Dr. Rishi Rajan Sahay, Assistant Professor, Shaheed Sukhdev College of Business Studies, University of Delhi.**

This project work is undertaken as part of our **certificate course in Data Analytics and Business Intelligence** and is submitted in partial fulfilment of the requirements for **the award of the certificate of Data Analytics and Business Intelligence at Shaheed Sukhdev College of Business Studies, University of Delhi**.

We affirm that the research and findings presented in this project are genuine. All sources of information and data have been acknowledged appropriately.

We also declare that any help received in carrying out this project and preparing the report has been duly acknowledged.

Abstract/Executive Summary

This report presents an analysis of the factors influencing credit card eligibility. The dataset comprises various attributes such as demographic details, financial information, and employment status.

The objective is to understand the key determinants of credit card eligibility and provide insights for financial institutions to enhance their credit assessment.

INTRODUCTION

Credit card eligibility analysis is a critical component for financial institution in determine the creditworthiness of potential customers. Credit cards offer a convenient and flexile payment method, but they also come with inherent risks for the issuing institutions. Ensuring that credit cards are issued to reliable and creditworthy individuals is essential to mitigate these risks.

The primary objective of this analysis is to identify the factors that significantly influence an individual's eligibility for credit card. By understanding these factors financial institutions can refine their credit assessment processes, enhance their decision-making capabilities, and reduce the risks of defaults.

This report presents a detailed analysis based on dataset comprising various attributes of individuals, including demographic details, financial status, employment information, and asset ownership. The dataset allows us to examine a wide range of factors that may impact credit card eligibility, such as age, income, employment stability, and ownership of assets like cars and properties.

RESEARCH OBJECTIVE

The primary objective of this project is to evaluate the performance of Logistic Regression, Random forest, Decision tree and SVM algorithms in predicting credit card eligibility of individual based on employment status, demographic details, and financial information. By comparing the accuracy, precision, F1 score, and confusion matrices of these classification models, we aim to identify the most effective algorithm for predicting credit card eligibility and provide valuable insights for financial institutions.

METHODOLOGY

DATASET

We took this dataset from 'Kaggel.com' in a csv file format & this was uploaded just 12 days ago. Thus, the data is still new to make a project. It has 9709 rows and 20 columns. Following is a picture of the same:

