Finance_Project_Report_23118031

Gaurav Patni

June 2025

1 Introduction

This project, undertaken as part of the Finance Club's Summer 2025 initiative, addresses the challenge of credit card default prediction using advanced machine learning and risk-based techniques. Leveraging a rich dataset of historical credit card behavior, the goal is to build a robust classification model that flags potential defaulters in advance.

2 Exploratory Analysis of Demographic Features

2.1 Categorical Demographic Features

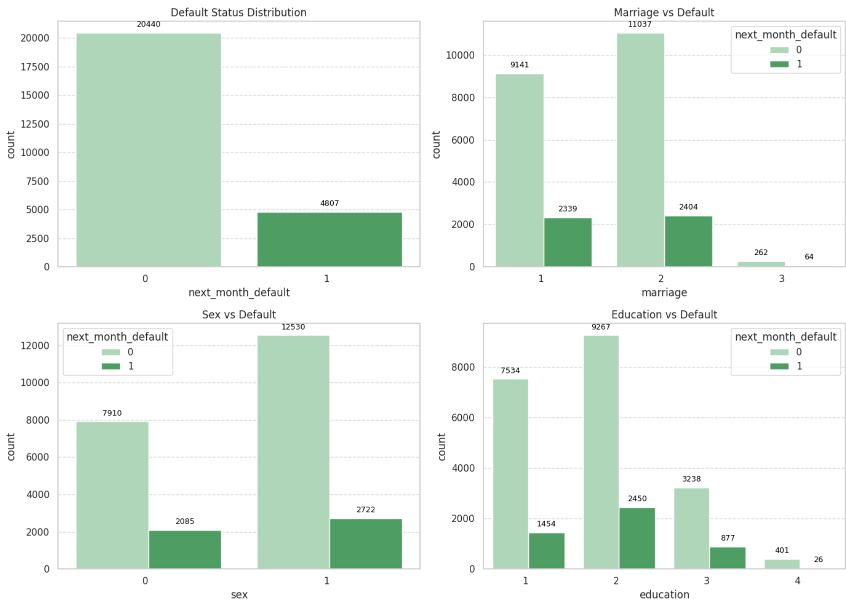

Figure 1: Distribution of Default Status Across Categorical Demographic Features

From Figure 1, we observe:

- **Marriage:** Single individuals show a slightly higher default rate compared to married customers.
- **Sex:** Minimal difference in default behavior between males and females.
- **Education:** Individuals with higher education have slightly lower default rates.

These insights suggest that demographic factors contribute to default likelihood, but may be secondary to behavioral features.

2.2 Payment History Features

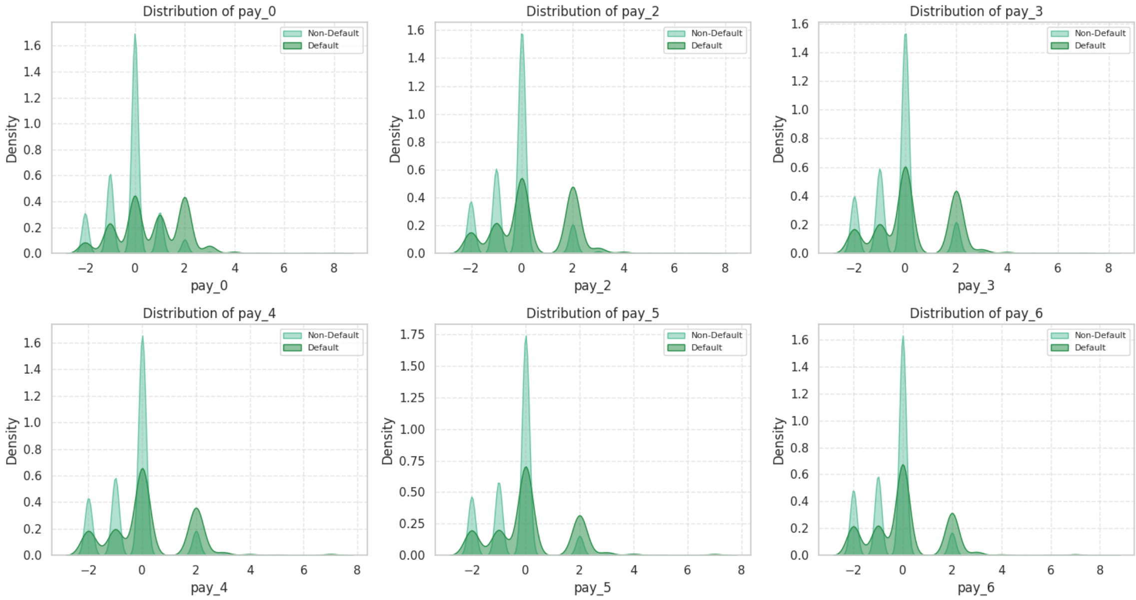

Figure 2: Distribution of Payment Status (`PAY_0` to `PAY_6`) by Default Status

- Non-defaulters (green) cluster around 0 and negative values—indicating on-time or early payments.
- Defaulters (red) peak at positive values (1, 2, or more), showing delayed payments.
- `PAY_0` shows the clearest separation between defaulters and non-defaulters.

