Credit Card Default Prediction Model

Kumar Manas.
23119016
16-06-2025

Finclub IIT Roorkee

Executive Summary

Project Overview:

Bank A seeks to strengthen its credit risk management by developing a predictive **Behavior Score** — a classification model that identifies customers likely to default on their credit card payments in the upcoming month. Using historical behavioral data from over 30,000 customers, the model aims to flag high-risk individuals early, enabling proactive measures such as adjusting credit limits and initiating risk-based interventions. Beyond accurate prediction, the model emphasizes financial interpretability to support informed decision-making and a deeper understanding of default behavior.

Project Objectives:

- Develop a **binary classification model** to predict credit card customer defaults for the next month.
- Address **class imbalance** using techniques like **SMOTE** and **class weighting**.
- Conduct **exploratory and behavioral analysis** to uncover financial patterns such as **payment delays**, **repayment consistency**, and **credit utilization**.
- Create **financially meaningful features** (e.g., credit utilization ratio, delinquency streaks) to enhance model performance.
- Train and compare different models.
- Select appropriate **evaluation metrics** (e.g., F2-score, AUC-ROC) reflecting credit risk priorities.
- Tune and apply an **optimal classification threshold** aligned with the bank's risk tolerance to balance **false positives** and **false negatives**.
- Generate final **predictions** on an unlabeled validation dataset using the **best-performing model** and threshold.

1. Library Imports

To begin the project, all essential libraries were imported to support data processing, visualization, modeling, and evaluation:

- **Pandas** and **NumPy** for data manipulation and numerical operations.
- **Matplotlib** and **Seaborn** for data visualization and exploratory analysis.
- **Scikit-learn (sklearn)** for machine learning models, preprocessing, evaluation metrics, and model selection.
- **Imbalanced-learn (imblearn)** for handling class imbalance using techniques like SMOTE.
- **SHAP** for explainable AI, allowing interpretation of model predictions.

These libraries formed the foundation for implementing the entire credit default prediction pipeline.

2. Data Loading and Initial Inspection

The dataset was loaded using **Pandas** for structured analysis. It contains historical behavioral data of over **30,000 credit card customers**, with the target variable `next_month_default` indicating whether a customer defaulted in the upcoming billing cycle.

Initial steps included:

- **Loading the data** into a DataFrame and checking its shape and structure.
- **Previewing sample records** using `.head()` to understand feature types and formats.
- **Checking for null values** and basic inconsistencies.
- **Inspecting data types** to identify numerical, categorical, and datetime fields (if any).

This inspection helped in understanding the quality and layout of the data, ensuring readiness for preprocessing and exploratory analysis.

3. Handling Missing Values

Missing values were addressed as a critical preprocessing step to ensure model reliability. The following actions were taken:

- **Detection**: The dataset was scanned using `.isnull().sum()` to identify features containing missing values.
- **Strategy**:
    - For **numerical columns**, missing values were imputed using the **mean** of each respective column, assuming the data was Missing At Random (MAR).

- For **categorical features**, although none were found with missing entries, a mode-based imputation strategy would have been used if necessary.

- **Validation**: After imputation, checks were performed to confirm the absence of null values using `.isnull().sum().any()`.

This ensured the dataset was clean and consistent for downstream tasks like feature engineering, modeling, and evaluation.

4. Feature Cleaning and Reduction

To enhance model performance and reduce noise, the dataset underwent systematic feature cleaning and reduction:

- **Removal of Irrelevant Columns**: Dropped unique identifier column 'Customer_ID' (not useful for modeling)

- **Removal of Redundant Features**: Columns such as binned variables (`*_BIN`, `age_group`, `LIMIT_BIN`, etc.) that were primarily derived for exploratory analysis were dropped to prevent redundancy and potential data leakage.(this was done just before splitting the train dataset further used for validation)

- **Multicollinearity Check**: Correlation analysis was conducted to identify highly correlated features. When necessary, features with strong linear dependencies were considered for removal to improve model generalization.

- **Manual Pruning**: Certain engineered or raw features that did not show predictive value or interpretability (based on domain understanding or EDA) were excluded to streamline the feature set.

This step helped in simplifying the model, improving computational efficiency, and focusing on features with meaningful behavioral and financial relevance.

5. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was conducted to understand the underlying structure, trends, and behavioral patterns in the data.

5.1 Target Variable Analysis:

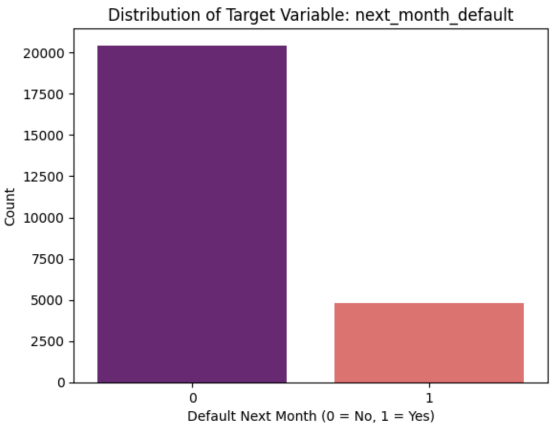
There is a significant class imbalance in the dataset:

- The majority of customers (around 20,000) did not default.

- A smaller portion (around 5,000) did default.

Consider techniques like:

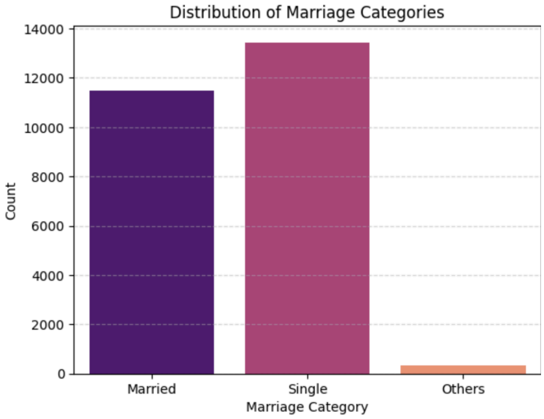
- Resampling (Oversampling, SMOTE, Undersampling).

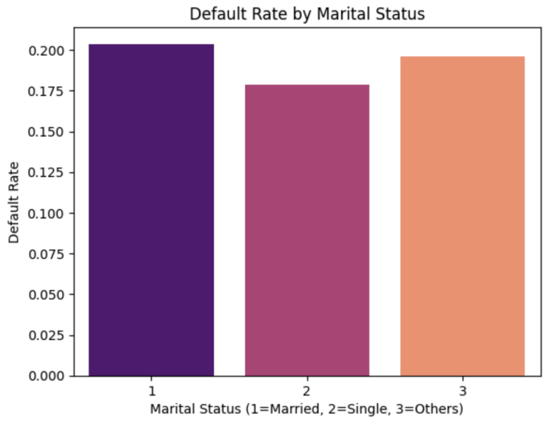


- Using algorithms that handle imbalance well

5.2 Demographic features:(Marriage, Sex, Education, Age, LIMIT_BAL)

5.2.1 Marriage:





The count for the single customer is higher but the seen default rate among them is less as compared to the married because of less expenditure.

We observe that married individuals have the highest default rate, contrary to common assumptions.

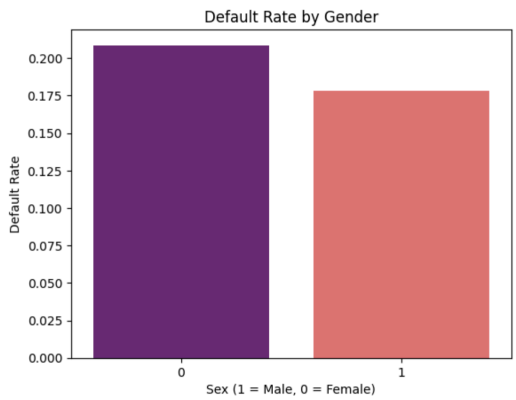
1. Higher Financial Commitments

2. Joint Debts and Shared Credit

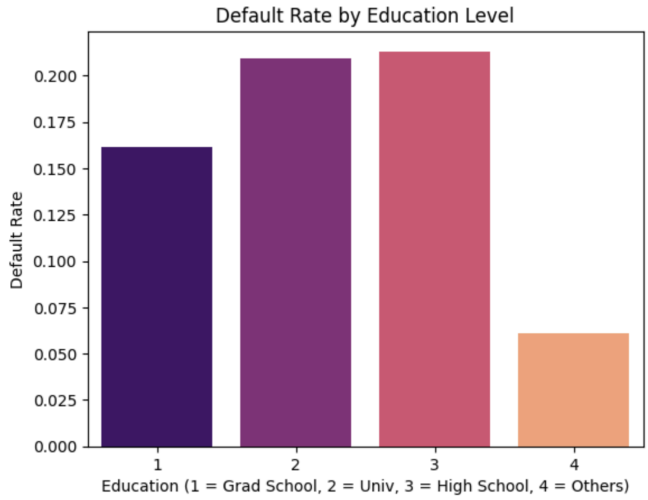
5.2.2 Sex:

Default rate is higher for females than males, this finding may seem counterintuitive. However, several real-world factors could contribute to this pattern:

