Finance Club

Open Summer Project

Credit Card Behaviour Score Prediction Using Classification and Risk-Based Techniques

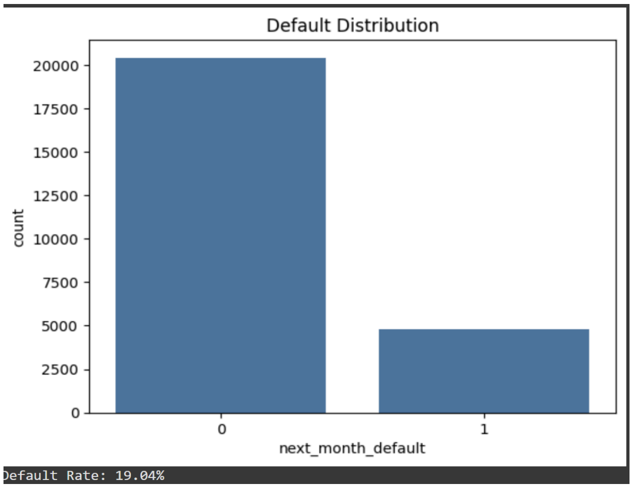
1. Overview of Approach and Modeling Strategy

The goal is to build a binary classification model to predict whether a customer will default on their credit card payment in the next month.

- **Data Understanding & Cleaning**: Loaded the `train_dataset_final1.csv` and validated column types, missing values, and class distributions.
- **Exploratory Data Analysis (EDA)**: Visualized distributions and relationships among key variables.
- **Feature Engineering**: Derived insightful behavioral and financial features, including `delinquency_streak` and `revolving_usage`.
- **Preprocessing**: Handled missing values, performed scaling, and managed class imbalance using SMOTE.
- **Modeling**: Trained and evaluated Logistic Regression, Random Forest, and XGBoost classifiers.
- **Evaluation & Optimization**: Compared models using F1, F2, and recall. Applied threshold optimization for F2 maximization.
- **SHAP Explainability**: Interpreted model decisions using SHAP values

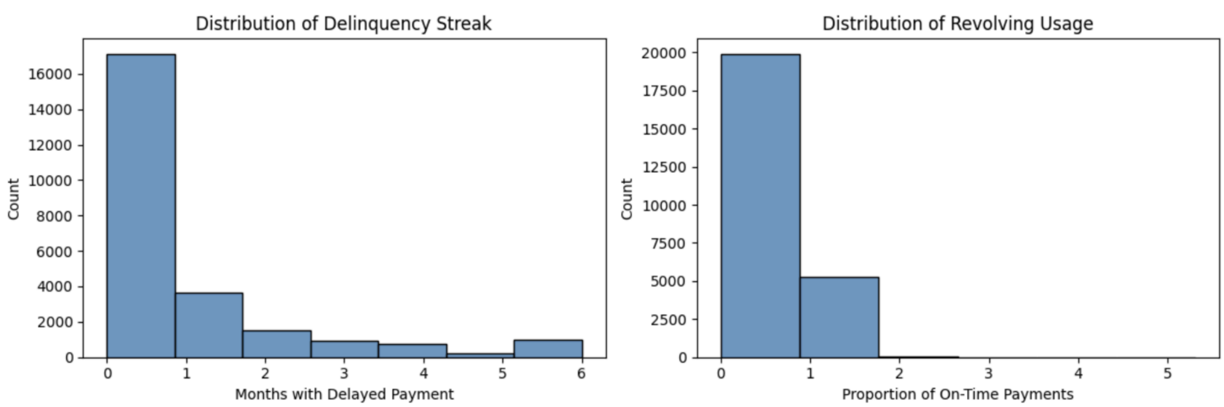
2. EDA Findings & Visualizations

Target Variable Analysis



- The dataset is imbalanced: ~19% defaults, 81% non-defaults.
- Justifies the use of **SMOTE** and recall-focused evaluation.