These trends confirm that recent payment behavior is a strong indicator of future default risk.

2.3 Credit Limit vs Default Status

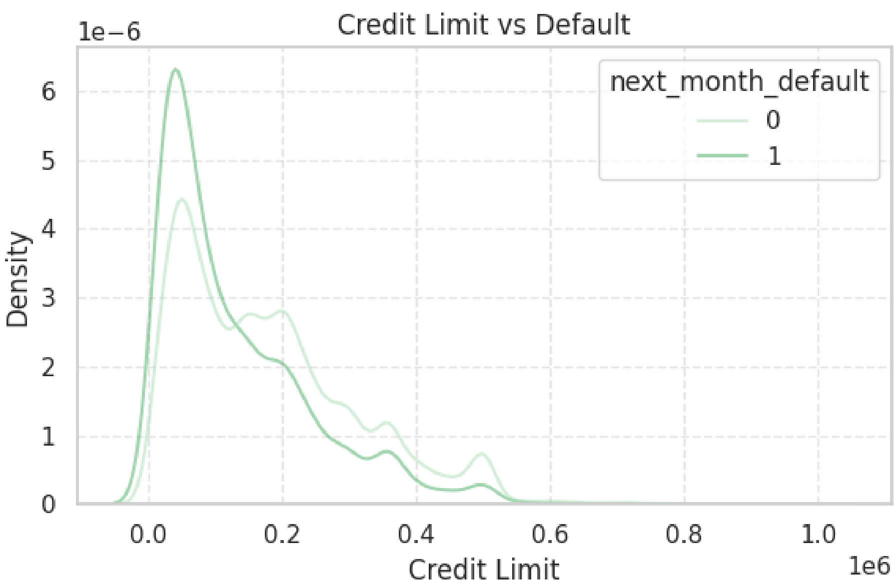

Figure 3: Distribution of Credit Limits by Default Status

Figure 3 shows that both defaulters and non-defaulters have skewed credit limits toward the lower end. However, the proportion of defaulters is significantly higher among customers with lower limits. This supports the conclusion that **credit limit is a strong predictor of default**.