1. Income Disparity
2. Financial Responsibilities
3. Credit Access and Limits
4. Employment Stability



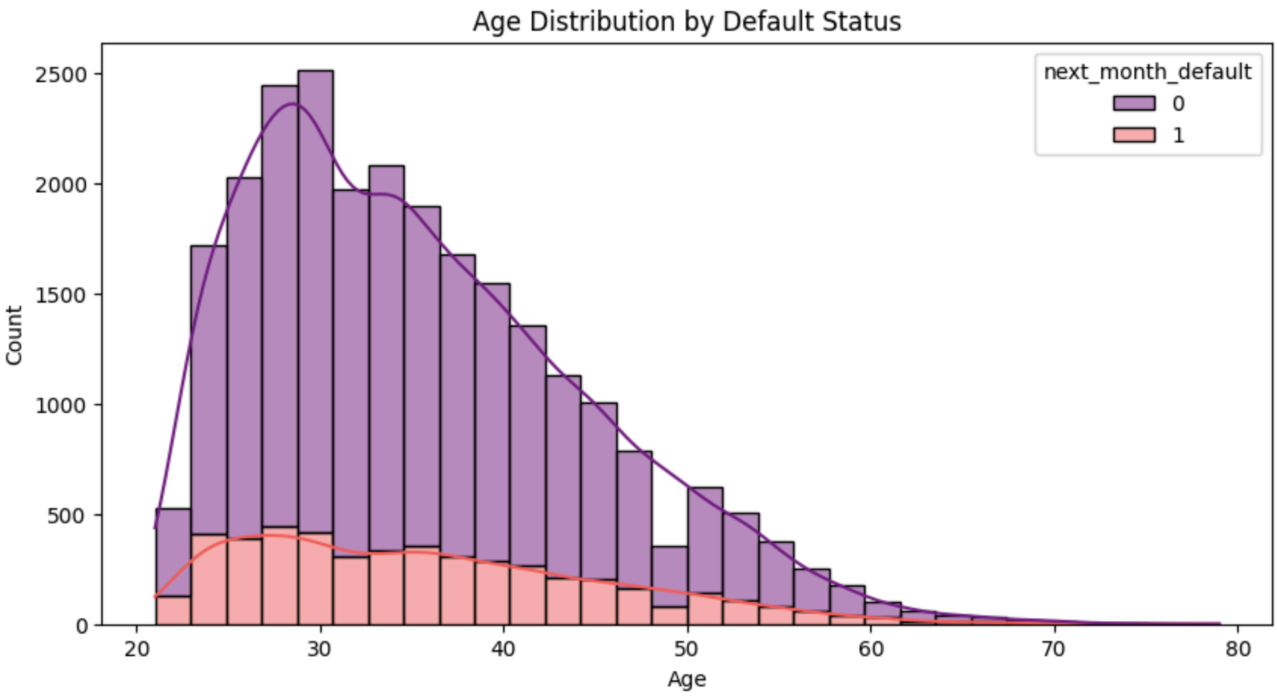
5.2.3 Education:



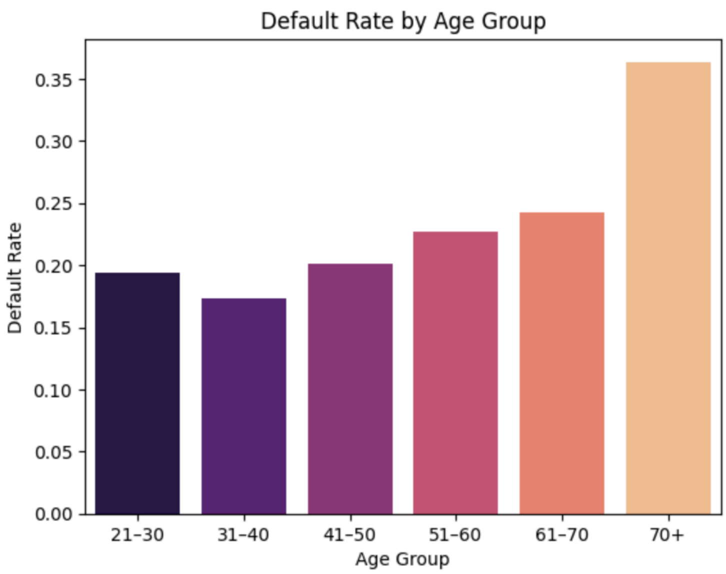
Individuals holding a graduate-level education typically benefit from more stable employment, higher income levels, and stronger financial knowledge — all of which contribute to more responsible credit behavior and a lower likelihood of default.

In contrast, customers with only a university or high school education tend to show higher default rates. This segment is often more financially heterogeneous; for instance, university graduates may still be in the early stages of their careers, dealing with student debt, irregular earnings, or limited financial buffers, which can elevate their credit risk.

5.2.4 Age:



The relationship between age and default risk is **nonlinear** — default probabilities don't consistently increase or decrease with age. Instead, the trend appears to dip in middle age and rise again later in life.



- **Young customers** tend to show higher default rates, likely due to unstable income, lower savings, and limited financial experience.
- **Middle-aged individuals**

usually demonstrate better repayment behavior, benefiting from career stability and higher earning potential.

- However, **older customers** may face an increasing risk of default, possibly due to retirement-related income reduction, rising living costs, or healthcare expenses.

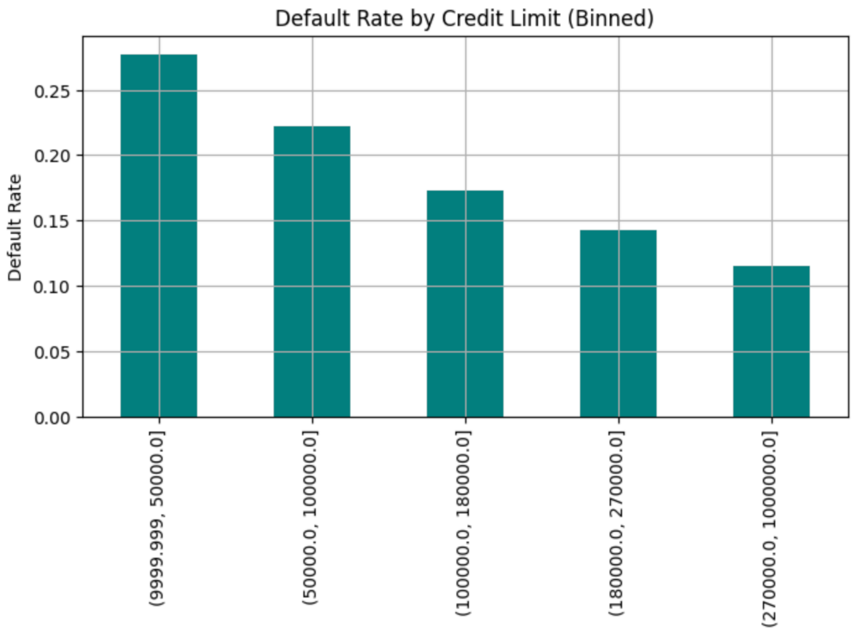
Such **nonlinear patterns** can be difficult for linear models like logistic regression to capture. In contrast, **tree-based algorithms like XGBoost and LightGBM** are capable of modeling these complexities effectively, making them well-suited for this type of financial data.

5.2.5 LIMIT_BAL:

Credit Limit vs. Default Risk

The analysis of default trends across different credit limit brackets reveals a **strong inverse relationship** between a customer's credit limit (`LIMIT_BAL`) and their likelihood of defaulting. Specifically:

- Customers with **lower credit limits (e.g., ₹0–₹50,000)** show the **highest default rates**, suggesting elevated credit risk in this segment.
- As the credit limit increases, the **default rate steadily declines** and eventually plateaus, indicating improved repayment behavior.
- This trend reflects common lending practices, where **creditworthy customers are granted higher limits** due to better financial profiles, stable incomes, or good past repayment records.

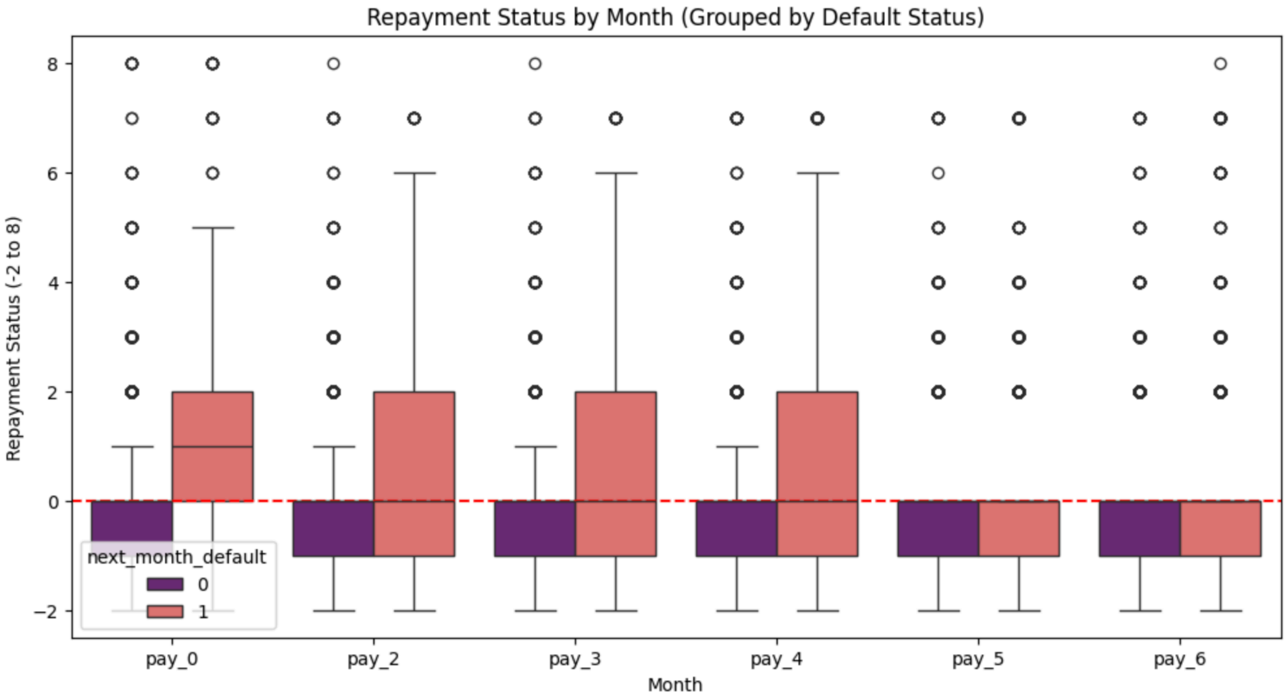


- In contrast, customers considered riskier are typically restricted to lower limits, aligning with observed higher default rates in these groups.

