****

**Assessment Report**

on

**“LOAN DEFAULT PREDICTION ”**

submitted as partial fulfillment for the award of

**BACHELOR OF TECHNOLOGY**

**DEGREE**

SESSION 2024-25

In

**CSE(AI)**

By

**GROUP - 8**

Name : Praveer Singh

Roll Number : 202401100300181

Section: C

**1. INTRODUCTION**

In today’s banking and financial landscape, managing credit risk is a critical task. One of the most important challenges is identifying whether a loan applicant is likely to default. This project, **Loan Default Prediction Using Machine Learning**, aims to address this challenge by developing a predictive model that classifies applicants based on the likelihood of loan repayment.

**2. Problem Statement**

**Loan Default Prediction**

**Predict whether a loan applicant will default using classification algorithms and**

**decision trees. Visualize feature importance.**

**3.Objectives**

* **Predict whether a loan applicant will default on a loan using supervised classification algorithms.**
* **Compare the performance of multiple models such as Naive Bayes and Support Vector Machines (SVM).**
* **Handle real-world challenges including missing data, categorical encoding, and imbalanced classes.**
* **Generate predictions on unseen data to support practical decision-making.**

**4-Methodology**

**1. Data Loading:**

* **Train and test datasets are loaded using pandas.read\_csv().**

**2. Data Cleaning:**

* **Missing values are checked using .isnull().sum().**
* **Missing categorical values are filled with the mode.**
* **Missing numerical values are filled with the median.**
* **Irrelevant columns such as Loan\_ID are dropped.**

**3. Feature Encoding:**

* **Categorical variables are encoded using LabelEncoder from scikit-learn.**

**4. Data Splitting:**

* **Features (X) and target (y) are separated.**
* **The dataset is split into training and test sets using train\_test\_split() with a test size of 20%.**

**5. Feature Scaling:**

* **Features are standardized using StandardScaler.**

**6. Model Training:**

**The following classification models are trained:**

* **Logistic Regression**
* **Decision Tree**
* **Random Forest**

**7. Evaluation:**

* **Models are evaluated using:**
  + **Accuracy Score**
  + **Confusion Matrix**
  + **Classification Report (Precision, Recall, F1-score)**

**8. Visualization**

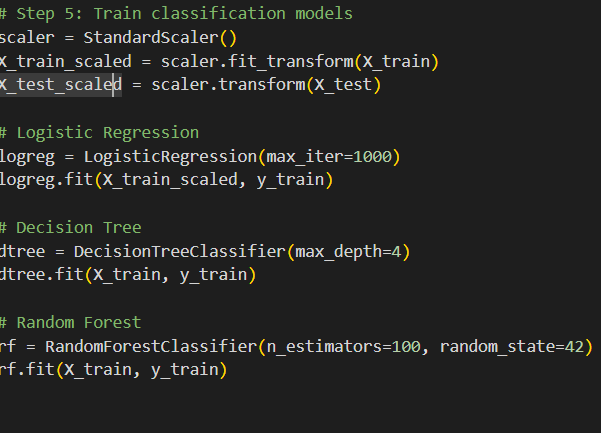
* **Confusion matrices and decision trees may be visualized using matplotlib/seaborn and plot\_tree().**