2.4 Continuous Numerical Features Overview

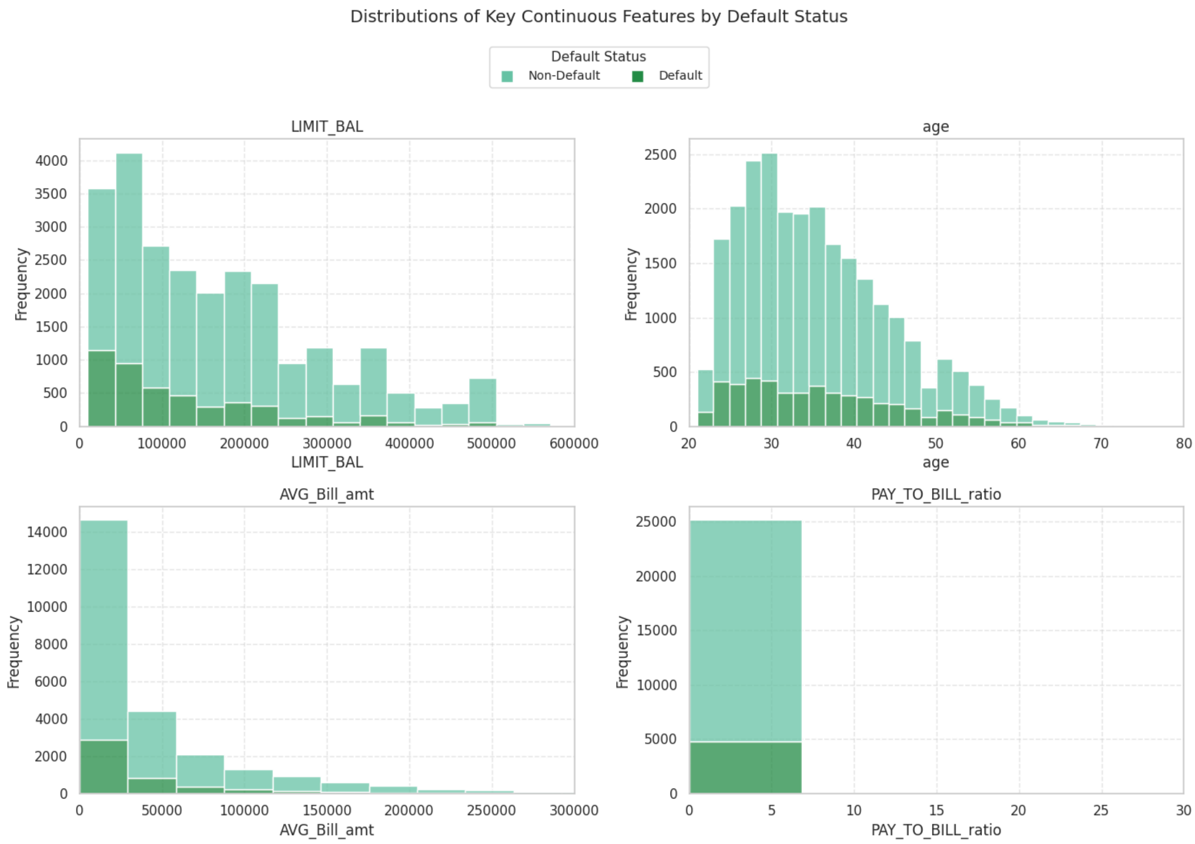

Figure 4: Distributions of Key Continuous Features

Figure 4 shows the distributions of key continuous variables:

- **LIMIT_BAL:** Defaulters are concentrated at lower credit limits; higher credit limits are associated with reduced default risk.
- **Age:** Customers aged 25–35 exhibit a higher likelihood of default compared to older individuals.
- **AVG_Bill_amt:** Defaulters generally have lower average bill amounts; higher billing activity appears linked to better repayment behavior.
- **PAY_TO_BILL_ratio:** Defaulters often show very low or zero repayment ratios, indicating poor financial discipline.

3 Feature Engineering

To enhance the model's ability to capture customer behavior and credit risk, a comprehensive set of features were engineered from the raw dataset. These features reflect credit utilization, delinquency trends, payment discipline, and financial volatility over a six-month period. The engineered features are categorized as follows:

3.1 Credit Utilization Features

- **credit_utilization_ratio:** Average monthly bill amount divided by the credit limit, indicating how much credit is being used on average.
- **max_monthly_utilization:** Highest bill amount over any month divided by the credit limit, capturing peak usage behavior.

3.2 Delinquency Features

- **num_delinquent_months:** Number of months with delayed payments (`PAY_x` ¿ 0).
- **max_delinquency:** Maximum delay level across six months.
- **avg_delinquency:** Mean of the `PAY_x` values, reflecting typical payment behavior.
- **ever_delinquent:** Binary indicator (1/0) showing whether the customer has ever been delinquent.
- **delinquency_trend:** Difference between `PAY_0` and `PAY_6`, highlighting improving or worsening behavior over time.

3.3 Payment Behavior Features

- **total_bill_amt:** Sum of all bill amounts over six months.
- **total_pay_amt:** Sum of all payments made over six months.
- **avg_pay_amt:** Average monthly payment amount.
- **underpaid_months:** Number of months where the customer paid less than the billed amount.
- **overpaid_months:** Number of months with payments exceeding the bill.

3.4 Volatility Features

- **bill_amt_std:** Standard deviation of bill amounts, capturing spending inconsistency.
- **pay_amt_std:** Standard deviation of payment amounts.
- **bill_amt_range:** Range between maximum and minimum bill values.
- **pay_amt_range:** Range between maximum and minimum payments.

3.5 Discipline Indicators

- **zero_spending_months:** Count of months where no bill was generated.
- **zero_payment_months:** Count of months where no payment was made.
- **fully_paid_months:** Number of months where the full bill was paid or exceeded.

These engineered features were designed to capture customer-level behavioral patterns and risk signals that are not directly evident in the raw features. Their inclusion significantly improves the model's capacity to differentiate between defaulters and non-defaulters, especially in recall-focused scenarios.