Own_car	Own_property	Work_phone	Phone	Email	Unemployed	Num_children	Num_family	Account_length	Total_income	Age	Years_employed	Income_type	Education_type	Family_status	Housing_type	Occupation_type	Target
1	1	1	0	0	0	0	2	15	427500	32.8685736	12.43557363	Working	Higher education	Civil marriage	Rented apartment	Other	1
1	1	0	0	0	0	0	2	29	112500	58.7938151	3.104786546	Working	Secondary / secondary special	Married	House / apartment	Security staff	0
0	1	0	1	1	0	0	1	4	270000	52.3214029	8.353354278	Commercial associate	Secondary / secondary special	Single / not married	House / apartment	Sales staff	0
0	1	0	0	0	1	0	1	20	283500	61.504343	0	Pensioner	Higher education	Separated	House / apartment	Other	0
1	1	1	1	1	0	0	2	5	270000	46.193967	2.105450488	Working	Higher education	Married	House / apartment	Accountants	0
1	1	0	0	0	0	0	2	17	135000	48.6745108	3.269060966	Commercial associate	Secondary / secondary special	Married	House / apartment	Laborers	0
1	0	0	0	0	0	0	2	25	130500	29.2107299	3.019911429	Working	Incomplete higher	Married	House / apartment	Accountants	1
0	1	0	1	0	0	0	2	31	157500	27.4639452	4.021985393	Working	Secondary / secondary special	Married	House / apartment	Laborers	1
0	1	0	0	0	0	1	2	44	112500	30.0293641	4.435409351	Working	Secondary / secondary special	Single / not married	House / apartment	Other	0
1	1	0	0	0	0	3	5	24	270000	34.741302	3.184185849	Working	Secondary / secondary special	Married	House / apartment	Laborers	0
0	1	0	0	0	0	1	3	39	405000	32.4222948	5.519620526	Commercial associate	Higher education	Married	House / apartment	Managers	0
1	1	0	1	0	0	0	2	43	112500	56.1325695	12.18368618	Commercial associate	Secondary / secondary special	Married	House / apartment	Drivers	0
1	1	0	0	0	0	2	4	39	135000	43.1521523	8.687378933	Working	Secondary / secondary special	Married	House / apartment	Laborers	0
0	1	0	0	0	0	1	3	24	211500	44.3869484	19.43640184	State servant	Secondary / secondary special	Civil marriage	House / apartment	Core staff	0
1	1	0	1	0	0	0	2	10	360000	45.6409098	14.68613319	Commercial associate	Secondary / secondary special	Married	House / apartment	Security staff	1
0	1	0	0	1	0	2	4	21	126000	33.9801639	4.854309123	Commercial associate	Higher education	Married	House / apartment	Managers	0
0	1	0	0	0	1	0	1	40	315000	55.2673908	0	Pensioner	Secondary / secondary special	Widow	House / apartment	Other	0
0	1	0	0	0	0	0	1	11	247500	46.5882256	3.687960738	Commercial associate	Higher education	Separated	Rented apartment	Core staff	0
0	1	0	0	0	0	0	1	7	297000	42.4895788	8.854391261	Commercial associate	Secondary / secondary special	Single / not married	Rented apartment	Laborers	0
0	0	0	1	0	0	0	2	43	157500	37.3505274	13.26789736	Commercial associate	Higher education	Married	House / apartment	High skill tech staff	1
0	1	1	1	0	0	0	2	12	135000	42.3937521	3.854973066	Working	Secondary / secondary special	Married	House / apartment	Drivers	0
0	1	0	0	0	0	1	3	45	225000	28.2579382	7.688042876	Working	Secondary / secondary special	Married	House / apartment	Other	0
1	1	0	0	0	0	0	2	4	135000	48.6936761	4.687296796	Working	Secondary / secondary special	Married	House / apartment	Laborers	0
0	1	0	1	0	0	0	2	15	157500	43.8804356	2.433999329	Working	Secondary / secondary special	Married	House / apartment	Other	1
0	1	0	0	0	1	0	2	49	112500	61.1073465	0	Pensioner	Secondary / secondary special	Married	House / apartment	Other	0
1	1	0	1	0	0	0	2	10	112500	57.1510709	6.518956584	State servant	Secondary / secondary special	Married	House / apartment	Drivers	0
1	1	1	1	0	0	0	1	30	270000	45.399974	3.269060966	Working	Secondary / secondary special	Separated	House / apartment	Cleaning staff	0
1	1	0	0	1	0	0	2	4	270000	29.0984757	2.266987002	Working	Higher education	Married	House / apartment	Managers	0
1	1	0	0	0	0	0	2	22	157500	27.1792029	2.266987002	Working	Secondary / secondary special	Married	House / apartment	Drivers	0
0	1	1	1	0	0	0	2	40	166500	51.817628	17.18310438	Working	Secondary / secondary special	Married	House / apartment	Laborers	0
0	1	0	0	1	0	2	4	24	216000	42.2842358	8.520366606	State servant	Higher education	Married	House / apartment	Other	0
1	1	0	0	1	0	0	2	21	270000	54.3775711	29.68712568	Commercial associate	Secondary / secondary special	Married	House / apartment	Laborers	0
1	1	0	0	0	0	0	2	17	225000	39.5627563	7.603167758	Commercial associate	Secondary / secondary special	Married	House / apartment	Laborers	1
0	1	0	0	0	0	0	1	13	315000	28.8958706	3.269060966	Working	Secondary / secondary special	Single / not married	House / apartment	Accountants	0

