

PROJECT REPORT

Loan Status Prediction Using Machine Learning

Submitted By: Shaurya Gundre

Internship Domain: Artificial Intelligence

Date: March 2026

Abstract

This project focuses on predicting loan approval status using machine learning techniques. The dataset contains applicant details such as income, education, credit history, loan amount, and property area. The goal of this project is to build a classification model that can predict whether a loan application will be approved or rejected.

Data preprocessing techniques were applied to handle missing values and convert categorical variables into numerical format. Logistic Regression was used as the machine learning algorithm for classification. The model achieved an accuracy of 78.86% on the testing dataset.

The results demonstrate how machine learning can assist financial institutions in automating loan approval decisions and improving efficiency.

1. Introduction

Machine Learning is a branch of Artificial Intelligence that enables systems to learn from data and make predictions or decisions without being explicitly programmed.

In the banking and financial sector, loan approval is an important and sensitive process. Banks must evaluate applicant information such as income, credit history, employment status, and property details before approving a loan.

Manual evaluation can be time-consuming and prone to human errors. Therefore, machine learning models can help automate the loan approval process, making it faster and more reliable.

2. Problem Statement

The objective of this project is to build a machine learning model that predicts whether a loan application will be approved or not based on applicant information.

Target Variable:

Loan_Status (Approved / Not Approved)

The problem is a binary classification problem.

3. Dataset Description

The dataset used in this project contains 614 loan application records with 13 features.

Important Features:

- Gender
- Married
- Dependents
- Education
- Self_Employed
- Applicant Income
- Co-applicant Income
- Loan Amount
- Loan_Amount_Term
- Credit_History
- Property_Area
- Loan_Status (Target Variable)

Total Records: 614

Total Features: 13

The target variable Loan_Status indicates whether the loan was approved or rejected.

4. Data Preprocessing

Before training the model, data preprocessing was performed.

Steps involved:

1. Checked for missing values in the dataset.
2. Missing categorical values were filled using the Mode (most frequent value).
3. Missing numerical values were filled using the Median.
4. Categorical variables were converted into numerical format using encoding techniques.
5. The dataset was split into training and testing sets using an 80:20 ratio.

Training Size: 80%

Testing Size: 20%

Data preprocessing is an important step to ensure the model performs accurately.

5. Machine Learning Model Used

Logistic Regression was used for this project.

Logistic Regression is a classification algorithm used when the target variable is binary. It predicts the probability of a class and assigns the output as either 0 or 1.

Reasons for choosing Logistic Regression:

- Suitable for binary classification
- Simple and efficient
- Performs well on structured datasets
- Easy to interpret

6. Model Training

The following steps were performed:

1. Imported required libraries such as Pandas, NumPy, and Scikit-learn.
2. Split the dataset using train_test_split.
3. Trained the Logistic Regression model using the training data.
4. Evaluated the model using the testing data.

The model was trained successfully without major errors.

7. Model Evaluation

The performance of the model was evaluated using:

- Accuracy Score
- Confusion Matrix
- Precision
- Recall
- F1-Score

Model Accuracy: 78.86%

Confusion Matrix Results:

- True Positives: 79
- True Negatives: 18
- False Positives: 25
- False Negatives: 1

The model performs well in predicting approved loans, while there is some scope for improvement in predicting rejected loans.

8. Results and Discussion

The Logistic Regression model achieved an accuracy of approximately 79%.

The results indicate that credit history and applicant income play significant roles in loan approval prediction.

Although the model performs well, it can be improved further by:

- Trying advanced algorithms
- Performing hyperparameter tuning
- Applying feature engineering techniques

9. Conclusion

This project successfully demonstrates the use of machine learning for loan status prediction.

The Logistic Regression model achieved an accuracy of 78.86%, showing that machine learning can assist banks in making faster and more accurate loan approval decisions.

The project highlights the importance of data preprocessing and model evaluation in building reliable machine learning systems.

10. References

- Scikit-learn Documentation
- Python Pandas Documentation
- Machine Learning course materials