Credit Card Behavior Score Prediction – Final Report

Introduction

In the modern financial ecosystem, credit cards are a cornerstone of consumer transactions, providing convenient short-term credit. However, they carry the inherent risk of customer default, where individuals fail to repay their monthly dues, posing significant financial risks to banks and credit institutions. This project aimed to develop a binary classification model to predict whether a customer will default in the next billing cycle based on historical behavioral data. The target variable, `next_month_default`, is defined as:

$$\text{next\_month\_default} = \begin{cases} 1 & \text{Customer defaults next month} \\ 0 & \text{No default occurs} \end{cases}$$

The objective was to enable financial institutions to proactively identify high-risk customers, thereby mitigating potential losses through early intervention.

Objective and Approach

The primary goal was to build a robust model to predict credit card defaults, prioritizing the identification of defaulters to minimize financial losses. We employed a data-driven approach, leveraging historical customer data to uncover patterns associated with default risk. The process involved thorough data preprocessing, exploratory data analysis(EDA), feature engineering, and model training with financially meaningful evaluation metrics. By addressing class imbalance and tuning classification thresholds, we ensured the model aligned with the business need to prioritize catching defaulters over avoiding false positives.

Data Preprocessing

The dataset comprised 25,247 rows and 27 columns. To prepare it for modeling, we performed the following steps:

- Dropped the `Customer_ID` column, as it lacked predictive value.
- Removed duplicate rows to ensure data integrity.

- Imputed missing values in the `age` column with the median, due to the column's right skewed distribution.
- Corrected discrepancies in categorical columns:
  - `marriage`: Replaced invalid value 0 with 1 (Single).
  - `education`: Replaced invalid values 0, 5, and 6 with 4 (Others).

These steps ensured a clean dataset, free of inconsistencies and ready for analysis.

Exploratory Data Analysis (EDA)

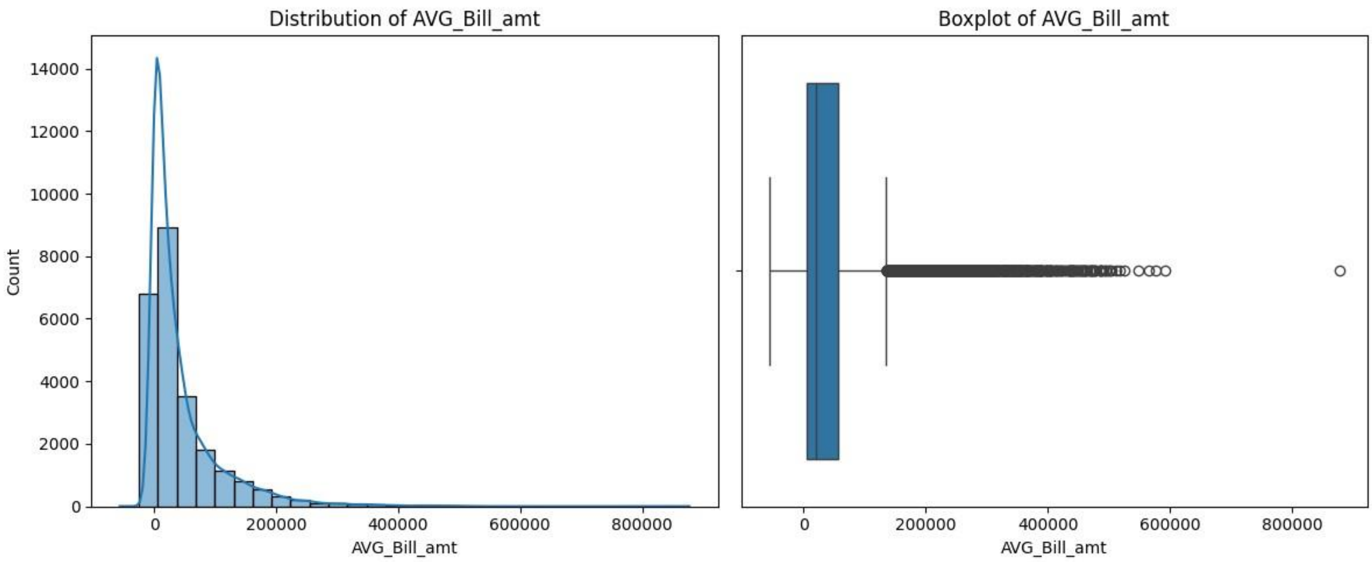
To understand the dataset and uncover patterns related to default risk, we conducted a comprehensive EDA, including univariate and bivariate analyses of demographic and financial features, as well as payment behavior.

