Predicting Default of Credit Card Clients

Priti Sagar, Devdeepsinh Zala, Dennis Zhuang

Problem Statement

1. Credit card defaults cause major financial losses to banks
2. High-risk customers need to be identified before issuing or renewing credit
3. A predictive ML model that prevent the risk

Why This Problem Matters

- Millions of customers increase banks' exposure to default risk.
- Traditional models rely on manual judgment and have limitations.
- Machine Learning captures non-linear patterns in the payment data.
- Banks can make better data-driven decisions using these models.

Dataset Source

Source

- UCI Machine Learning Repository
- 30,000 observations and 25 features

Features

- Demographics (SEX, EDUCATION, MARRIAGE, AGE)
- Credit Limit (LIMIT_BAL)
- Repayment status for 6 months (PAY_0…PAY_6)
- Bill amounts for 6 months (BILL_AMT1…BILL_AMT6)
- Previous payment amounts (PAY_AMT1…PAY_AMT6)

Data Preprocessing

Features

- No missing values after cleaning
  - Dropped ID column
- Categorical: SEX, EDUCATION, MARRIAGE, PAY_0–PAY_6
  - Nominal and ordinal encoding
- Numeric: Dominate the dataset (Bill amounts, Credit Limit, etc.)

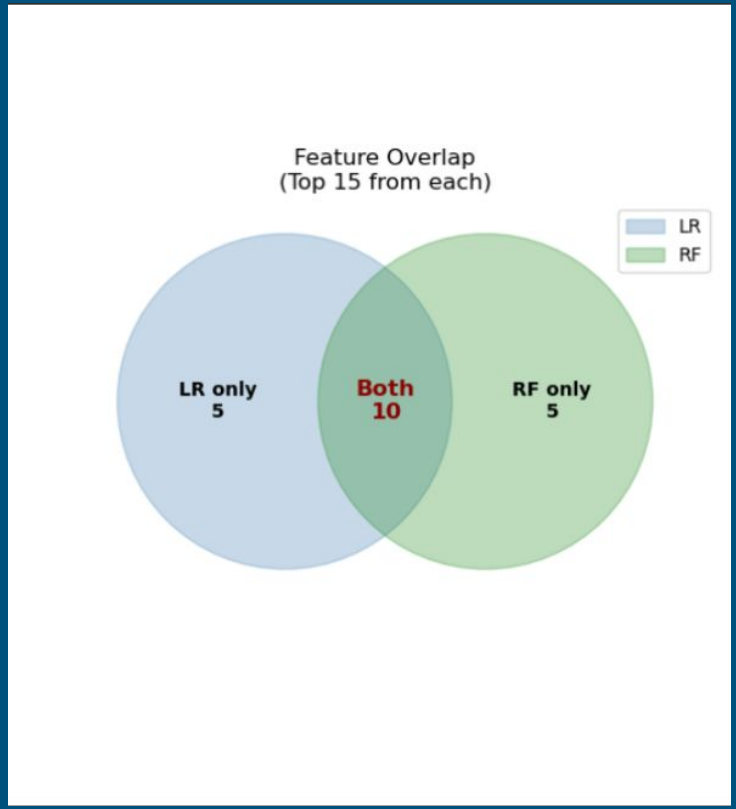
Data Preprocessing

Class Imbalance 78%-22%

- Class Weights

Feature importance

- 25 features => 10 features based on feature importance from Logistic Regression and Random Forest



Prior/Related work

Yeh & Lien (2009) – *“The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients”*

- Compared Logistic Regression, Decision Trees, Neural Networks, and SVMs on this exact dataset.
- Found that Neural Networks achieved the best probability-of-default accuracy.

Wahab et al. (2024) — *“A Comparative Study of Machine Learning Models for Credit Card Default Prediction”*

- Confirm that tree ensembles (Random Forest, XGBoost) generally outperform linear models.
- Logistic Regression remains popular due to interpretability.