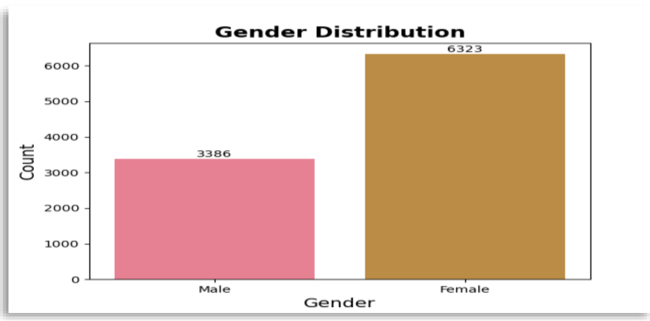
DATA DESCRIPTION

1. **ID**: An identifier for each individual (customer).
2. **Gender**: The gender of the individual.
3. **Own-car**: A binary feature indicating whether the individual owns a car.
4. **Own-property**: A binary feature indicating whether the individual owns a property.
5. **Work-phone**: A binary feature indicating whether the individual has a work phone.
6. **Phone**: A binary feature indicating whether the individual has a phone.
7. **Email**: A binary feature indicating whether the individual has provided an email address.
8. **Unemployed**: A binary feature indicating whether the individual is unemployed.
9. **Number-children**: The number of children the individual has.
10. **Number-family**: The total number of family members.
11. **Account-length**: The length of the individual's account with a bank or financial institution.
12. **Total-income**: The total income of the individual.
13. **Age**: The age of the individual.
14. **Years-employed**: The number of years the individual has been employed.
15. **Income-type**: The type of income (e.g., employed, self-employed, etc.).
16. **Education-type**: The education level of the individual.
17. **Family-status**: The family status of the individual.
18. **Housing-type**: The type of housing the individual lives in.
19. **Occupation-type**: The type of occupation the individual is engaged in.
20. **Target:** The target variable for the classification task, which indicates whether the individual is eligible for a credit card or not (e.g., Yes/No, 1/0).

EXPLORATORY DATA ANALYSIS (EDA)

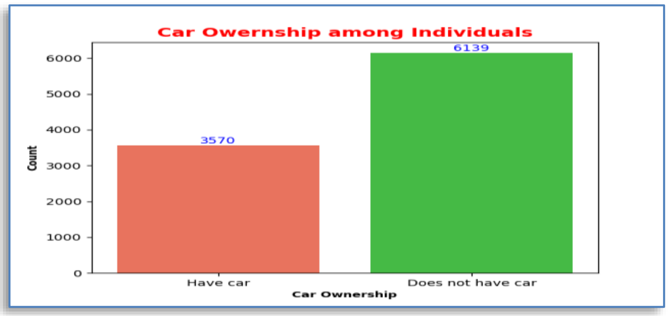
1. **Gender Distribution**:

   This graph shows that there are a more female than male in this dataset.



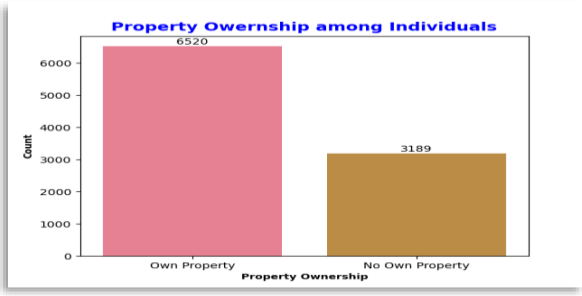
2. **Count of Car Ownership:**

This is barplot show count of how many
individuals have their own car or not.
This graph show that 6139 individuals doesn’t have there own car.



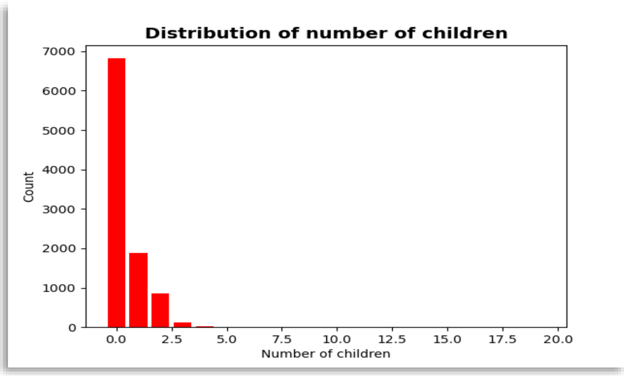
3. **Count of Property Ownership:**

This is a barplot show the count of how many
individuals have their own property or not.
This graph show that 6520 individuals have there own property.



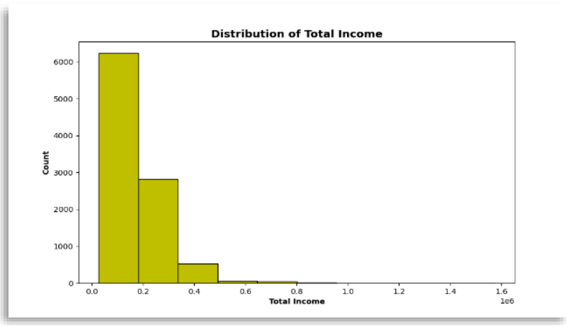
4. **Distribution of number of children:**

This barplot show that number of children among individuals. This graph show that most individuals have no children.