4 Insights from Feature Engineered Columns

Credit Utilization Features

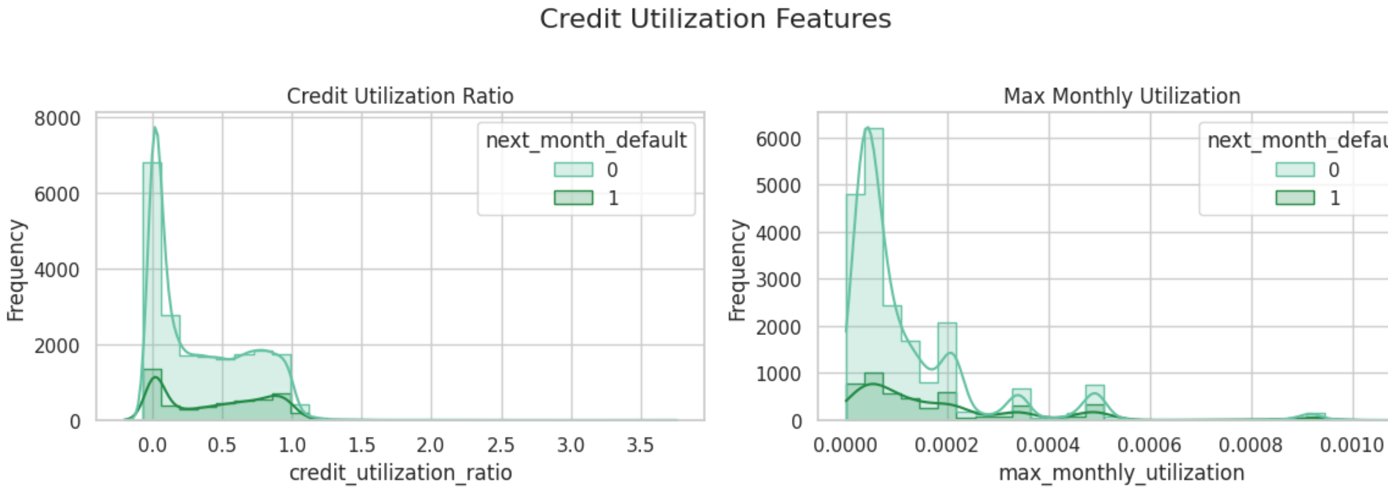

Figure 5: Distribution of Credit Utilization Ratio and Max Monthly Utilization by Default Status

Two features—*credit utilization ratio* and *max monthly utilization*—were analyzed to assess how customers use their available credit.

Findings:

- Defaulters typically exhibit higher credit utilization, often nearing or exceeding their credit limits.
- The distribution of max monthly utilization is heavily skewed toward lower values, but defaulters show a slightly heavier tail.

Delinquency Features

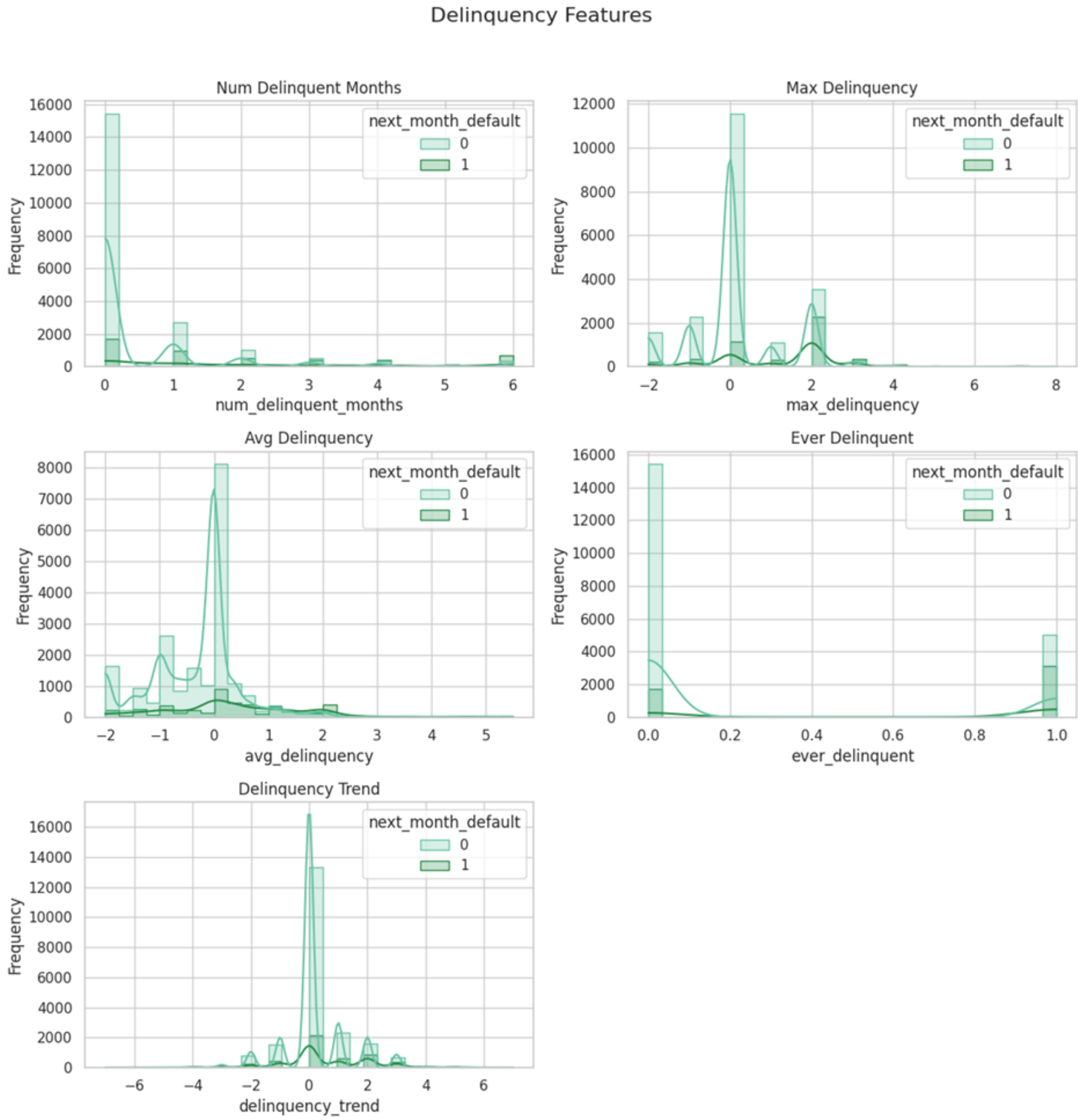

Figure 6: Distribution of Various Delinquency Features Grouped by Default Status

Delinquency-related behavior was assessed using features such as the number of delinquent months, average and maximum delinquency, binary ever-delinquent flags, and recent trends.