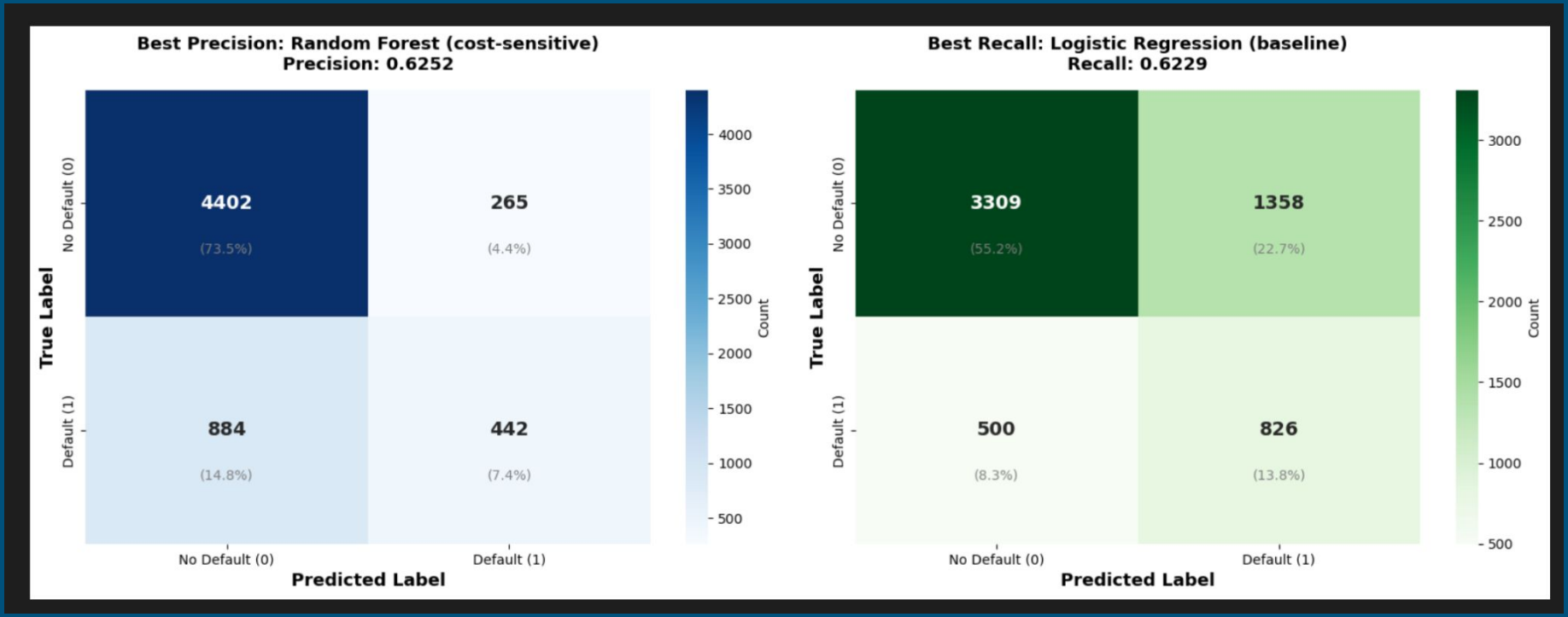
Our approach - Structured ML Pipeline

- Data Preparation
  - Clean, validate
- Model Training
  - Logistic Regression - Baseline
  - Decision Tree
  - SVM
  - Ensemble Methods
  - Random Forest
- Evaluations
  - Accuracy
  - Precision
  - Recall
  - AUC-ROC Curve
  - Confusion Matrix