Target Variable: `next_month_default`

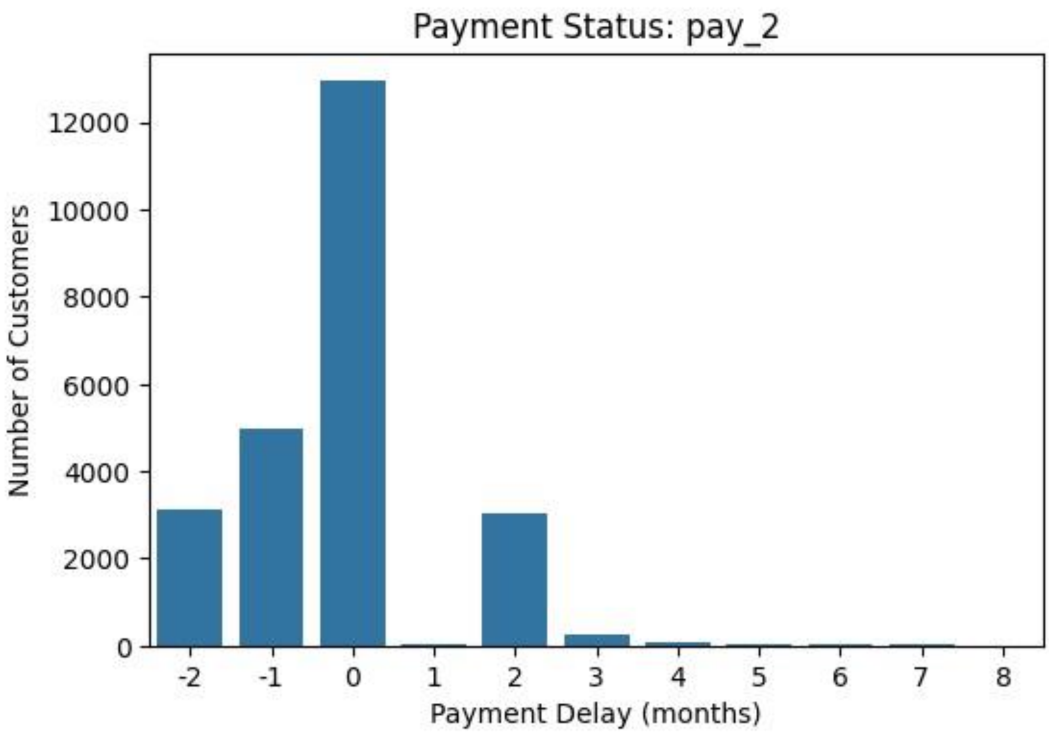
The target variable was highly imbalanced, with a majority of customers labelled as nondefaulters (0). This imbalance necessitated techniques like SMOTE to ensure models would not be biased toward the majority class.

