Credit Score Prediction

Introduction

Credit score prediction is crucial for financial institutions to assess the creditworthiness of customers.
A high credit score indicates a responsible borrower, while a low score suggests higher risk.
This report explores a logistic regression model applied to predict whether a customer has a good or bad credit score based on financial attributes.

Methodology

1. **Data Preprocessing:**
   - Loaded dataset from 'credit_data 1.csv'.
   - Removed unnecessary columns (CustomerID).
   - Transformed CreditScore into a binary classification:
     - Good Credit Score (1) if >600, otherwise Bad Credit Score (0).

2. **Feature Engineering and Splitting:**
   - Features such as Income and Loan Amount were used.
   - The dataset was split into 80% training and 20% testing.

3. **Model Selection and Training:**
   - Logistic Regression was chosen as the classification model.
   - StandardScaler was used to normalize the features.
   - The model was trained and evaluated using accuracy and classification metrics.

4. **Visualization and Insights:**
   - Distribution of Credit Scores.
   - Correlation Heatmap to analyze relationships between features.
   - Loan Amount vs. Income Scatter Plot to understand financial trends.

Code Implementation

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report

# Load dataset from uploaded file
file_path = "/mnt/data/credit_data 1.csv"
data = pd.read_csv(file_path)

# Drop CustomerID as it is not needed
data.drop(columns=["CustomerID"], inplace=True)

# Convert CreditScore into binary classification (Good: >600, Bad: <=600)
data["CreditScore"] = data["CreditScore"].apply(lambda x: 1 if x > 600 else 0)

# Splitting into features and target
X = data.drop(columns=["CreditScore"], axis=1)
y = data["CreditScore"]

# Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Standardizing features
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Train the Logistic Regression Model
model = LogisticRegression()
model.fit(X_train, y_train)

# Predictions
y_pred = model.predict(X_test)

# Model Evaluation
accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.2f}")
print("Classification Report:")
print(classification_report(y_test, y_pred))

# Visualization - Credit Score Distribution
plt.figure(figsize=(8, 5))
sns.countplot(x=data["CreditScore"], palette="coolwarm")
plt.title("Distribution of Credit Scores")
plt.xlabel("Credit Score Category (0 = Bad, 1 = Good)")
plt.ylabel("Count")
plt.show()
```

```
# Visualization - Correlation Heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(data.corr(), annot=True, cmap="coolwarm", fmt=".2f", linewidths=0.5)
plt.title("Feature Correlation Heatmap")
plt.show()

