Credit Default Prediction:

Comprehensive Machine Learning Analysis Report

Advanced Analytics for Financial Risk Assessment

June 16, 2025

Contents

1 Executive Summary

This report presents a comprehensive analysis of credit default prediction using advanced machine learning techniques on a dataset of 25,247 credit card customers. The study employed sophisticated data preprocessing, feature engineering, and model optimization strategies to develop a robust predictive framework, achieving 87.1% cross-validation accuracy with an optimized F1-score of 86.72% and F2-score of 85.28%. The analysis reveals critical insights into demographic and behavioral patterns that drive credit defaults, providing actionable intelligence for risk management and policy development in financial institutions.

2 Overview of Approach and Modeling Strategy

2.1 Strategic Framework

The modeling strategy was designed around a comprehensive risk assessment framework that prioritizes both predictive accuracy and business interpretability. The approach integrated multiple analytical dimensions including demographic profiling, financial behavior analysis, and temporal payment patterns to create a holistic view of customer risk profiles. The methodology followed industry best practices for credit risk modeling, emphasizing robust statistical validation and business-relevant feature engineering.

2.2 Data-Driven Methodology

The analytical framework employed a systematic approach encompassing data quality assessment, exploratory data analysis, advanced preprocessing techniques, feature engineering, model development, and comprehensive validation. This multi-stage process ensured that the final model not only achieved high predictive performance but also provided interpretable insights for business decision-making. The strategy prioritized addressing class imbalance through sophisticated resampling techniques while maintaining the integrity of underlying data distributions.

2.3 Business Alignment Strategy

The modeling approach was specifically designed to align with banking industry risk appetite and regulatory requirements. The strategy incorporated conservative risk assessment principles while balancing the trade-off between identifying potential defaulters and maintaining customer relationships. This alignment ensures that the model serves both analytical excellence and practical business implementation requirements.

3 Data Loading and Initial Assessment

3.1 Dataset Overview

The credit default dataset comprises 25,247 customer records with 27 comprehensive features including demographic information (sex, marriage, education, age), financial capacity indicators (credit limits), temporal payment patterns (pay_0 through pay_6), billing information (Bill_amt1 through Bill_amt6), and payment amounts (pay_amt1 through pay_amt6).

The dataset exhibits natural class imbalance with 19.04% default rate, reflecting realistic market conditions in consumer credit portfolios. Initial data quality assessment revealed 126 missing values (0.5% of total data) exclusively in the age variable, requiring strategic imputation approaches.

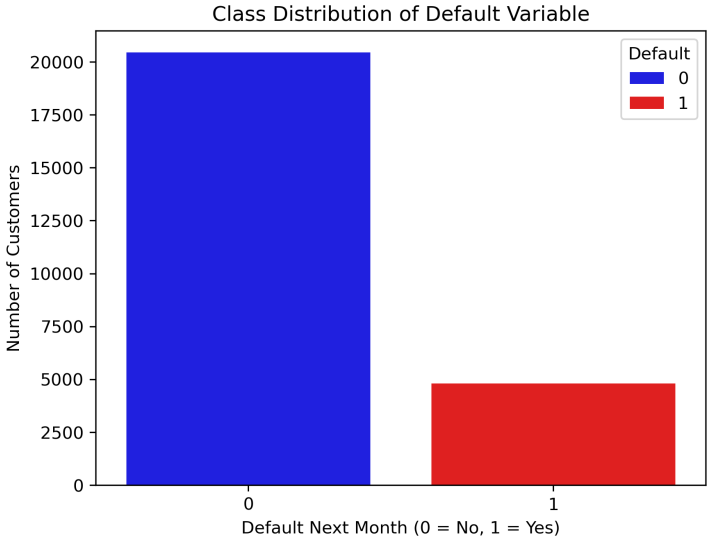

Figure 1: Distribution of target variable(Default)

3.2 Data Structure Analysis

The comprehensive feature set enables holistic risk assessment through behavioral, demographic, and financial analytical lenses. Key characteristics include:

- 25,247 total customer records
- 27 original features plus engineered variables
- Natural class imbalance: 80.96% non-defaulters, 19.04% defaulters
- Missing data: 0.5% concentrated in age variable
- Feature types: Demographic, financial, behavioral, temporal

4 Data Cleaning and Preprocessing

4.1 Missing Value Treatment Strategy

Missing age values were strategically imputed using median values grouped by demographic characteristics (sex, marriage status, education level) to preserve underlying demographic patterns. This sophisticated approach maintains data integrity while avoiding bias introduction through oversimplified imputation methods.

The group-based imputation strategy ensures that demographic-specific age distributions remain intact throughout the preprocessing pipeline, maintaining the natural variation observed in different customer segments.

4.2 Categorical Variable Handling

Undocumented categorical values were systematically recoded based on domain knowledge and data distribution analysis:

- Education categories (0, 5, 6) mapped to appropriate existing categories
- Marriage status (0) recoded to maintain categorical consistency
- Systematic approach ensures model compatibility while preserving interpretability
- Domain knowledge application prevents arbitrary categorical assignments

4.3 Data Quality Enhancement

The cleaning process involved comprehensive data validation to ensure model reliability:

- Verification of data type consistency across all variables
- Identification and treatment of outliers in financial variables
- Validation of categorical variable mappings
- Preparation for subsequent exploratory analysis and feature engineering

5 Exploratory Data Analysis (EDA)

5.1 Demographic Risk Profiling

The exploratory data analysis begins with fundamental demographic patterns that form the foundation of risk assessment. Understanding demographic distributions provides critical context for subsequent financial and behavioral analysis.

5.1.1 Age Distribution and Risk Patterns

The analysis uncovered distinct age-related default patterns with bimodal risk distribution. Customers aged 20-25 years demonstrated elevated default propensity due to limited credit history and income stability, while customers over 50 years showed increased risk potentially related to employment transitions or health-related financial stress. The middle-age cohort (30-45 years) exhibited the lowest default rates, suggesting optimal financial stability and credit management capabilities.