Findings:

- **Number of Delinquent Months:** Most non-defaulters have zero delinquent months. In contrast, defaulters frequently show one or more.
- **Max and Avg Delinquency:** Defaulters tend to exhibit higher maximum and average delinquency values, indicating more severe repayment delays.

- **Ever Delinquent:** A significant share of defaulters have experienced at least one delinquency event, unlike non-defaulters who predominantly remain at zero.
- **Delinquency Trend:** Non-defaulters exhibit flat or improving (negative) trends, while defaulters often show increasing or volatile trends.

Conclusion: Delinquency patterns—both in quantity and trajectory—are strong predictors of future default risk.

Payment Behavior Features

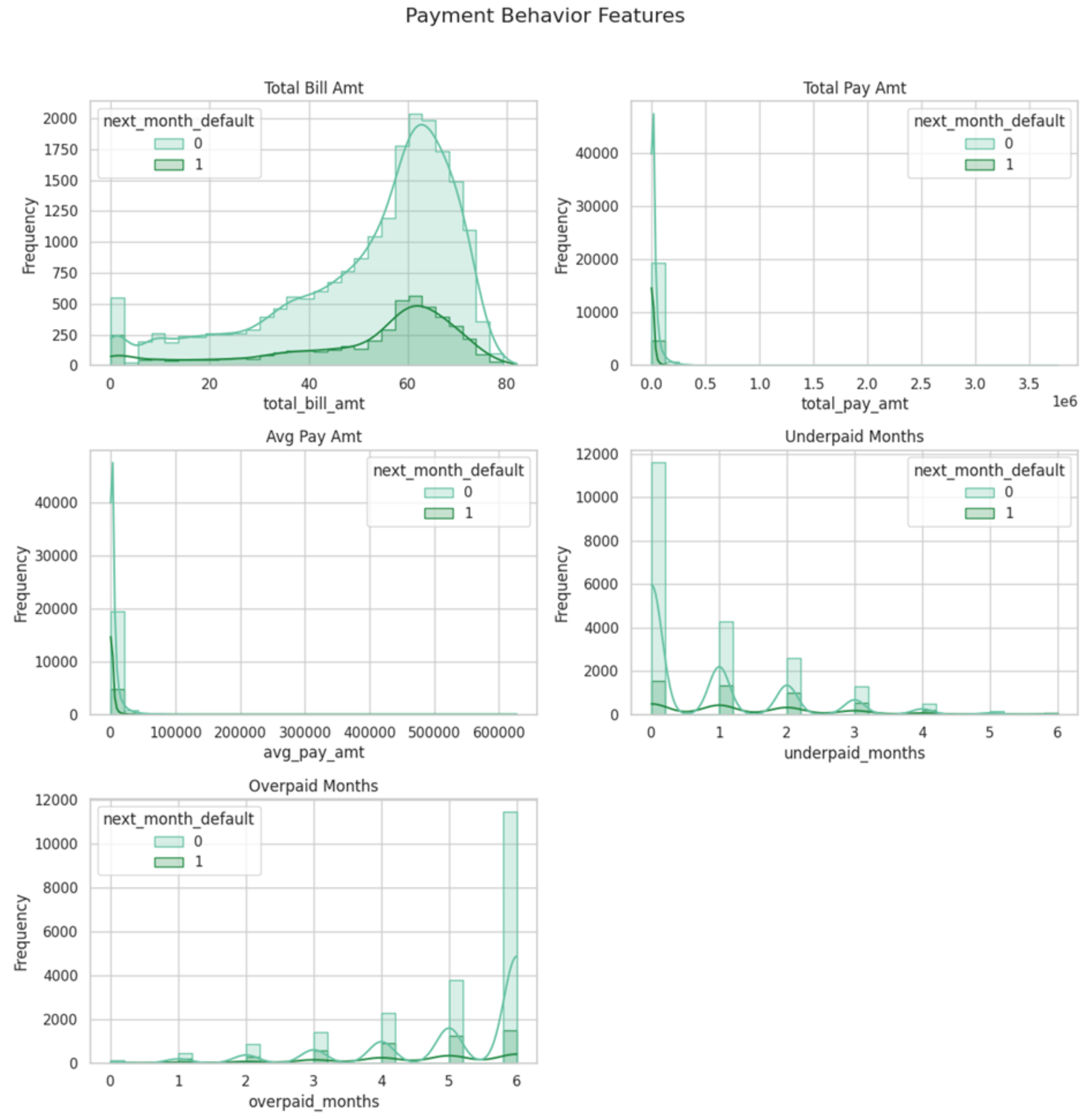

Figure 7: Distribution of Payment Behavior Features by Default Status

This section evaluates how users manage their billing and repayment over time, highlighting key trends across both defaulters and non-defaulters.