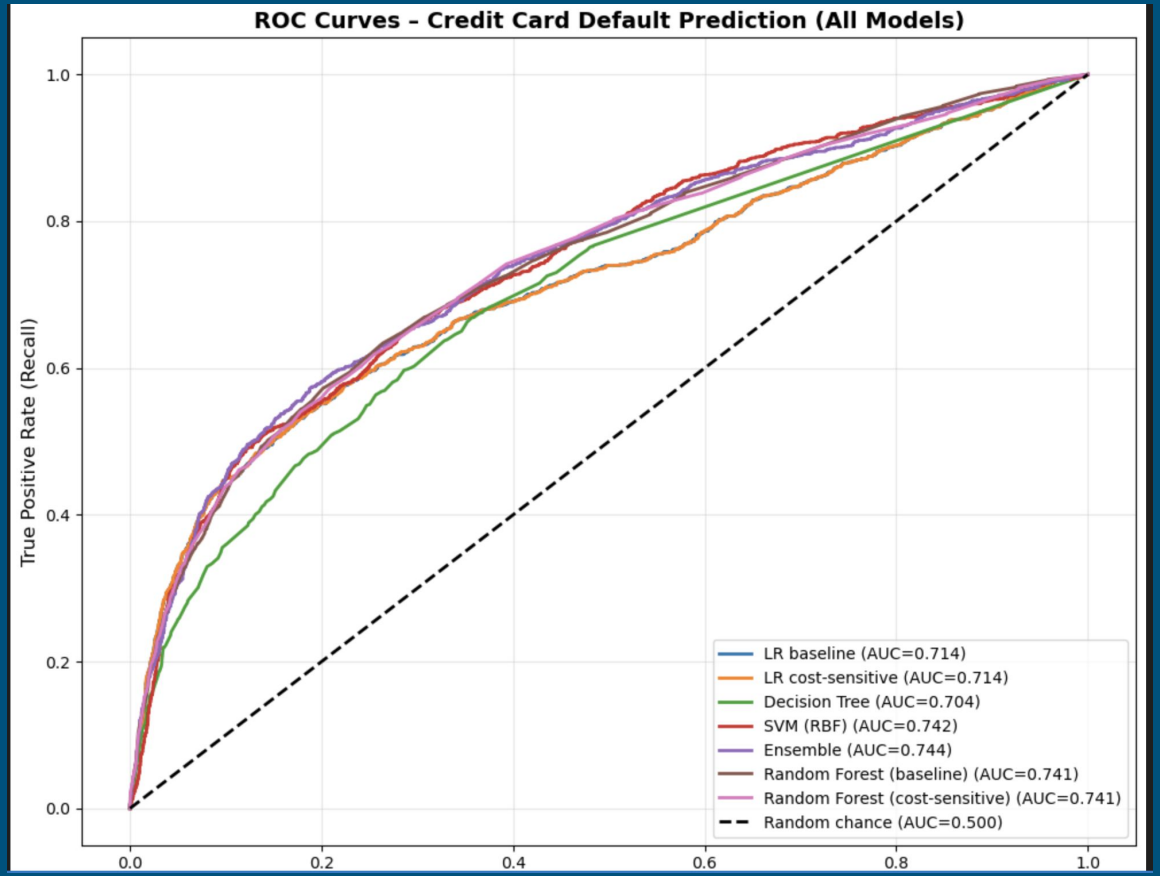
Model Comparison Results

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Comment
Logistic Regression	69.00%	37.82%	62.30%	47.07%	71.61%	Interpretable baseline model
Logistic Regression (Cost-sensitive)	76.20%	46.66%	52.03%	49.18%	71.36%	Using class weights increases sensitivity to defaults, improving recall at cost of accuracy but it aligns with financial risk management regulations and explainability
Decision Tree (Cost-sensitive)	69.67%	37.83%	56.94%	45.46%	70.42%	Captures non-linear relationships and produces balanced recall, but prone to overfitting, which limits AUC performance.
SVM	78.48%	51.39%	50.23%	50.80%	74.20%	Offers the best trade-off among single models—strong recall and the highest AUC—effective for complex decision boundaries in tabular data
Ensemble LR_cost Decision Tree SVM	80.06%	56.14%	45.17%	60.06%	74.35%	Combines strengths of all models; achieves the best overall balance of precision, recall, and AUC. Improved default detection while maintaining stability
Random Forest	80.83%	62.52%	33.33%	43.48%	74.07%	Lower recall implies higher missed defaults which is a bad sign for company

Confusion Matrix



Model Comparison Results



Business Implications

- Conservative Approach: Requires High Precision (>60%).
  - Random forest
- Aggressive Approach: Requires High Recall (>60%).
  - Logistic Regression
- Balanced Approach: F1-Score of 50-60%, ROC-AUC >0.70.
  - Ensemble (Logistic Regression, Decision Tree, SVM)

Challenges and resolution

- Data Issues: Class imbalance (addressed with cost-sensitive methods and class weights).
- Feature Engineering: Finding important features using Logistic regression and Random Forest
- Model Optimization: Hyperparameter tuning to improve model quality.
- Interpretability: Financial institutions require explainable models, a key consideration for our final ensemble choice.

Outcome and future extensions

Outcome

- Cleaned dataset
- Data report
- Machine Learning Pipelines
- Comparison of multiple models

Future Thoughts:

- Using different financial dataset test the generalizability
- Using time-series analysis for the Payment history data
- Incorporate SHAP explainability

Thank you

- Priti Sagar
  - pp693@drexel.edu
- Devdeepsinh Zala
  - dkz27@drexel.edu
- Dennis Zhuang
  - dz374@drexel.edu