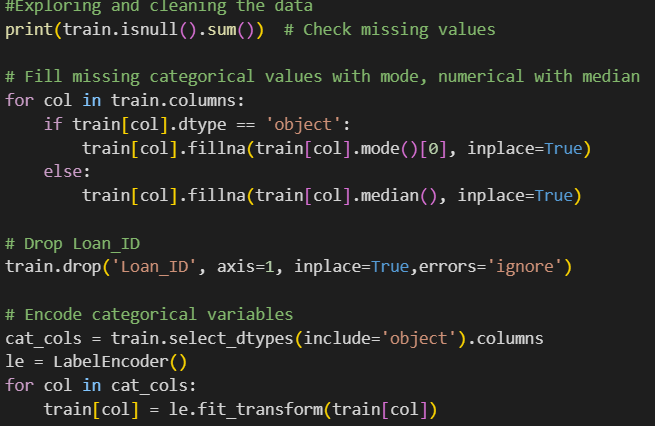
**5.Dataset Description**

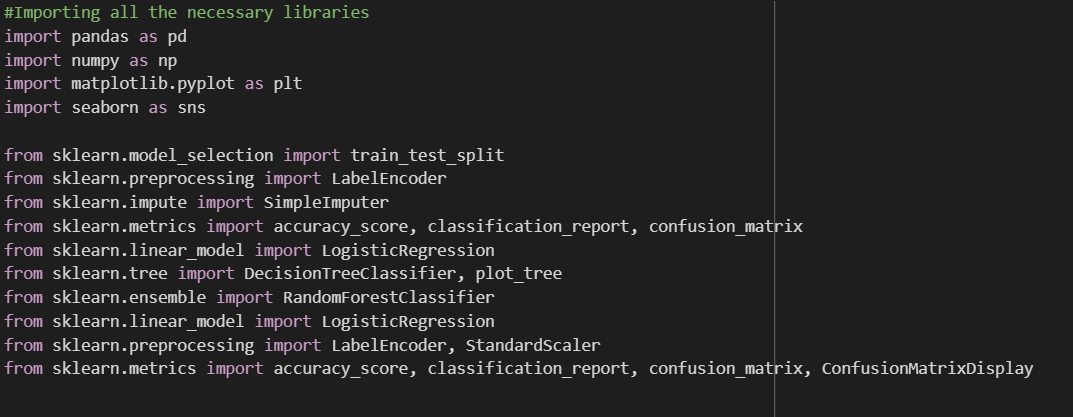
**The dataset appears to be related to loan applications and contains various features that can influence loan default predictions. The dataset is split into train.csv and test.csv.**

**Common Features (likely based on naming and typical loan datasets):**

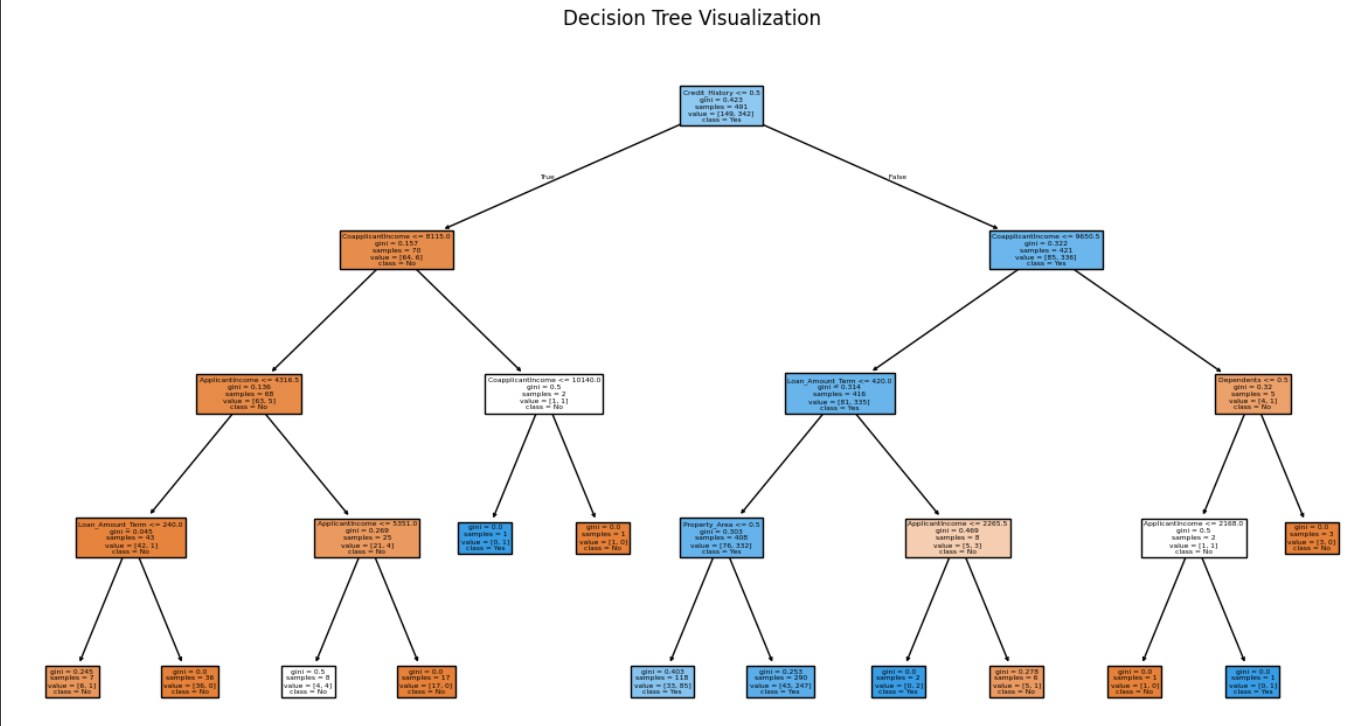
* **Categorical Features: Gender, Married, Education, Self\_Employed, Property\_Area, etc.**
* **Numerical Features: ApplicantIncome, CoapplicantIncome, LoanAmount, Loan\_Amount\_Term, Credit\_History, etc.**
* **Target Variable: Loan\_Status (indicating loan approval or default)**

