G H Patel College of Engineering & Technology

(The Charutar Vidhya Mandal (CVM) University)
Bakrol, Anand

COMPUTER ENGINEERING DEPARTMENT

Project Report

On

Loan Approval Prediction

Submitted By

Name of Student: Yash Sodhaparmar
Enrollment Number: 12202130501070

Name of Student: Devansh Srivastav
Enrollment Number: 12202130501074

Artificial Intelligence and Machine Learning (202046702)
A.Y. 2024-25 EVEN TERM

❖ OBJECTIVE

The primary objective of this project is to develop a machine learning-based system that can predict whether a loan application should be approved or rejected. In the financial industry, especially in banking, loan approvals are critical decisions influenced by numerous applicant attributes such as income, credit score, employment status, and asset ownership.

This system aims to automate and streamline the loan approval process using supervised machine learning techniques. It enhances decision-making accuracy, reduces processing time, and minimizes human bias. The system also forms the foundation for integrating intelligent financial analytics in real-time applications such as banking portals or loan management systems.

❖ DATABASE USED

The dataset used in this project is a structured loan application dataset with a total of **4,269 entries**. It includes information about individuals applying for loans and contains both numerical and categorical features such as:

- **Demographic & Financial Information**:
  - no_of_dependents, income_annum, loan_amount, loan_term, cibil_score
- **Asset Ownership**:
  - residential_assets_value, commercial_assets_value, luxury_assets_value, bank_asset_value
- **Categorical Features**:
  - education, self_employed
- **Target Variable**:
  - loan_status – A binary label (1 = Approved, 0 = Rejected)

The dataset was preprocessed to handle missing values, encode categorical data, and normalize numerical values to ensure optimal model performance.

❖ MODEL CHOSEN

Several machine learning models were trained and evaluated to determine the best performer for this binary classification task:

- **Logistic Regression**
- **Support Vector Machine Classifier**
- **Random Forest Classifier**

Among these, the **Random Forest Classifier** delivered the best results in terms of overall performance, interpretability, and generalization. It is an ensemble learning technique that aggregates multiple decision trees to reduce variance and avoid overfitting.

Hyperparameter tuning was applied with specific attention to:

- Number of estimators
- Max depth of trees
- Class weight handling (to address class imbalance)

❖ PERFORMANCE METRICS

The model performance was evaluated using standard classification metrics, which provide deeper insights than accuracy alone:

- **Accuracy**: 97.89%
- **Precision**: 0.98
- **Recall**: 0.98
- **F1-Score**: 0.98
- **ROC-AUC Score**: ~0.98

These metrics demonstrate a strong predictive performance and make the model suitable for real-world banking applications where false positives and false negatives both carry financial risk.

❖ CHALLENGES & LEARNINGS

🔻 Challenges:

1. **Class Imbalance**: The dataset was skewed with far more rejected applications than approved ones. Techniques like stratified sampling and class weighting helped mitigate this.
2. **Categorical Encoding**: Columns like education and self_employed needed label encoding, which was handled using LabelEncoder.
3. **Missing and Dirty Data**: Some columns contained inconsistent formatting and missing values which had to be cleaned and standardized.
4. **Model Overfitting**: Simpler models overfit on training data; hence ensemble models like Random Forest and XGBoost were more effective.
5. **Deployment Testing**: Integrating the model with a Streamlit-based web app required real-time input handling, scaling, and prediction pipeline integration.

🔻 Learnings:

- Gained deeper understanding of data preprocessing, especially when dealing with real-world financial data.
- Understood the trade-offs between different ML models in terms of interpretability and performance.
- Learned the importance of using the right metrics in imbalanced classification tasks.
- Acquired hands-on experience in deploying a predictive model as a user-friendly web application.

❖ CONCLUSION

This project successfully developed and deployed a machine learning model capable of predicting loan eligibility with high accuracy and reliability. By leveraging historical applicant data and using well-tuned ensemble methods, the system can assist banks in making faster, more consistent, and less biased loan approval decisions.

The project enhanced practical knowledge of ML development, evaluation, and deployment and stands as a foundation for more advanced AI-powered financial decision-making systems.