



# PREDICTING CREDIT CARD DEFAULTS

A Data Science Project for  
Risk Management

*Nicole Otieno*





# Overview

- Goal: Identify customers likely to default on credit card payments.
- Stakeholder: Commercial bank's Credit Risk Management Team.
- Approach: Predictive modeling using historical payment and demographic data.
- Outcome: Actionable insights to reduce financial risk.





# Business Problem

- Defaulting customers cause financial losses to banks.
- Early identification allows:
- Adjusting credit limits
- Offering payment plans
- Prioritizing monitoring of high-risk customers



# The Dataset

- Source: UCI Credit Card Default dataset (Kaggle)
- 30,000 customers, 25 features:
- Demographics: Age, Sex, Education, Marriage
- Credit/Bill info: Credit limit, Billing amounts, Past payments
- Target: Default next month (1 = default, 0 = no default)





# Modeling

- Iterative modeling process:
  1. Baseline Logistic Regression
  2. Tuned Logistic Regression
  3. Random Forest Classifier
- Metrics used: Recall (priority), Precision, F1-score, Accuracy
- Why: High recall ensures risky customers are identified



# Model Evaluation

Model	Recall	Precision	F1-Score	Accuracy
Baseline Logistic Reg	0.65	0.57	0.61	0.81
Tuned Logistic Reg	0.72	0.6	0.66	0.82
Random Forest	0.7	0.59	0.64	0.83

simplified model



# Key Insights

- Customers with late past payments are most likely to default.
- Lower credit limits increase default risk.
- Demographic factors (age, education, marriage) influence risk less than payment history.
- Random Forest confirms payment history as most important features.



# Limitations

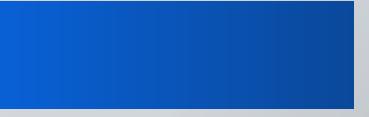
- Some risky customers may still be missed.
- Only historical payment and demographic data used; other behaviors not captured.
- Customer behavior changes over time → models need periodic updates.
- More complex models (Random Forest) are harder to interpret.





# Recommendations

- Deploy tuned logistic regression to monitor high-risk customers.
- Collect additional features (behavioral, external financial) for future improvements.
- Periodically retrain the model to capture changing customer behavior.
- Consider threshold tuning to optimize recall vs precision based on business priorities.



---

# THANK YOU

---