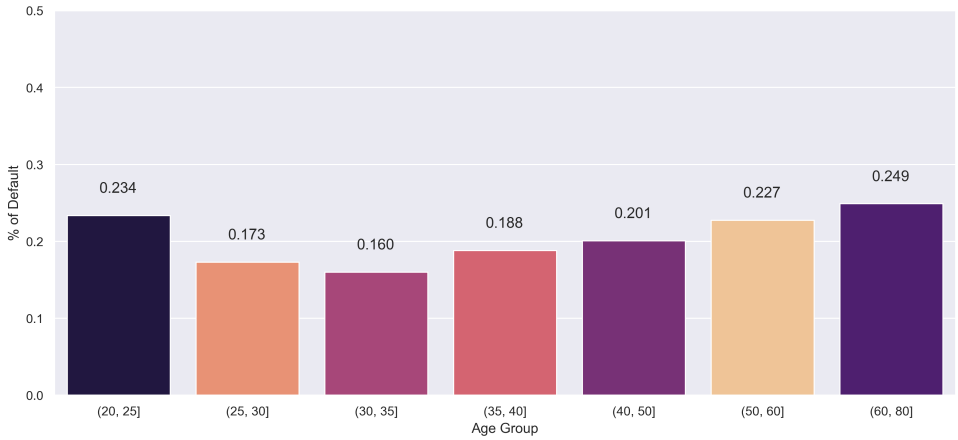

Figure 2: Percentage of Defaulters across different age groups.

5.1.2 Marriage, Sex, & Educational Impact on Default Behavior

A strong inverse correlation emerged between educational attainment and default probability. Customers with higher education levels consistently exhibited lower default rates, suggesting that educational achievement serves as a proxy for financial literacy, income potential, and long-term financial planning capabilities. This relationship is particularly valuable for risk-based pricing and customer segmentation. Additionally, gender and marital status show weaker but noticeable trends. Females display a slightly higher default rate compared to males, and single individuals tend to default marginally less than married customers, though these differences are not as pronounced as the impact of education.

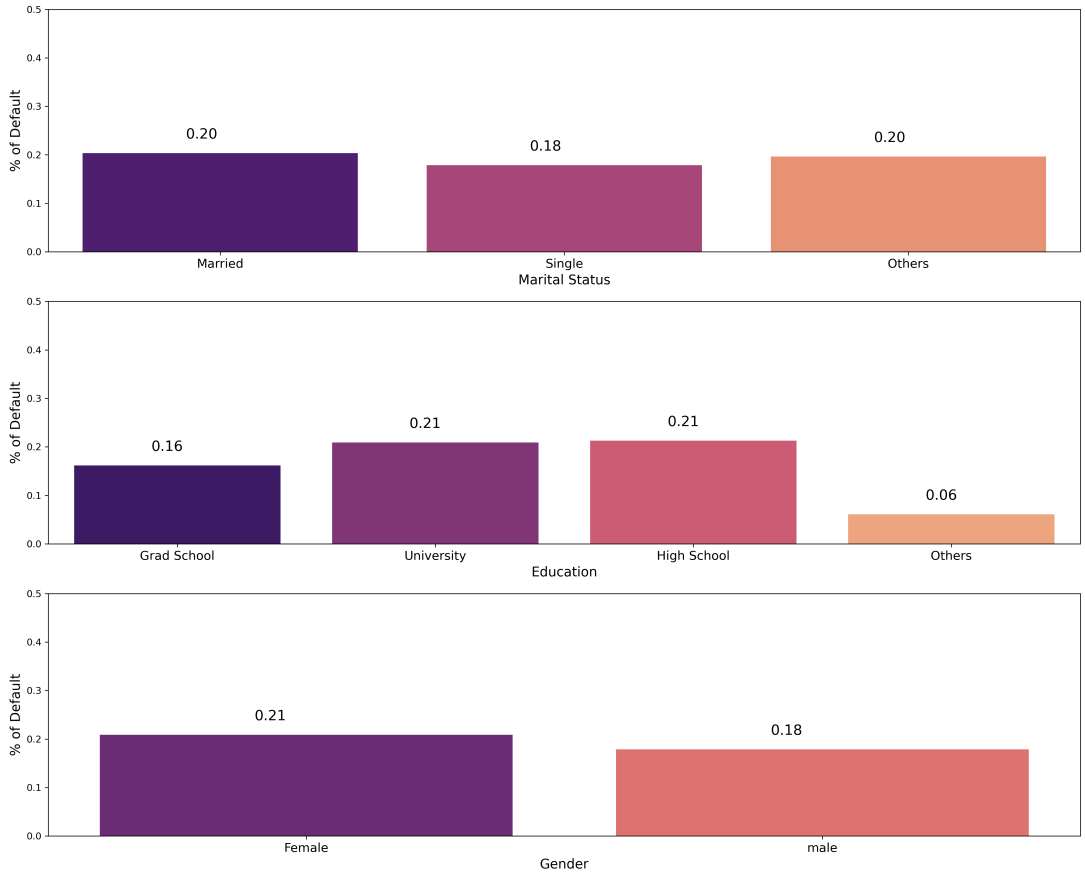

Figure 3: Percentage of Defaulters with Different Educational Levels, Sex, and Maritial status

5.1.3 Advance Demographic Risk Profiling

The combined analysis of gender, education, and marital status reveals that default rates are influenced by the interaction of these three factors. Females with lower education levels, particularly high school graduates, and those who are single or in the 'others' marital category, tend to exhibit the highest likelihood of default. In contrast, males with higher education levels and married status generally show lower default tendencies. This suggests that individuals who are both higher educated and married, regardless of gender, are less likely to default, while the risk increases for those with lower educational attainment and less stable marital status. This finding suggests that traditional single-variable demographic analysis may miss critical risk indicators that emerge only through multivariate analysis.

