Loan Approval Prediction System

Title of the Project

Loan Approval Prediction System

Objective

The objective of this project is to build a machine learning model that predicts whether a loan application will be approved or rejected based on applicant details.

The system helps financial institutions make quick and data-driven decisions, reducing manual efforts and improving accuracy.

Dataset Used

The dataset used for this project contains various features related to loan applications, such as:

- Number of Dependents

- Education Level

- Self-Employment Status

- Annual Income

- Loan Amount Requested

- Loan Term

- CIBIL Score

- Residential, Commercial, and Luxury Asset Values

- Bank Asset Value

- Loan Approval Status (Target Variable)

Model Chosen

Several machine learning models were implemented and compared to determine the best-performing one:

1. Linear Regression - For baseline analysis

2. Random Forest Classifier - For handling non-linear relationships

3. K-Nearest Neighbors (KNN) - For pattern-based classification

4. Decision Tree Classifier - For rule-based decision-making

5. Support Vector Machine (SVM) - For classification with optimal hyperplanes

After evaluating the models using various performance metrics, the best-performing model was selected and saved using joblib for future predictions.

Performance Metrics

To compare the models, the following metrics were used:

- Accuracy: Measures overall correctness of predictions.

- Precision: Ensures correctness of positive (approved) loan predictions.

- Recall: Measures how well actual approvals were identified.

- F1 Score: Harmonic mean of precision and recall.

- Confusion Matrix: Displays the distribution of true/false positives and negatives.

Challenges & Learnings

Challenges:

- Handling Categorical Variables: Proper encoding of education and self-employment status.

- Feature Scaling: Required normalization for optimal model performance.

- Imbalanced Data: Addressed using resampling techniques.

- Hyperparameter Tuning: Needed to optimize model accuracy and avoid overfitting.

Learnings:

- Importance of feature engineering in improving model accuracy.

- The effectiveness of ensemble methods (Random Forest performed the best).

- The significance of scaling and encoding for diverse datasets.

- How to deploy a Streamlit-based UI for real-time loan prediction.

This project successfully developed a machine learning-based Loan Approval Prediction System that can be utilized for better financial decision-making in the banking sector.