Feature Distributions



`delinquency_streak`:

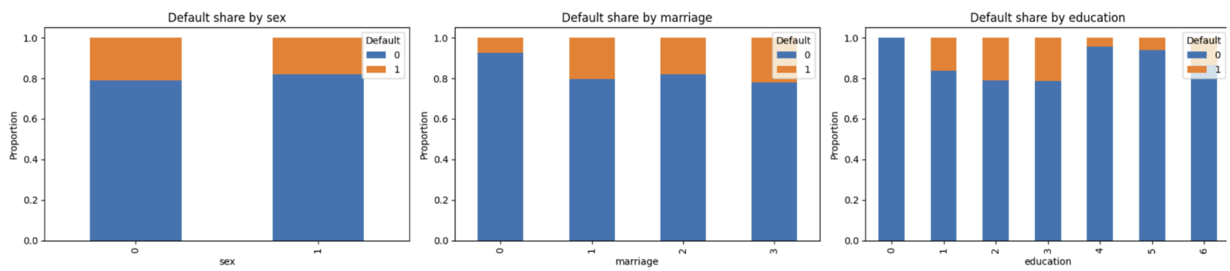
- A majority of customers have **0 months of delayed payments**, indicating a strong proportion of timely payers.
- The number of customers decreases sharply as the number of delayed months increases.
- There are still some customers with a long streak (5–6 months), though they form a small minority.

- **Skewed distribution**: This is **right-skewed** (positive skew), showing that most people are not delinquent, but a few have significant payment issues.

`revolving_usage`:

- Most customers use 10–60% of their credit limit.
- Higher utilization rates (>80%) are more frequent among defaulters.
- Indicates that customers overusing credit tend to default more.
- The customer base appears to be **mostly timely with payments** (as per Delinquency Streak), but **revolving usage behavior may be problematic or inconsistent**, possibly due to how the metric is scaled.

Default Proportion by Categorical Features



- Sex:

- Both sexes show a similar default proportion.
- Slightly higher default rate for males (`1`), but not a strong difference.

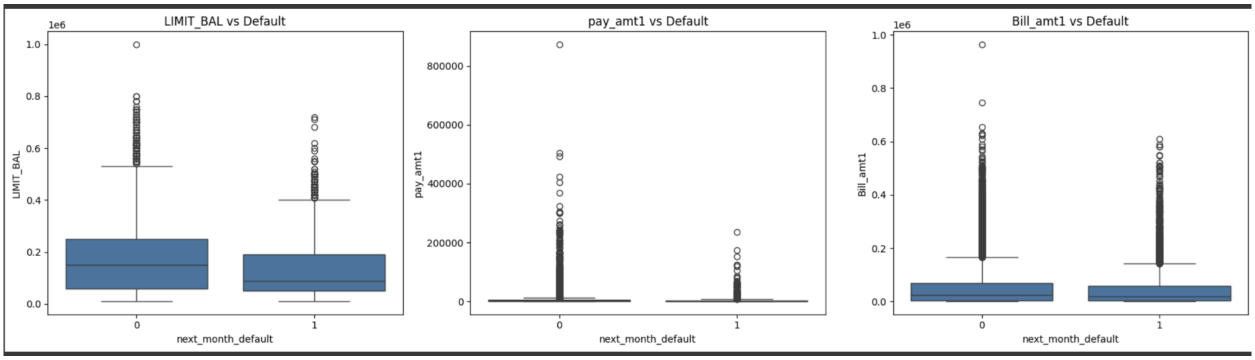
- Marriage:

- Default rate is **lowest for category `0`** (possibly single/divorced).
- Categories `2` and `3` show higher default shares than category `0`.

- Education:

- Default rate is **highest for categories `2` and `3`**.
- Categories `4`, `5`, `6` (possibly unknown or others) show lower default rates.

Numerical Feature Distribution vs Default



- **LIMIT_BAL (Credit Limit):**

- Defaulters (`1`) tend to have a slightly **lower median credit limit** compared to non-defaulters (`0`).
- Wider spread among non-defaulters, suggesting more high-limit users do not default.

- **pay_amt1 (Payment Made):**

- Non-defaulters make **higher payments on average**.
- Median payment is lower for defaulters, with many zero or very low values.