Findings:

- **Total Bill Amount:** Non-defaulters typically exhibit a bell-shaped distribution with most values concentrated in the mid-range. Defaulters are fewer in number and skewed slightly left, indicating lower bill amounts.

- **Total Pay Amount:** Payment values for both groups are right-skewed, but defaulters tend to pay less overall compared to non-defaulters. The peak near zero is more pronounced for defaulters, suggesting poor repayment behavior.
- **Average Pay Amount:** Similar to total payment trends, defaulters generally report lower average monthly payments. The long tail for non-defaulters indicates a few customers consistently making higher payments.
- **Underpaid Months:** A large number of defaulters have underpaid in multiple months, with higher frequency at values above two. Non-defaulters mostly cluster around zero or one underpaid month.
- **Overpaid Months:** Non-defaulters dominate the higher end of overpaid months (e.g., 5–6 months), suggesting a proactive repayment behavior. In contrast, defaulters seldom overpay.

Conclusion: Defaulters tend to have lower bills, pay less, underpay more frequently, and rarely overpay—clear indicators of elevated default risk.

Volatility Features

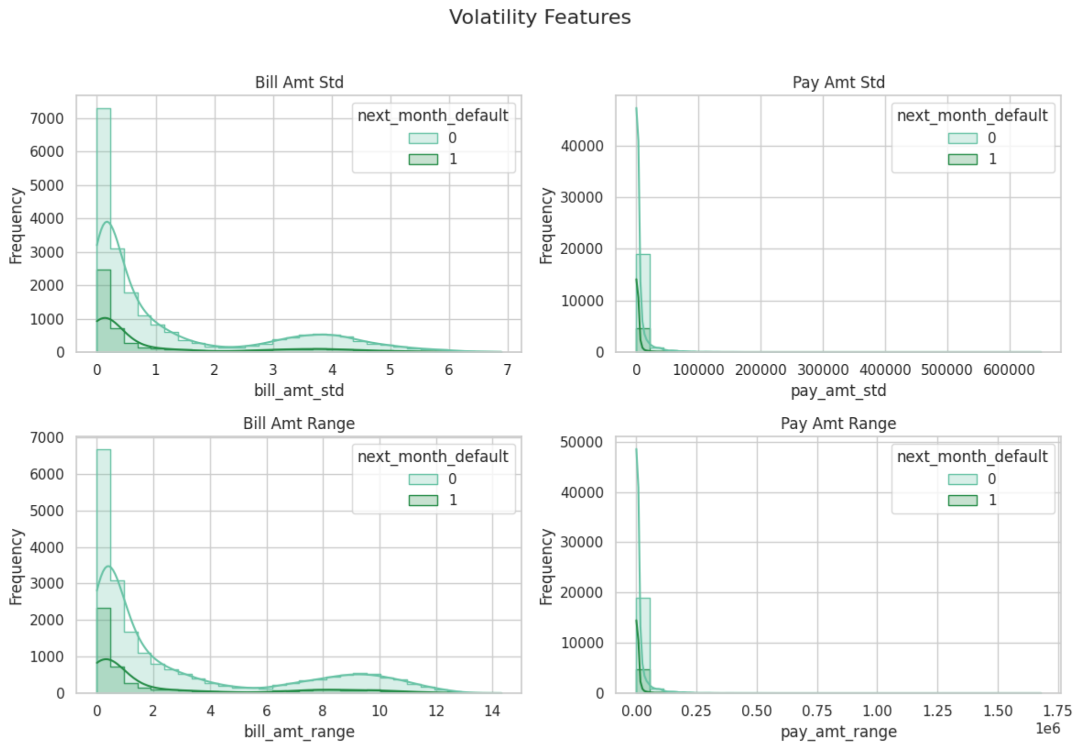

Figure 8: Distribution of Volatility Features by Default Status

Volatility in billing and repayment behaviors can signal financial instability, which in turn may correlate with default risk. Four volatility-related features were analyzed: standard deviation and range of both bill and pay amounts.

Findings:

- **Bill Amount Std & Range:** Non-defaulters display a wider spread in both standard deviation and range of bill amounts, suggesting variable usage patterns. Defaulters, in contrast, are more concentrated near zero, indicating consistently low or flat billing behavior.
- **Pay Amount Std & Range:** Payment volatility (both std and range) also tends to be lower for defaulters. While most users show low payment variation, non-defaulters demonstrate a long tail, likely representing responsible customers with flexible but timely repayments.

Conclusion: Low volatility in billing and repayment might point toward stagnant or constrained financial activity, commonly associated with defaulters. Conversely, higher variability may reflect responsible and dynamic credit behavior.

