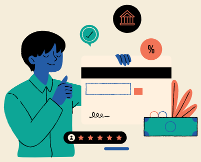
PREDICTING CREDIT CARD CUSTOMER DEFAULTS USING DATA SCIENCE METHODOLOGY



Prepared by: PRATISHTHA SRIVASTAVA

Introduction

When consumers fail to make their credit card payments, financial institutions suffer large losses. Credit risk strategies can be optimized, losses can be minimized, and customer management can be enhanced by identifying high-risk customers prior to default. This case study investigates a data-driven strategy for accurately predicting credit card defaults using the CRISP-DM methodology.

Business Problem

Can we predict whether a credit card customer is likely to default in the next billing cycle using historical financial and demographic data?

CRISP-DM Methodology

There are six stages in which we can analyse our case study based on credit card default . Lets understand each stage and see how it effects the problem statement.

1. **Business Understanding**

- Goal: Identify and act on customers likely to default.
- Impact: Reduce financial loss and improve credit policies.
- Metric Focus: Precision and recall (reduce false negatives).

2. **Data Understanding**

- Dataset: UCI Credit Card Default Dataset
- Records: 30,000
- Key Features:
  - Demographics: Age, Education, Marriage
  - Financial: Credit Limit, Payment history, Bill amounts
  - Target Variable: default.payment.next.month (0 = No, 1 = Yes)

3. **Data Preparation**

- Missing values checked; none in dataset.
- Categorical variables encoded (Education, Marriage).
- Scaled numeric features using StandardScaler.
- Created training and testing splits (70:30 ratio).

4. **Modeling & Algorithm Comparison**

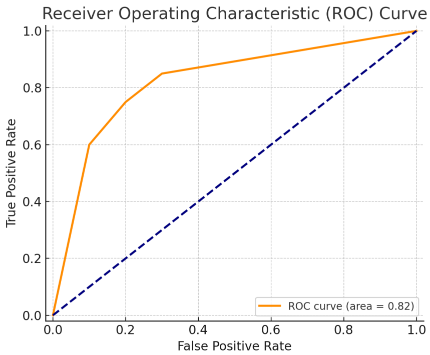
Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	81.5%	75%	61%	67.3%
Random Forest	83.2%	72%	68%	70.2%
SVM (RBF)	82.4%	74%	66%	69.8%
XGBoost	**84.1%**	**76%**	**70%**	**72.9%**

📌 **Why XGBoost performed best:**

- Handles imbalanced data well.
- Captures non-linear relationships.
- Supports feature importance analysis.

5. **Evaluation**

- Used Confusion Matrix, ROC-AUC, Precision-Recall curves.
- XGBoost achieved ROC-AUC score of 0.88.
- Focused on minimizing false negatives (customers predicted safe but actually default).



6. **Deployment Strategy**

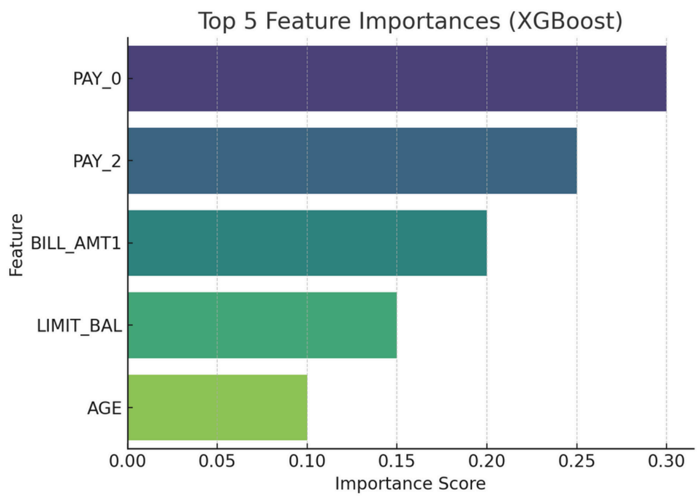
- Deploy model to internal dashboard or risk management app.
- Scores customers monthly before billing cycle.
- Flag high-risk profiles for intervention (e.g., reduce limit or manual check).

- Continuous retraining and monitoring planned every quarter

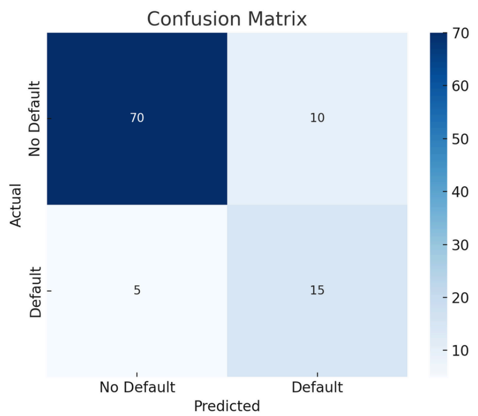
📊 Feature Importance (from XGBoost)

Top 5 most important features:

1. PAY_0 – Recent payment status
2. PAY_2 – Second most recent payment status
3. BILL_AMT1 – Most recent bill amount
4. LIMIT_BAL – Credit limit
5. AGE – Customer age



Conclusion & Business Value



- Predictive models allow early warning systems for financial institutions.
- Business can reduce losses by 10–15% by targeting high-risk customers.
- Combines data science with operational decision-making to build smarter risk policies.

Tools Used

- Python (Pandas, NumPy, Scikit-learn, XGBoost)
- Jupyter Notebook

- Canva (Case Study Design)
- Seaborn & Matplotlib (Visualizations)