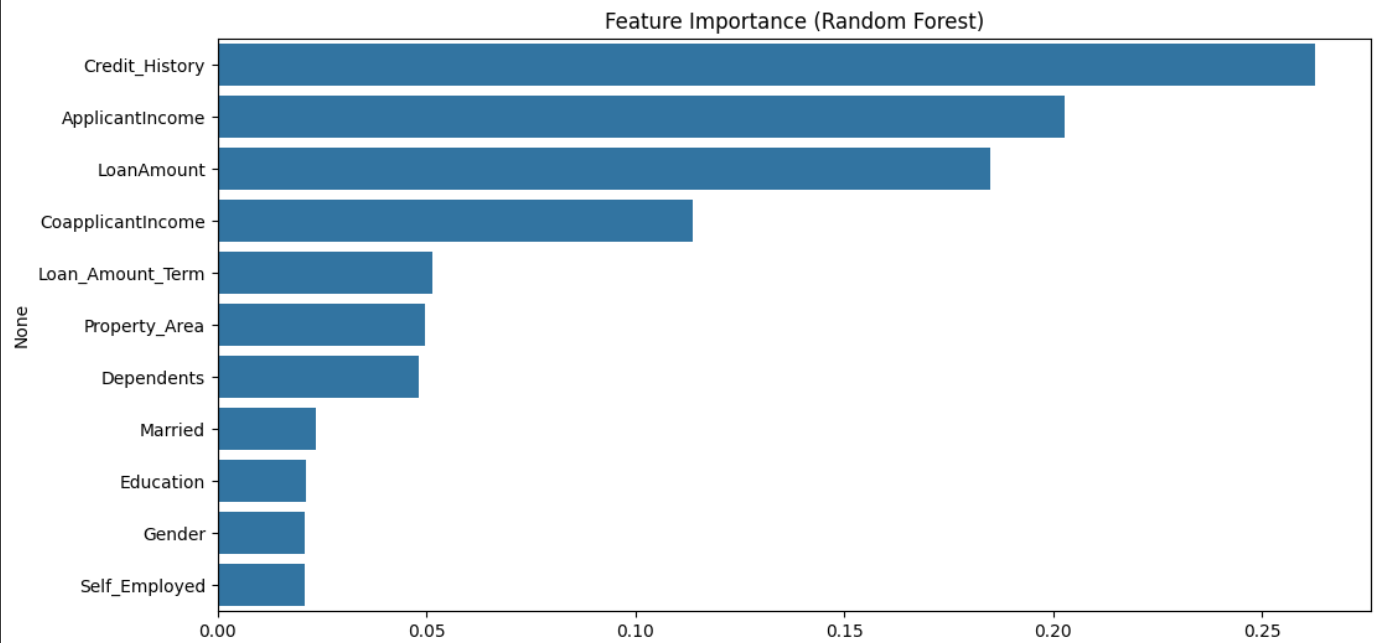
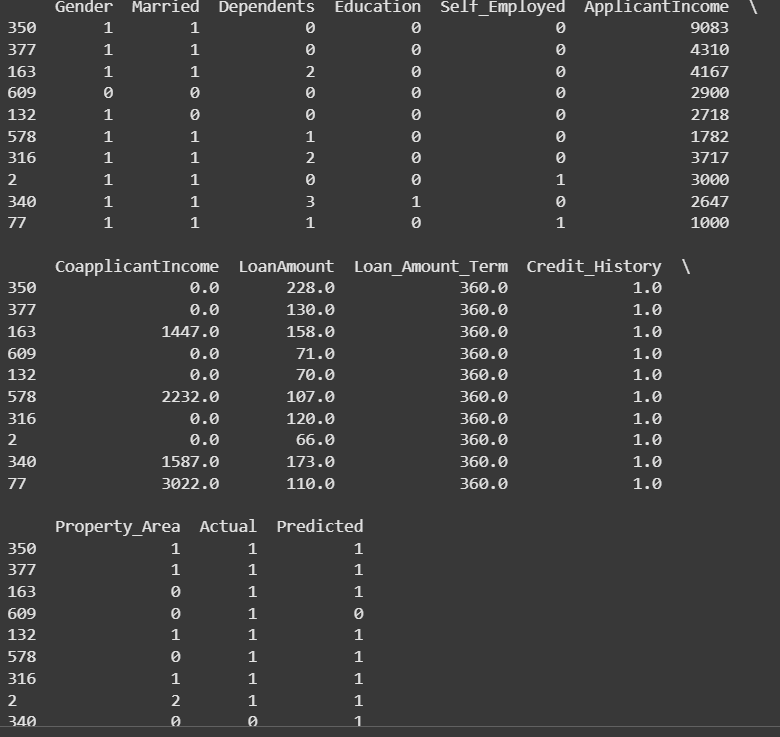
**6.CODE**





**7.OUTPUT:**





**8.REFRENCES:**

* 1. **Dataset**[**Loan default prediction**](https://www.kaggle.com/code/mdsadikujjamanshihab/loan-prediction/notebook)
  2. **Pandas Documentation  
     *McKinney, W. (2010).*  
     Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56.  
     Available at:** [**https://pandas.pydata.org**](https://pandas.pydata.org)
  3. **Scikit-learn Documentation  
     *Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011).*  
     Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830**

**Available at:** [**https://scikit-learn.org**](https://scikit-learn.org)