# **Credit Card Behaviour Score Prediction**

# The goal of this project is to build a binary classification model to predict whether a credit card customer will default on their payment in the next billing cycle . This forward-looking Behaviour Score enables the Bank to improve credit risk management by identifying potential defaulters early, supporting strategic interventions and exposure management. **Exploratory Data Analysis (EDA) & Financial Insights**

### **Key Variables & Patterns:**

# Target Imbalance: Only ~22% of customers are defaulters, indicating strong class imbalance that must be handled carefully.

# Payment History: PAY\_0 to PAY\_6 reveal that customers who default tend to have more payment delays and higher delinquency levels.

# Box Plots: Defaulters have lower payment amounts (Pay\_amt1 to Pay\_amt6), higher average bills, and elevated utilization ratios.

# Utilization Ratio: Strongly correlated with default risk; serves as a proxy for financial stress.

# Repayment Trends: KDE plots of avg\_pay show clear separation between defaulters and non-defaulters.

# Categorical Analysis: Singles show slightly higher default risk; education and gender are weak predictors.

### **Visualizations:**

# Count plots of default status, marital status, gender, education.

# Correlation heatmap and scatter plots (e.g., avg\_bill vs avg\_pay).

# Box plots of financial variables by default status.

# Time series bar plots of average payment delay.

# KDE and histograms of financial features.

# The EDA uncovered real behavioral patterns and financial stress signals, not just statistical trends.

# 

## **Feature Engineering:**

## To make the model financially interpretable and predictive, I created:

# avg\_bill, avg\_pay: 6-month average billing and repayment values.

# utilization: average bill / credit limit.

# pay\_to\_bill\_ratio: repayment to bill ratio over 6 months.

# num\_delinquent, max\_delinquency: counts of overdue payments and worst delay.

# These features simulate real-world credit scoring logic used in underwriting and monitoring.

## **Class Imbalance Handling**

# To address default class underrepresentation:

# Applied SMOTE (Synthetic Minority Oversampling Technique) after standard scaling.

# Balanced the dataset while preserving minority class characteristics.

## **Model Development & Performance**

# I evaluated multiple models:

# Logistic Regression (baseline, interpretable)

# Random Forest (robust ensemble)

# XGBoost (high-performing gradient boosting)

### **Evaluation Metrics:**

# ROC AUC: Separability of classes; good for imbalanced data.

# F2 Score: Prioritized metric — emphasizes recall over precision, which aligns with Bank A’s objective to minimize missed defaulters.

# XGBoost delivered the best trade-off between AUC and F2.

## **Threshold Tuning for Business Impact**

# Rather than using a default 0.5 probability threshold, I:

# Optimized the threshold to maximize F2 score, improving recall.

# Helped the bank detect more defaulters early, with acceptable false positives.

## **Final Prediction & Submission**

# Predictions were generated on an unlabeled validation dataset using the optimized model and threshold.

## **Conclusion**

# This project demonstrates how machine learning can support credit risk management with interpretable, business-aligned insights. By engineering behavior-aware features, handling imbalance, and tuning thresholds, I created a model that does more than predict, it enables actionable early warning for Bank.

# 