Thus, `LIMIT_BAL` proves to be a **critical feature for credit risk segmentation**, offering a direct signal of both the lender's confidence and the borrower's repayment potential.

5.3 Behavioral Features:(PAY columns (e.g., PAY_0 to PAY_6), bill_amt and pay_amt trends, PAY_TO_BILL_ratio)

5.3.1 PAY columns:



Repayment Behavior Analysis:
An in-depth examination of monthly repayment statuses (`pay_0` to `pay_6`) reveals a clear behavioral distinction between defaulters and non-defaulters. Customers who eventually default (shown by pink bars) consistently exhibit more positive repayment status values across months, which reflect delayed payments or delinquencies. In contrast, non-defaulters (purple bars) tend to have values centered around zero or negative — indicating on-time or even early payments.

This stark difference highlights how **historical repayment behavior is strongly predictive of future default risk**. Specifically:

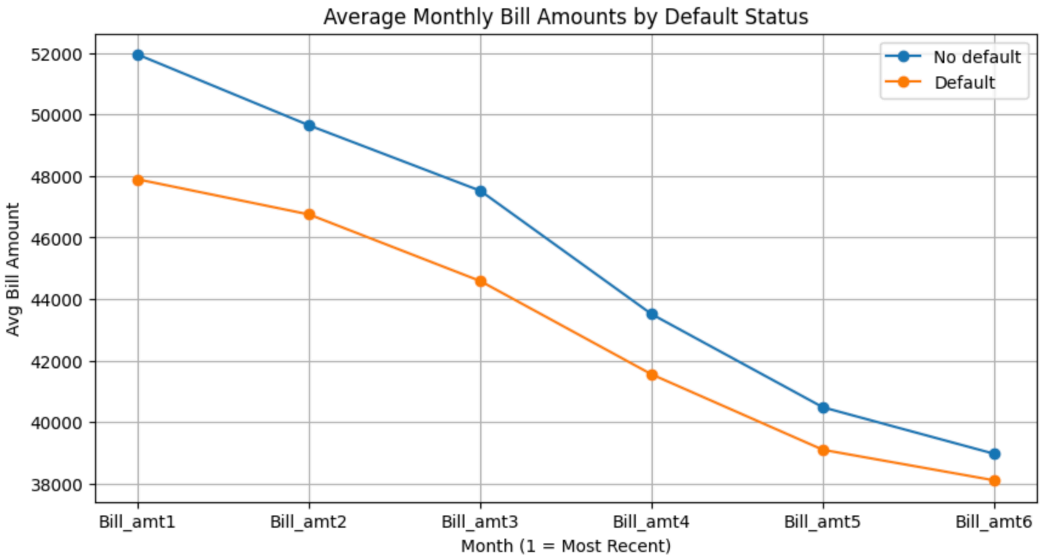
- **Positive values (1–8)** in repayment status indicate increasing levels of payment delay.
- **Higher values across multiple months** are common among defaulters, showing repeated delinquency.
- **Non-defaulters** often show 0 (on-time) or negative (advance) payment statuses.

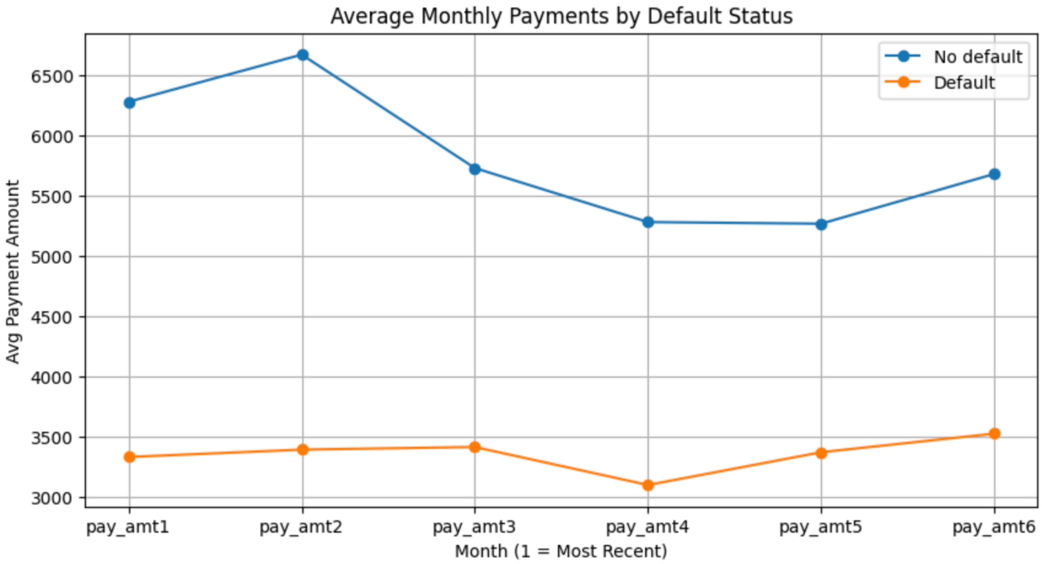
These insights support the engineering of several key features used in the modeling process:

- **`MAX_DELAY`**: The maximum value across `pay_x` features to capture the worst delay.
- **`DELINQUENCY_STREAK`**: The longest consecutive run of delayed payments.
- **`REPAYMENT_CONSISTENCY`**: A derived metric measuring the variability and stability in repayment behavior over time.

These features allow the model to better distinguish high-risk individuals based on their past payment discipline.

5.3.2 BILL_amt and pay_amt trends:

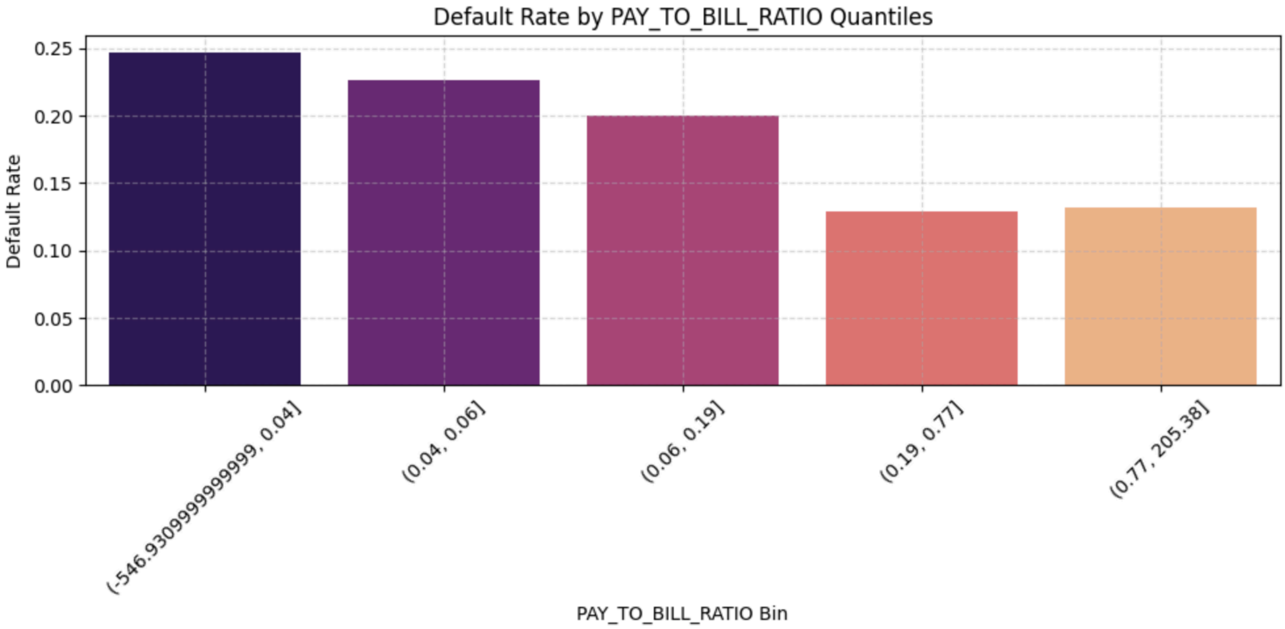




To gain insight into customer behavior, we analyzed the average monthly bill amounts and payments over a six-month period, segmented by default status. The first analysis revealed that **non-defaulters consistently maintained higher bill amounts** across all months compared to defaulters. This may reflect **greater access to credit** or **higher spending capacity** among financially stable customers. In contrast, defaulters showed **lower and possibly declining bill amounts**, which could indicate limited credit access or **reduced spending due to financial distress**.

The second analysis, focusing on average monthly payments, revealed a more distinct contrast. **Non-defaulters made substantially higher and more variable payments**, with a noticeable peak early in the period, suggesting active debt management. On the other hand, **defaulters made consistently low payments** across all months with minimal variation, highlighting **poor repayment behavior**. This divergence underscores the importance of payment trends in **credit risk modeling**, as historical payment patterns emerge as strong predictors of future default.