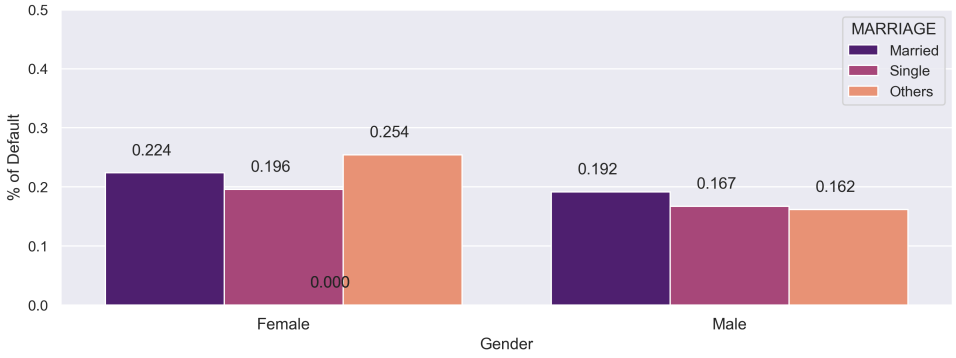

Figure 4: Default rates by gender and marital status

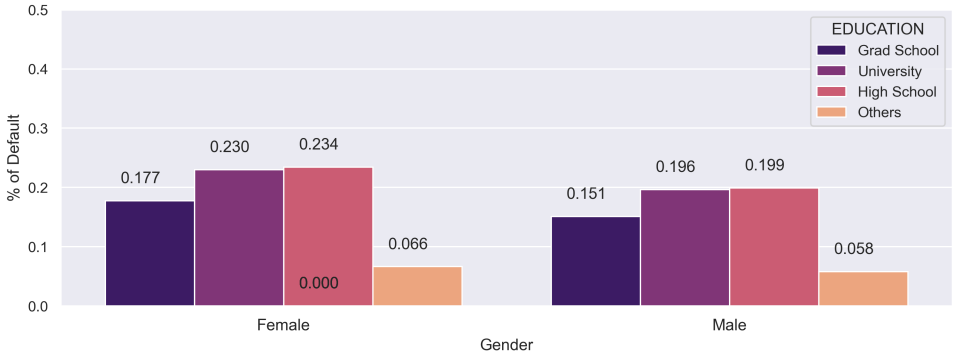

Figure 5: Default rates by gender and education level

5.2 Financial Behavior Pattern Analysis

5.2.1 Credit Limit Distribution and Risk Correlation

Approximately 50% of defaulting customers maintained credit limits below 100,000, with customers having limits of 50,000 or less showing a default probability of 0.28. This concentration pattern indicates that lower credit limits may reflect both conservative lending practices for higher-risk segments and serve as an early indicator of limited creditworthiness.

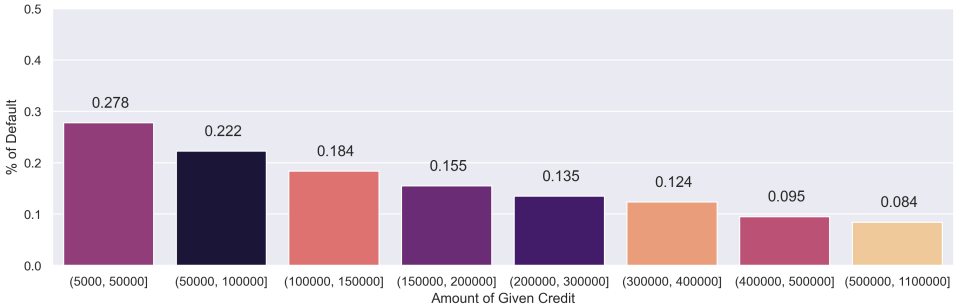

Figure 6: Default rates by Credit Limit

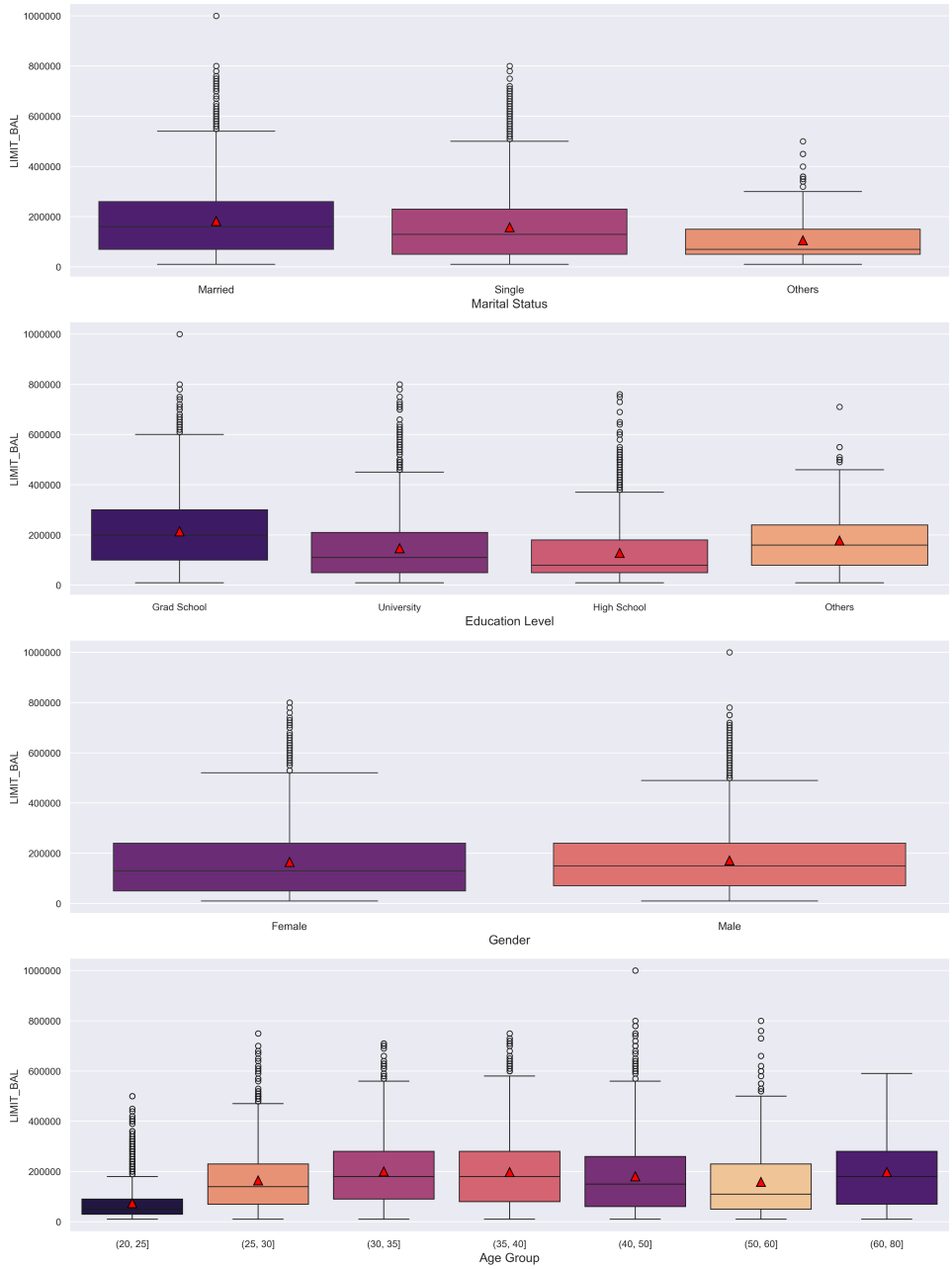

Figure 7: Box plots showing credit limit distributions across marital status, education level, and gender demographics