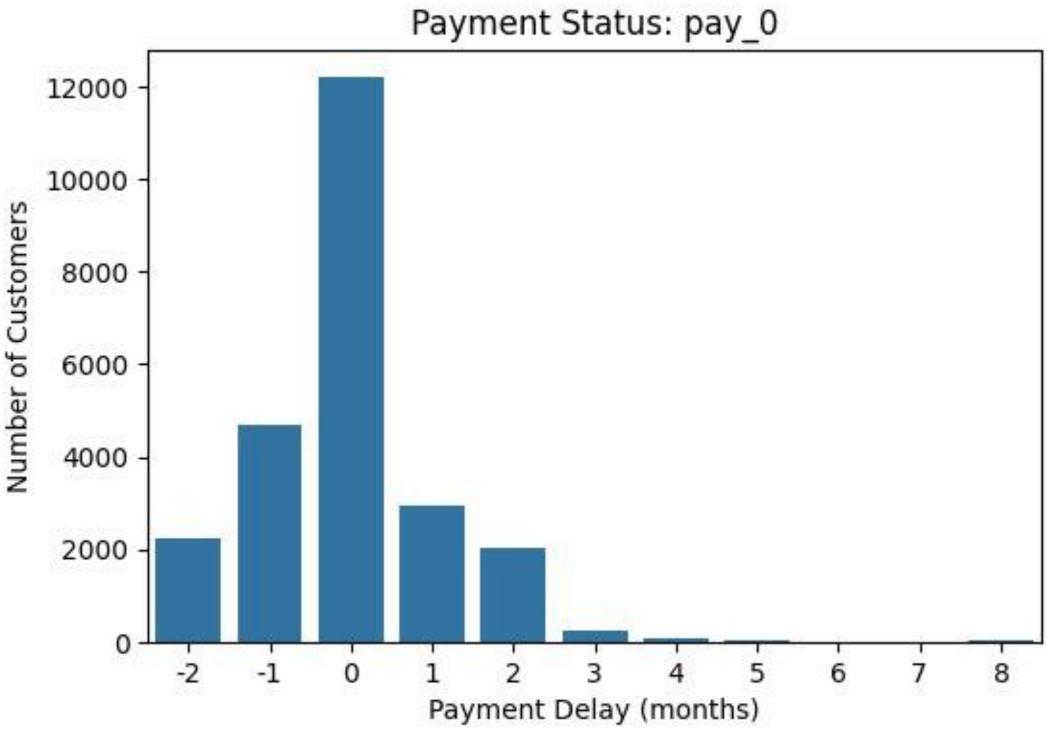
Univariate Analysis

- **Demographic Features**:
  - `Age`: Right-skewed, with most customers in their late 20s to early 40s. Missing values were imputed with the median.
  - `Sex`: Balanced distribution, with no strong correlation to default.
  - `Education`: Predominantly university or graduate-educated customers. Invalid values (0, 5, 6) were recoded as "Others."
  - `Marriage`: Most customers were married, followed by single. Invalid value 0 was recoded as "Others."
- **Financial Features**:
  - `LIMIT_BAL`: Right-skewed with outliers; log transformation applied for stability.
  - `AVG_Bill_amt` and `total_pay_amt`: Highly skewed, with most customers showing low usage but some exhibiting extreme values.



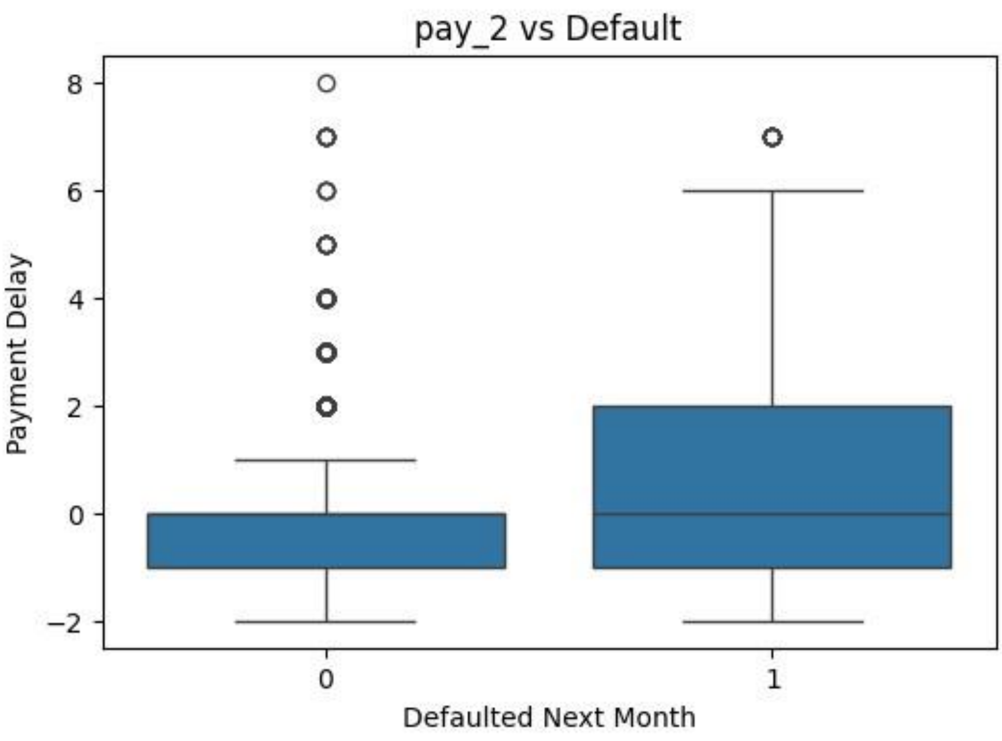
- **Payment Status (`PAY_0` to `PAY_6`)**:
  - Most customers had values of 0 (partial payments) or -1 (fully paid on time).
  - A notable cluster had value 2 (2-month delay), indicating moderate delinquency.



