Loan Default Prediction Project

AMAN SHARMA

Project Overview

- **Objective:**
- Develop a machine learning model to predict loan default risks.
- Provide insights to help financial institutions make data-driven decisions.

- **Dataset:**
- **Records:** 10,000 borrowers
- **Features:** Age, Income, Loan Amount, Credit Score, Interest Rate, etc.
- **Target Variable:** Loan_Status (Default or Non-Default)

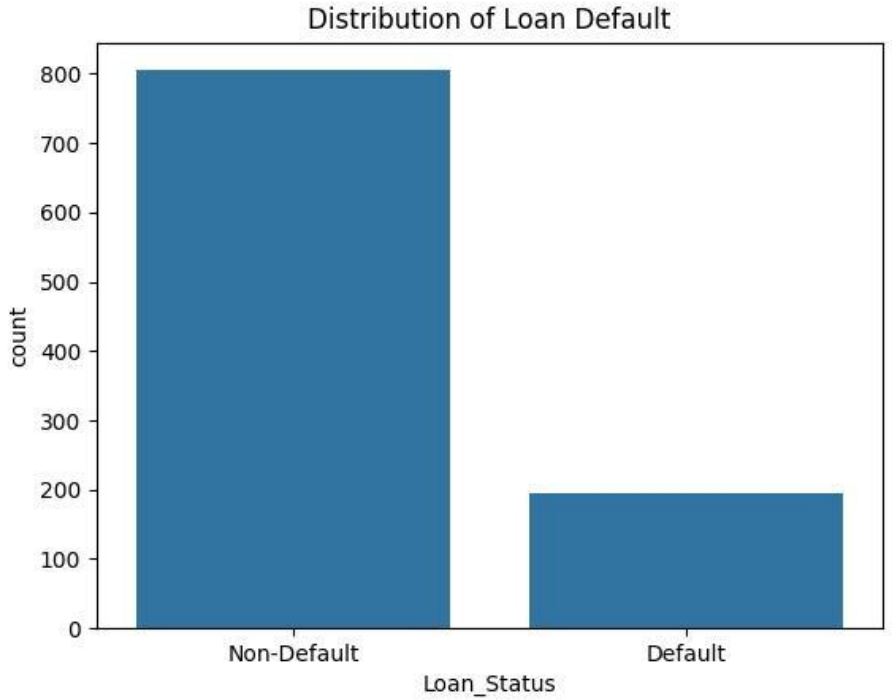
Problem Statement

- **The Challenge:**
- Loan defaults lead to significant financial losses.
- Traditional credit scoring methods are not always reliable.
- **The Solution:**
- Use machine learning to build a predictive model that identifies high-risk borrowers.
- Improve decision-making and reduce financial risk.

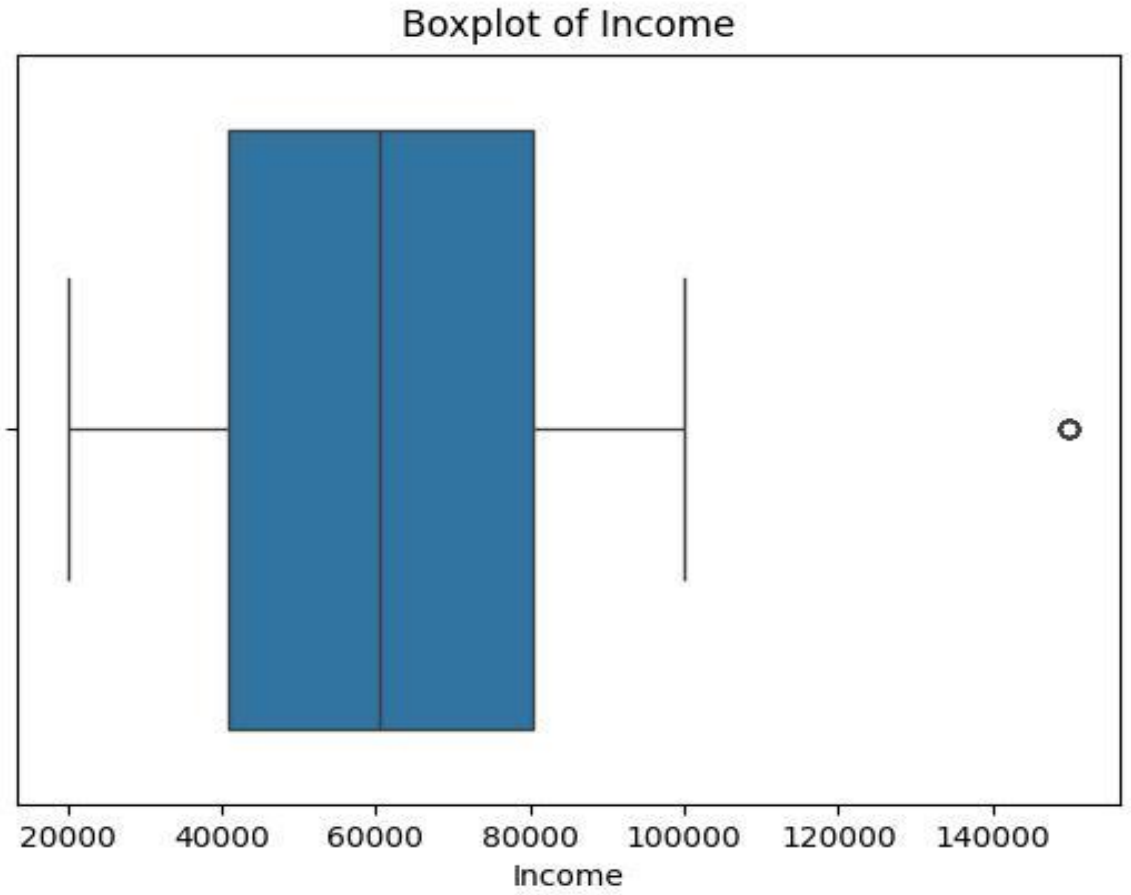
Exploratory Data Analysis (EDA)