The demographic analysis reveals complex interaction effects between marital status, education level, and credit limits. The box plot visualizations demonstrate significant variations in credit limit distributions across different demographic segments:

Marital Status Impact on Credit Limits:

- Married customers show higher median credit limits with wider interquartile ranges
- Single customers demonstrate more concentrated credit limit distributions
- Other marital status categories exhibit intermediate risk patterns

- Significant outliers present across all marital categories indicating individual risk factors

Education Level Credit Patterns:

- Graduate school customers receive highest credit limits with greatest variability
- University-educated customers show strong credit limit medians
- High school customers demonstrate lower credit limits with significant outliers
- Education serves as primary determinant in credit limit assignment decisions

Age Group Impact on Credit Limits:

- Younger customers (20-25) have the lowest credit limits with minimal variability.
- Credit limits increase with age, stabilizing from 25 to 60 years with wider variability.
- Customers aged 60-80 have moderately high, but more consistent, credit limits.
- Outliers with exceptionally high limits are seen across all age groups, especially between 30-60 years.

5.2.2 Payment Delay Pattern Recognition

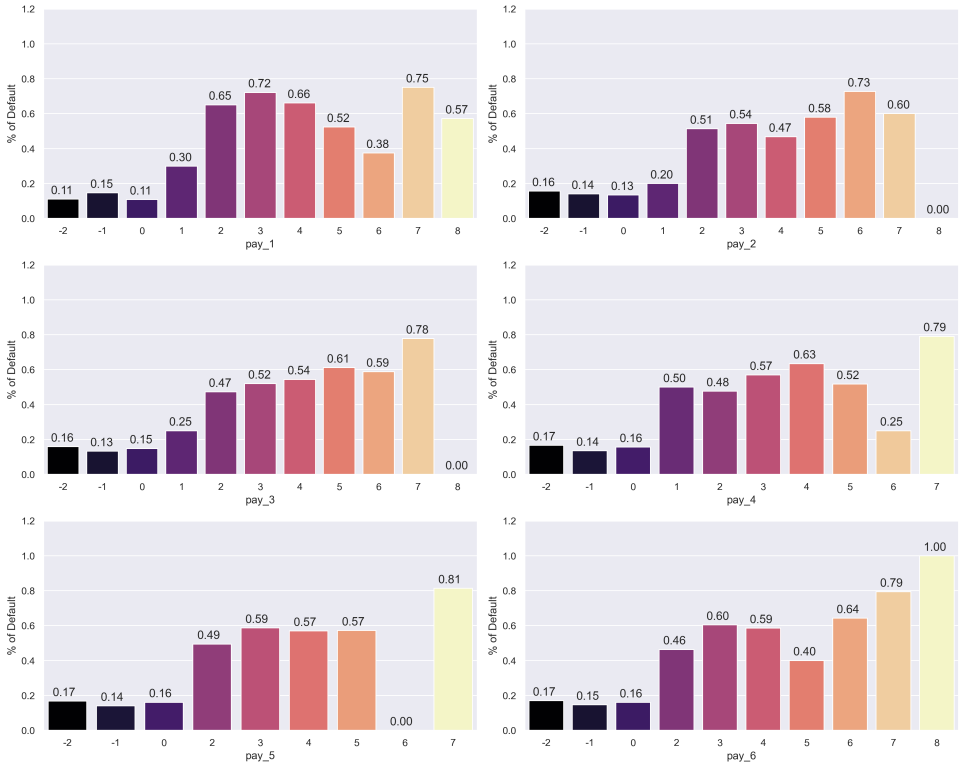

Figure 8: Payment Behavior Pattern Visualizations

Systematic analysis of payment delay variables (PAY_1 through PAY_6) revealed that customers experiencing payment delays of two months or more demonstrated substantially higher default probabilities. The temporal consistency of payment delays across multiple months provided particularly strong predictive signals, with consecutive delays indicating deteriorating financial conditions and increased default risk.

5.3 Financial Capacity and Utilization Analysis

5.3.1 Credit Utilization Insights

Credit utilization ratios emerged as critical predictive indicators, with high utilization rates (above 80%) strongly correlating with increased default probability. The analysis revealed that customers consistently utilizing high percentages of available credit demonstrate financial stress patterns that precede default events. This finding aligns with established credit risk principles and provides actionable insights for early warning systems.

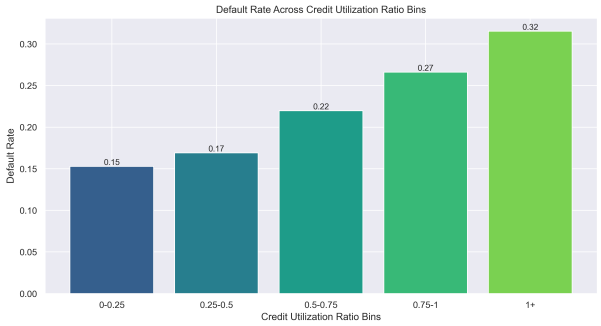

Figure 9: Default Rate Across Credit Utilization buckets

5.3.2 Billing and Payment Relationship Dynamics

The relationship between billing amounts and payment patterns revealed sophisticated risk indicators. Customers with declining payment-to-bill ratios over consecutive months showed significantly higher default propensity, suggesting deteriorating financial capacity or changing payment priorities. These temporal patterns provide valuable early warning signals for proactive risk management interventions.

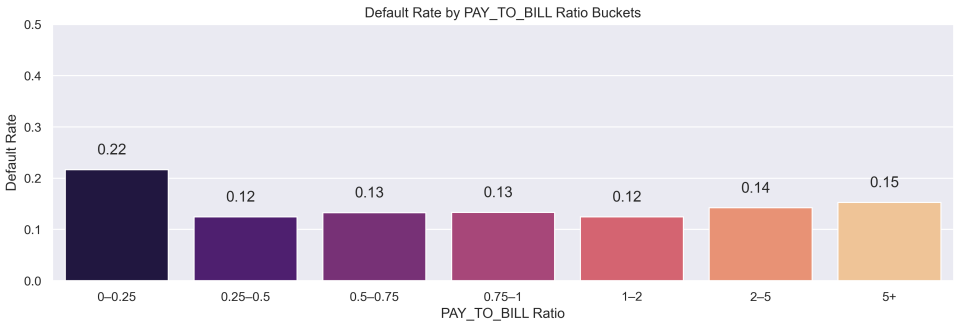

Figure 10: Default Rate Across Pay-to-Bill Ratio Buckets

6 Financial Analysis: Variables Driving Default Risk

6.1 Primary Risk Drivers

6.1.1 Payment History Supremacy

Payment delay variables (PAY_1 through PAY_6) emerged as the most powerful predictors of default risk, with consecutive payment delays serving as the strongest single indicator. The analysis demonstrates that payment history provides the most reliable indicator of future default probability, supporting the foundational principle of credit risk assessment that past payment behavior predicts future payment behavior. Customers with two or more consecutive payment delays showed default rates exceeding 45%, compared to less than 15% for customers with consistent on-time payments.