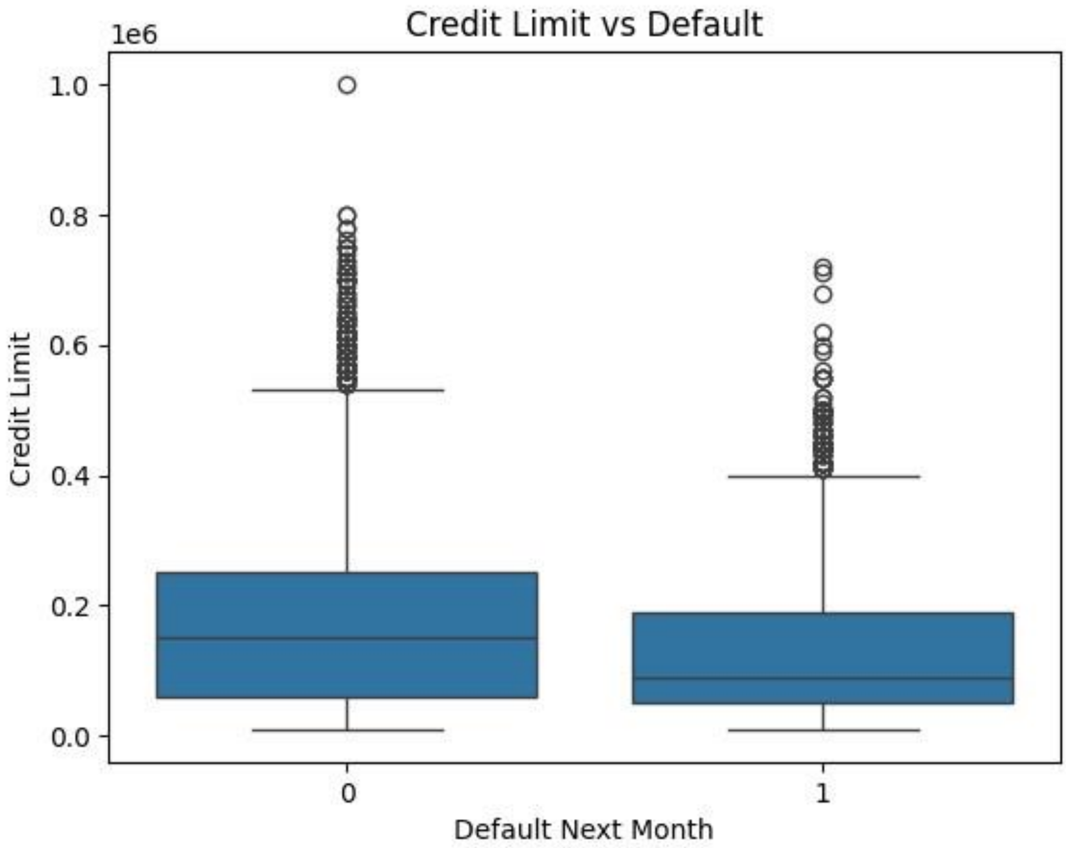


Bivariate Analysis

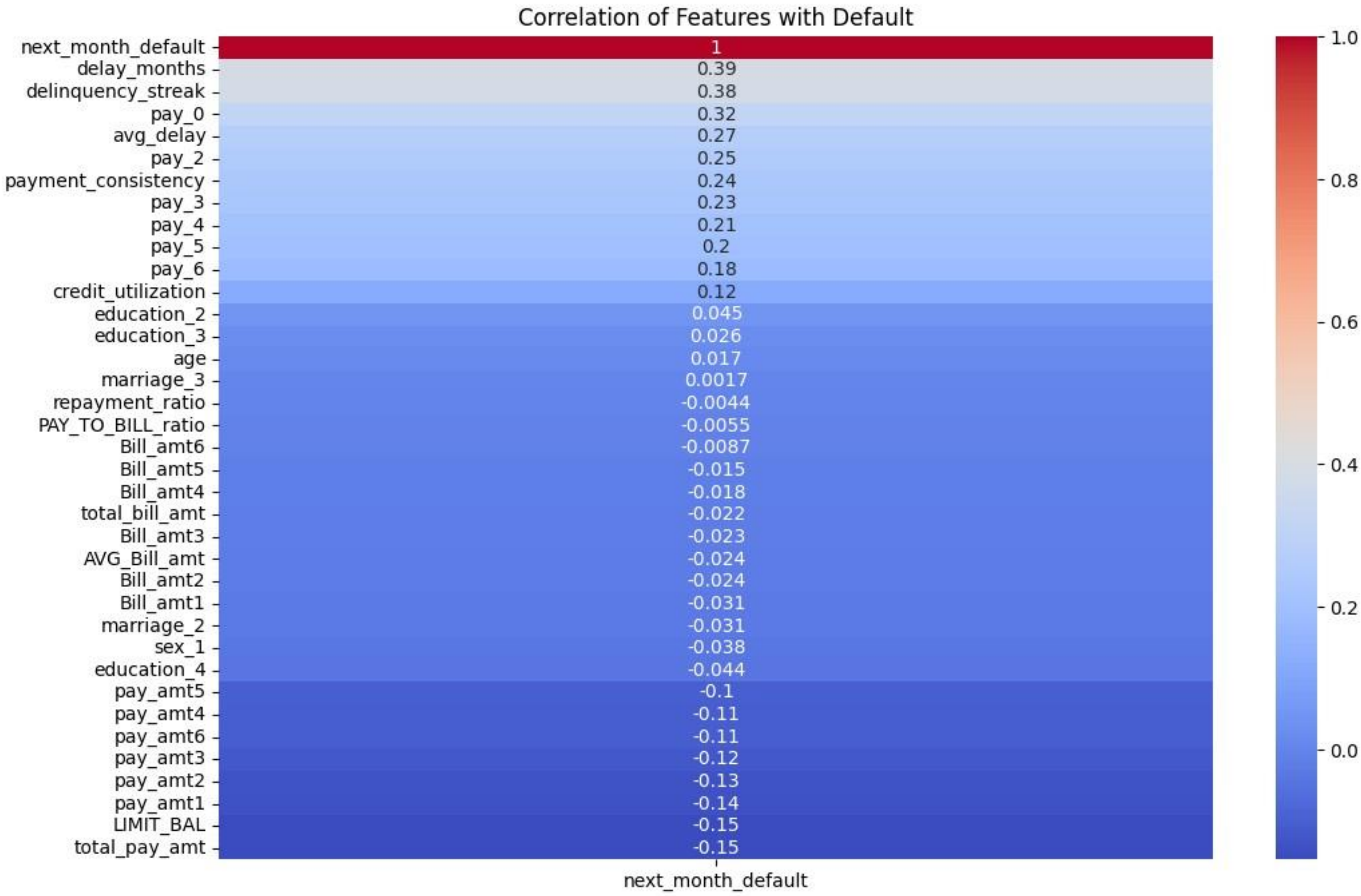
- **Payment Delays vs. Default**: Defaulters exhibited higher delays, particularly in recent months (`PAY_0` to `PAY_4`). The effect was less pronounced for `PAY_5` and `PAY_6`, suggesting decayed influence over time.