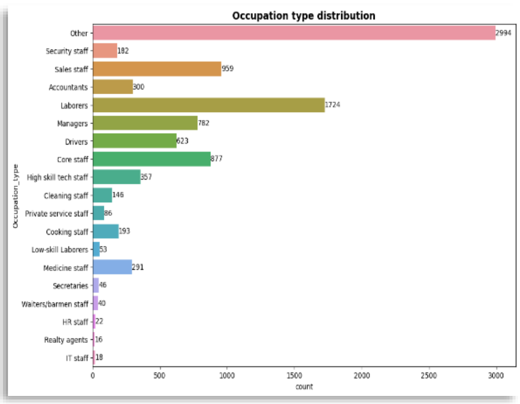
5. **Distribution of Total Income:**

This a histogram show the income distribution among individuals.



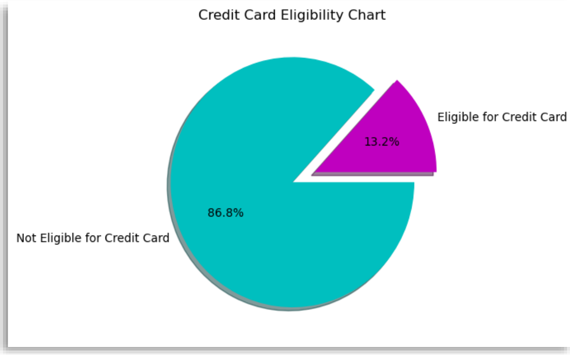
6. **Count of Occupation Type:**

This a countplot show the count of each occupation type. This graph show that most occupation type is other and laborers.



7. **Percentage of card eligibility:**

This a pie chart show that the number of percentage of individual eligibility for credit card. This graph show that 86.8% individuals are not eligible for credit card.



CLASSIFICATION ALGORITHMS

Predicting the individuals credit card eligibility. In this data analysis project, we explore the efficency of four classification algorithms—Random Forest, Logistic Regression, Support Vector Machine (SVM), and Decision Tree—in evaluating and predicting the eligibility.

LOGISTIC REGRESSION

Logistic Regression is a widely used statistical method for binary classification tasks. Despite its name, logistic regression is a linear model that predicts the probability of an instance belonging to a particular class. It employs the logistic function (sigmoid function) to map the output of a linear combination of input features to a probability score between 0 and 1. Logistic regression provides interpretable results, as it allows for estimating the effect of each predictor variable on the probability of the outcome.

SUPPORT VECTOR MACHINE (SVM)

Support Vector Machine is a powerful supervised learning algorithm for both classification and regression tasks. SVM aims to find the optimal hyperplane that separates different classes in the feature space while maximizing the margin between the classes. By transforming the input features into a higher-dimensional space using kernel functions, SVM can efficiently handle non-linear decision boundaries. SVM offers flexibility in choosing different kernel functions such as linear, polynomial, and radial basis function (RBF), allowing for versatile classification tasks.

RANDOM FOREST

Random Forest is a popular machine learning algorithm used for both classification and regression tasks. It works by creating a “forest” of decision tress during training, where each tree is built using a random subset of the features and a subset of the training data.

During prediction, the algorithm aggregates the predictions of each tree to make a final prediction. This aggregation can be done by averaging the predictions for regression tasks or using voting for classification tasks.

Random Forest is known for its high accuracy, scalability, and ability to handle large datasets with high dimensionality. It also provides a measure of feature importance, which can be useful for understanding the underlying data and feature selection.