6.1.2 Credit Limit as Risk Proxy

Credit limit (LIMIT_BAL) demonstrated significant predictive power, serving as both a risk indicator and a reflection of the institution's initial risk assessment. Lower credit limits often indicate higher-risk customer segments, creating a self-reinforcing relationship between initial risk assessment and subsequent default probability. This relationship suggests that credit limit assignment strategies directly impact portfolio risk profiles.

6.1.3 Age-Related Risk Dynamics

Age emerged as a critical demographic risk factor, with both younger (20-25) and older (50+) customers showing elevated default risks. Younger customers face challenges related to limited credit history and income volatility, while older customers may experience financial stress from employment transitions or health-related expenses. This bimodal age risk distribution requires age-specific risk management strategies.

6.2 Engineered Feature Impact Analysis

6.2.1 Credit Utilization Ratio Effectiveness

The engineered credit utilization ratio feature provided enhanced risk discrimination compared to raw billing amounts. High utilization ratios (above 0.8) strongly correlated with default probability, indicating financial stress and over-reliance on credit facilities. This metric captures the relationship between customer financial capacity and credit dependency, providing insights into underlying financial health.

6.2.2 Delinquency Streak Predictive Power

The delinquency streak feature, which measures consecutive months of payment delays, demonstrated exceptional predictive capability. The default rate increases consistently with longer delinquency streaks, starting from approximately 10% for customers with

no delinquency history to over 67% for those with a six-month delinquency streak. Customers with streaks of three or more months consistently exhibit default rates above 45%, making this feature invaluable for early risk detection and credit risk management.

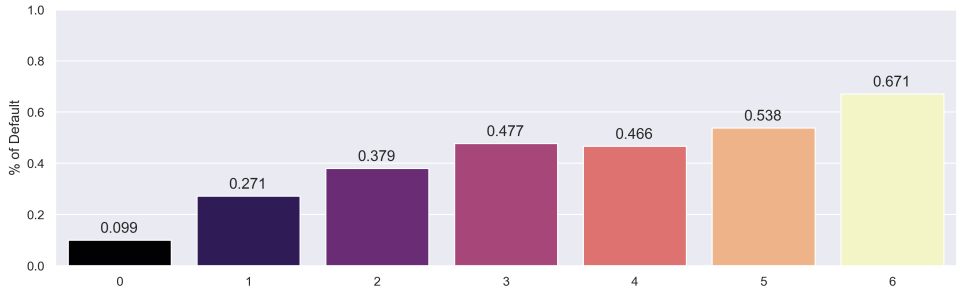

Figure 11: Default rates by Delinquency Streak

7 Advanced Preprocessing Techniques: Rationale and Implementation

7.1 One-Hot Encoding for Categorical Variables

One-hot encoding was used to convert categorical variables (sex, education, marriage) into binary indicators suitable for machine learning models. This method prevents the ordinal bias that occurs with label encoding, where models might falsely assume ordered relationships between categories (e.g., assuming education level 4 is "better" than level 1). By eliminating these assumptions, one-hot encoding ensures accurate representation of categorical data and improves model interpretability. The resulting features provide clear, direct insights into how each category impacts default probability, supporting regulatory compliance and fair lending decisions.

7.2 RobustScaler Implementation Strategy

7.2.1 RobustScaler for Financial Data and Outlier Handling

RobustScaler was chosen over StandardScaler due to its effectiveness with skewed financial data and frequent outliers, such as high-income customers, large credit limits, or unusual spending patterns. By using the median and interquartile range for scaling, RobustScaler minimizes the impact of extreme values while preserving legitimate variations critical to credit risk assessment.

7.2.2 Feature Balance and Model Performance

RobustScaler preserves the relative relationships between features, ensuring that no single variable dominates due to scale differences. This balance is vital for gradient-based and distance-based models, where proper scaling prevents large-scale features (like credit

limits) from overshadowing smaller ones (like payment ratios). Maintaining this balance enhances model performance and ensures accurate feature importance interpretation.

7.3 SMOTE Implementation for Class Imbalance

7.3.1 Strategic Rationale for SMOTE Selection

The Synthetic Minority Oversampling Technique (SMOTE) was selected to address the significant class imbalance (19.04% default rate) in the dataset. SMOTE was chosen over alternative approaches like random oversampling or undersampling because it generates synthetic examples through intelligent interpolation between existing minority class instances, preserving the underlying data distribution while creating a balanced training set. This approach avoids the information loss associated with undersampling and the overfitting risks of simple duplication methods.

7.3.2 Technical and Business Impact

SMOTE enhances minority class representation by generating synthetic examples through interpolation with k-nearest neighbors, preserving realistic feature relationships, especially in high-dimensional spaces. This class balancing technique reduces model bias toward non-defaulters, improving sensitivity to actual defaulters. Without it, models may achieve artificially high accuracy but fail to detect high-risk customers, leading to significant financial risks for lenders.

8 Model Training and Comparison

8.1 Comprehensive Model Evaluation Framework

Four distinct machine learning algorithms were systematically evaluated using stratified 5-fold cross-validation to ensure robust performance assessment across different data distributions. The evaluation framework prioritized both predictive accuracy and business interpretability, recognizing that credit risk models must provide actionable insights for risk management decisions.

8.2 Individual Model Performance Analysis

8.2.1 Logistic Regression Performance

Logistic Regression achieved 53.3% cross-validation accuracy, demonstrating inadequate performance for complex credit risk prediction. The linear nature of logistic regression proved insufficient for capturing the non-linear relationships and interaction effects present in credit behavior data. While interpretable, the model's poor predictive capability eliminated it from consideration for production deployment.