- **Credit Limit vs. Default**: Customers with lower credit limits were more likely to default, with the trend becoming clearer after log scaling.



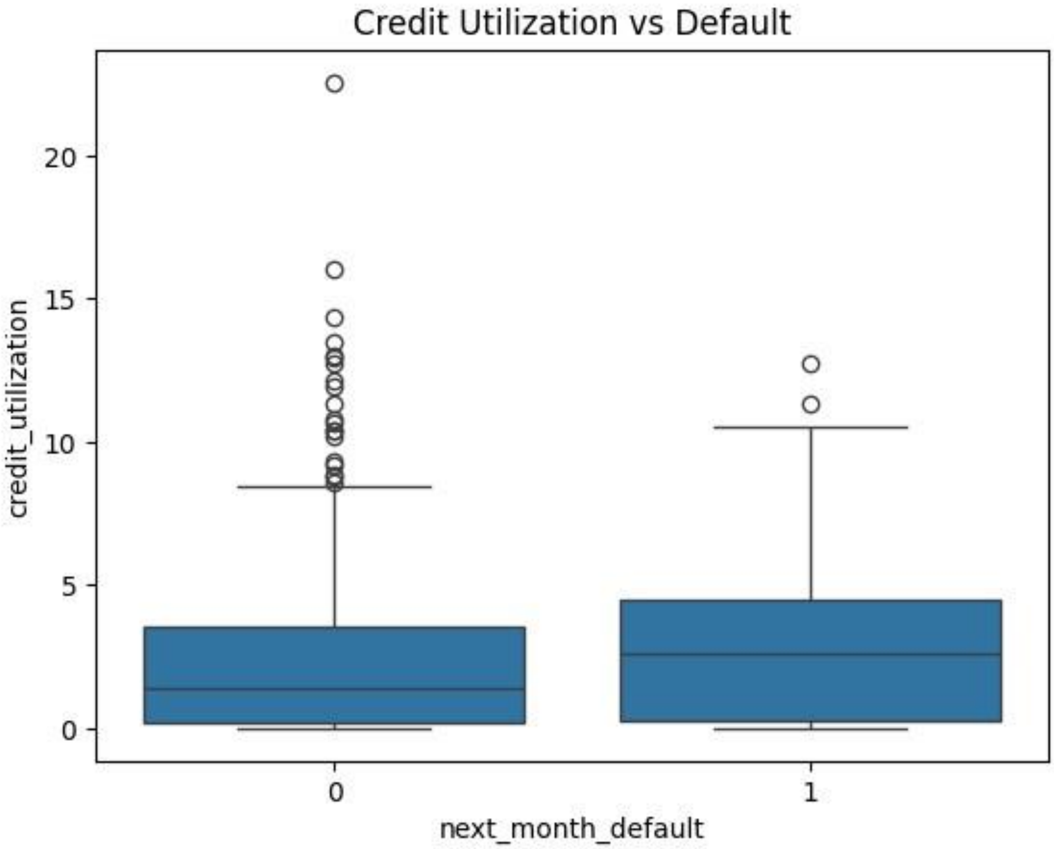
- **CorrelationAnalysis**: Thetargetvariableshowedstrongpositivecorrelationswith`delay_months` (0.39) and payment status features (`PAY_0` to `PAY_6`). Negative correlations were observed with `LIMIT_BAL`, `repayment_ratio`, and bill amounts, indicating that better financial standing reduces default risk.



