Loan Approval Prediction Using Logistic Regression

Project Title:

Loan Approval Prediction Using Logistic Regression

Overview:

This project aims to develop a logistic regression model to predict loan approval based on customer characteristics using the Loan Prediction Dataset from Kaggle. The model will help financial institutions make informed decisions by identifying key factors influencing loan approval.

Objectives:

1. To predict loan approval status using logistic regression.
2. To identify significant independent variables affecting loan approval.
3. To evaluate the model's performance using appropriate metrics.

Dataset:

- **Source:** Kaggle - Loan Prediction Dataset
- **Features:** Gender, Marital Status, Education, Income, Loan Amount, Credit History, etc.
- **Target Variable:** Loan Approval (Approved/Not Approved)

Methodology:

1. **Data Preprocessing:**
    - Handle missing values and outliers.
    - Encode categorical variables.
2. **Variable Selection:**
    - Select at least four independent variables: Gender, Income, Loan Amount, Credit History.
3. **Model Development:**
    - Build a logistic regression model to predict loan approval.
4. **Model Evaluation:**
    - Assess model performance using accuracy, precision, recall, and ROC-AUC curve.

Tools and Software:

- Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

Conclusion:

This project will provide valuable insights into the factors influencing loan approval, aiding financial institutions in making data-driven decisions. The logistic regression model will serve as a predictive tool to enhance loan approval processes.