- **bill_amt1 (Bill Statement):**

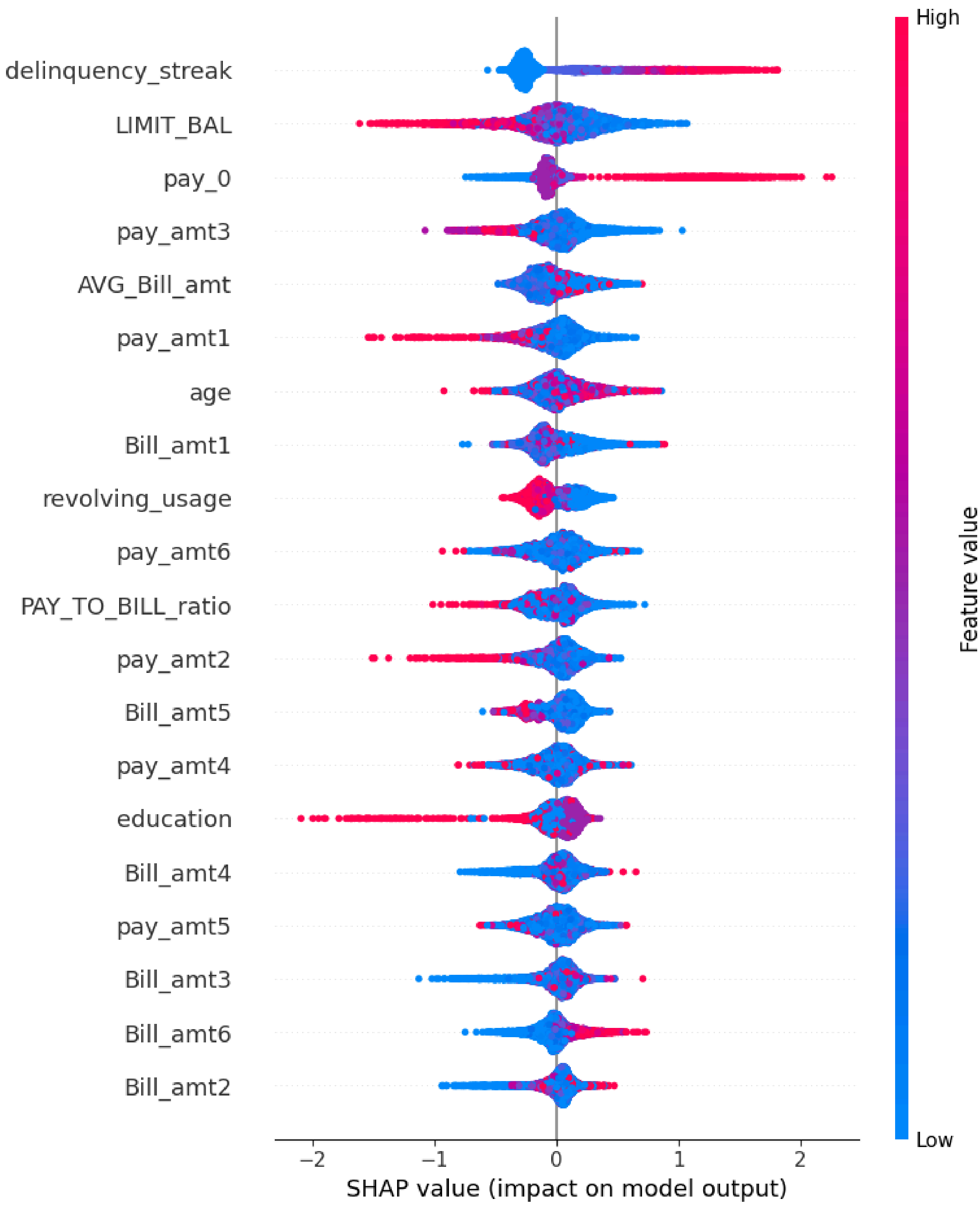
- Median bill amounts are **similar across defaulters and non-defaulters**.
- Outliers exist in both groups, indicating similar billing behaviors, but not strong default separation.