DECISION TREE

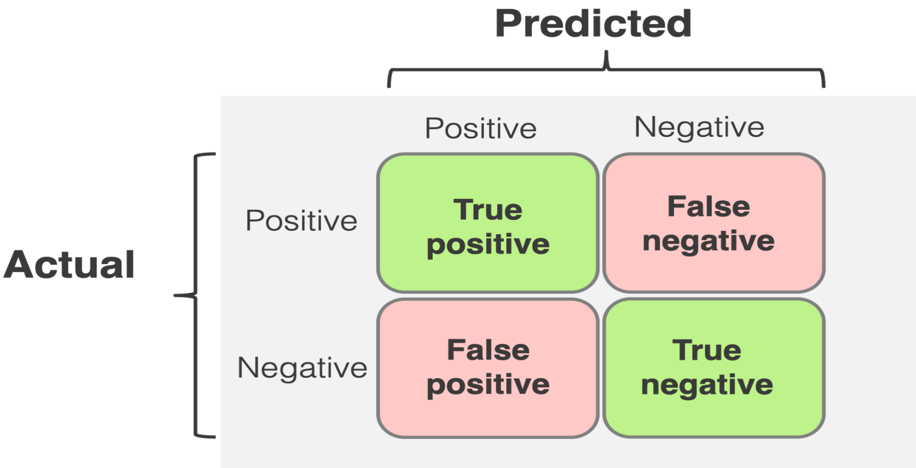
A Decision Tree is a machine learning algorithm used for classification and regression tasks. It models decisions and their possible consequences as tree-like structure of nodes and branches.

- **Root Node:** Represents the entire dataset and is split into two or more homogeneous sets.
- **Decision Nodes:** Intermediate nodes that represent the outcome of a test on a feature, leading to further splits.
- **Leaf Nodes:** Terminal nodes that represent the final decision or classification.

The algorithm splits the data based on feature values to create branches, aiming to maximize the separation between different classes or outcomes. This process continues until the algorithm reaches a stopping criterion, such as a maximum depth or a minimum number of samples per leaf. Decision Trees are easy to interpret and visualize but can be prone to overfitting, especially with complex datasets.

CONFUSION MATRIX

A confusion matrix is a performance evaluation tool used in classification tasks to visualize the performance of a machine learning model. It presents a summary of the predictions made by the model compared to the actual ground truth across different classes.



COMPONENTS OF CONFUSION MATRIX

1) True Positive (TP): The instances that were correctly predicted as positive (or belonging to the target class) by the model.

2) True Negative (TN): The instances that were correctly predicted as negative (or not belonging to the target class) by the model.

3) False Positive (FP): The instances that were incorrectly predicted as positive by the model when they actually belong to the negative class. Also known as Type I error.

4) False Negative (FN): The instances that were incorrectly predicted as negative by the model when they actually belong to the positive class. Also known as Type II error.

INTERPRETATION

o **Accuracy:** The overall correctness of the model, calculated as (TP + TN) / (TP + TN + FP + FN).

o **Precision**: The proportion of true positive predictions out of all positive predictions made by the model, calculated as TP / (TP + FP).

o **Recall** (Sensitivity): The proportion of true positive predictions out of all actual positive instances, calculated as TP / (TP + FN).

o **F1 Score:** The harmonic mean of precision and recall, providing a balance between the two metrics, calculated as 2 * (Precision * Recall) / (Precision + Recall).

MODEL DEVELOPMENT

- Before starting with the algorithms, we have to convert the categorical data into dummy variable i.e. (True/False).
- After that we will convert this data into integer format i.e. (0 to 1). After that observe the independent and dependent variable in our dataset.
- We then split the data using Sklearn library into Training data (80% of dataset) and Test data (20% of dataset).