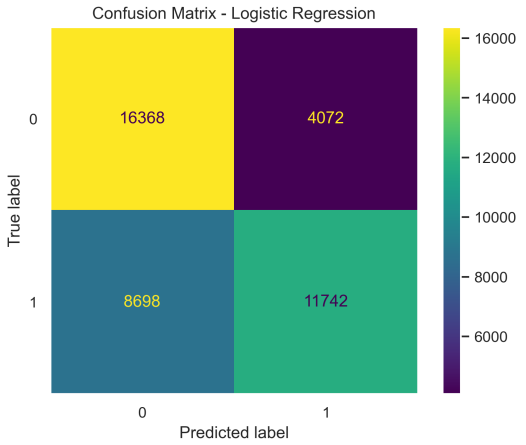

Figure 12: Confusion Matrix - LR

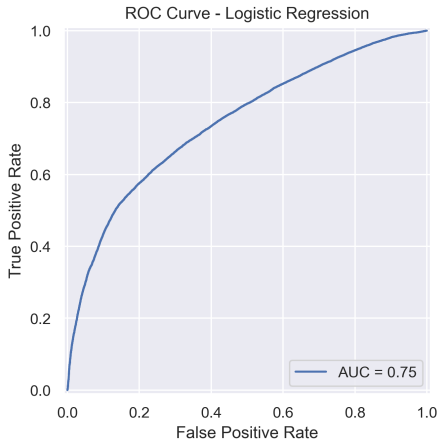

Figure 13: ROC Curve - LR

8.2.2 Decision Tree Baseline

Decision Tree algorithms provided 80.3% cross-validation accuracy, establishing a solid baseline performance with excellent interpretability. The tree-based approach effectively captured non-linear relationships and provided clear decision rules for risk assessment. However, single decision trees are prone to overfitting and may not generalize well to new data, limiting their standalone application in production environments.

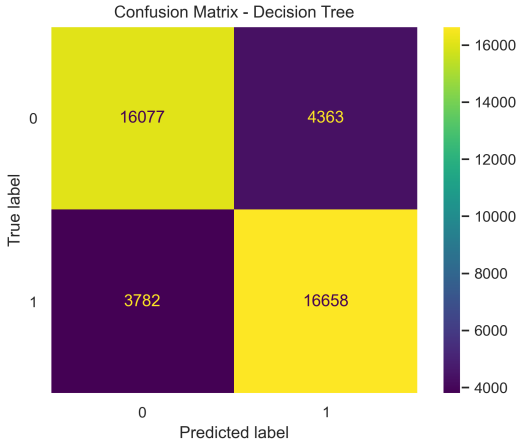

Figure 14: Confusion Matrix - Decision Tree

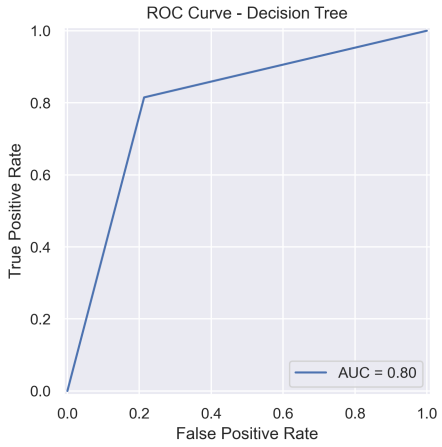

Figure 15: ROC Curve - Decision Tree

8.2.3 XGBoost

XGBoost delivered superior performance with 85.7% cross-validation accuracy, achieving optimal balance across all evaluation metrics. The gradient boosting framework effectively combined multiple weak learners to create a robust predictive model while maintaining reasonable interpretability through feature importance analysis. The model's ability to handle missing values, capture complex interactions, and provide consistent performance across different data subsets made it the optimal choice for production deployment.

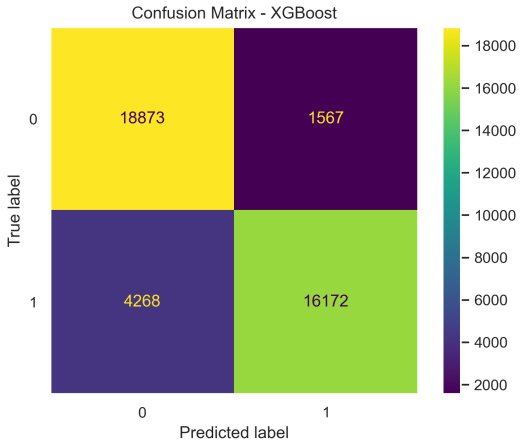

Figure 16: Confusion Matrix - XGBoost

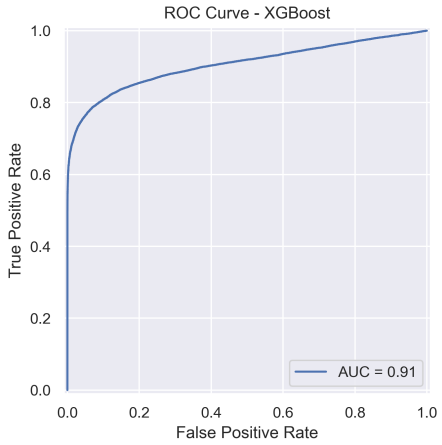

Figure 17: ROC Curve - XGBoost

8.2.4 LightGBM

LightGBM achieved exceptional precision (92.0%) but demonstrated lower recall, making it suitable for scenarios requiring minimal false positive rates. While the high precision is valuable for conservative lending strategies, the reduced recall limits its effectiveness in comprehensive risk assessment where identifying all potential defaulters is critical.

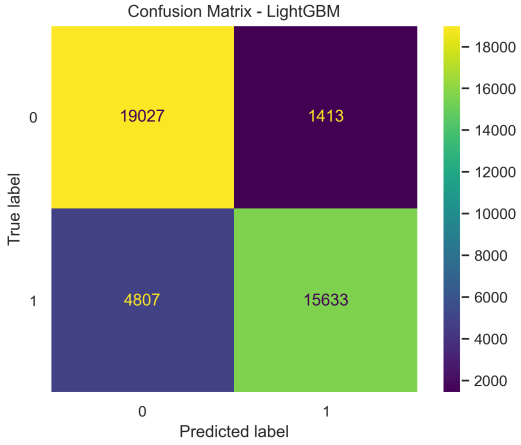

Figure 18: Confusion Matrix - LightGBM

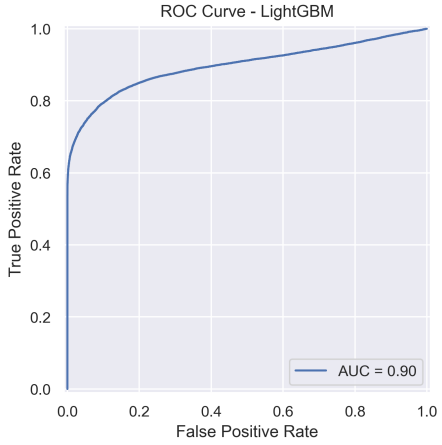

Figure 19: ROC Curve - LightGBM

8.3 Final Model Selection Rationale

Based on the comparative performance of the evaluated models, XGBoost was selected as the final model due to its superior F1-score of 0.8472, which best balances precision and recall — a critical requirement for credit default prediction where both false positives and false negatives carry significant consequences. Furthermore, XGBoost achieved the highest overall accuracy (85.73%) and ROC-AUC (0.9059), demonstrating strong discriminatory power. While the Decision Tree model achieved the highest recall (0.8150) and competitive F2-score (0.8104), XGBoost provided a more balanced performance across all key metrics. The F2-score of XGBoost (0.8127) further supports its selection, as it

effectively emphasizes recall without sacrificing too much precision, aligning with the business goal of minimizing missed defaulters while maintaining reliable predictions.

