Credit Card Default Risk Prediction

Objective

To predict credit card default risk using historical customer and payment data, enabling data-driven credit risk management decisions.

Data

- 30,000 credit card clients in Taiwan (2005)
- Demographics, credit limits, bill statements, and repayment history
- Target: default in the following month (22% default rate)

Approach

- Cleaned and standardized demographic categories
- Engineered financial risk features (payment delays, bill/payment aggregates, ratios)
- Trained and evaluated three models: Logistic Regression, Random Forest, XGBoost
- Evaluated using recall, precision, F1-score, and ROC-AUC

Key Results

- Logistic Regression achieved highest recall (70%) but low precision
- Random Forest achieved highest precision (65%) but low recall
- **XGBoost balanced recall (53%) and precision (48%) with strong ROC-AUC (0.75)**

Key Drivers of Default

- Recent payment delays
- Low payment-to-bill ratios
- Lower credit limits combined with poor repayment history

Recommendation

Push XGBoost as a risk scoring model to prioritize high-risk customers for proactive credit interventions. Adjust thresholds depending on the tolerance for false positives.