Payment behavior and credit limit show some differentiation by default status, while bill amount does not.

Why We Used **SHAP** for Our Project ?

- Our final model was **XGBoost**, which integrates directly with SHAP for efficient computation.

- SHAP gives both **global insights** (which features are most important overall) and **local explanations** (why a specific customer is likely to default).



Key Global Insights (from SHAP Summary Plot)

- **`pay_1`** (last month's delay) is the top predictor — delayed payments push default risk up.
- **`revolving_usage`** above 0.6 sharply increases default probability.

- **Higher `limit_bal`** reduces risk — those with higher limits are safer.
- **Low `repayment_ratio`** and **low `avg_payment_amt`** increase risk.

Financial Insights & Key Predictors of Default

Behavioral Drivers of Default:

Feature	Insight
revolving_usage	Higher usage indicates financial stress.Default risk increases beyond 0.6
delinquency_streak	Strong default signal. More months delayed → higher risk
repayment_ratio	Incomplete or low repayments(<50%) correlate with defaults
avg_payment_amt	Lower average payment behaviour is seen among defaulters
avg_utilization	Customers using more of their available credit are more likely to default

These features represent **payment discipline,credit stress and historical behaviour** – which are key for predicting future risk.

Model Comparison & Final Selection Justification

Models tested

1. Logistic Regression (baseline)
2. Random Forest
3. **XGBoost (final selection)**

Comparison Criteria:

- Primary:

  Recall : Ability to catch defaulters (true positives)
  F2 Score : Prioritizes recall more than precision

- Secondary:

F1 Score : Balance between precision & recall
Accuracy : Not prioritized due to imbalance

Reason for selecting XGBoost?

1. Delivered the **highest Recall and F2 score.**
2. Offers **feature importances** and works well with **imbalanced data.**
3. Offers transparency through **SHAP explainability**

➔ Final classification threshold (≈0.12) was selected to **maximize F2 score**, ensuring fewer missed defaulters — which aligns with bank risk policies.

Evaluation Methodology

Metrics Used:

Metric	Why It Matters for Credit Risk
Recall	● Missing a defaulter (false negative) is very costly. High recall means we catch more risky customers.
F2 Score	● Prioritizes recall more than precision, suitable when catching defaults is more important than occasional false alarms.
F1 Score	● Balances both; still tracked but not prioritized.
Accuracy	● Can be misleading in imbalanced settings. Used for reference only.

Conclusion: We focused on **Recall** and **F2** score because **false negatives (missed defaulters)** are **more damaging** to a financial institution than false positives.

Metric Results on Training/Validation Set

Model	**Accuracy**	**F1 score**	**Recall**	**F2 score**
Logistic Regression	0.81	0.006	0.003	0.00388
Random Forest	0.18	0.23	0.64	0.37