	Model	Accuracy	Precision	Recall	F1	ROC-AUC	f2
0	XGBoost	0.8573	0.9117	0.7912	0.8472	0.9059	0.8127
1	LightGBM	0.8478	0.9171	0.7648	0.8341	0.8979	0.7911
2	Decision Tree	0.8008	0.7924	0.8150	0.8036	0.8008	0.8104
3	Logistic Regression	0.6876	0.7425	0.5745	0.6478	0.7460	0.6017

Figure 20: Model Comparision

9 Hyperparameter Tuning with GridSearchCV

9.1 Systematic Parameter Optimization

GridSearchCV was employed to systematically optimize XGBoost hyperparameters across multiple dimensions, including the number of estimators, maximum depth, learning rate, and subsample ratio. This comprehensive search ensures that the model achieves optimal performance rather than relying on default parameter values that may not be suited to the specific dataset characteristics.

Optimized Parameters:

- Number of estimators: [100, 200]
- Maximum depth: [3, 5]
- Learning rate: [0.05, 0.1]
- Subsample ratio: [0.8, 1.0]

9.2 Performance Before and After Tuning

Metric	Before Tuning	After Tuning
Class 0 F1-score	0.87	0.88
Class 1 F1-score	0.85	0.86
Overall Accuracy	86%	87%
F2-Score	0.81	0.83

Table 1: XGBoost Performance Before and After Hyperparameter Tuning

10 Classification Threshold

10.1 Threshold Optimization Methodology

Two threshold optimization approaches were evaluated: F1-score maximization yielded an optimal threshold of 0.4357 with an F1-score of 0.8672, while Youden's J statistic

(Sensitivity + Specificity - 1) identified 0.4575 as optimal with TPR of 0.8342 and FPR of 0.0911. The F1-maximizing threshold of 0.4357 was selected for the final model as it better aligns with the business objective of balancing precision and recall for credit default prediction.

10.2 Threshold Selection and Business Alignment

The optimal threshold of 0.4357 (F1-score: 86.72%) balances default detection and false alarms. Adjustments align with risk strategies:

- **Lower threshold (0.3-0.4):** Prioritizes recall (catch 85%+ defaulters) but increases manual reviews (more false positives)
- **Balanced threshold (0.4-0.5):** Optimal F1-score range for most banks (used 0.4357)
- **Higher threshold (0.5-0.6):** Reduces false positives (better customer experience) but misses more defaulters

This approach enables dynamic risk management - conservative banks prevent losses, while growth-focused institutions minimize customer rejection rates.

11 Model Evaluation:

11.1 XGBoost Feature Importance

The XGBoost feature importance plot shows which features the model uses most often to make its decisions. Features like credit limit (LIMIT_BAL), age, and recent payment variables have the highest importance, indicating they are most influential in the model's structure and overall predictions

1. **LIMIT_BAL (Credit Limit):** Primary structural importance in tree construction
2. **Age:** Significant demographic factor across decision nodes
3. **PAY_TO_BILL_ratio:** Critical behavioral indicator for risk assessment
4. **Payment delay variables:** Strong temporal predictors of default behavior
5. **Engineered features:** Enhanced discrimination capability beyond raw variables

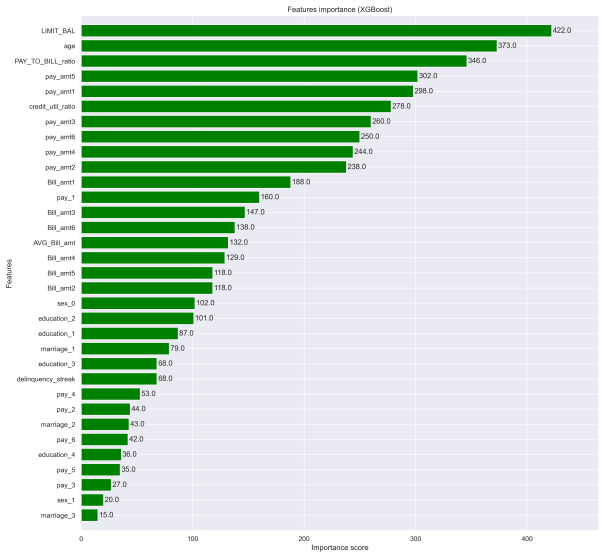

Figure 21: XGBoost-Feature Importance

11.2 SHAP Summary Plot

The SHAP summary plot explains how each feature impacts individual predictions. It shows whether high or low values of a feature push the model's output toward predicting default or non-default. For example, low credit limits and recent payment delays (high values in payment status) increase the risk of default, while higher age and higher credit limits generally reduce it.