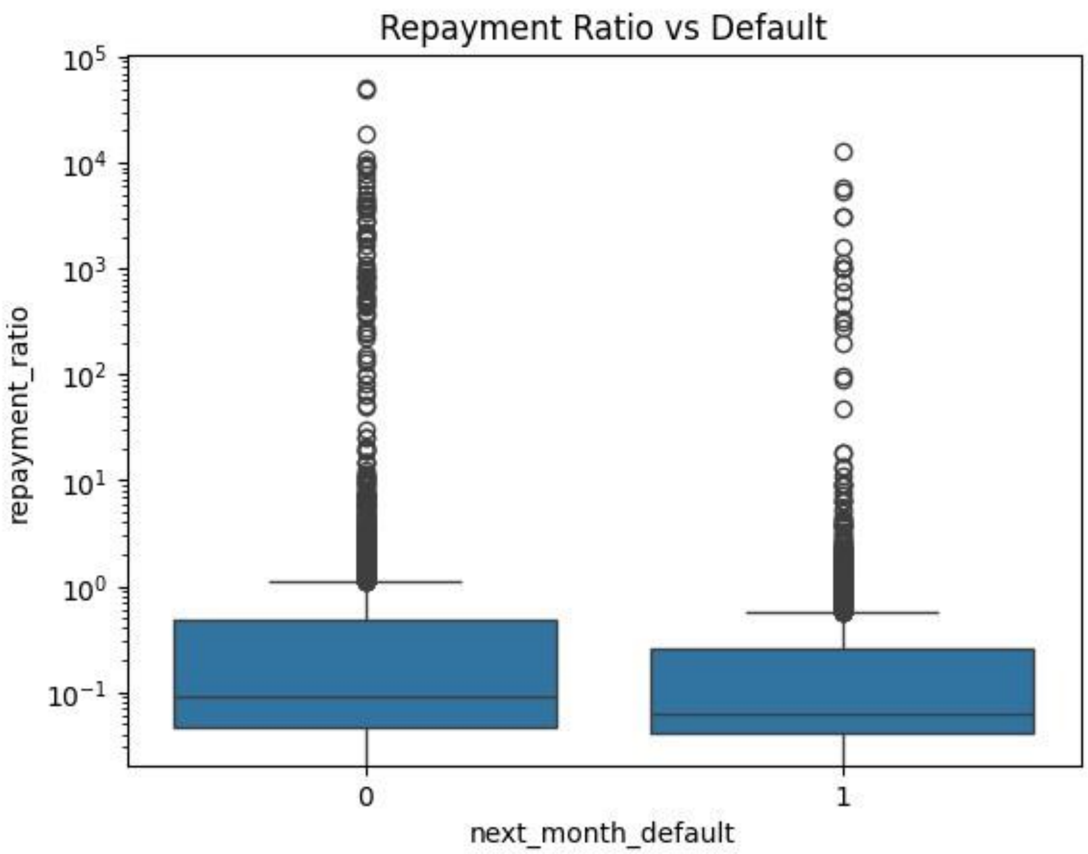
Financial Behavior Insights

The EDA revealed key financial behaviors associated with default risk:

- **Credit Utilization**: Defaulters typically had higher credit utilization (bill amount relative to credit limit), indicating over-reliance on credit. However, some non-defaulters also showed high utilization, suggesting it is a risk indicator but not definitive.



- **Overdue Behavior**: Customers with more delayed months were significantly more likely to default, with non-defaulters dominating lower delay counts.
- **Repayment Consistency**: Defaulters exhibited lower repayment ratios (total payments divided by total bills), while non-defaulters demonstrated greater consistency in repaying dues, aligning with financial discipline.



These insights guided feature engineering to capture financially relevant patterns.

Feature Engineering

To enhance the model's ability to predict defaults, I engineered the following financially meaningful features:

- `credit_utilization`: Ratio of total bill amount to credit limit, capturing reliance on credit.
- `delay_months`: Count of months with payment delay greater than 0, summarizing overdue behavior.
- `repayment_ratio`: Ratio of total payments to total bills, indicating repayment consistency.
- `avg_delay`: Average delay across `PAY_0` to `PAY_6`, quantifying payment tardiness.
- `payment_consistency`: Measure of payment stability, reflecting reliability in meeting obligations.
- `delinquency_streak`: Longest consecutive streak of delayed months, capturing persistent delinquency.