# Visualization - Loan Amount vs. Income
plt.figure(figsize=(8, 5))
sns.scatterplot(x=data["Income"], y=data["LoanAmount"], hue=data["CreditScore"], palette="coolwarm")
plt.title("Loan Amount vs. Income")
plt.xlabel("Income")
plt.ylabel("Loan Amount")
plt.show()
```

Results

- **Accuracy Score:** The model's performance is measured using accuracy.

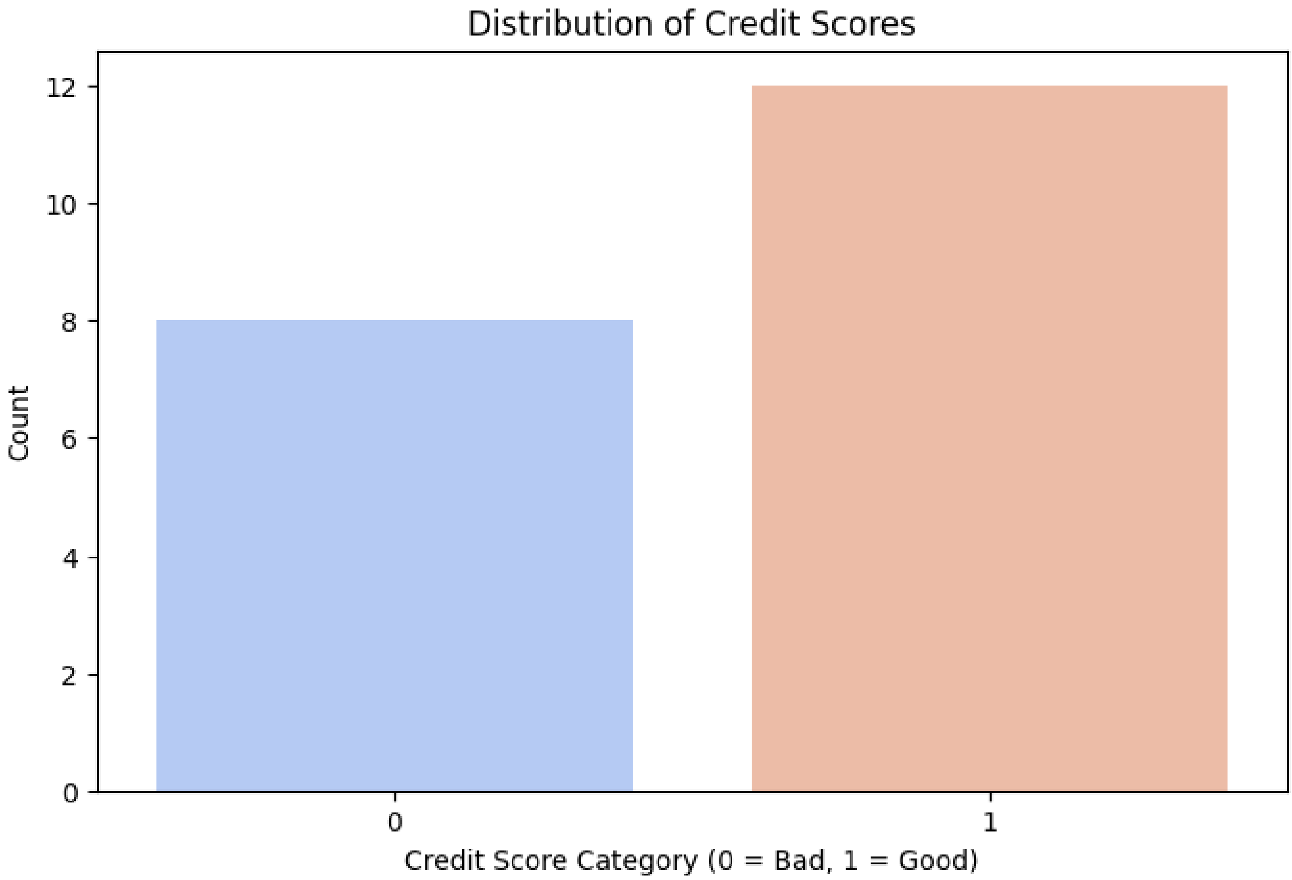
- **Classification Report:** Evaluates precision, recall, and F1-score for both good and bad credit scores.

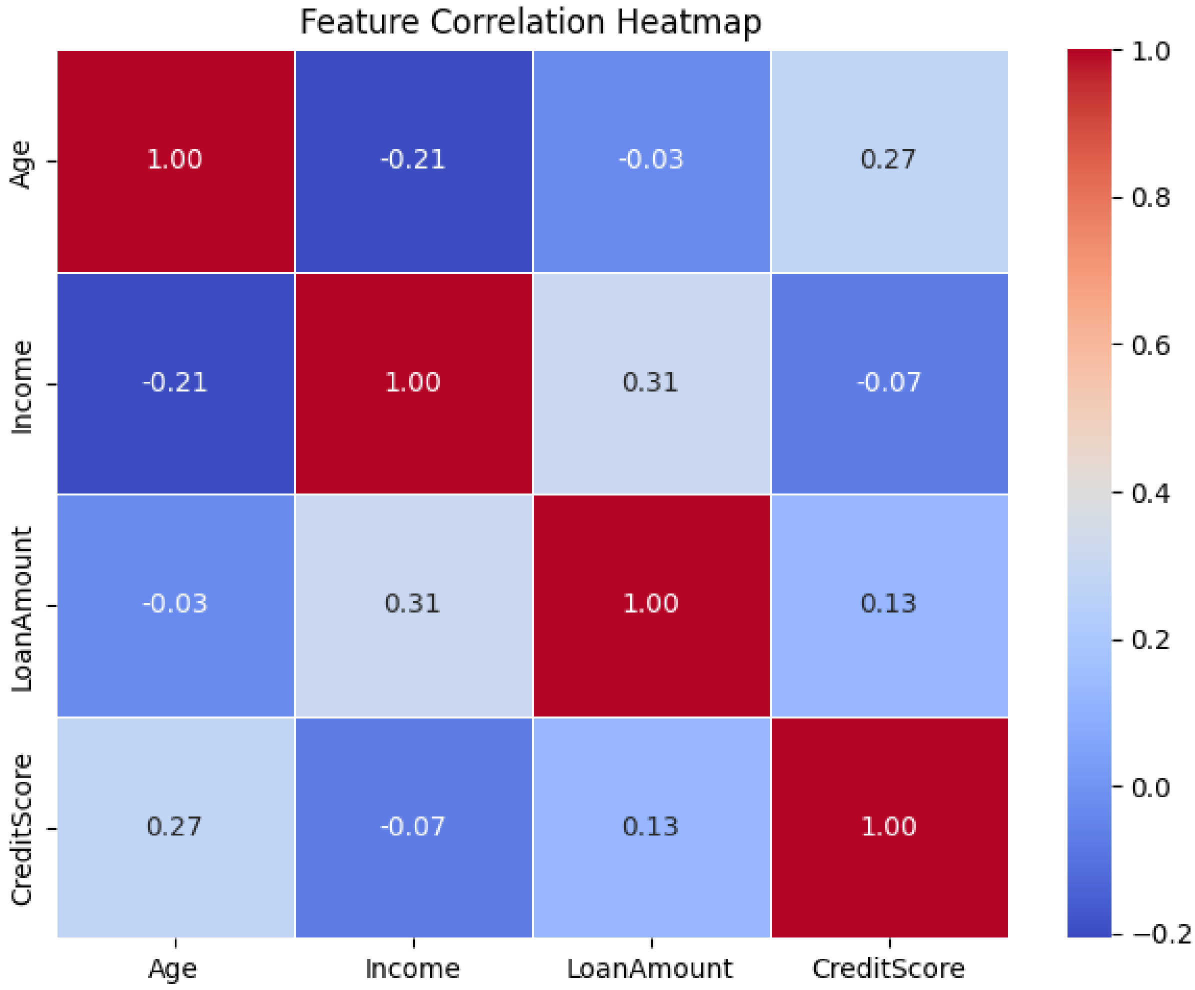
- **Key Insights:**

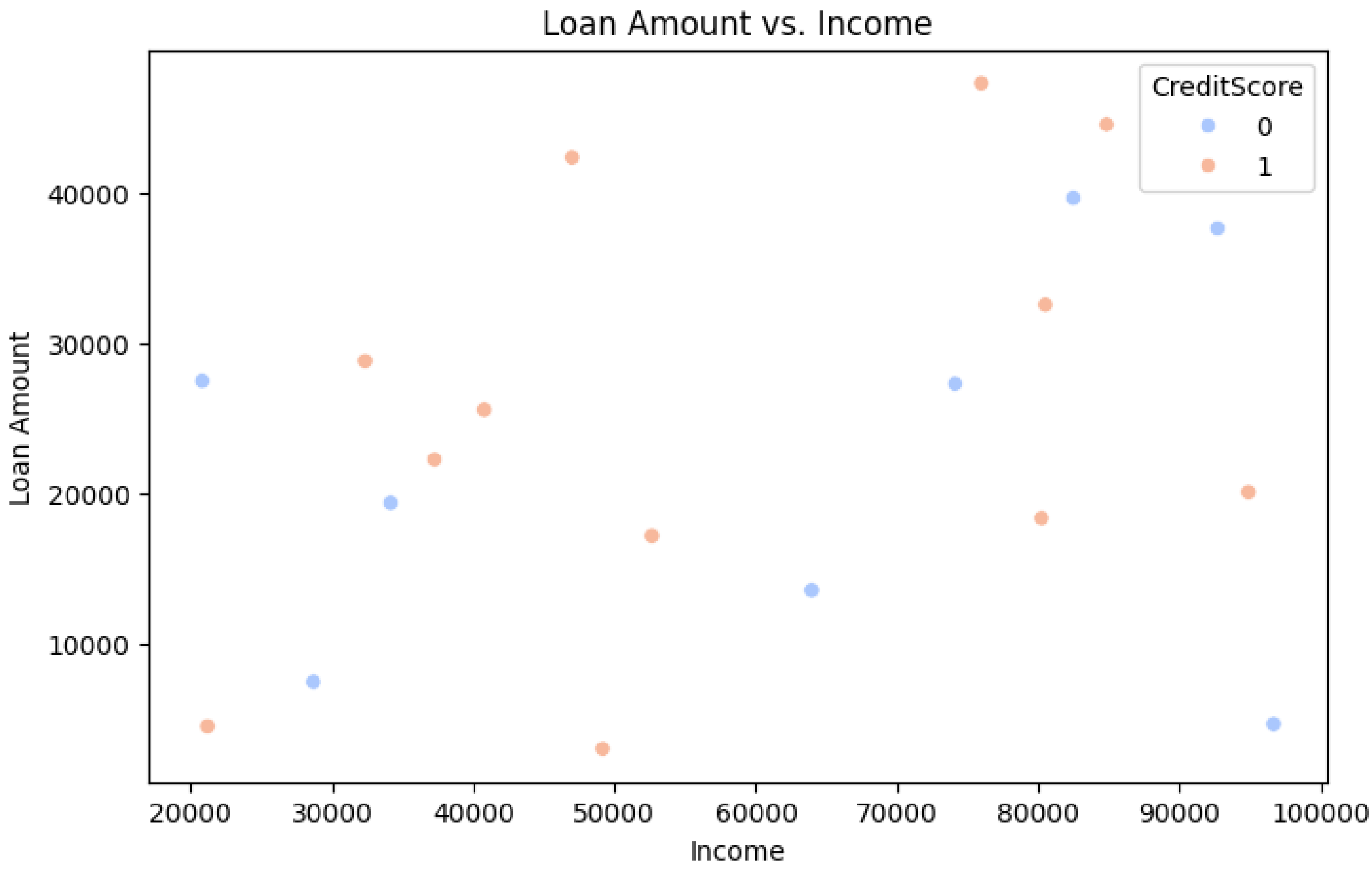
  - The correlation heatmap shows relationships between income, loan amount, and credit scores.

  - Credit score distribution helps understand class balance.

  - Financial trends indicate how income impacts loan amount and creditworthiness.




Feature Correlation Heatmap
Age
Income
LoanAmount
CreditScore
1.00
-0.21
-0.03
0.27
-0.21
1.00
0.31
-0.07
-0.03
0.31
1.00
0.13
0.27
-0.07
0.13
1.00
Age
Income
LoanAmount
CreditScore
1.0
0.8
0.6
0.4
0.2
0.0
−0.2


Loan Amount vs. Income
CreditScore
0
1
40000
30000
20000
10000
Loan Amount
20000
30000
40000
50000
60000
70000
80000
90000
100000
Income