Loan Applicant Default Report

Introduction

This project aims to predict whether a loan applicant will default using logistic regression. Loan default prediction is critical for financial institutions to reduce risk and make informed lending decisions. We use a dataset that includes information about applicants such as income, loan amount, credit history, and more.

Objective

To build and evaluate a logistic regression model that predicts loan default based on applicant data.

Dataset Description

The dataset includes the following features: - Gender, Married, Education, Self-employed, Property Area (categorical)- Applicant Income, Loan Amount, Credit History (numerical)- Loan Status (target variable - binary: 1 for default, 0 for no default)

Data source: Kaggle Loan Prediction Dataset

Steps Involved

1. Data Loading and Exploration

 2. Handling Missing Values

3. Encoding Categorical Variables

4. Feature Selection and Splitting Data

5. Model Building using Logistic Regression

6. Model Evaluation using Accuracy, Confusion Matrix, ROC Curve, and AUC

Model Summary

We used logistic regression to model the binary classification problem. The model predicts whether an applicant will default based on selected features

Evaluation Metrics

Accuracy Score: Measures overall correctness

- Confusion Matrix: Shows True/False Positives and Negatives

- Classification Report: Precision, Recall, F1-Score

- ROC Curve & AUC: Measures model's ability to distinguish between classes

Conclusion

The logistic regression model provides a useful baseline for predicting loan default. Features such as credit history, loan amount, and income significantly impact the probability of default. Further improvements could include using more advanced models and additional features like credit score