Loan Default Prediction Project – Model Analysis Report

1■■ Objective

The goal of this project is to predict whether a borrower will default on a loan based on various financial and demographic factors.

2■■ Dataset Overview

Source: loan_data.csv
Total Records: 9,578
Target Variable: not.fully.paid (1 = default, 0 = repaid)

Main Features Used:
credit.policy, int.rate, installment, log.annual.inc, dti, fico, days.with.cr.line, revol.bal, revol.util, inq.last.6mths, purpose.

3■■ Model Used

Algorithm: Decision Tree Classifier (Scikit-learn)
Decision Trees are easy to interpret and visualize, making them ideal for financial decision models.

4■■ Model Performance

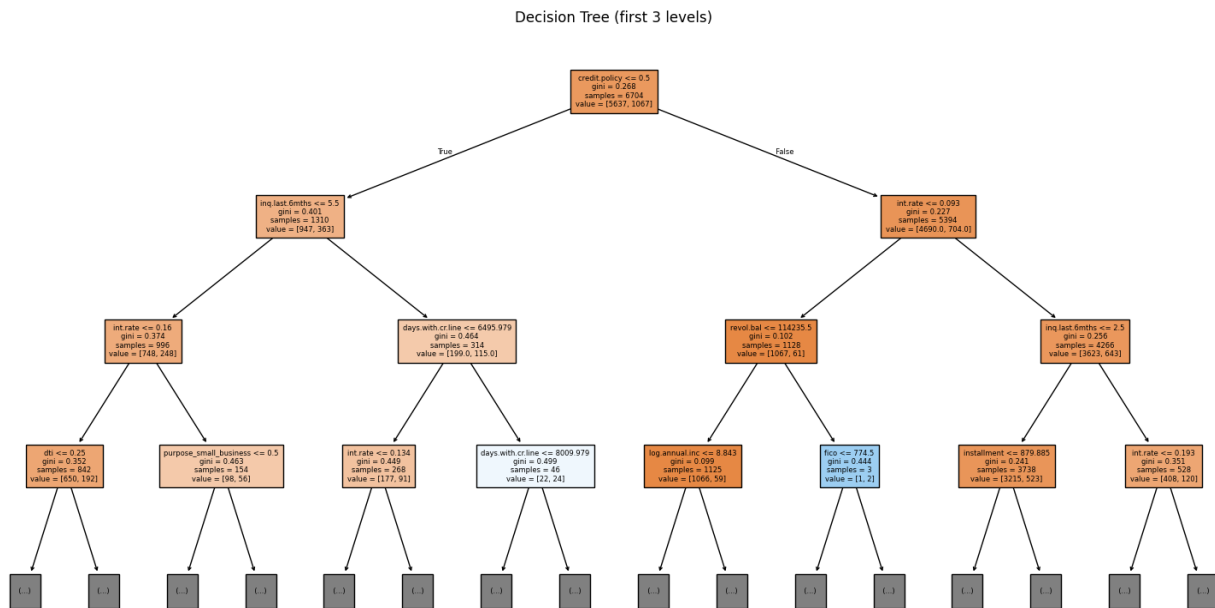**Accuracy:** 0.74
Precision (defaults): 0.20
Recall (defaults): 0.21
F1-score (defaults): 0.20

Interpretation: The model correctly predicts 74% of total cases but struggles to detect all defaults (low recall).
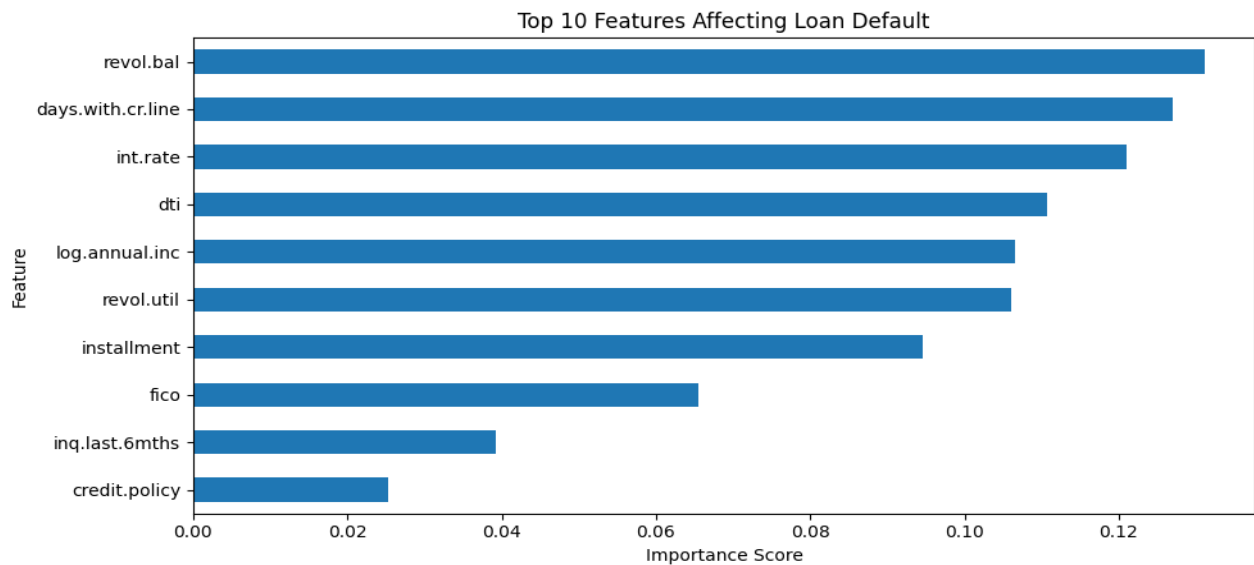
5■■ Decision Tree Visualization

The decision tree helps visualize how credit policy, interest rate, and balance affect default risk.

Decision Tree (first 3 levels)

6■■ Feature Importance

Top 10 features influencing loan default prediction.

Top 10 Features Affecting Loan Default

7■■ Recommendations

- Focus on applicants with high revol.bal and int.rate — apply stricter risk controls.
- Offer financial literacy or repayment plans for high-risk customers.
- Balance the dataset (e.g., SMOTE) in future versions to improve recall.
- Try Random Forest or XGBoost for better accuracy.

8■■ Conclusion

The Decision Tree model provides useful interpretability for understanding loan default behavior. While the accuracy is good (74%), recall can be improved for better detection of defaulters. The visual insights strongly align with real-world credit risk factors, proving the model's analytical value.