

Credit Risk Scoring (Loan Approval Prediction) Project

Project Overview

This project aims to **automate loan approval decisions** using machine learning by analyzing applicant data such as income, credit history, loan amount, and other key features. The goal is to **reduce manual processing time, minimize human bias, and improve approval accuracy** while maintaining transparency in decision-making.

Business Problem Addressed

Banks and financial institutions face challenges in **quickly and accurately evaluating loan applications**. Manual processing is **time-consuming, subjective, and prone to errors**.

This project solves this by:

- ✓ **Automating approval/rejection decisions** using ML models
- ✓ **Reducing bias** by relying on data-driven insights
- ✓ **Improving efficiency** by speeding up the decision process

Key Performance Indicators (KPIs)

1. **Model Accuracy** – How often the model predicts correctly (target: >85%)
2. **Precision & Recall** – Minimizing false approvals (high risk) and false rejections (lost customers)
3. **Feature Importance** – Understanding which factors most impact approvals (e.g., credit history, income)
4. **Deployment Readiness** – Ensuring the model works in real-world scenarios

Technical Approach

1. Data Preprocessing

- Handled missing values in Credit_History and LoanAmount
- Encoded categorical variables (Gender, Married, Education)
- Engineered new features like **Debt-to-Income Ratio**

2. Model Selection & Training

Tested multiple algorithms:

- **Logistic Regression** (Baseline)

- **Random Forest** (Robust & interpretable)
- **XGBoost** (Best performance for structured data)
- **KNN & SVM** (Benchmark comparisons)

3. Hyperparameter Tuning

Used **RandomizedSearchCV** to optimize:

- **XGBoost:** learning_rate, max_depth, n_estimators
- **Random Forest:** max_depth, min_samples_split
- **Logistic Regression:** C (regularization strength)

4. Model Evaluation

- **Best Model: XGBoost (~89% accuracy)**
- **Key Findings:**
 - **Credit History** is the strongest approval predictor
 - **Higher Income** applicants get approved more often
 - **Loan Amount vs. Income Ratio** impacts rejection rates

5. Model Deployment (Future Work)

- **API Integration** (Flask/FastAPI) for real-time predictions
- **Monitoring** for model drift & performance decay

Conclusion

This project successfully **automates loan approvals** while maintaining **high accuracy and fairness**. Next steps include **A/B testing against manual underwriting** and **deploying the model in a cloud environment** for scalability.

Final Deliverables:

- ✅ Trained ML models (XGBoost, Random Forest, Logistic Regression)
- ✅ Feature importance & business insights
- ✅ Python scripts for preprocessing, training, and predictions