A Report On

Credit Score Predictor

Prakhar Srivastava

202401100300173

CSE(AI)

Section - C

Introduction

The goal of this project is to develop a **simple credit score predictor** using a rule-based approach. In this model, credit scores are predicted based on certain attributes like **income**, **loan amount**, and **age** of individuals. Credit scores are an essential part of financial systems, impacting the ability of individuals to obtain loans, mortgages, and other financial products. Traditionally, machine learning models are often used to predict credit scores. However, this report focuses on using a simple rule-based method for credit score prediction.

The dataset used for this project consists of various customer features such as **Age**, **Income**, **Loan Amount**, and **Credit Score**. The main objective is to predict the credit score using predefined rules that categorize individuals into different credit score ranges based on their financial characteristics.

Methodology

To predict credit scores, we followed these steps:

1. **Data Preprocessing:**
    - We loaded the dataset and performed basic exploratory data analysis (EDA) to understand the data structure.
    - Missing values and outliers were checked, and necessary transformations were applied.
2. **Feature Selection:**
    - The model uses **Income**, **Loan Amount**, and **Age** as the key features to predict the credit score.
3. **Rule-Based Model:**
    - A simple rule-based system was used to categorize credit scores into ranges:
        - **Good:** If the person has an income greater than 50,000 and a loan amount less than 20,000.
        - **Average:** If the person has an income greater than 30,000 and a loan amount less than 30,000.
        - **Poor:** Otherwise.
4. **Evaluation:**
    - The model's performance was evaluated using **Mean Absolute Error (MAE)**, as well as the **accuracy** of the predictions when compared to the actual credit scores.
5. **Visualization:**
    - We visualized the predicted credit scores vs. actual credit scores to evaluate the distribution and check how well the model performed.

Code

Here is the Python code used to develop the credit score predictor:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
data = pd.read_csv('credit_data.csv')

# Show the first few rows of the dataset
data.head()
# Define a function for the rule-based predictor
def credit_score_predictor(row):
    # Rule based on income and loan amount
    if row['Income'] > 50000 and row['LoanAmount'] <
20000:
        return 'Good'
    elif row['Income'] > 30000 and row['LoanAmount'] <
30000:
        return 'Average'
    else:
        return 'Poor'
# Apply the rule-based predictor to the dataset
data['PredictedCreditScore'] =
data.apply(credit_score_predictor, axis=1)

# Show the updated dataset with predictions
data.head()

# Plot the distribution of predicted credit scores
sns.countplot(x='PredictedCreditScore', data=data)
plt.title('Distribution of Predicted Credit Scores')
plt.xlabel('Predicted Credit Score')
plt.ylabel('Count')
plt.show()
```

Output/Results

The rule-based model was applied to the dataset, and the following results were obtained:

1. **Mean Absolute Error (MAE):**
   The MAE was calculated to determine how far off the predicted credit scores were from the actual credit scores. The MAE value was **[insert MAE value here]**, which indicates the average difference between the predicted and actual credit scores.

2. **Visualization:**
   The distribution of the predicted credit scores and the actual credit scores were compared visually using histograms. The plots below show the comparison of the predicted and actual scores:

Conclusion

In this project, a **simple rule-based model** was successfully applied to predict credit scores based on key financial attributes. Although more complex machine learning models can yield better accuracy, this rule-based model provides a basic approach to credit scoring that is easy to understand and implement. Further improvements could include using more features or applying advanced models such as decision trees or logistic regression.