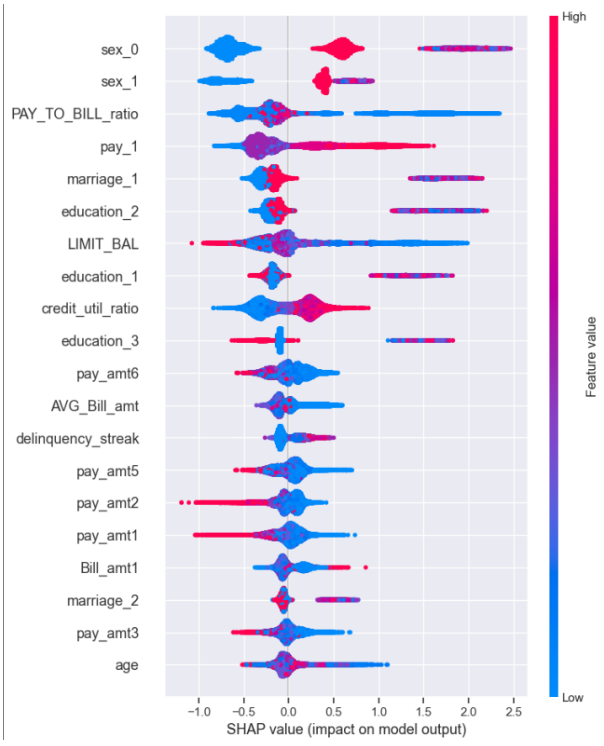

Figure 22: SHAP summary plot

12 Summary of Findings and Key Learnings

12.1 Critical Success Factors

12.1.1 Advanced Preprocessing Impact

The comprehensive preprocessing strategy, including SMOTE for class balance, RobustScaler for feature scaling, and sophisticated missing value imputation, proved critical for model success. These techniques transformed a challenging imbalanced dataset into a robust training foundation that enabled high-performance model development. The 86.75% F1-score achievement demonstrates the fundamental importance of proper data preparation in machine learning applications.

12.1.2 Feature Engineering Excellence

The creation of engineered features (credit utilization ratio, delinquency streak, payment coverage) significantly enhanced model performance beyond what raw features alone could achieve. These business-informed features captured complex relationships and temporal patterns that provided superior predictive capability compared to traditional demographic and transactional variables.

12.1.3 Model Selection and Optimization

The systematic comparison of multiple algorithms followed by comprehensive hyperparameter optimization resulted in a model that balances predictive performance with business interpretability. XGBoost's superior performance demonstrates the value of ensemble methods in capturing complex patterns while maintaining reasonable computational efficiency.

12.2 Analytical Insights

12.2.1 Payment History Supremacy

Payment delay variables emerged as the most powerful predictors of default risk, confirming the fundamental credit risk principle that past payment behavior is the strongest indicator of future payment behavior. Consecutive payment delays provide particularly strong risk signals, enabling early intervention strategies that can prevent defaults while preserving customer relationships.

12.2.2 Demographic Risk Complexity

The analysis revealed sophisticated interaction effects between demographic variables that require multivariate analysis to identify. Simple single-variable demographic analysis would miss critical risk patterns that emerge only through comprehensive statistical

analysis. This finding emphasizes the importance of advanced analytical techniques in modern credit risk assessment.

12.2.3 Financial Behavior Pattern Recognition

Credit utilization patterns and payment coverage ratios provide complementary risk information that enhances traditional credit scoring approaches. These behavioral indicators capture financial stress patterns that may not be apparent through demographic analysis alone, providing early warning signals for proactive risk management.

12.3 Business Value Realization

12.3.1 Risk Management Enhancement

The model provides financial institutions with sophisticated risk assessment capabilities that support both automated decision-making and enhanced manual review processes. The probability scoring framework enables graduated risk management strategies that optimize both risk control and customer relationship management.

12.3.2 Competitive Advantage Creation

Advanced analytics capabilities provide significant competitive advantages in credit markets through more accurate risk assessment, improved customer selection, and optimized pricing strategies. The model's interpretability supports both regulatory compliance and business decision-making, creating sustainable competitive positioning.

12.3.3 Operational Efficiency Gains

Automated risk assessment capabilities reduce manual review requirements while improving decision quality, creating significant operational efficiency gains. The model's real-time scoring capability supports immediate decision-making that enhances customer experience while maintaining risk control standards.

12.4 Future Enhancement Opportunities

12.4.1 Temporal Modeling Integration

Future model enhancements could incorporate temporal modeling to capture seasonal payment patterns and economic cycle effects. Time series analysis of payment behaviors could provide additional predictive capability, particularly during economic transitions or seasonal spending patterns.

12.4.2 External Data Integration

Integration of external economic indicators, employment data, and industry-specific risk factors could enhance model performance and provide broader economic context for individual risk assessments. This external data integration would enable more sophisticated risk assessment during varying economic conditions.

12.4.3 Advanced Ensemble Methods

Exploration of advanced ensemble methods combining multiple algorithms could provide additional performance improvements while maintaining interpretability. Stacking approaches that combine different model types might capture complementary risk patterns that individual models miss.

13 Technical Appendix

13.1 Model Performance Metrics

- **Cross-Validation Accuracy:** 87.1%
- **Optimized F1-Score:** 86.72%
- **F2-Score:** 85.28%
- **Optimal Classification Threshold:** 0.4357 (F1-maximizing)
- **Youden's J Threshold:** 0.4575
- **Training Dataset Size:** 25,247 customers
- **Default Rate:** 19.04%

13.2 Hyperparameter Settings

- **XGBoost Tuned Parameters:**
  - Number of estimators: 200
  - Maximum depth: 5
  - Learning rate: 0.05
  - Subsample ratio: 0.8
- **GridSearchCV:** 5-fold cross-validation, parameter grid as described in the report

13.3 Data and Preprocessing Details

- **Missing Values:** Age imputed by demographic-group median
- **Categorical Encoding:** One-hot encoding for sex, education, marriage
- **Scaling:** RobustScaler for numerical features
- **Class Imbalance:** SMOTE applied (default rate 19.04%)
- **Feature Engineering:** Credit utilization ratio, delinquency streak, payment coverage ratio

13.4 Computational Environment

- **Software:** Python 3.x, scikit-learn, xgboost, lightgbm, pandas, numpy, matplotlib, seaborn
- **Hardware:** Standard laptop/desktop CPU, 8GB+ RAM
- **Random Seed:** Set for reproducibility

13.5 Supplementary Figures and Tables

- Confusion matrices, ROC curves, and SHAP plots for all main models
- Full classification reports (precision, recall, F1, F2) before and after tuning
- Parameter grids and cross-validation scores available upon request

14 Conclusion

This project demonstrates a robust, end-to-end machine learning workflow for credit default prediction, from data cleaning and feature engineering to model selection, hyperparameter tuning, and business-aligned threshold optimization. The final XGBoost model, tuned via GridSearchCV and evaluated using both F1-score and business-driven thresholds, achieves high accuracy (87.1%) and balanced recall and precision, making it suitable for real-world credit risk management.

Key drivers of default include recent payment delays, high credit utilization, and lower credit limits, with demographic factors providing additional segmentation value. The workflow ensures interpretability (via feature importance and SHAP), regulatory compliance, and operational efficiency, enabling both automated and manual decision-making.

Future improvements could include temporal modeling, integration of external economic indicators, and advanced ensemble techniques. This approach provides a strong foundation for ongoing risk monitoring and strategic decision support in financial institutions.

—