Correlation analysis confirmed these features had stronger relationships with `next_month_default` than raw features like `BILL_AMT` and `PAY_AMT`, which were dropped to reduce redundancy. Categorical variables (`education` and `marriage`) were onehot encoded to prepare for modelling.

Data Splitting and Class Imbalance Handling

The dataset was split into 80% training and 20% testing sets to evaluate model performance. To address the class imbalance in `next_month_default`, we applied SMOTE (Synthetic Minority Oversampling Technique) to the training set, generating synthetic examples of defaulters. This ensured the model learned patterns from both classes without bias toward the majority nondefaulter class.

Model Building

Four machine learning models were trained to predict defaults:

- Logistic Regression
- XGBoost
- Random Forest
- LightGBM

Each model was trained on the preprocessed and SMOTE-balanced training data, with engineered features and one-hot encoded categorical variables.

Evaluation Metrics

Given the financial context, we prioritized metrics that align with the goal of minimizing losses from defaults:

- **Precision (for class 1)**: Proportion of predicted defaulters who actually defaulted. High precision reduces false positives, avoiding unnecessary restrictions on non-risky customers.
- **Recall (for class 1)**: Proportion of actual defaulters correctly identified. High recall is critical, as false negatives (missed defaulters) result in direct financial losses.
- **F1 Score**: Harmonic mean of precision and recall, balancing both metrics.
- **F2 Score**: Weighted harmonic mean, emphasizing recall (with □=2) to prioritize catching defaulters over avoiding false positives.

- **AUC-ROC**: Measures the model's ability to distinguish between defaulters and nondefaulters, providing a comprehensive view of performance across thresholds.

The F2 score was particularly emphasized, as missing defaulters (false negatives) has a higher financial cost than incorrectly flagging non-defaulters (false positives).

Model Performance

The models were evaluated on the test set, with performance metrics reported at their respective optimal thresholds:

LightGBM and XGBoost achieved the highest F2score (0.607) , making it the best-performing model for this task.
We use **XGBoost** as our best model because its high score of F2 and Recall(1), which makes it suitable for our task,

Model	Threshold	Precision (1)	Recall (1)	F1 Score	F2 Score / AUC
Logistic Regression	Default	0.40	0.64	0.49	0.57 / –
XGBoost	0.11	0.28	0.86	0.42	0.607 / 0.76
Random Forest	0.13	0.27	0.84	0.41	0.590 / 0.75
LightGBM	0.18	0.33	0.76	0.46	0.607 / 0.77

Table 1: Model performance metrics on the test set.

Threshold Tuning

The default classification threshold of 0.5 was unsuitable for the financial context, as it did not prioritize recall sufficiently. To optimize the threshold, we:

- Evaluated F1 and F2 scores across a range of thresholds using the test set.
- Selected a threshold of approximately 0.11, which maximized the F2 score, emphasizing recall to minimize false negatives.

This choice aligns with the business priority of catching as many defaulters as possible, as false negatives lead to financial losses, while false positives cause only minor inconveniences (e.g., unnecessary credit reviews). The tuned threshold improved recall for class 1 (defaulters), ensuring better risk mitigation.

Conclusion

This project successfully developed a predictive model for credit card default risk, with XGBoost emerging as the top performer due to its high F2 score and Recall(1). Through careful preprocessing, EDA, and feature engineering, we created financially meaningful features like `delay_months`, `repayment_ratio`, and `credit_utilization`, which enhanced the model's predictive power. By addressing class imbalance with SMOTE and tuning the classification threshold to 0.11, we ensured the model prioritized identifying defaulters, aligning with the financial goal of minimizing losses. The insights and model can assist financial institutions in proactively managing credit risk, improving decision-making, and reducing default-related losses.