Discipline Indicators

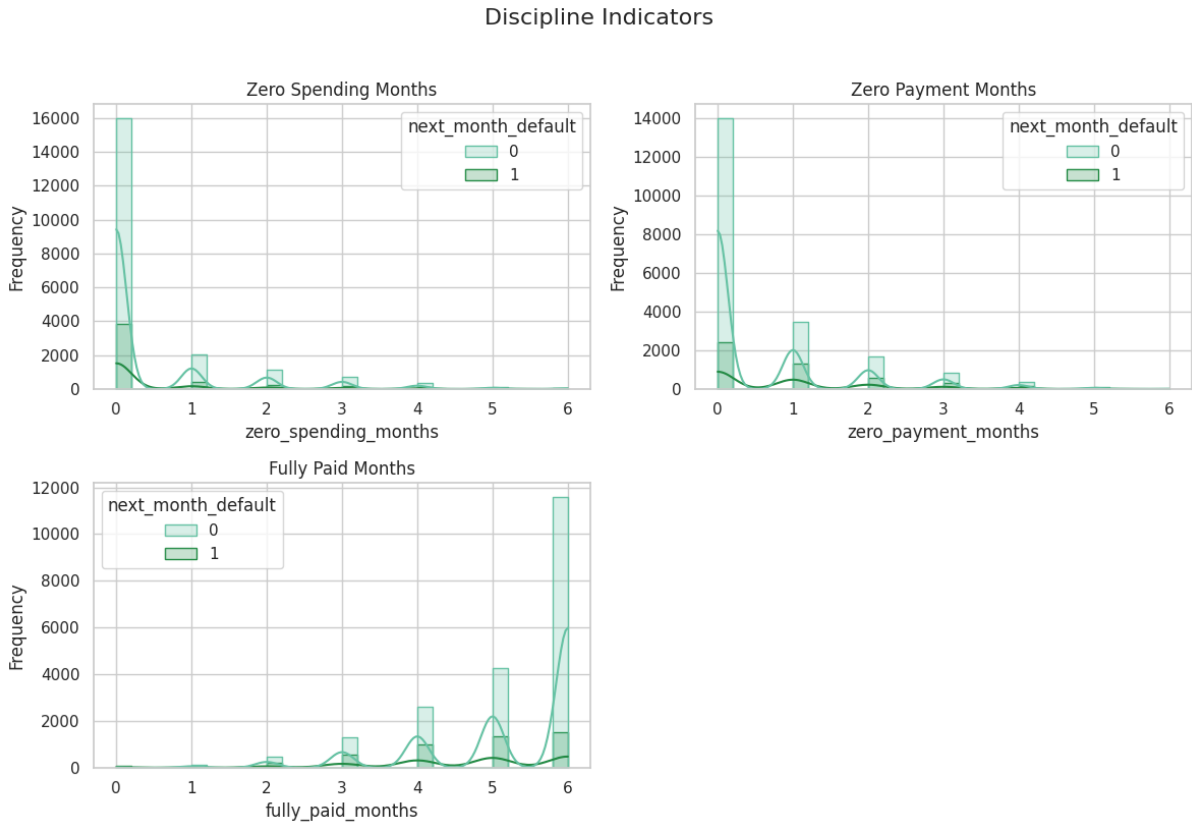

Figure 9: Distribution of Discipline Features by Default Status: Zero Spending Months, Zero Payment Months, and Fully Paid Months

Three discipline-related behavioral features were explored: *zero spending months*, *zero payment months*, and *fully paid months*.

Findings:

- Most customers had few to no months with zero spending or zero payment, but defaulters had higher frequencies in these categories.
- The *fully paid months* feature shows a strong separation: non-defaulters tend to fully pay more often, particularly in all six months.
- A clear behavioral signal emerges—individuals with consistently good repayment habits are less likely to default.

5 Handling Class Imbalance with SMOTENC

The dataset exhibited a significant imbalance between defaulters and non-defaulters, with the majority class (non-default) dominating the training data. To address this imbalance and prevent model bias toward the majority class, we applied **SMOTENC (Synthetic Minority Oversampling Technique for Nominal and Continuous features)**.

SMOTENC is an extension of the traditional SMOTE algorithm that supports mixed-type data—numerical and categorical. It synthesizes new samples for the minority class by interpolating

6 Objective

The objective of this analysis is to identify the best-performing model for predicting credit card defaults, focusing on maximizing the **F2-score**. In this context, the F2-score emphasizes **recall** over precision, aligning with the goal of minimizing false negatives (i.e., undetected defaulters).

7 Evaluation Criteria

Models are evaluated using the following metrics:

- **Accuracy**: overall correctness
- **Precision**: true positive rate among predicted positives
- **Recall**: true positive rate among actual positives
- **F1 Score**: harmonic mean of precision and recall
- **F2 Score**: weighted harmonic mean favoring recall

8 Model Evaluation and Threshold Optimization

We evaluate four models—**Logistic Regression**, **XGBoost**, **Random Forest**, and a **Neural Network (LSTM)**—under two thresholding conditions:

- **Default Threshold (0.5)**: Traditional binary classification cutoff
- **F2-Optimized Threshold**: Threshold adjusted to maximize the F2-score

8.1 Importance of Recall in Credit Risk

In imbalanced settings like credit default prediction, false negatives (missing a real defaulter) are far costlier than false positives. While the F1-score treats both errors equally, the F2-score addresses this by placing higher weight on recall:

$$F_\beta = (1+\beta^2)\cdot\frac{Precision\cdot Recall}{\beta^2\cdot Precision + Recall}, \quad where \beta = 2$$

This aligns better with real-world objectives of financial institutions to **minimize undetected defaults**.