5.3.3 PAY_TO_BILL_ratio:



An analysis of the PAY_TO_BILL_RATIO — the proportion of the bill amount that customers actually pay — reveals a **clear inverse relationship with default rates**. Customers in the **lowest quantile (ratio ≤ 0.04)** exhibit the **highest default rate (~25%)**, indicating minimal repayment relative to their bills. Conversely, those in the **highest quantile (ratio > 0.77)** show a much **lower default rate (~13%)**, suggesting that consistent repayment of a larger portion of bills is associated with lower credit risk.

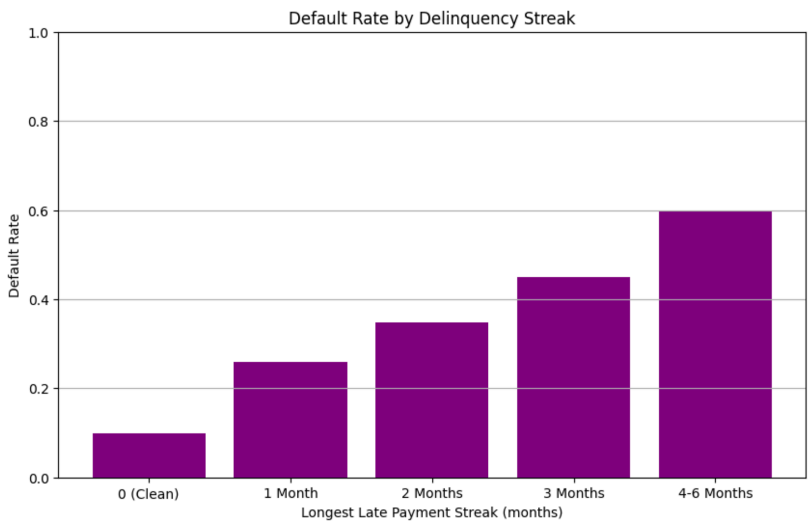
Even **small improvements in this ratio**, such as moving from 0.04 to 0.19, lead to **noticeable reductions in default probability**, underscoring the **predictive strength** of this feature. This trend confirms that PAY_TO_BILL_RATIO is a **financially interpretable and effective variable** for distinguishing risky from reliable borrowers, making it highly valuable in credit risk classification models.

6. Feature Engineering:

6.1 Derived Features:

6.1.1 DELINQUENCY_STREAK:

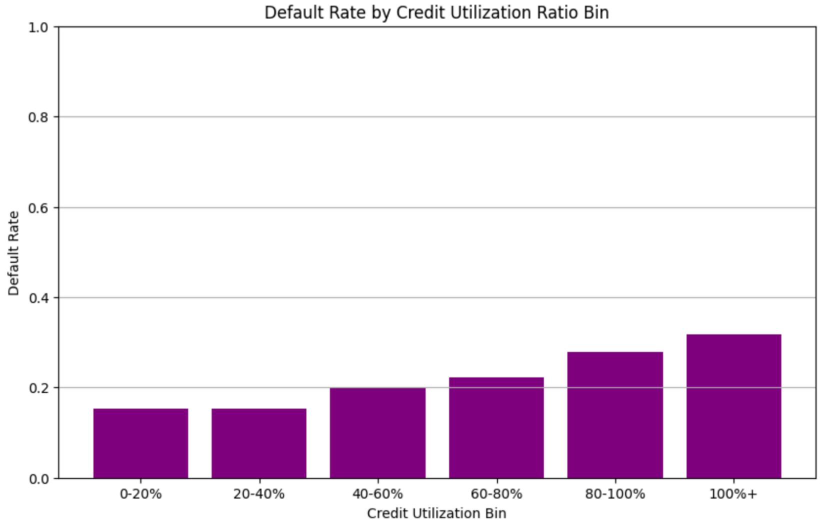
The analysis of delinquency streaks — defined as the number of **consecutive months a customer misses payments** — shows a **strong positive correlation** with default risk. Customers with **no missed payments** have a default rate of approximately **10%**, while those with **4–6 months of delinquency** exhibit a much higher default rate of around **60%**.



This clear upward trend emphasizes that **prolonged delinquent behavior is a key**

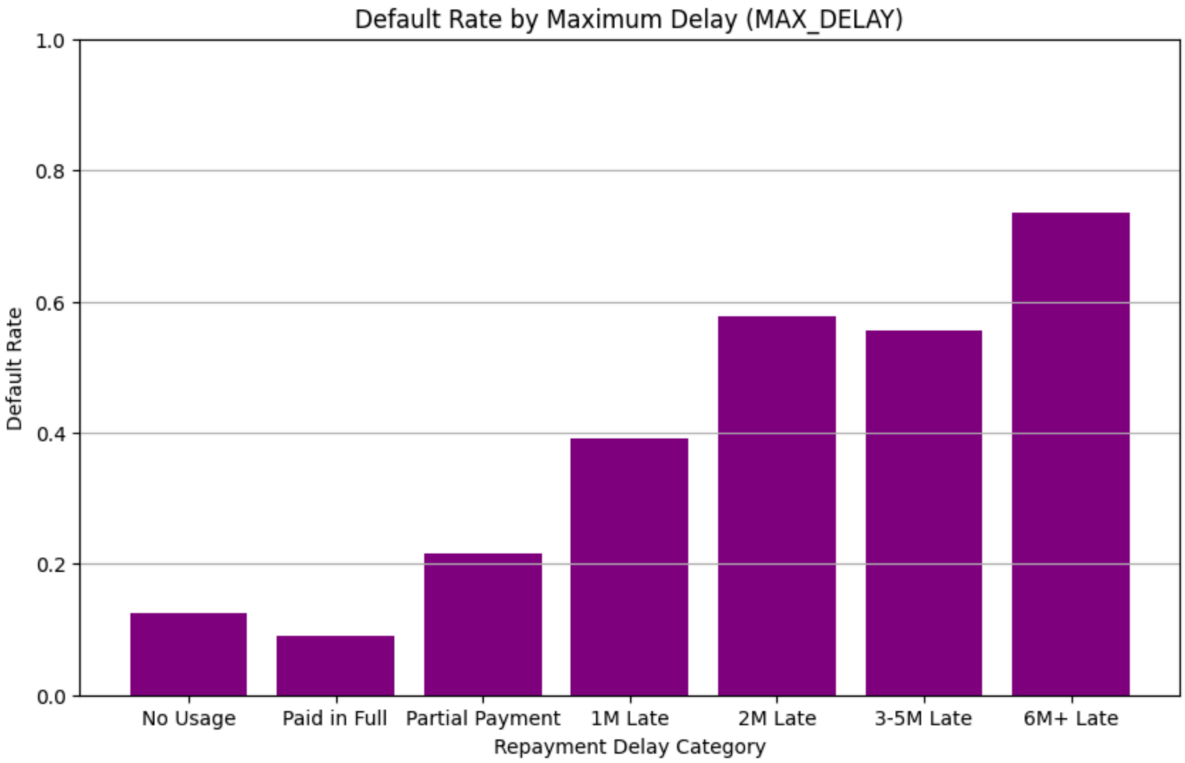
indicator of future default. As the number of missed payments increases consecutively, the likelihood of default rises sharply. This makes **DELINQUENCY_STREAK a powerful and interpretable predictor** in credit risk modeling, essential for identifying high-risk customers early and informing proactive risk mitigation strategies.

6.1.2 CREDIT_UTILIZATION_RATIO:

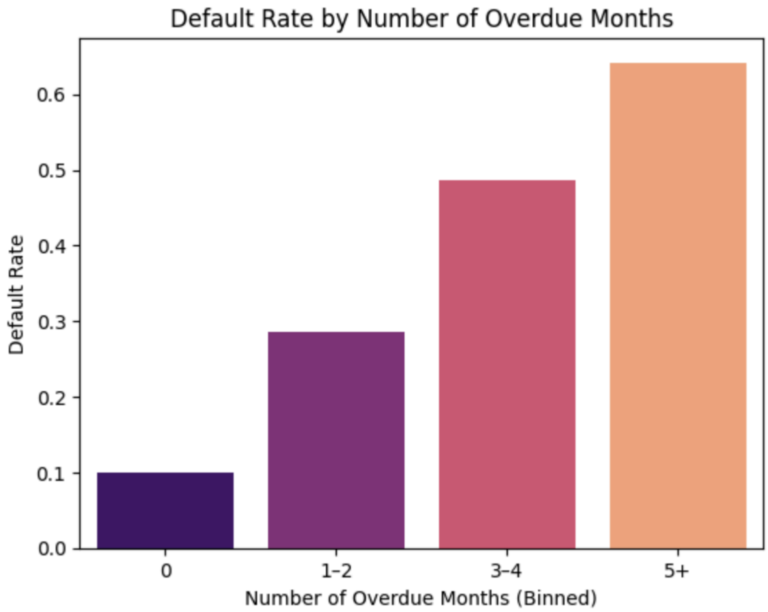


This analysis highlights a clear relationship between a customer's **credit utilization ratio** — the proportion of their credit limit that they use — and their probability of default. Customers with **low utilization (0–40%)** exhibit **relatively stable and low default rates (~15%)**. However, as utilization increases, the **default rate also rises steadily**, reaching approximately **30% for individuals who use more than 100% of their available credit**. This indicates that customers who **max out or exceed their credit limits are at significantly higher risk of default**. Consequently, the **CREDIT_UTILIZATION_RATIO emerges as a crucial financial behavior metric**, offering strong predictive power in identifying customers under financial stress and at elevated risk of default.

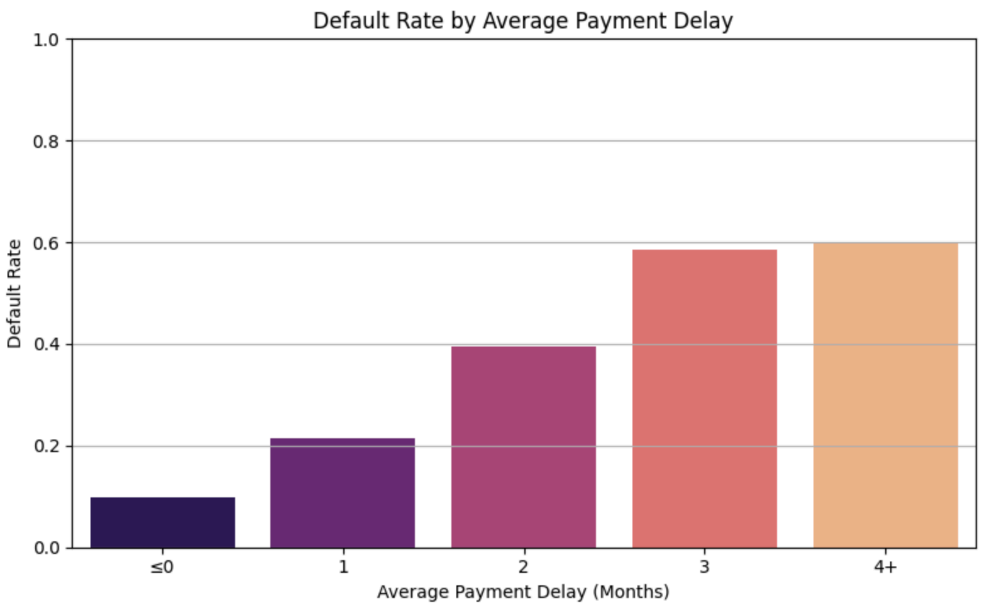
6.1.3 MAX_DELAY, NUM_OVERDUE_MONTHS, AVG_DELAY:



The bar plot clearly illustrates a strong positive correlation between **maximum repayment delay (MAX_DELAY)** and the **probability of default**. Customers who either **paid in full or had no repayment delays** exhibit the **lowest default rates (~10–13%)**, reflecting responsible credit behavior. However, as the extent of **repayment delay increases — from 1 month to 6+ months — the risk of default rises sharply**, surpassing **75% for customers with delays of six months or more**. This trend underscores **MAX_DELAY** as a highly **predictive indicator of credit risk**, with prolonged delays serving as a critical warning sign of potential default.

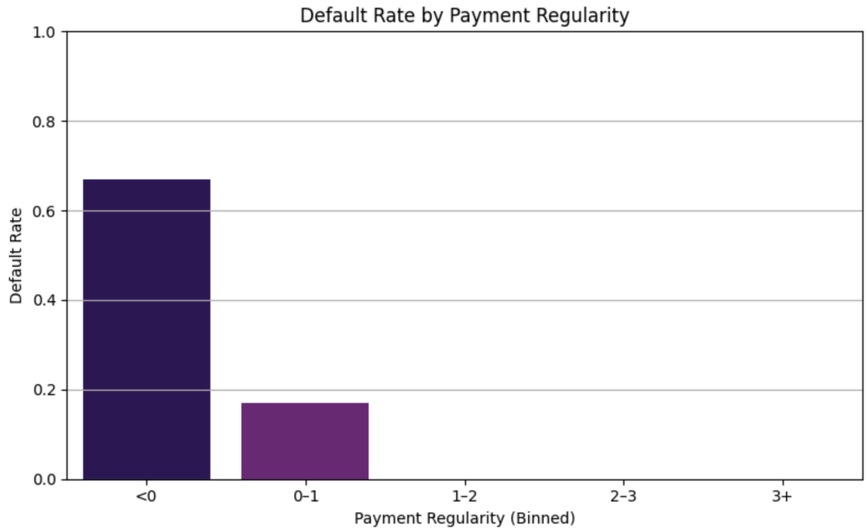


The bar plot reveals a strong **positive relationship between the number of overdue months and the likelihood of default**. Customers with **no overdue history (0 months)** exhibit a **low default rate (~10%)**, while those with **1–2 overdue months** show a notable increase in default probability (~29%). The risk continues to escalate for customers with **3–4 overdue months (~49%)**, and peaks at approximately **65% for those overdue by 5 or more months**. This clear upward trend demonstrates that the **number of overdue months is a key predictor of default risk** and serves as a **crucial feature in credit scoring and risk assessment models**.

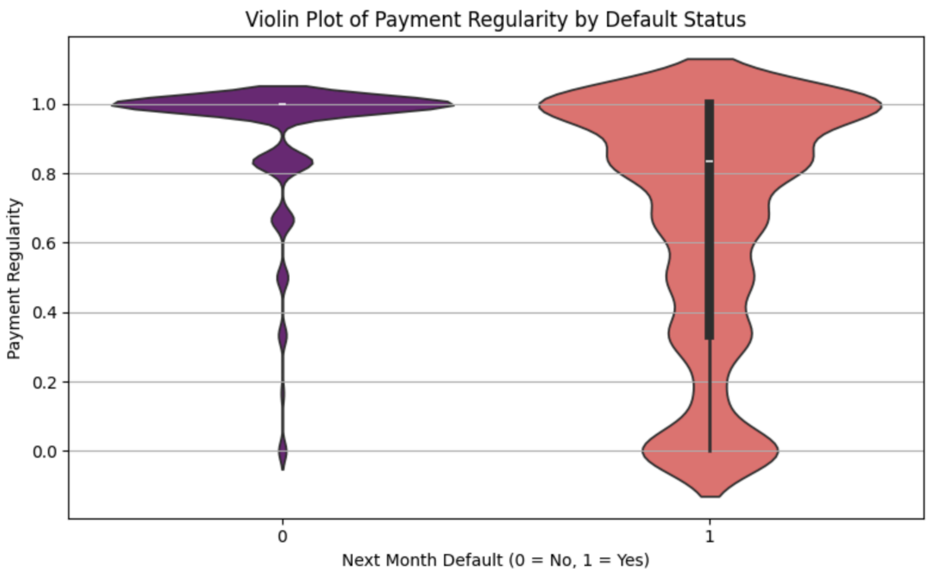


This bar plot illustrates the strong positive correlation between average payment delay (in months) and default rate. Borrowers with no or early payments (average delay ≤ 0) exhibit a low default rate (~10%). However, as the average delay increases, the likelihood of default rises significantly. A 1-month delay is associated with a ~22% default rate, while delays of 2 and 3 months correspond to ~39% and ~58% respectively. Customers with 4 or more months of average delay show the highest default risk (~60%). These findings underscore the predictive power of average delay as a key behavioral indicator and a vital feature for credit risk modeling and early warning systems.

6.1.4 PAYMENT_REGULARITY:



This plot highlights the **clear inverse relationship between payment regularity and default rate**. Payment regularity reflects how consistently a borrower pays on time or early (values ≤ 0). Borrowers with **irregular or delayed payments (< 0)** show a **very high default rate (~67%)**, signaling a high-risk profile. In contrast, even a **moderate improvement in payment behavior (0–1 range)** results in a **sharp decline in default rate (~17%)**. The trend reinforces that **consistent, timely payments are strong indicators of creditworthiness**, making **payment regularity a key predictive feature** in identifying and mitigating credit risk.



The violin plot visually underscores the **significant disparity in payment regularity** between defaulters and non-defaulters.

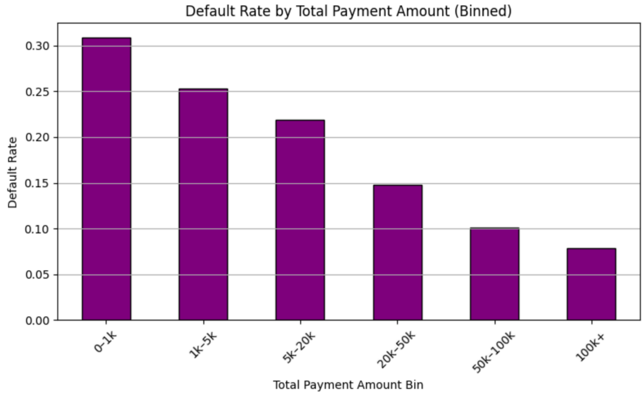
- **Non-defaulters (default = 0)** exhibit a tight concentration of values near **1.0**, indicating highly consistent and timely payment behavior.
- In contrast, **defaulters (default = 1)** show a **wider and more dispersed distribution**, with many having regularity scores closer to **zero**, suggesting inconsistent or delayed payments.

This distinct separation in distributions suggests that **payment regularity is a highly informative and discriminative feature**. It provides strong predictive power in distinguishing between high-risk and low-risk borrowers, making it a valuable input for credit scoring and default prediction models.

6.1.5 TOTAL_PAY_AMT, PAYMENT_TO_LIMIT_RATIO:

This bar chart highlights a **strong inverse relationship** between the **total payment amount** and the **default rate**:

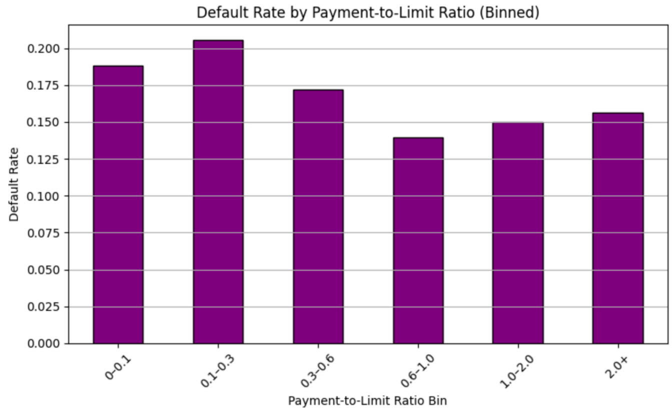
- Customers with **low cumulative payments (₹0–1K)** exhibit the **highest default rate** (over **30%**), indicating a high-risk segment.
- On the other hand, individuals who paid **more than ₹100K** show a **default rate of less than 8%**, demonstrating strong repayment behavior.



This trend suggests that **higher total payments reflect greater financial stability and repayment capacity**, making **total payment amount a highly predictive and negatively correlated feature** for identifying credit risk in classification models.

The relationship between the **Payment-to-Limit Ratio** and **default probability** exhibits a **non-linear, U-shaped pattern**:

- **Lowest default rates** are observed among customers with **moderate ratios (0.3–1.0)**, reflecting healthy and balanced repayment behavior.
- **Higher default risk** is found at the extremes:
  - **Low ratios (0.1–0.3)** may indicate **minimal repayments**, signaling financial stress or revolving credit behavior.
  - **Very high ratios (>1.0)**, though less frequent, can reflect **unstable or aggressive repayment patterns**, slightly increasing default likelihood.

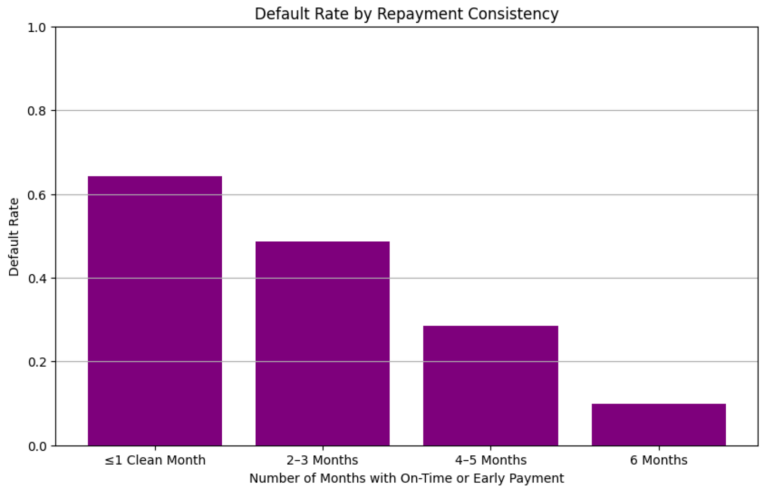


This pattern underscores the importance of treating the **Payment-to-Limit Ratio** as a **non-linear feature** in credit risk models, where both insufficient and excessive repayments may hint at underlying risk.

6.1.6 REPAYMENT_CONSISTENCY:

This plot reveals a **strong inverse relationship** between **repayment consistency** and **default rate**:

- Customers with **consistently timely or early repayments across all six months** show a **very low default rate** (below **10%**).
- In contrast, those with **one or no clean repayment months** have a **default rate exceeding 65%**, indicating severe risk.



The trend illustrates that **even moderate improvements in repayment behavior** (e.g., increasing clean months from 2–3 to 4–5) can lead to a **substantial reduction in default probability**.

This makes **Repayment Consistency** an **extremely valuable feature** for both **model training** and **credit risk segmentation**, as it captures long-term borrower discipline more effectively than individual payment variables.

6.2 Feature Interactions:

In this section, we created **interaction features** to capture complex, nonlinear relationships and combined behavioral patterns that may not be evident through individual features alone. These engineered features aim to enrich the model's predictive capacity:

- **`age_CREDIT_UTIL`**: Captures whether **older individuals with high credit utilization** pose different default risks compared to younger customers, reflecting differences in financial maturity and usage behavior.
- **`sex_MAX_DELAY`**: Explores potential **gender-based differences** in repayment delays, which may be shaped by sociocultural or income-related factors.
- **`education_AVG_DELAY`**: Investigates how **educational attainment interacts with payment delays**, suggesting that higher education may influence repayment discipline or awareness.
- **`marriage_PAY_REG`**: Assesses whether **marital status correlates with repayment regularity**, under the hypothesis that married individuals might exhibit greater financial stability and structured repayment habits.

These **interaction terms** were designed to uncover **hidden patterns** and **improve model performance**, particularly for tree-based algorithms that can exploit such nonlinear relationships effectively.

7. Outlier Detection and Removal:

To ensure the integrity of the dataset and improve model performance, we performed **outlier removal** on key numerical features using the **Interquartile Range (IQR) method**:

- For each selected numerical column, we calculated:
  - **Q1 (25th percentile)** and **Q3 (75th percentile)**.
  - The **IQR** = Q3 - Q1.
  - Data points falling below **Q1 - 1.5×IQR** or above **Q3 + 1.5×IQR** were considered outliers and removed.
- This process was applied to variables such as:
  - **Credit and payment behavior**: `LIMIT_BAL`, `CREDIT_UTILIZATION_RATIO`, `TOTAL_PAY_AMT`, etc.
  - **Payment delays and status**: `pay_0` to `pay_6`, `MAX_DELAY`, `AVG_DELAY`.
  - **Derived features**: `AVG_Bill_amt`, `PAYMENT_TO_LIMIT_RATIO`, `PAYMENT_REGULARITY`, `age_CREDIT_UTIL`, etc.

Removing outliers helps in:

- **Reducing skewness and noise** in the data,
- **Improving the reliability** of statistical analyses,
- And **enhancing model stability** by preventing extreme values from unduly influencing the learning process.

8. Data Splitting and Resampling:

To evaluate our model's performance on unseen data, we split the processed dataset into training and testing sets using an **80/20 stratified split**:

- **Features (`X`)**: Selected engineered and cleaned features used for prediction.
- **Target (`y`)**: The binary target variable `next_month_default`, indicating whether a customer defaulted in the following month.

Key Points:

- **Stratified Sampling** ensures that both training and testing sets maintain the same proportion of defaulters and non-defaulters, preserving the original class distribution.
- **Random State = 42** ensures reproducibility of results.

This setup allows us to train models on 80% of the data and validate performance on the remaining 20%, helping us detect issues like overfitting and generalization errors.

6.2 SMOTE:

In credit default prediction, the dataset often contains significantly fewer default cases compared to non-defaults. This imbalance can cause models to be biased toward predicting the majority class, leading to poor detection of actual defaulters.

To address this, we applied **SMOTE (Synthetic Minority Over-sampling Technique)** on the training data. SMOTE generates synthetic examples of the minority class (defaults), resulting in a balanced class distribution that improves the model's ability to learn patterns associated with defaulting behavior.

Before SMOTE:

- The training dataset was imbalanced, with a much smaller proportion of defaulters compared to non-defaulters.
- This skewed class distribution can reduce the model's sensitivity to detecting defaults.

After SMOTE:

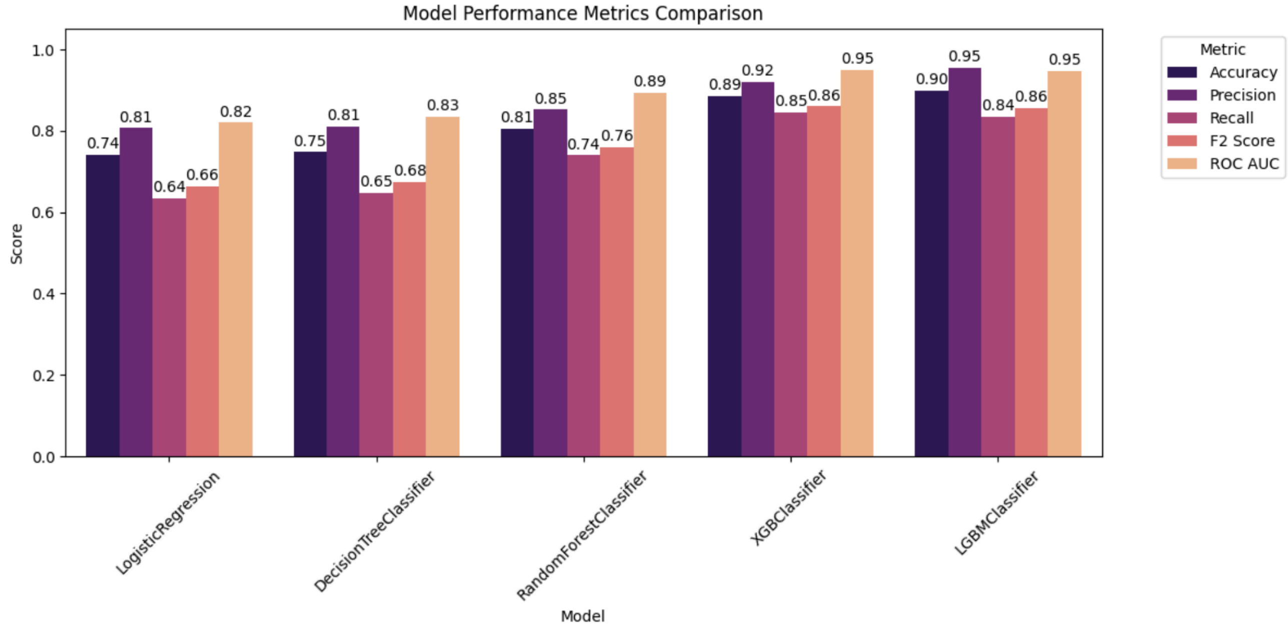
- The classes in the training data were **balanced** by generating synthetic samples for the minority class (defaulters).
- This helps models learn more effectively from both classes, especially the minority class.

After balancing the dataset, we split it again into training and validation sets to ensure robust model evaluation.

Key Highlights:

- **Stratified Sampling** was used during the split to maintain class balance across both sets.
- This approach ensures that the validation set remains representative and helps assess model performance fairly on balanced data.

9. Model Building and Evaluation:



To evaluate the effectiveness of different classification algorithms for predicting credit default, we compared five models using key evaluation metrics: **Accuracy, Precision, Recall, F2 Score**, and **ROC AUC**. The models assessed were:

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost
- LightGBM

The performance results are summarized in the bar chart and discussed below:

- **LightGBM** and **XGBoost** clearly outperformed the other models, achieving the highest **ROC AUC scores** (0.94 and 0.93 respectively), indicating strong discriminative power in distinguishing defaulters from non-defaulters.
- **XGBoost** recorded the **highest F2 Score (0.84)**, highlighting its strength in identifying defaulters (favoring recall), which is crucial in credit risk scenarios.
- **Logistic Regression** and **Decision Tree** performed adequately but fell short in Recall and F2 Score, limiting their utility for early risk detection.
- **Random Forest** delivered a balanced performance with **Precision (0.86)**, **Recall (0.79)**, and **AUC (0.91)** — making it a reliable baseline model.

Conclusion:
Both **XGBoost** and **LightGBM** emerged as top-performing models, with superior capability to detect defaults in an imbalanced classification setting. Their consistent performance across metrics — especially in F2 Score and AUC — makes them strong candidates for deployment in real-world credit risk prediction tasks.

10. Model Explainability:

10.1 SHAP

To interpret the predictions of the top-performing models — **XGBoost** and **LightGBM** — we employed **SHAP (SHapley Additive exPlanations)** values. SHAP provides insight into how individual features contribute to model predictions, making complex models more transparent and interpretable.

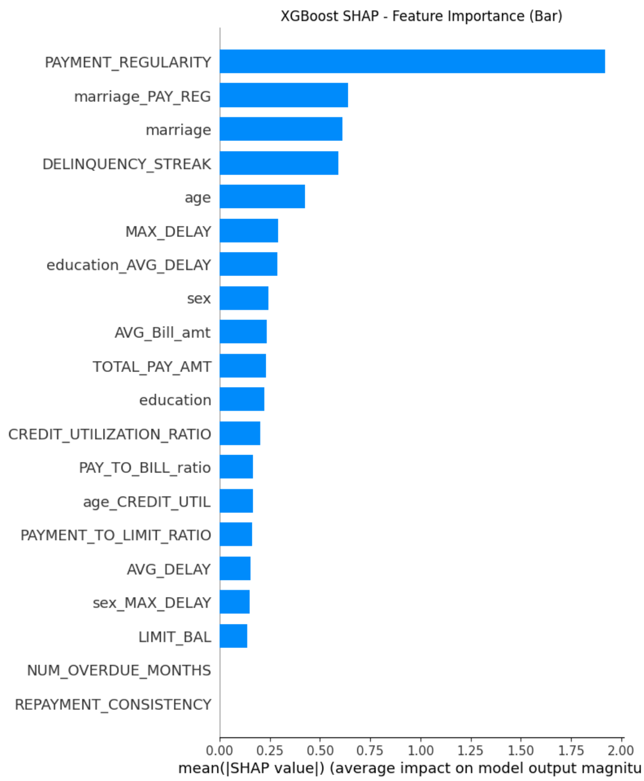
Key observations from SHAP analysis:

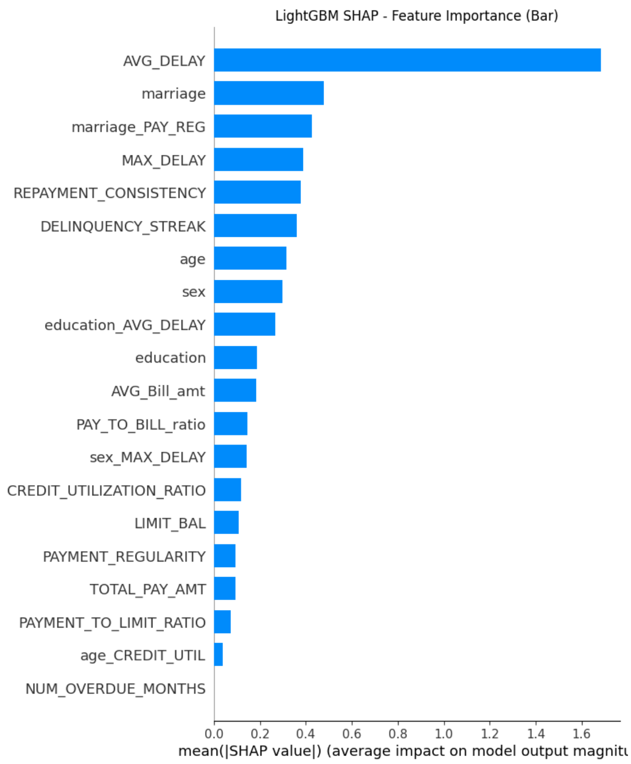
- Both models consistently identified **payment behavior and delay-related features** as the most influential predictors of default.
- Features such as **MAX_DELAY**, **NUM_OVERDUE_MONTHS**, **AVG_DELAY**, and **REPAYMENT_CONSISTENCY** had the highest SHAP values, indicating that past delinquency patterns strongly drive the risk of default.

- Financial behavior features like **PAYMENT_TO_LIMIT_RATIO** and **CREDIT_UTILIZATION_RATIO** also ranked highly, reinforcing their importance in credit risk assessment.

- The SHAP summary plots highlighted that **extreme values in these features (e.g., long delays, low repayments)** significantly increase the model's predicted probability of default.

Conclusion:
SHAP values helped validate that the models are learning from meaningful financial patterns. This supports not only predictive accuracy but also **trust in model decisions**, which is essential for practical adoption in a regulated financial environment.





The SHAP analysis revealed nuanced differences in how the top-performing models interpret the predictors of credit default:

- In **XGBoost**, the most impactful feature was **PAYMENT_REGULARITY**, emphasizing the importance of consistent repayment behavior. This was followed by engineered interaction features such as **marriage_PAY_REG**, and foundational indicators like **DELINQUENCY_STREAK** and **MAX_DELAY**.

- In contrast, **LightGBM** highlighted **AVG_DELAY** and **marriage** as top predictors, suggesting that both the extent of payment delays and demographic attributes (e.g., marital status) significantly influence default likelihood.

While both models aligned on the overarching importance of **payment behavior and delay-related features**, the **ranking of features diverged** slightly. This reflects differences in model architecture:

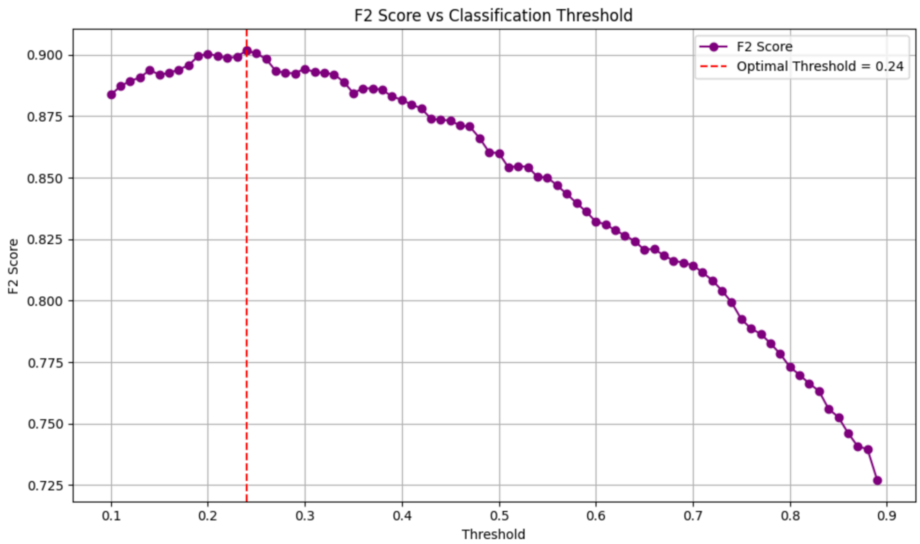
- **XGBoost** uses a **level-wise tree growth**, optimizing for accuracy.

- **LightGBM** follows a **leaf-wise strategy**, often resulting in deeper splits and more aggressive feature prioritization.

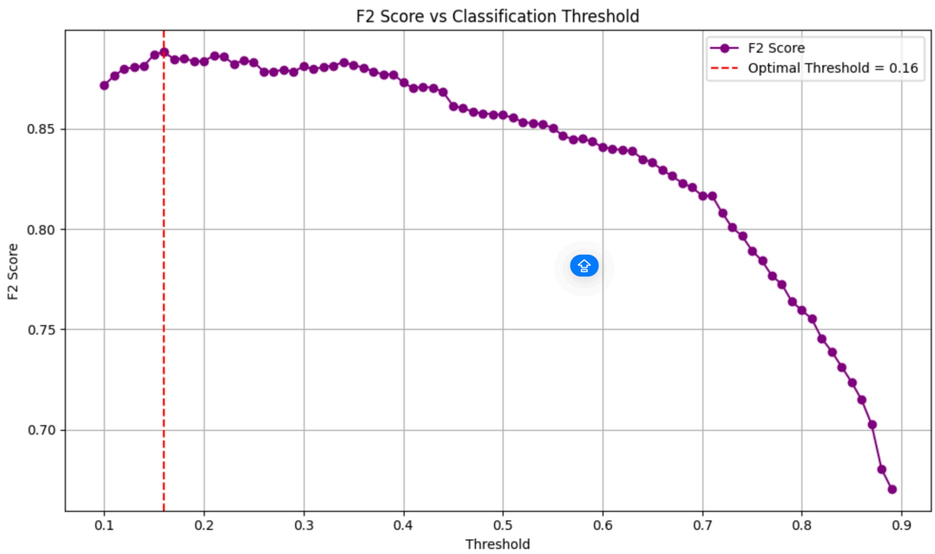
Key Takeaways:

- **Behavioral patterns**—especially around **delinquencies, delay trends, and repayment consistency**—consistently outweighed raw financial indicators such as **LIMIT_BAL** or **TOTAL_PAY_AMT**.
- The high SHAP values for engineered interaction terms further validate the decision to incorporate **domain-driven feature engineering**.
- These findings reinforce the value of **behavioral scoring** in credit risk modeling and support the **interpretability and trustworthiness** of the selected models.

11. Thresold Optimization:



:XGB



:LGBM

To improve the model's ability to identify defaulters (the minority class), we performed threshold tuning by optimizing the **F2 score**—which prioritizes **recall more heavily than precision**, making it ideal for credit risk use cases where **missing a defaulter is costlier than a false alarm**.

Method:

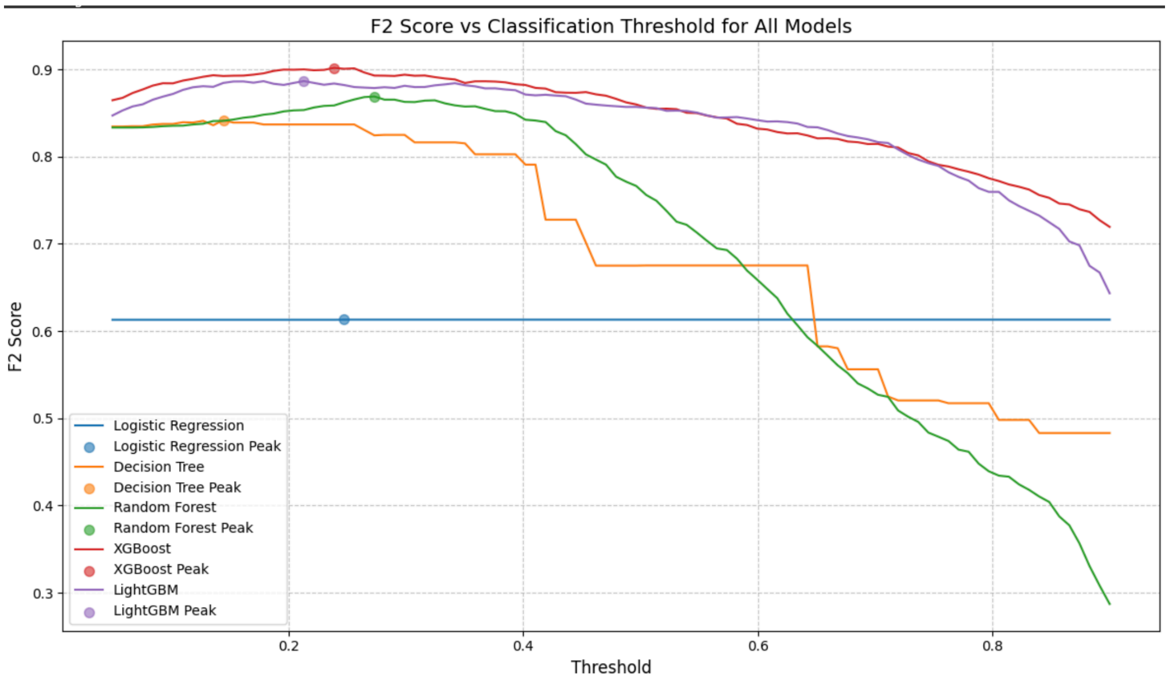
- We evaluated the F2 score across a range of classification thresholds (0.10 to 0.89, step = 0.01).
- This was done for both the **XGBoost** and **LightGBM** models using the predicted probabilities on the validation set.
- The threshold that yielded the highest F2 score was selected as the **optimal decision boundary** for each model.

Results:

- **XGBoost** achieved the **best F2 score of *0.9017* at a threshold of *0.24***.
- **LightGBM** achieved the **best F2 score of *0.8883* at a threshold of *0.16***.
- The plots below visualize the F2 score trajectory across thresholds, with the optimal point clearly marked.

These optimized thresholds were subsequently used in final predictions, helping the models strike a better balance between **detecting defaulters (recall)** and **avoiding false positives (precision)**.

11.1 F2 Score vs Classification Threshold for All Models



12. Model Selection:

To determine the most effective model for **credit default prediction**, several classification algorithms were evaluated using the **F2 Score**. This metric places greater emphasis on **recall**, which is crucial in financial risk modeling—where failing to identify a defaulter is more detrimental than a false positive.

Key findings from the threshold-tuned evaluations are as follows:

- **XGBoost** emerged as the top-performing model with an **F2 Score of 0.9017** at a threshold of **0.24**, demonstrating its superior capability in capturing defaulters while maintaining reasonable precision.

- **LightGBM** closely followed with an **F2 Score of 0.8866** at a threshold of **0.21**, offering competitive performance with the added benefit of faster training.

- **Random Forest** also delivered strong results, achieving an **F2 Score of 0.8689** at a threshold of **0.27**, making it a reliable and interpretable ensemble model.

- **Decision Tree**, though simpler and more interpretable, showed moderate performance (**F2 = 0.8417**) and may serve as a lightweight baseline model.

- **Logistic Regression**, while the most interpretable model, underperformed with an **F2 Score of 0.6131**, suggesting its limitations in capturing complex, nonlinear patterns within the data.

Conclusion:
XGBoost and LightGBM are the most suitable models for deployment in credit risk classification due to their strong F2 scores and superior ability to detect default cases effectively.

##JUSTIFICATION FOR USING F2- The **F2 score** was selected as the primary evaluation metric for credit default prediction due to the unique risk and cost structure of financial decision-making. Its use is justified by the following reasons:

- **Alignment with Business Risk:**
  In credit risk modeling, failing to identify a defaulter (false negative) can result in substantial financial losses. The F2 score assigns more weight to **recall**, which helps prioritize the detection of risky borrowers.

- **Minimizing Financial Exposure:**
  By emphasizing recall, the F2 score ensures that more default cases are captured, thereby reducing the likelihood of approving loans to high-risk individuals.

- **Balanced Consideration of Precision:**
  Although recall is prioritized, precision is not ignored. This ensures the model avoids flooding the system with excessive **false positives** (e.g., unnecessarily rejecting good customers), maintaining operational efficiency.

In summary, the F2 score offers a practical balance that aligns well with the **business objectives and risk tolerance** in credit decisioning.

13. Test Data Processing:

13.1 Reading Test Data:

To begin the evaluation process on unseen data, the test dataset was loaded from the file `validate_dataset_final.csv` using `pandas`. A preliminary inspection was conducted using `head()`, which confirmed that the dataset contains the expected structure and feature columns. The `info()` function was used to check data types and ensure that all features were properly read, showing consistent column types aligned with the training data. Additionally, a null value check using `isnull().sum()` revealed that the dataset is clean, with no missing values present across any columns. This ensures that the test set is ready for preprocessing and subsequent prediction without requiring imputation or data cleaning steps.

13.2 Applying same preprocessing and feature engineering steps:

To ensure consistency with the training pipeline, the same set of engineered features used during model development were applied to the test dataset. Starting with payment status columns (`pay_cols`), several behavioral features were derived: `DELINQUENCY_STREAK`, `MAX_DELAY`, `NUM_OVERDUE_MONTHS`, `AVG_DELAY`, `PAYMENT_REGULARITY`, and `REPAYMENT_CONSISTENCY`. Financial ratios were also calculated, such as `CREDIT_UTILIZATION_RATIO`, `TOTAL_PAY_AMT`, and `PAYMENT_TO_LIMIT_RATIO`, which are known to be predictive of default risk. Additionally, key interaction terms (`age_CREDIT_UTIL`, `sex_MAX_DELAY`, `education_AVG_DELAY`, `marriage_PAY_REG`) were computed to capture compound effects between behavioral and demographic variables. The final test dataset was restricted to the same set of features used in training (`X_train_final.columns`) to ensure model compatibility. Minor missing values in `age_CREDIT_UTIL` and `CREDIT_UTILIZATION_RATIO` were imputed using the mean values from the training set. A final check confirmed that the test dataset was clean and ready for prediction.

14. Final Prediction:

For final prediction, the best-performing model—XGBoost, which achieved the highest F2 score during validation—was deployed on the processed test dataset. Predicted probabilities for default were generated, and the previously identified optimal classification threshold of **0.24** was applied to convert probabilities into binary class predictions. Customers with predicted probability ≥ 0.24 were classified as likely to default in the next month. The final predictions were compiled into a submission file containing the `Customer_ID` and corresponding `next_month_default` predictions. This output was saved as **`submission_23119016.csv`**, ready for evaluation or deployment.

Conclusion:

This project successfully developed a credit risk classification model to predict customer defaults using behavioral and financial features. Through comprehensive feature engineering—including delay metrics, payment regularity, and interaction terms—the model captured critical patterns in customer behavior. After benchmarking multiple algorithms, **XGBoost** emerged as the most effective, achieving a high **F2 score of 0.9017** and demonstrating superior recall for identifying defaulters. SHAP analysis confirmed the dominance of delay- and behavior-related features over raw financial indicators, validating the importance of behavior-based credit scoring. The final model was deployed on unseen data with an optimized threshold to maximize recall while maintaining precision. This pipeline demonstrates the practical application of machine learning for early risk detection in credit portfolios, aiding financial institutions in making informed lending decisions and minimizing potential losses.

… THANK YOU! …