



Department of Computer Science and Engineering (Data Science)

NAME: **PARNA MISHRA**

SAP ID: **60009220201**

COURSE NAME: **Machine Learning - I Laboratory**

BATCH: **D2-1**

Mini Project

Task 4

(Model)

Title of Your Project: Predicting Loan Approval

Aim of the Project:

The primary objective of this project is to build a predictive model to determine the approval status of a loan application. This model will assist banks and financial institutions in making decisions regarding loan approvals based on applicant data, thereby reducing financial risks and enhancing decision-making efficiency.

Data Description:

The dataset comprises several attributes associated with loan applicants, as follows:

- **Loan_ID**: A unique identifier for each loan application (Categorical)
- **Gender**: Male or Female (Categorical)
- **Married**: Indicates if the applicant is married (Yes/No) (Categorical)
- **Dependents**: Number of dependents reliant on the applicant (Categorical)
- **Education**: Education level of the applicant (Graduate/Not Graduate) (Categorical)
- **Self_Employed**: Whether the applicant is self-employed (Yes/No) (Categorical)
- **ApplicantIncome**: Monthly income of the applicant (Numerical)
- **CoapplicantIncome**: Monthly income of the coapplicant (Numerical)
- **LoanAmount**: Total loan amount requested (Numerical)
- **Loan_Amount_Term**: Term of the loan in months (Numerical)
- **Credit_History**: Credit history meets guidelines (0/1) (Categorical)
- **Property_Area**: Urban, Semi-Urban, Rural (Categorical)
- **Loan_Status**: Loan approved (Y/N) (Categorical)



Department of Computer Science and Engineering (Data Science)

Data Preprocessing:

Handling Missing Values:

- Missing values in categorical data (like Gender, Married, Self_Employed) are imputed using the mode of the respective columns.
- Missing numerical values (such as LoanAmount) are imputed using median or mean, depending on the distribution of the data to avoid skewness.

Outlier Treatment:

- Outliers can significantly impact the model's performance, especially in regression models. Techniques such as logarithmic transformation or capping and flooring are used on skewed data such as ApplicantIncome and LoanAmount to normalize their distributions.

Feature Engineering:

- New features can be derived from existing data, enhancing the predictive power of the model. For example, TotalIncome might be created by summing ApplicantIncome and CoapplicantIncome.
- The Loan to Income ratio might be another derived feature to assess the repayment capacity of an applicant.

Encoding Categorical Variables:

- Categorical variables are converted into a format that can be provided to ML models. Techniques such as one-hot encoding or label encoding are used depending on the algorithm requirements.

Scaling Features:

- Numerical features are scaled using standardization or normalization to ensure that no variable dominates another due to its scale, improving the stability and performance of the learning algorithm.



Department of Computer Science and Engineering (Data Science)

Data Modeling:

The project evaluated multiple machine learning models including Logistic Regression, Decision Trees, Random Forest, and SVM. Each model was trained on the dataset and their hyperparameters were optimized using techniques like GridSearchCV to ensure optimal performance.

Performance Evaluation:

Model performances were compared using metrics such as accuracy, precision, recall, and F1 score. The models' ability to accurately predict loan approval was further analyzed using confusion matrices. The Random Forest model performed the best in terms of accuracy and overall balance between sensitivity and specificity, making it the model of choice for predicting loan approval.



Department of Computer Science and Engineering (Data Science)

FLASK DEPLOYMENT :

← → ↻ 127.0.0.1:5000 🔍 ☆ 📄 | 📄 🖨 ⋮

Loan Approval Prediction Form

Gender (1 for Male, 0 for Female):

Married (1 for Yes, 0 for No):

Dependents (Number):

Education (1 for Graduate, 0 for Not Graduate):

Self Employed (1 for Yes, 0 for No):

Applicant Income:

Coapplicant Income:

Loan Amount:

Loan Amount Term:

Credit History (1 for Good, 0 for Bad):

Property Area (0 for Rural, 1 for Urban, 2 for Semi-Urban):



Department of Computer Science and Engineering (Data Science)

Loan Approval Prediction Form

Gender (1 for Male, 0 for Female):

Married (1 for Yes, 0 for No):

Dependents (Number):

Education (1 for Graduate, 0 for Not Graduate):

Self Employed (1 for Yes, 0 for No):

Applicant Income:

Coapplicant Income:

Loan Amount:

Loan Amount Term:

Credit History (1 for Good, 0 for Bad):

Property Area (0 for Rural, 1 for Urban, 2 for Semi-Urban):



Department of Computer Science and Engineering (Data Science)

Loan Approval Prediction Form

Loan Not Approved

[Back to Form](#)



Department of Computer Science and Engineering (Data Science)

Loan Approval Prediction Form

Gender (1 for Male, 0 for Female):

Married (1 for Yes, 0 for No):

Dependents (Number):

Education (1 for Graduate, 0 for Not Graduate):

Self Employed (1 for Yes, 0 for No):

Applicant Income:

Coapplicant Income:

Loan Amount:

Loan Amount Term:

Credit History (1 for Good, 0 for Bad):

Property Area (0 for Rural, 1 for Urban, 2 for Semi-Urban):



Department of Computer Science and Engineering (Data Science)

Loan Approval Prediction Form

Loan Approved

[Back to Form](#)



Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)



Department of Computer Science and Engineering (Data Science)