Loan eligibility Prediction

# Submitted by:

Preeti Maan Shah (102217100)

BE Third Year CSE

Submitted to:

Dr. Anjula Mehto

Assistant Professor



## Computer Science and Engineering Department Thapar Institute of Engineering and Technology, Patiala

November 2024

TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| S. No | Topic | Page No. |
| 1 | Introduction or Project Overview | 3 |
| 2 | Problem Statement | 4 |
| 3 | Overview of the Dataset used | 5 |
| 4 | Project workflow | 6 |
| 5 | Results | 7 |
| 6 | Conclusion | 8 |

**Introduction or Project Overview**

Loan Eligibility Prediction is a classification problem where the goal is to determine whether a loan should be approved or not based on various applicant details such as income, credit history, employment status, and other factors. This project is critical for financial institutions to streamline their decision-making processes and reduce risks.

Key objectives:

* Automate loan approval decisions.
* Minimize manual errors in the evaluation process.
* Improve efficiency and customer experience.

**Problem Statement**

Financial institutions often face challenges in assessing the eligibility of applicants for loans. Traditional methods involve manual evaluations, which are time-consuming and prone to inconsistencies.

Problem Statement: To develop a machine learning model that predicts the loan approval status based on applicant details, thereby enabling financial institutions to make quicker and more reliable decisions.

**Overview of the Dataset used**

The dataset includes applicant information and the corresponding loan approval status. Below are the key features:

* **Gender**: Male or Female
* **Marital Status**: Married or Not Married
* **Dependents**: Number of dependents (e.g., 0, 1, 2, etc.)
* **Education**: Graduate or Not Graduate
* **Self-Employed**: Self-employed status (Yes or No)
* **ApplicantIncome**: Monthly income of the applicant
* **CoapplicantIncome**: Monthly income of the co-applicant (if any)
* **LoanAmount**: Loan amount requested
* **Loan\_Amount\_Term**: Loan repayment term in months
* **Credit\_History**: Credit history (1 for good credit, 0 otherwise)
* **Property\_Area**: Property location (Urban, Rural, Semiurban)
* **Loan\_Status**: Target variable (Y = Approved, N = Not Approved)

**Key Observations:**

* Missing values were identified and analysed, with their percentage calculated for each column.
* Visualizations were created to understand the distribution of categorical variables like Gender, Marital Status, and Education.

# Project Workflow

The workflow for the project was as follows:

1. **Data Collection**:
   * The dataset containing applicant details and loan approval status was loaded and inspected.
2. **Data Preprocessing**:
   * Missing values were analysed and handled.
   * Categorical variables were converted into numerical form using encoding.
   * Numerical features were normalized to ensure consistent scaling.
3. **Data Visualization**:
   * Count plots were created for categorical variables like Gender, Dependents, and Property Area to explore data distributions.
4. **Data Splitting**:
   * The data was split into training and testing sets, with 80% used for training and 20% for testing.
5. **Model Building**:
   * A **DecisionTreeClassifier** was implemented as the primary model for prediction.
   * Hyperparameter tuning was performed using **GridSearchCV** to optimize the model.
6. **Model Evaluation**:
   * The model's performance was evaluated using metrics such as:
     + Accuracy Score
     + Precision, Recall, and F1-Score
     + Confusion Matrix
7. **Prediction**:
   * The model was tested on new data to predict loan eligibility.

**Results**

**Model Performance**:

* **Accuracy Score**: The model achieved an accuracy of approximately 0.71 for Naive Bayes and 0.8571 for Random Forest

**Visualization**:

* A **Confusion Matrix** was generated to visualize the model's predictions, displaying the count of correct and incorrect classifications.
* Key features like Credit History, Applicant Income, and Loan Amount were identified as the most influential in predicting loan eligibility.

**Sample Predictions**:

* Examples of loan eligibility predictions for specific applicants were generated to demonstrate the model's application.

**Conclusion**

The Loan Eligibility Prediction project successfully demonstrates the application of machine learning to automate loan evaluation processes. The **DecisionTreeClassifier** provided reliable predictions based on applicant data, achieving high accuracy and reducing manual intervention.

**Key Takeaways**:

* Credit history, income levels, and loan amounts were found to be the most significant factors in loan approval decisions.
* Automating the process minimizes manual errors and improves decision-making speed and reliability.

**Future Work**:

* Explore alternative models like Logistic Regression or Random Forest for better performance.
* Use a larger and more diverse dataset to improve generalizability.





