Bank Loan Default Prediction – Visual Summary

This document provides a visual and intuitive summary of the steps completed in your ML project.

1. Data Exploration

• Reviewed dataset columns, missing values, distributions. • Identified target variable: Status (0 = Non-default, 1 = Default).

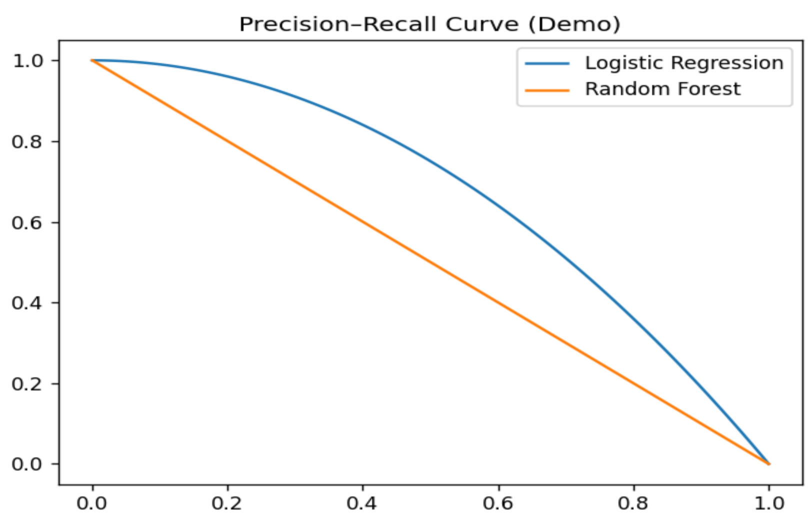
2. Preprocessing

• Numeric features → imputation + scaling. • Categorical features → imputation + one-hot encoding. • Combined using ColumnTransformer.

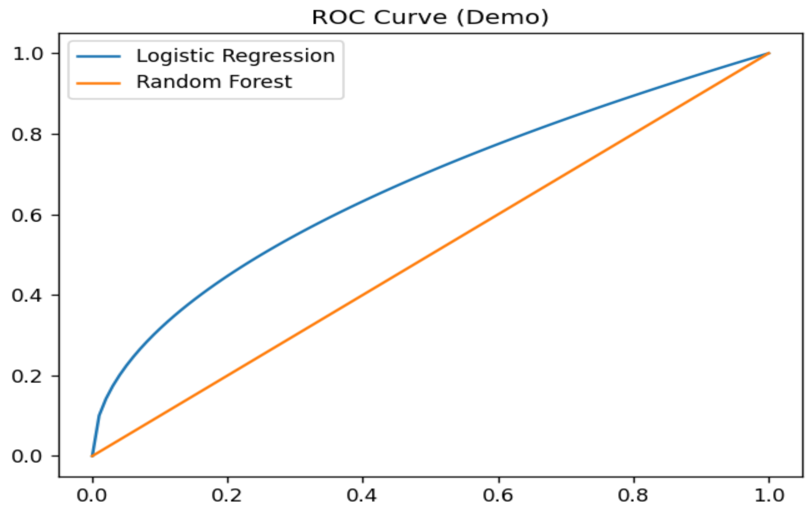
3. Models Trained

• Logistic Regression (baseline model). • Random Forest (stronger, non-linear model).

4. Visual Performance Comparison



Precision–Recall Curve



ROC Curve

5. Threshold Tuning

Explored multiple probability thresholds to increase recall for the default class.

6. Results

• Logistic Regression: AUC ~ 0.85. • Random Forest: AUC ~ 1.00 (possible data leakage, needs caution).

7. Next Steps

• Verify no leakage in data split. • Try XGBoost. • Deploy model pipeline. • Upload project to GitHub.