RESULTS OF CLASSIFICATION ALGORITHMS

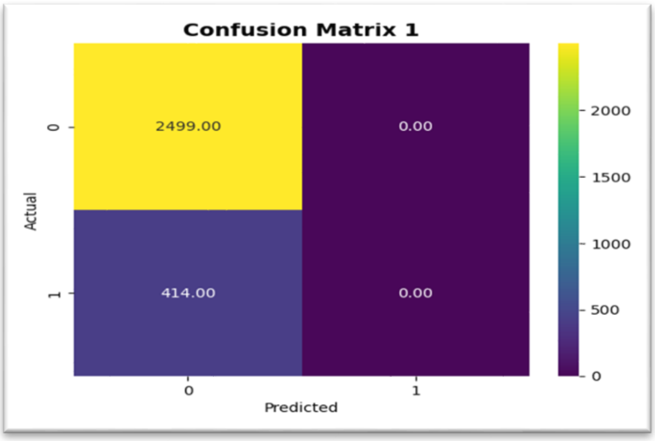
LOGISTIC REGRESSION

I. Accuracy Score : 0.8578

II. Precision Score : 0.0

III. Recall Score : 0.0

IV. F1 Score : 0.0



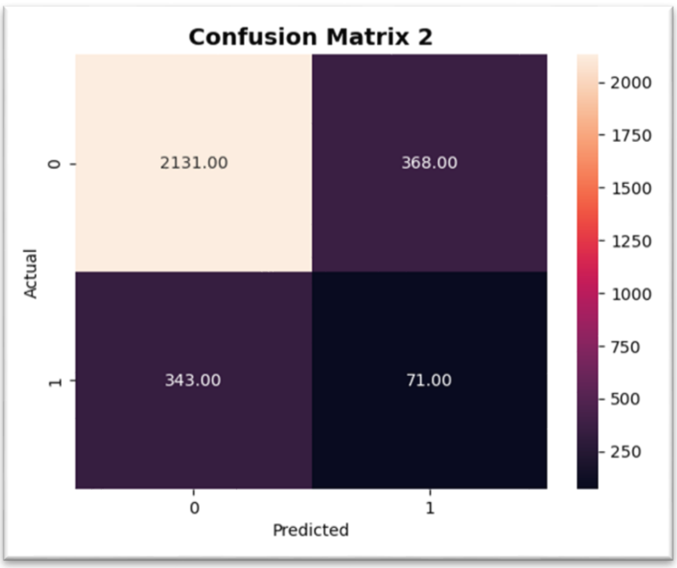
DECISION TREE

I. Accuracy Score: 0.755

II. Precision Score: 0.511

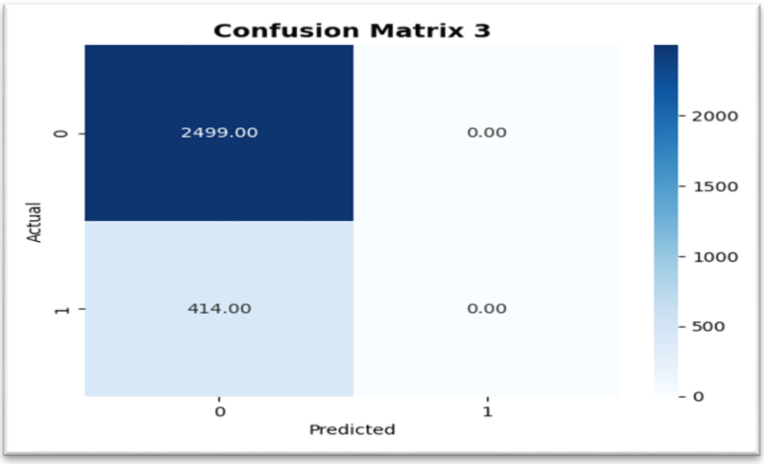
III. Recall Score: 0.5121

IV. F1 Score: 0.511



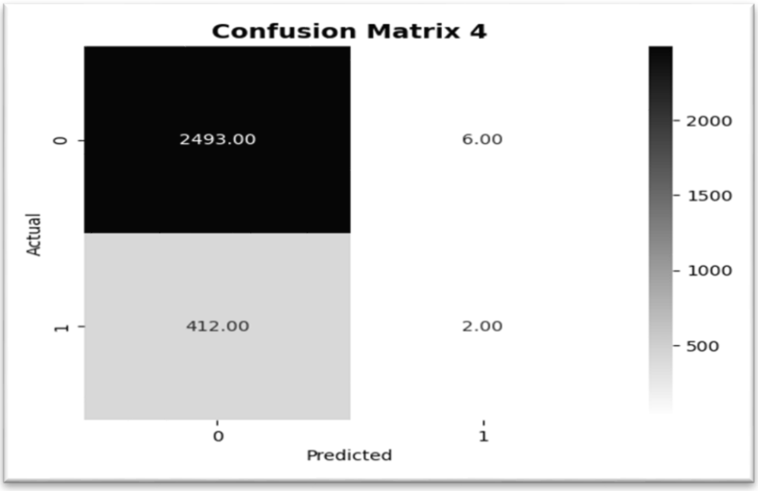
SUPPORT VECTOR MACHINE

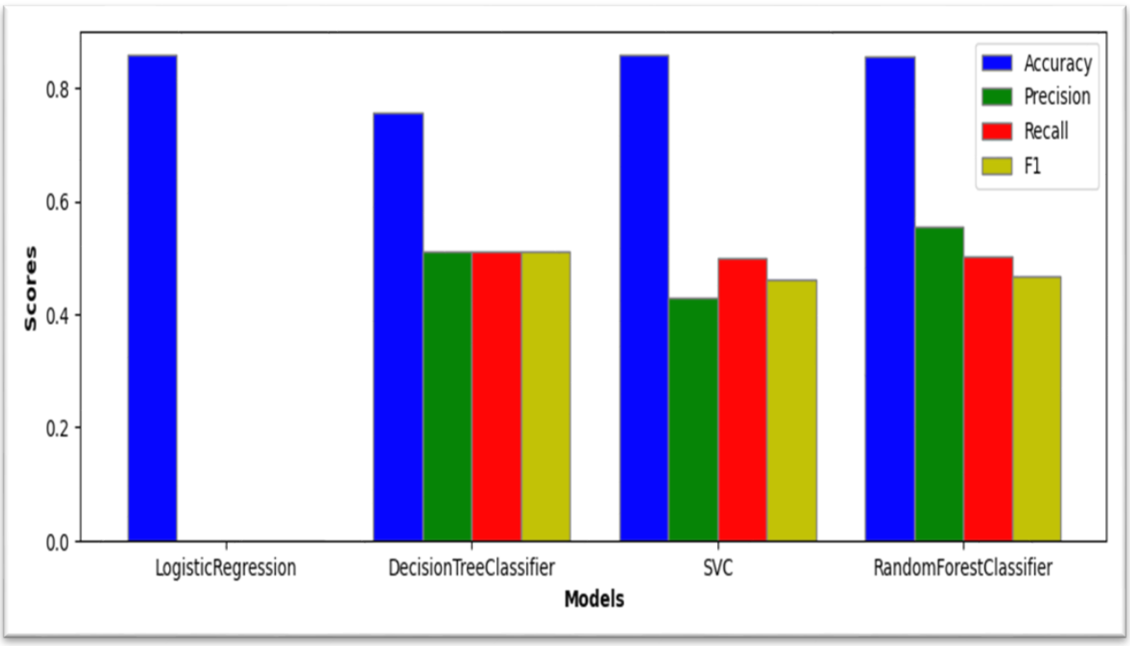
I. Accuracy Score: 0.857
II. Precision Score: 0.428
III. Recall Score: 0.5
IV. F1 Score: 0.461



RANDOM FOREST

I. Accuracy Score: 0.8565
II. Precision Score: 0.554
III. Recall Score: 0.501
IV. F1 Score: 0.466





CONCLUSION

The dataset reveals several key insights: There are more females than males, and 6139 individuals do not own a car. Property ownership is slightly higher, with 6520 individuals owning property. Most individuals have no children. Income distribution varies widely, and the most common occupations are 'other' and 'laborers'. Notably, 86.8% of individuals are not eligible for a credit card, highlighting potential financial constraints within the dataset. Overall, the data suggests significant economic challenges for the majority of individuals.

Based on the evaluation metrics, the Decision Tree model achieves balanced performance across precision, recall, and F1 scores, indicating its effective handling of both false positives and false negatives. Although Logistic Regression and SVM show high accuracy, their precision and recall are notably lower, especially for Logistic Regression, which yields zero scores for precision, recall, and F1. Random Forest performs well with good accuracy and relatively higher precision but falls short in recall and F1 scores compared to the Decision Tree.
Given the comprehensive balance of metrics, the Decision Tree appears to be the best model for our project, offering robust and reliable performance.

BIBLOGRAPHY

- https://www.kaggle.com/datasets/rohit265/credit-card-eligibility-data-determining-factors
- https://www.geeksforgeeks.org/understanding-logistic-regression/#what-is-logistic-regression
- https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.airtel.in%2Fblog%2Fcredit-card%2Fsecured-credit-card-meaning-how-it-works%2F&psig=AOvVaw38g_v04YcXdExn0-tEOCg5&ust=1717163403663000&source=images&cd=vfe&opi=89978449&ved=0CBIQjRxqFwoTCJCGqqfCtYYDFQAAAAAdAAAAABAE