9 Model Performance with F2-Optimized Thresholds

Table 1: Model Performance (F2-Optimized Thresholds)

Model	Accuracy	F2 Score	Precision	Recall	F1 Score
Neural Network	**0.648**	**0.6120**	0.32	0.79	0.46
Random Forest	0.562	0.6063	0.28	**0.85**	0.42
Logistic Regression	0.572	0.5768	0.28	0.79	0.41
XGBoost	0.446	0.5912	0.24	**0.91**	0.39

10 Metric Insights

- **Accuracy**: Decreases after threshold optimization due to favoring recall.
- **Precision**: Drops as the model becomes more sensitive to positives.

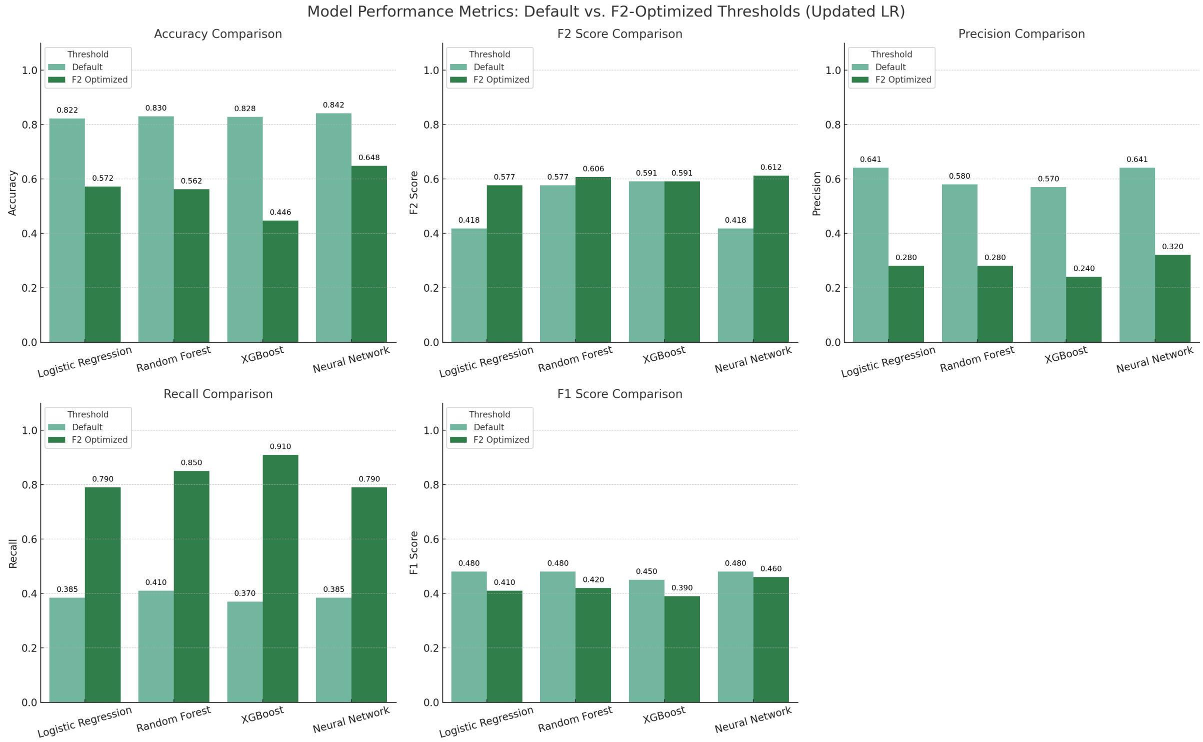

Figure 10: Performance Comparison of Models Using Default vs. F2-Optimized Thresholds

- **Recall**: Increases in all models — mission accomplished.
- **F2 Score**: Shows significant improvement, validating optimization.

11 Model Comparison and Recommendation

Neural Network is the top-performing model with the highest F2-score (0.6120), offering strong recall and reasonable trade-offs.

Random Forest is a strong second, achieving excellent recall (0.85) with moderate precision.

XGBoost leads in recall (0.91) but has very low precision and accuracy, making it better suited for use in generating high-risk watchlists rather than direct decisioning.

Logistic Regression offers simplicity and interpretability but underperforms in F2 compared to neural approaches.

Final Recommendation

- **Best Model: Neural Network** – balances high recall with best overall F2.
- **Alternative: Random Forest** – simpler, interpretable, with strong recall.
- **Flagging Model: XGBoost** – use for high-recall alerting scenarios.