MAJOR PROJECT–2

LOAN APPROVAL PREDICTION USING CLASSIFICATION MODELS

1. INTRODUCTION

Loan approval is an important process in the banking sector. Financial institutions must evaluate loan applications carefully to reduce the risk of defaults. Manual evaluation can be time-consuming and inconsistent. Classification models help in predicting whether a loan should be approved or rejected based on applicant information.

This project focuses on predicting loan approval status using classification models by analyzing financial and personal attributes of loan applicants.

2. OBJECTIVE

The main objectives of this project are:

- To analyze loan applicant data

- To preprocess and clean the dataset

- To build classification models for loan approval prediction

- To compare model performance and select the best model

3. DATASET DESCRIPTION

The **Loan Approval Dataset** is used in this project. The dataset contains information related to applicants' income, loan details, credit score, assets, and employment status.

The dataset includes multiple records with both numerical and categorical attributes. The target variable is **loan_status**, which indicates whether a loan is approved or rejected.

4. ATTRIBUTE DESCRIPTION

The dataset contains the following attributes:

- **loan_id:** Unique identifier for each loan application

- **no_of_dependents:** Number of dependents of the applicant

- **education:** Educational qualification of the applicant

- **self_employed:** Employment status of the applicant

- **income_annum:** Annual income of the applicant

- **loan_amount:** Amount of loan applied for

- **loan_term:** Duration of the loan

- **cibil_score:** Credit score of the applicant

- **residential_assets_value:** Value of residential assets

- **commercial_assets_value:** Value of commercial assets

- **luxury_assets_value:** Value of luxury assets

- **bank_asset_value:** Value of bank assets

- **loan_status:** Loan approval status (target variable)

5. DATA PREPROCESSING

Data preprocessing was performed to prepare the dataset for classification. Column names were cleaned to remove extra spaces. Missing values were handled using suitable replacement methods. Categorical attributes were converted into numerical form so that classification models could be applied effectively.

6. MODEL BUILDING

Two classification models were implemented:

- Logistic Regression

- Random Forest Classifier

The dataset was split into training and testing sets using an 80:20 ratio.

7. MODEL EVALUATION

The models were evaluated using accuracy, precision, recall, and F1-score. Confusion matrix and ROC curve were also used to assess classification performance.

8. PERFORMANCE COMPARISON

Logistic Regression provided baseline classification performance. Random Forest Classifier achieved higher accuracy and better overall prediction results. Therefore, Random Forest was selected as the best model for loan approval prediction.

9. CONCLUSION

This project successfully demonstrated the use of classification models for predicting loan approval. Proper data preprocessing and model selection improved prediction accuracy. Random Forest Classifier produced more reliable results compared to Logistic Regression.