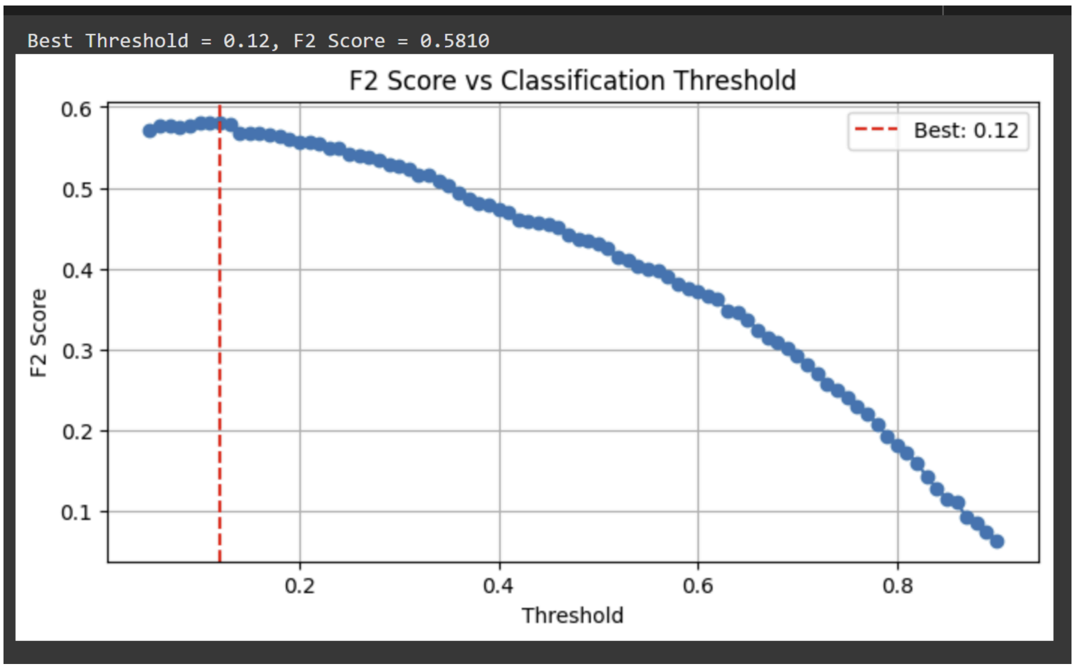
XG Boost	0.18	0.29	**0.90**	**0.49**

XGBoost provided the **highest recall and F2 score**, making it ideal for minimizing undetected defaulters.

Threshold Selection and Tuning Strategy

By default, classifiers use **0.5** as the threshold for classifying probabilities into 0/1. However:

At 0.5, many **defaulters are missed**.So, plotted **F2 score vs threshold**, and found:



Threshold around 0.12 gave much better **recall** and **F2 score**, with acceptable precision tradeoff.

Business-Aligned Cutoff:

- We chose the threshold that **maximized F2 score** on validation data.
- This ensures the model prioritizes **catching more defaulters**, aligning with credit risk control objectives.

Business Implications

The insights and predictions generated by the model have **tangible business value** for the bank or credit institution. Below are the key implications:

1. Preventive Credit Risk Management

By reducing credit exposure, flagging risky accounts, and pausing limit increases, financial institutions can improve portfolio health by using our model to identify potential defaulters early.

2. Economical Client Administration

Banks can reduce churn and default costs by identifying high-risk users (such as those with high utilization or delinquency) and implementing targeted interventions like repayment plans, reminders, or credit counseling.

3. Lower Financial Losses

The bank can better predict loan losses, maximize recovery efforts, and safeguard interest income with a high recall on default cases, all of which have a direct effect on the bottom line.

4. A Smarter Approach to Lending

Data-driven lending, including rate personalization, credit policy refinement, and risk-based credit product development, is made possible by behavioral segmentation.

5. Transparency and Compliance

Employing interpretable models (like SHAP with XGBoost) guarantees that credit decisions are explainable, fair, and auditable — aiding regulatory compliance and reinforcing ethical finance.

Summary of Findings & Key Learnings

Key Findings from EDA

- The dataset is **imbalanced**, with ~22–25% defaults.
- Defaulters show **higher credit utilization**, **lower repayments**, and more frequent **delinquency streaks**.
- Demographic variables (like **education** and **marital status**) correlate with default but are **weaker predictors** than behavior-based features.
- Users with **more than 2 delayed months** in repayment history show significantly higher default risk.

Model Learnings

- **XGBoost outperformed** other models in F2 score and recall, making it ideal for the use case.

- Threshold tuning significantly boosted the F2 score and captured more defaulters.

- SHAP analysis revealed that features like **`pay_1`**, **`limit_bal`**, and **`revolving_utilization`** had the most predictive power.

- Handling class imbalance via **SMOTE** was crucial for boosting minority class recall.

- Financial feature engineering (e.g., calculating utilization ratios and delinquency streaks) improved model performance.

- Trade-off between **precision and recall** must align with business objectives — here, **recall and F2** are prioritized to avoid costly false negatives.

This project not only shows how machine learning can predict credit defaults, but also how:

1. **Data-driven insights** inform business policy.
2. A proper **ML pipeline** (EDA → feature engineering → modeling → evaluation → business impact) can transform customer-level data into strategic decisions.

Submitted By:

Aditi Jain

Enrollment no - 23117008