Key Insights:

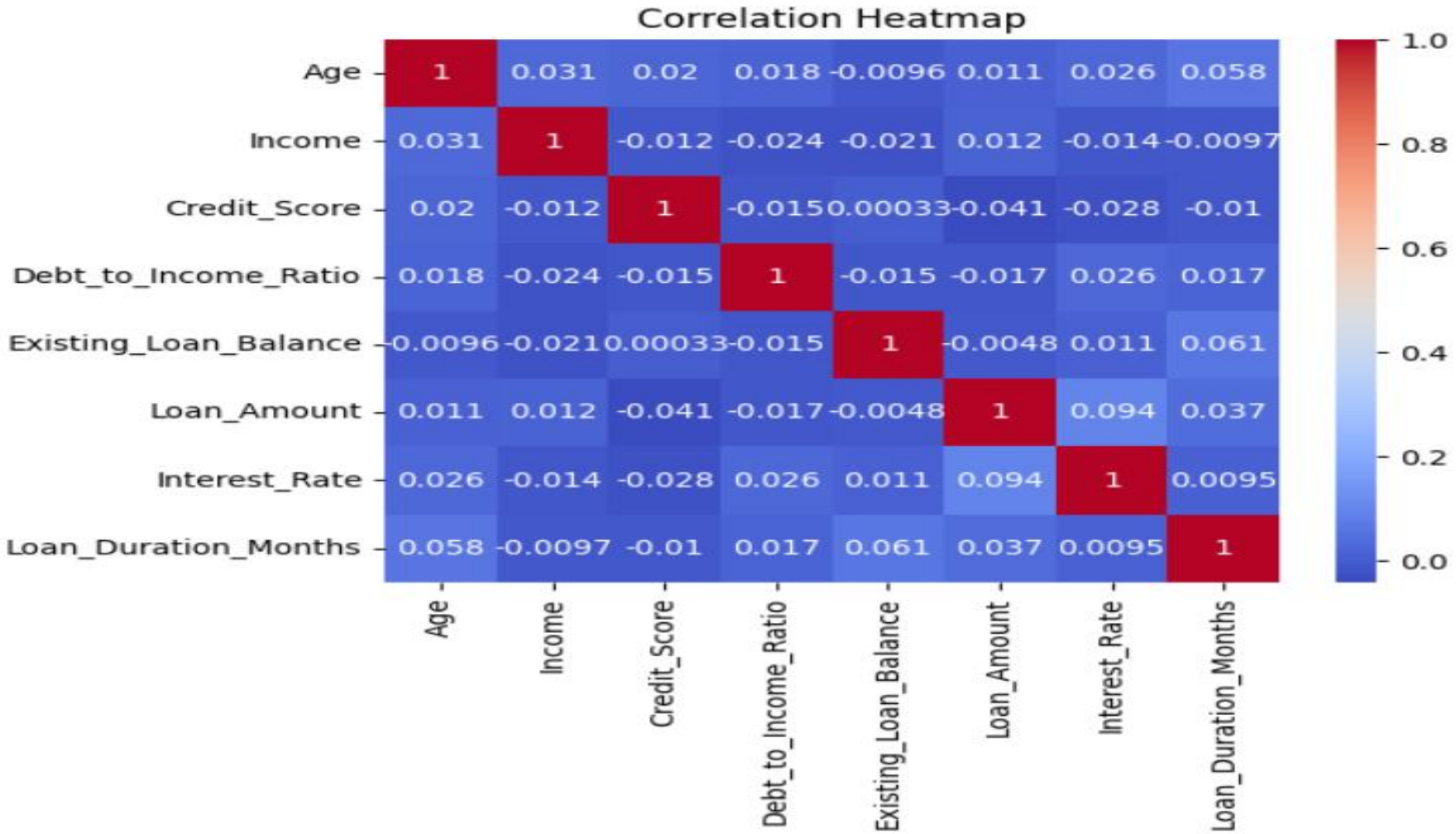
- Loan default rate is around **10%**, leading to class imbalance



- 'Income' shows significant outliers, as observed in the boxplot.



- Most features show weak correlations, indicating minimal multicollinearity, which reduces the need for feature elimination based on correlation alone.



Data Preprocessing

Steps Taken:

- Handled missing values in Gender and Employment Status.
- Encoded categorical variables.
- Scaled numerical features using MinMaxScaler.
- Addressed class imbalance using SMOTE .
- Remove Outliers using IQR method

Model Development

- **Approach:**
- Tested **Logistic Regression** and **Random Forest**.
- Selected **Random Forest** for the best performance.

Insights & Recommendations

- Lenders should **focus on high-risk borrowers** identified by the model.
- Adjust **interest rates dynamically** based on borrower risk.
- Implement **stricter approval criteria** for long-duration loans.

Model Evaluation

Performance Metrics:

Metric	Score
Accuracy	92%
Precision	89%
Recall	98%
F1-Score	93%
ROC-AUC	98%

Confusion Matrix: