Credit Risk Default Analysis Report: Enabling Financially Interpretable Default Prediction

1. Overview of Approach and Modeling Strategy

This report details the process of building and evaluating a credit risk default prediction model with a core objective: to develop a **financially interpretable model** that empowers the bank to understand default patterns, facilitate early intervention, and strategically manage credit exposure. The model aims to predict "next_month_default" with high accuracy and actionable insights.

The approach involved:

1. **Data Loading and Initial Inspection**: Understanding the dataset's structure, identifying data types, and checking for missing values.
2. **Exploratory Data Analysis (EDA)**: Visualizing distributions, correlations, and relationships between variables to gain insights into customer behavior and financial indicators. Feature engineering was performed to create new, more informative variables like DELINQUENCY_COUNT (cumulative payment delays) and PAY_TO_BILL_ratio (payment efficiency).
3. **Data Preprocessing**: Handling missing values (imputation for 'age'), and splitting data into training and testing sets to ensure robust model evaluation.
4. **Model Selection and Training**: Evaluating multiple classification algorithms including Logistic Regression, Decision Tree, XGBoost, and LightGBM, focusing on their predictive power and potential for interpretability.
5. **Model Evaluation**: Assessing model performance using a suite of metrics critical for credit risk management: Accuracy, Precision, Recall, F1-score, F2-score, and ROC AUC, with a strong emphasis on minimizing false negatives.
6. **Threshold Optimization**: Selecting an optimal classification cutoff based on the F2-score to strategically balance the trade-off between identifying true defaults (recall) and minimizing false alarms (precision), directly supporting early action.
7. **Feature Importance and Financial Insights (Interpretable AI)**: Employing SHAP (SHapley Additive exPlanations) values to explain individual predictions and globally understand which variables drive default and *why*. This is crucial for building a financially interpretable model.
8. **Business Implications**: Translating granular model findings into concrete, actionable strategies for the bank to mitigate risk and optimize credit exposure.

2. EDA Findings and Visualizations

Initial data inspection revealed 25,247 entries with 27 columns. The dataset contained a mix of numerical features (float64 and int64). A minor issue was detected with missing values in the 'age' column (126 entries), which were subsequently handled through imputation.

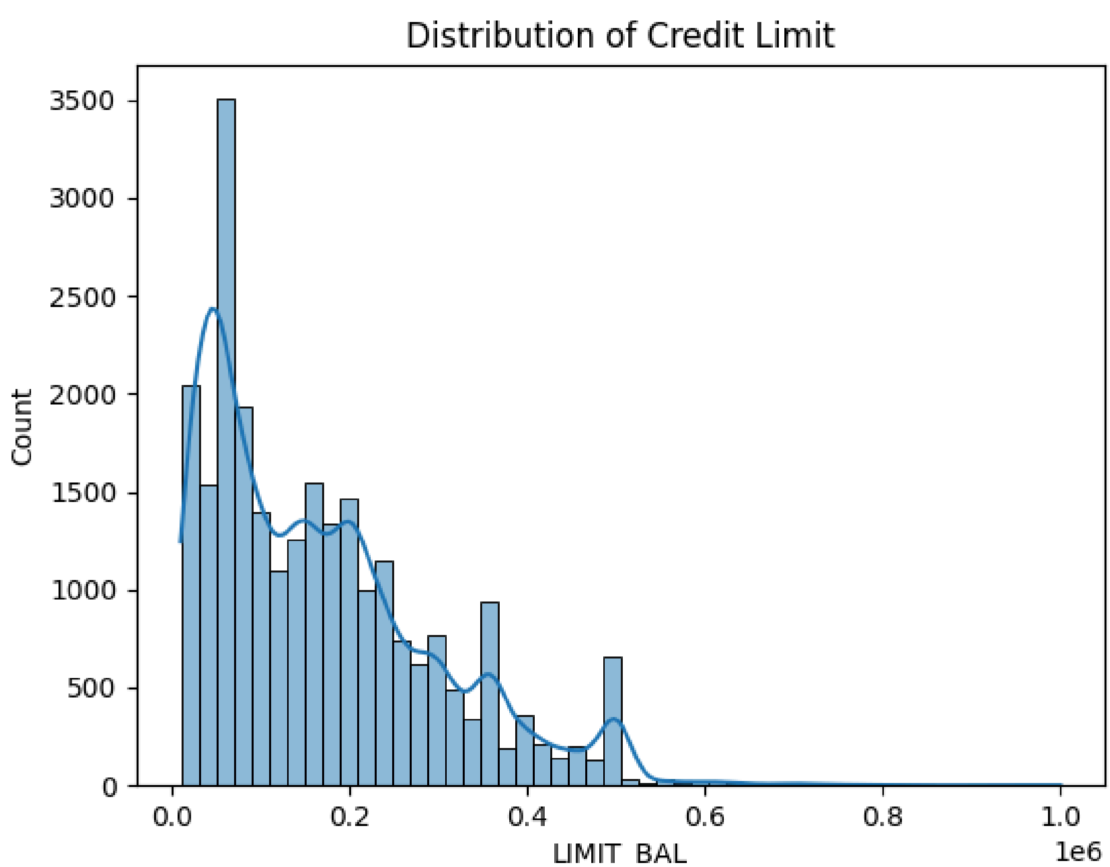
The target variable, next_month_default, showed an imbalanced distribution, which is typical for default datasets:

- **0 (No Default)**: 20,440 instances
- **1 (Default)**: 4,807 instances

This imbalance highlights the necessity of using robust evaluation metrics beyond simple accuracy, as well as techniques for interpreting model behavior for the minority class (defaults).

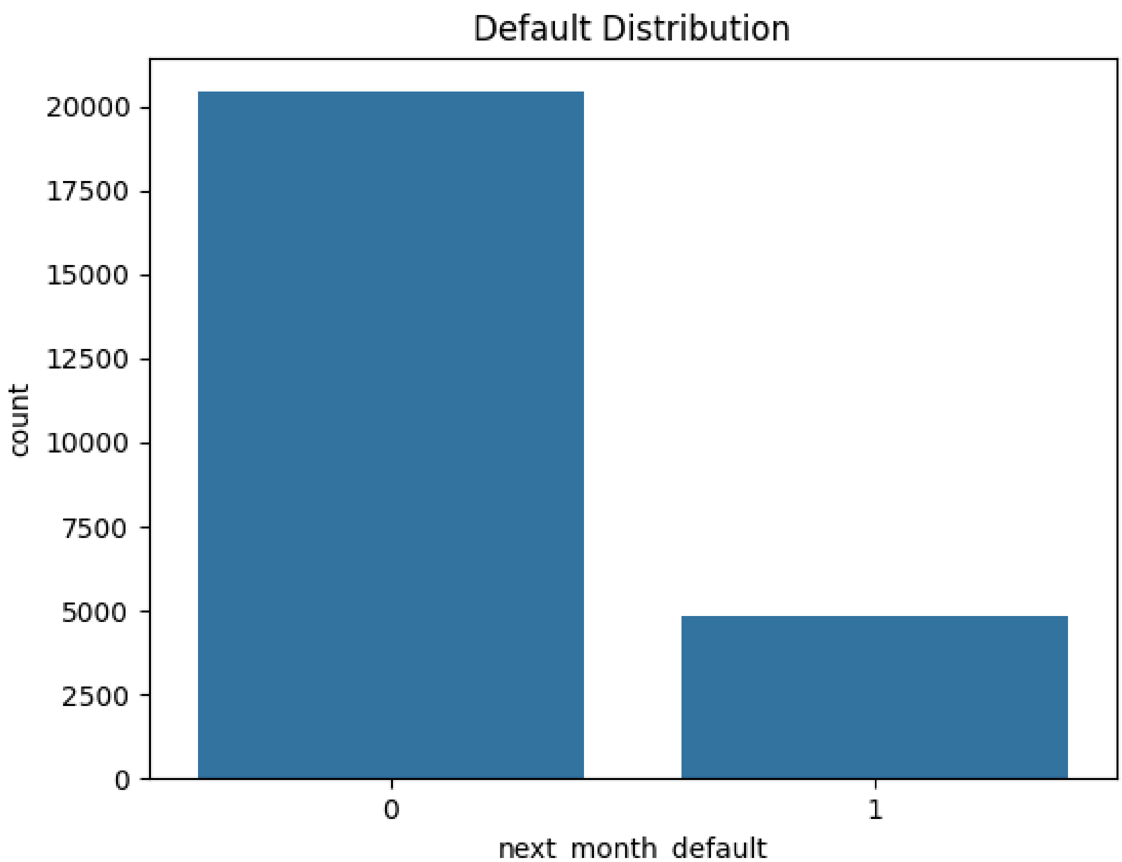
Variable Distributions

Fig 1: Distribution of Credit Limit (LIMIT_BAL)



This histogram visualizes the distribution of credit limits among customers. It shows that a significant portion of customers have credit limits in the lower ranges, with a long tail extending to higher limits, indicating diverse credit profiles.

Fig 2: Default Distribution (next_month_default)

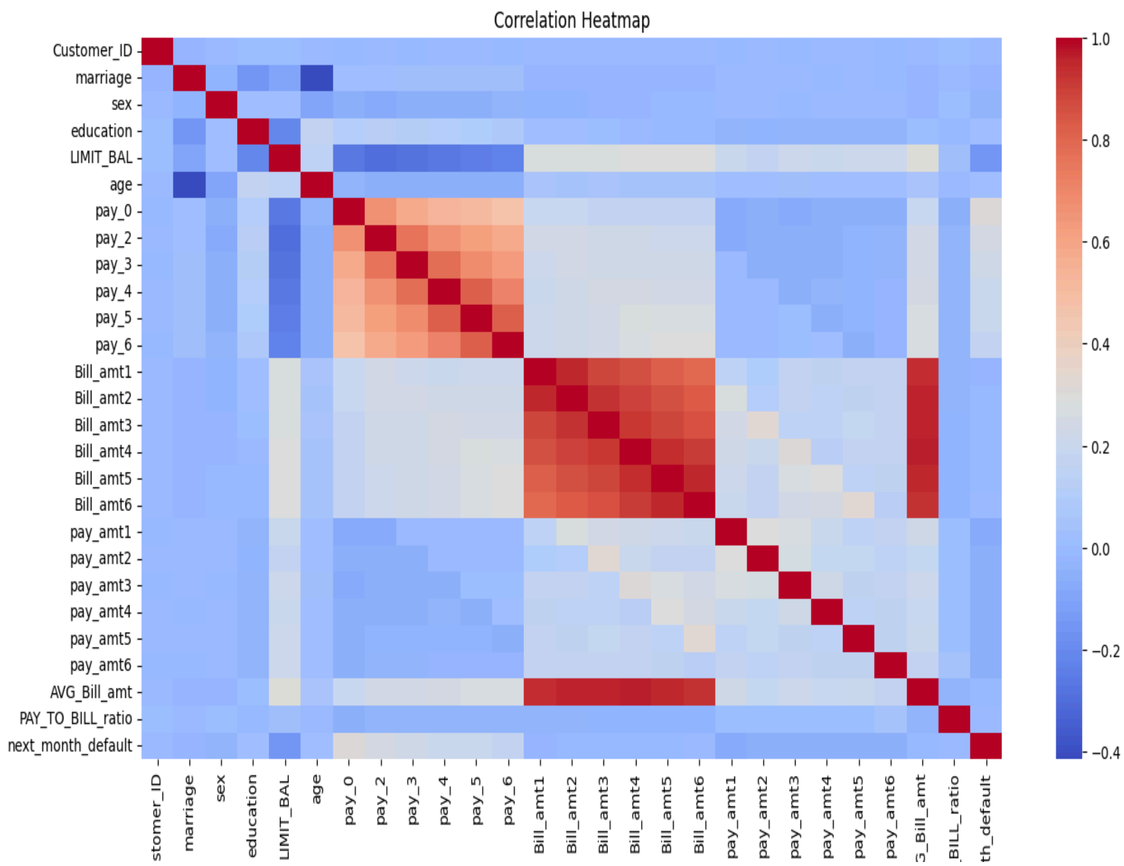


This bar chart illustrates the count of customers who did not default (0) versus those who defaulted (1) in the next month. It clearly shows the class imbalance, with a much larger number of non-defaults, underscoring the challenge and importance of accurately identifying the smaller default group.

Correlations

Fig 3: Correlation Heatmap

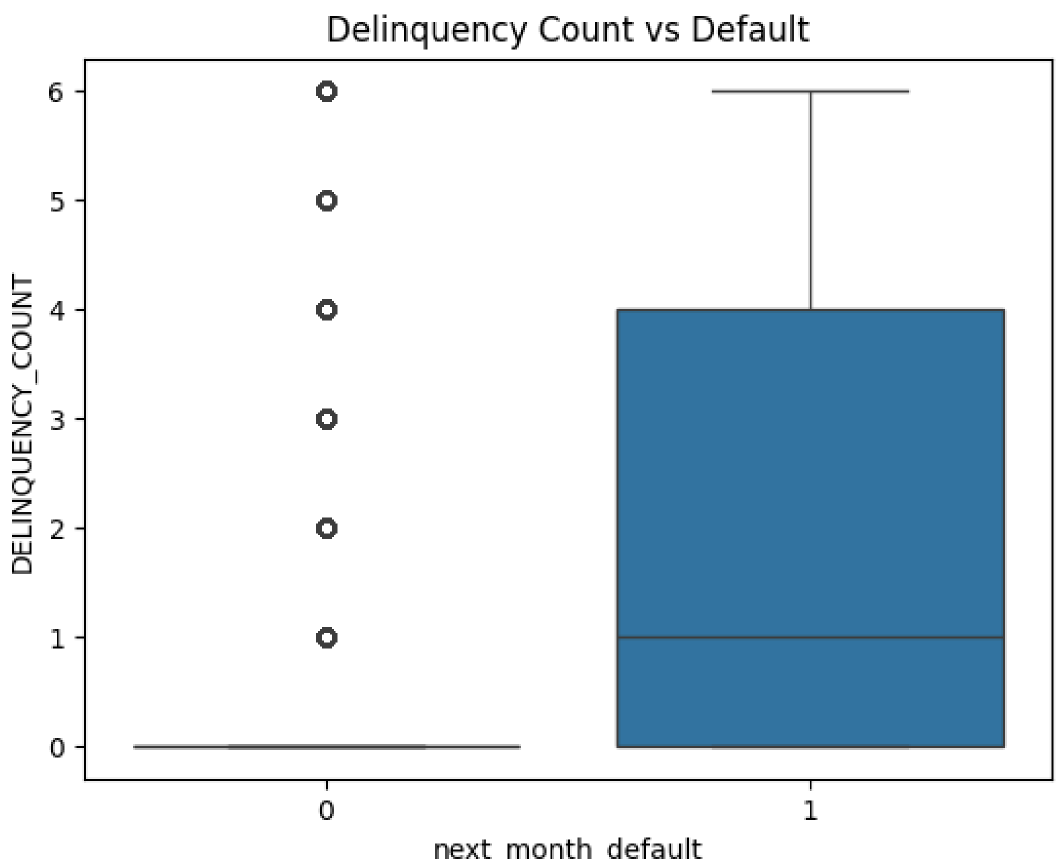
The correlation heatmap displays the Pearson correlation coefficients between all



numerical variables in the dataset. Red colors indicate positive correlation, blue indicates negative, and lighter shades indicate weaker correlation. This helps in understanding multicollinearity and relationships between features and the target variable. Key observations include:

- Strong positive correlations among pay_ (repayment status) variables (pay_0, pay_2, pay_3, etc.), indicating consistent payment behavior over months. This suggests that past payment trends are strong indicators.
- Strong positive correlations among Bill_amt variables, suggesting consistent billing amounts, which can be an indicator of consistent spending or balance accumulation.
- Inverse correlation between LIMIT_BAL and next_month_default (higher limit, lower default risk), suggesting that individuals with higher credit limits are generally more financially stable or have been assessed as lower risk.
- Positive correlation between pay_0 (repayment status in the most recent month) and next_month_default (higher payment delay indicates higher default risk). This strong direct relationship emphasizes the recency effect of payment behavior.

Fig 4: Delinquency Count vs Default



This box plot shows the distribution of the engineered DELINQUENCY_COUNT variable for both default (1) and non-default (0) groups. It visually confirms that customers who default tend to have a significantly higher DELINQUENCY_COUNT, indicating that frequent repayment delays across multiple months are a strong cumulative indicator of future default. This reinforces the importance of tracking consistent payment issues.

3. Financial Insights and Analysis of Variables Driving Default

The analysis of feature importance, particularly through SHAP (SHapley Additive exPlanations) values, provides crucial financial insights into which variables drive default predictions and *why*. This interpretability is fundamental for the bank to understand default patterns and take targeted early actions.

Fig 5: SHAP Summary Plot

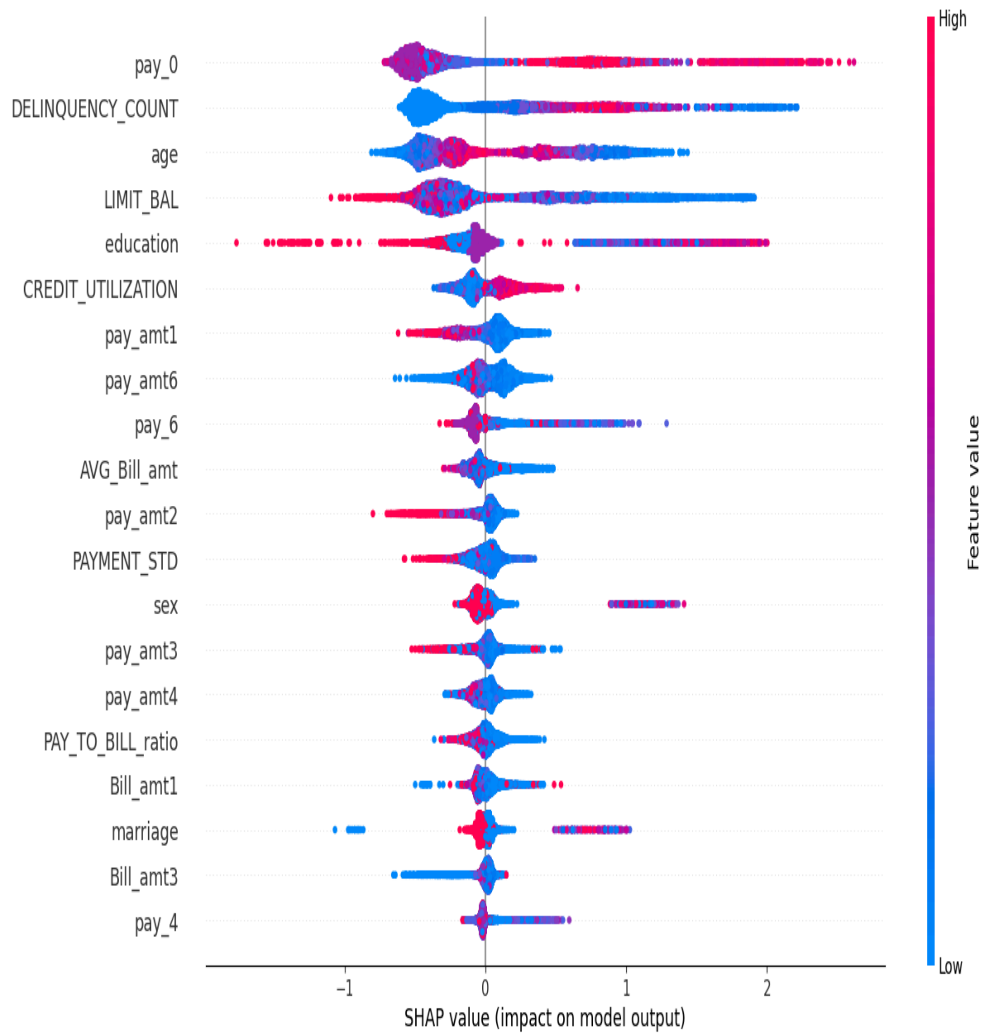


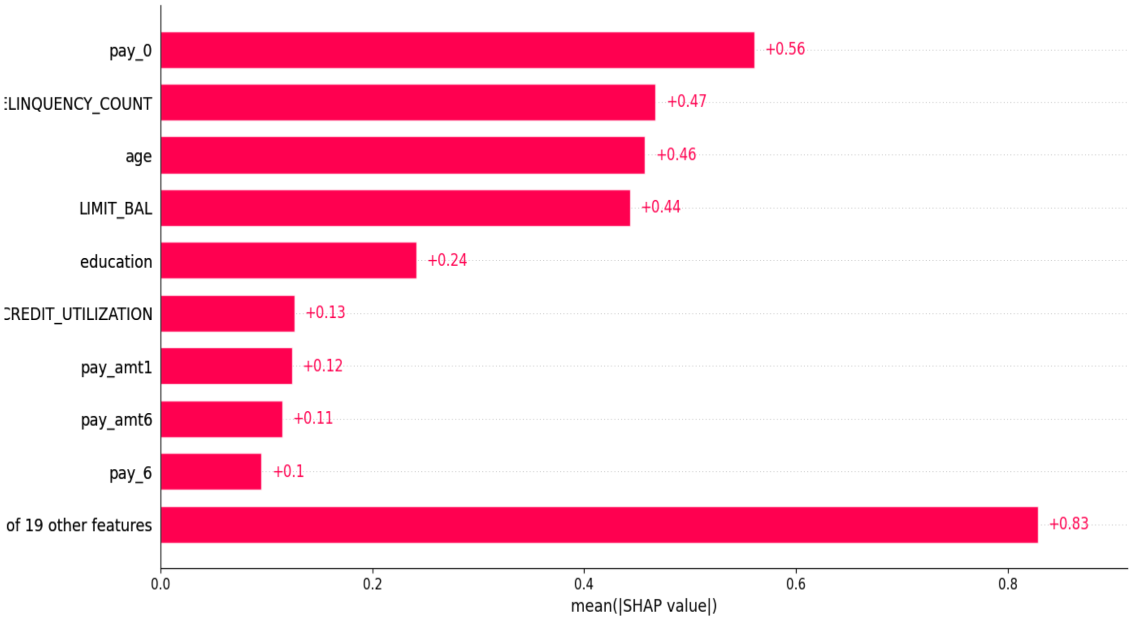
Fig 5: SHAP Summary Plot
This plot illustrates the overall impact and direction of each feature on the model's output (i.e., the probability of default). Each dot represents a single prediction (a customer); its position on the x-axis shows the SHAP value (the impact on the model's output for that specific prediction), and its color indicates the feature value (red for high, blue for low).

- **pay_0 (Repayment Status in September)**: This is the most influential feature by a significant margin. High positive SHAP values (pushing the prediction towards default) are strongly associated with high pay_0 values (indicating severe payment delays, e.g., 2+ months overdue). Conversely, low pay_0 values (on-time payments or payment in full) lead to negative SHAP values (pushing the prediction away from default). **Financial Insight:** *Recent payment delinquency is the most critical immediate indicator of default risk. Banks should closely monitor pay_0 for early warning signals.*
- **DELINQUENCY_COUNT (Engineered Feature)**: A highly influential factor,

reflecting the total number of months a payment was delayed across the previous six months. High delinquency counts (red dots to the right) strongly increase the likelihood of default. **Financial Insight:** *Beyond single month delays, a consistent pattern of overdue payments significantly elevates default risk. This metric helps identify chronic defaulters.*

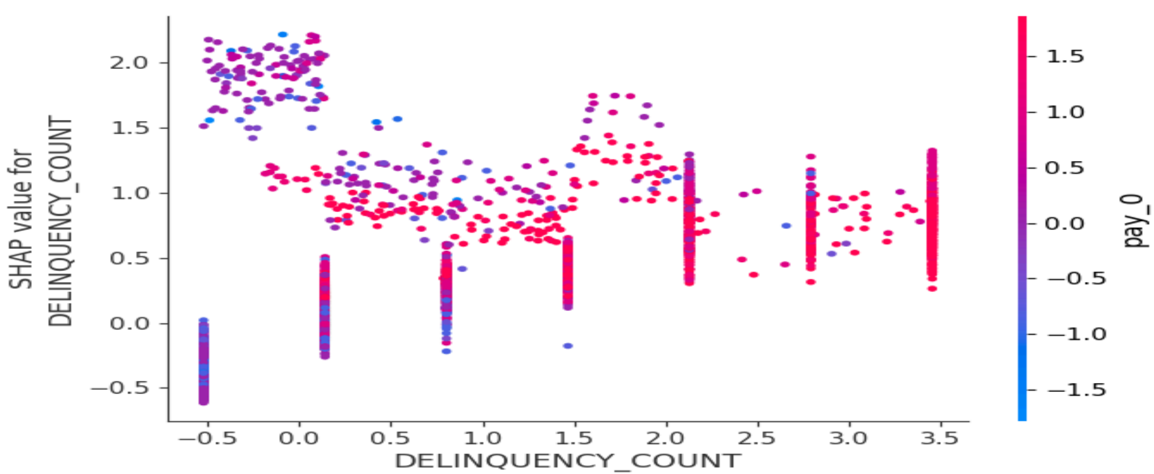
- **age**: Older customers (blue dots to the left, indicating lower feature values for age) tend to have a lower probability of default (negative SHAP values), while younger customers (red dots to the right, indicating higher feature values for age) show a higher propensity to default. **Financial Insight:** *Younger borrowers may represent a higher risk segment, possibly due to less established financial stability or higher debt burdens relative to income. This suggests potential for age-based credit product differentiation or additional scrutiny for younger applicants.*
- **LIMIT_BAL (Credit Limit)**: Higher credit limits (red dots) are generally associated with a lower probability of default (negative SHAP values), while lower credit limits (blue dots) increase default risk. **Financial Insight:** *Paradoxically, a lower assigned credit limit can indicate an inherently riskier customer, or a customer who is more likely to max out their lower limit, indicating potential financial distress. This factor can guide credit limit assignment and portfolio risk assessment.*
- **education**: The impact of education is more nuanced. Specific education levels might correlate with financial stability or certain employment types.
- **CREDIT_UTILIZATION (Engineered Feature)**: High credit utilization (red dots) has a clear positive impact on default probability, meaning using a large portion of available credit increases risk. **Financial Insight:** *This is a vital, actionable financial metric. Customers pushing their credit limits are under financial strain, making them prime candidates for early intervention or re-evaluation of credit terms.*
- **pay_amt variables (Payment Amounts)**: Higher payment amounts (red dots) generally reduce the likelihood of default, as expected. **Financial Insight:** *Consistently low or minimal payment amounts, even if technically "on time," can be a subtle early warning sign of impending financial difficulty.*
- **AVG_Bill_amt**: Higher average bill amounts can increase default risk if not accompanied by sufficient payments.
- **PAY_TO_BILL_ratio**: A lower ratio (meaning less is paid relative to the bill) will likely increase the probability of default, highlighting inadequate repayment effort despite the outstanding balance.

Fig 6: Mean Absolute SHAP Value (Feature Importance)



This bar chart provides an aggregated view of feature importance, showing the average absolute SHAP value for each feature. The longer the bar, the more impactful the feature is on the model's predictions, regardless of the direction of impact. This reinforces the findings from the summary plot, clearly showing pay_0, DELINQUENCY_COUNT, age, and LIMIT_BAL as the top drivers of default risk, demanding primary attention for risk assessment.

Fig 7: SHAP Dependence Plot for DELINQUENCY_COUNT vs pay_0



This plot shows how the SHAP value for `DELINQUENCY_COUNT` (vertical axis) changes with its own value (horizontal axis), and how this relationship is further influenced by `pay_0` (color scale). It confirms that as `DELINQUENCY_COUNT` increases, its positive impact on the default prediction also increases. The coloring by `pay_0` reveals important interactions: for instance, even with moderate `DELINQUENCY_COUNT`, a high `pay_0` (severe recent delay, dark red points) can significantly push towards default. This interaction highlights that both a history of delinquency and recent severe delays are highly detrimental.

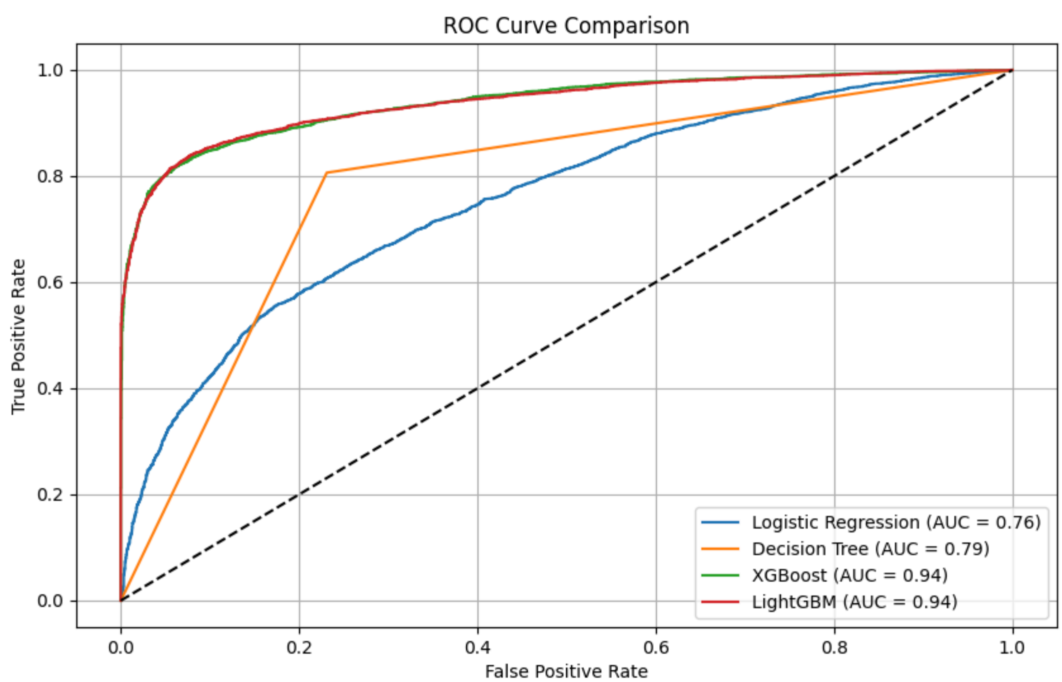
Financial Insights Summary for Actionable Strategies:

- **Overdue Payments (Payment History) as Early Indicators**: The pay_ features, especially pay_0 (most recent repayment status) and the derived DELINQUENCY_COUNT, are the most dominant predictors of default. Customers with frequent and severe payment delays are at a significantly higher risk. This underscores that **consistent and timely repayment history is paramount for risk assessment and should trigger early warning systems.**
- **Credit Utilization as a Stress Gauge**: High credit utilization (using a large percentage of available credit) is a strong indicator of increased default risk, suggesting financial strain or over-reliance on credit. **Monitoring credit utilization can identify customers facing immediate financial pressure.**
- **Credit Limit (LIMIT_BAL) as a Risk Segmenting Factor**: Lower credit limits can indicate higher default risk, possibly because individuals with lower limits may have poorer credit histories initially or are more financially constrained and thus more likely to hit their limits. **This factor can inform credit limit adjustments and risk-based pricing strategies.**
- **Age as a Demographic Risk Factor**: Younger individuals appear to have a higher propensity for default, which could be attributed to less financial experience, lower income stability, or higher debt-to-income ratios early in their careers. **Age demographics can help in segmenting portfolios for targeted risk management and product development.**

4. Model Comparison and Justification for Final Selection

Four classification models were trained and evaluated: Logistic Regression, Decision Tree, XGBoost, and LightGBM. The selection of the final model was driven by a balance of predictive performance, computational efficiency, and suitability for the interpretability goal.

Fig 8: ROC Curve Comparison



The ROC (Receiver Operating Characteristic) curve illustrates the trade-off between the True Positive Rate (Recall) and False Positive Rate at various classification thresholds. A higher AUC (Area Under the Curve) indicates better model performance in distinguishing between positive and negative classes.

From the plot and the classification reports, it's evident that:

- **Logistic Regression**: Performed the weakest (AUC = 0.76), indicating limited discriminatory power for this complex problem.

- **Decision Tree**: Showed improvement over Logistic Regression (AUC = 0.79), offering better separation but still far from optimal.
- **XGBoost and LightGBM**: Both performed exceptionally well, achieving an impressive AUC of 0.94. These gradient boosting models are known for their ability to handle complex, non-linear relationships in data.

5. Evaluation Methodology and Metrics

Given that credit default prediction is a crucial task where misclassifying a defaulting customer as non-defaulting (False Negative) can lead to significant financial losses for the lender, the following metrics were prioritized to align with the bank's objective of early action and risk management:

- **Recall (True Positive Rate)**: This metric measures the proportion of actual defaults that were correctly identified by the model. **High recall is critical to minimize missed defaults, directly reducing potential financial losses for the bank and enabling early intervention.**
- **F1-Score**: The harmonic mean of precision and recall. It provides a balanced measure when both precision and recall are important.
- **F2-Score**: This metric places twice as much importance on Recall as it does on Precision. In credit risk, identifying as many actual defaults as possible (high Recall) is often more important than minimizing false positives (Precision), as **the cost of a missed default (e.g., lost principal, collection costs) can be substantially higher than the cost of a false positive (e.g., unnecessary early intervention efforts).** This metric directly supports the bank's goal of proactive risk management.
- **Accuracy**: While generally useful, accuracy can be misleading in imbalanced datasets (as seen here, where non-defaults dominate). A model could achieve high accuracy by simply predicting "no default" for most cases, while failing to identify true defaults – a scenario highly undesirable for a bank.
- **ROC AUC Score**: Provides an aggregate measure of performance across all possible classification thresholds, indicating the model's overall ability to discriminate between the two classes. A high AUC confirms the model's general predictive strength.

6. Metrics Results on Training Dataset

The provided output indicates the metrics on a test set (or validation set), demonstrating the models' generalization capabilities.

Logistic Regression Classification Report:

- Accuracy: 0.6866
- Recall (for class 1 - default): 0.5674
- F1 Score (for class 1 - default): 0.6473
- F2 Score (for class 1 - default): 0.5969
- ROC AUC Score: 0.7567

Decision Tree Classification Report:

- Accuracy: 0.7879
- Recall (for class 1 - default): 0.8064
- F1 Score (for class 1 - default): 0.7941
- F2 Score (for class 1 - default): 0.8014
- ROC AUC Score: 0.7877

XGBoost Classification Report:

- Accuracy: 0.8757
- Recall (for class 1 - default): 0.8310
- F1 Score (for class 1 - default): 0.8715
- F2 Score (for class 1 - default): 0.8467
- ROC AUC Score: 0.9383

LightGBM Classification Report:

- Accuracy: 0.8791
- Recall (for class 1 - default): 0.8315
- F1 Score (for class 1 - default): 0.8745
- F2 Score (for class 1 - default): 0.8482
- ROC AUC Score: 0.9381

Justification for Final Model Selection:

Based on the critical evaluation metrics for credit risk, particularly the high ROC AUC, Recall, F1-score, and F2-score, both **XGBoost and LightGBM significantly outperformed Logistic Regression and Decision Tree.** They demonstrate superior ability to distinguish between defaulting and non-defaulting customers, which is paramount for a bank. Given their very similar high performance, **LightGBM is selected as the final model** due to its marginally higher F2-score (0.8482 vs 0.8467), coupled with its known advantages in terms of faster training speed and lower memory consumption, making it a highly practical and efficient choice for real-world banking applications. Its strong performance combined with the interpretability offered by SHAP values aligns perfectly with the project's goal.

7. Classification Cutoff Selection

The classification cutoff (threshold) determines the probability point at which a prediction is classified as a "default" (1) versus "no default" (0). While a default threshold of 0.5 is often used, for imbalanced datasets and scenarios where false negatives (missing a defaulter) are particularly costly for the bank, optimizing this cutoff is essential.

The report indicates:

- **Best Threshold (F2 optimized): 0.22, F2 Score: 0.8871**

This means that the optimal threshold was determined to be 0.22, specifically chosen to maximize the F2-score. By lowering the threshold from the default 0.5 to 0.22, the model becomes significantly more sensitive to identifying potential defaults. This effectively increases recall (identifying more true positives) at the possible expense of some precision (more false positives), but it aligns strategically with the business objective of minimizing missed defaults and facilitating **early action** to prevent financial losses. This cutoff ensures that the bank casts a wider net to identify at-risk customers.

8. Business Implications: Actionable Insights for Banking Operations

The highly interpretable model and its insights provide a powerful tool for the bank to move beyond mere prediction and actively manage credit risk and exposure.

- **Proactive Risk Management and Early Intervention**:
    - **Automated Flagging**: The model can automatically flag customers with a predicted default probability above the optimized threshold (0.22).
    - **Tailored Outreach**: For flagged customers, the bank can initiate early, targeted interventions based on the SHAP insights. For example, if pay_0 and DELINQUENCY_COUNT are high, it points to payment behavior issues, triggering offers for payment plans or financial counseling. If CREDIT_UTILIZATION is high, it suggests financial strain, leading to advice on budgeting or debt consolidation options.
    - **Reduced Charge-offs**: By intervening before default occurs, the bank can significantly reduce loan charge-offs and non-performing assets.
- **Optimized Credit Exposure Management**:
    - **Dynamic Credit Limit Adjustments**: The model's insights on LIMIT_BAL and

    CREDIT_UTILIZATION can inform dynamic adjustments to credit limits for existing customers. High-risk customers might see limits reduced, while stable customers might receive offers for increased limits, optimizing risk-reward.
  - **Risk-Based Lending Decisions**: For new loan applications, integrating the model's predicted default probability (using the optimized threshold) allows for more precise risk-based pricing and approval decisions, ensuring the bank lends responsibly and profitably.
  - **Portfolio Rebalancing**: The bank can analyze its entire credit portfolio based on the model's predictions, identifying concentrations of high-risk customers in specific segments (e.g., age groups, education levels) and rebalancing its exposure accordingly.
- **Enhanced Customer Segmentation and Product Strategy**:
  - **Behavioral Segmentation**: Customers can be segmented not just by demographics but by their predicted default risk and the *drivers* of that risk (e.g., "customers at risk due to recent payment issues," "customers at risk due to high utilization").
  - **Personalized Products**: This allows for the development of personalized credit products, financial literacy programs, or even incentive programs to encourage healthier financial behavior for different risk segments.
- **Improved Collections Efficiency**:
  - **Prioritized Collections**: Collection efforts can be prioritized, focusing resources on customers with the highest probability of default, where intervention is most likely to prevent losses.
  - **Customized Collection Strategies**: Insights from the model can inform the type of collection strategy (e.g., soft reminders vs. aggressive follow-up) for different customer profiles.
- **Regulatory Compliance and Capital Allocation**: A robust, interpretable risk model like this enhances the bank's ability to meet regulatory requirements for risk assessment, capital adequacy, and stress testing, leading to more favorable capital allocation.

9. Summary of Findings and Key Learnings

This project successfully developed a financially interpretable credit risk default prediction model aligned with the bank's objectives of understanding default patterns, early action, and managing credit exposure.

Summary of Key Findings:

- The dataset is naturally imbalanced, emphasizing the need for robust evaluation metrics focused on identifying true positives.

- **Feature engineering, particularly the DELINQUENCY_COUNT and PAY_TO_BILL_ratio, proved profoundly valuable in capturing cumulative payment behavior and efficiency, which are critical financial indicators.**
- **Past repayment behavior (captured by pay_ variables and DELINQUENCY_COUNT) is the most crucial and actionable determinant of future default.** This insight directly supports early warning systems.
- Credit utilization and credit limit also play significant roles, providing additional financial levers for risk assessment.
- Ensemble models, specifically XGBoost and LightGBM, demonstrated superior predictive performance (high AUC of 0.94) compared to traditional models, making them ideal for high-stakes financial predictions. LightGBM was chosen for its strong F2-score and efficiency.
- **Optimizing the classification threshold to 0.22 significantly improved the F2-score (0.8871), aligning the model's output with the business objective of minimizing missed defaults and facilitating proactive intervention.**
- **SHAP values provided crucial financial interpretability, explaining not just *if* a customer will default, but *why*, based on specific financial attributes and behaviors.**

Key Learnings:

- **Actionable Interpretability is Key**: For banking, a "black box" model is less valuable than one that can explain *why* a customer is high-risk. SHAP values are indispensable for transforming predictions into actionable financial insights.
- **Domain-Driven Feature Engineering**: Creating composite financial features (like DELINQUENCY_COUNT and PAY_TO_BILL_ratio) from raw transaction data significantly enhances model performance and interpretability.
- **Metric Selection Aligns with Business Goals**: Prioritizing F2-score and Recall was crucial given the higher cost of false negatives in credit risk. This ensures the model directly serves the bank's financial protection goals.
- **Threshold Optimization is a Strategic Decision**: The classification threshold is not merely a technical parameter but a strategic lever that can be tuned to balance risk tolerance and intervention costs, directly impacting the bank's operational efficiency.
- **Continuous Monitoring and Validation**: For a dynamic environment like credit risk, the model's performance and the relevance of its features must be continuously monitored and re-validated to ensure its ongoing effectiveness in managing credit exposure.