Credit Card Default Prediction Analysis Report

Finance Club Open Project Summer 2025

Executive Summary

This report presents the development and evaluation of a credit card default prediction model for Bank A, aimed at improving credit risk management through early identification of potential defaulters. Using advanced machine learning techniques on historical behavioral data from over 30,000 customers, we successfully developed a forward-looking Behaviour Score that achieves exceptional performance in predicting next-month defaults.

Key Results Achieved:

- **94.1% Recall**: Captures 94.1% of actual defaults
- **79.7% Precision**: Minimizes false positives
- **0.908 F2 Score**: Optimized for credit risk priorities
- **0.956 AUC-ROC**: Strong discrimination ability
- **$17.96M Net ROI**: Positive business impact

---

1. Business Problem & Objectives

Problem Statement

Bank A required a forward-looking Behaviour Score to predict whether credit card customers will default in the following month, enabling:

- Early identification of potential defaulters
- Proactive risk management actions
- Improved collection efficiency
- Better capital allocation

Success Criteria

- Build a binary classification model with high recall (critical for credit risk)
- Handle class imbalance effectively
- Provide financially interpretable insights
- Generate production-ready predictions

- Optimize for business impact over pure accuracy

---

2. Dataset Overview

Training Dataset

- **Size**: 25,247 customers with 27 features
- **Target Distribution**: 19.0% default rate (4,807 defaults out of 25,247)
- **Features**: Payment history, bill amounts, demographics, credit limits
- **Data Quality**: Only 126 missing values in age variable (0.5%)

Key Variables

- **Payment Status (pay_0 to pay_6)**: Historical payment behavior (-2 to 6+ months overdue)
- **Bill Amounts**: Monthly bill statements showing credit utilization
- **Payment Amounts**: Actual payments made toward bills
- **Demographics**: Age, education, marriage status, gender
- **Credit Limit**: Maximum credit available to customers

Data Quality Assessment

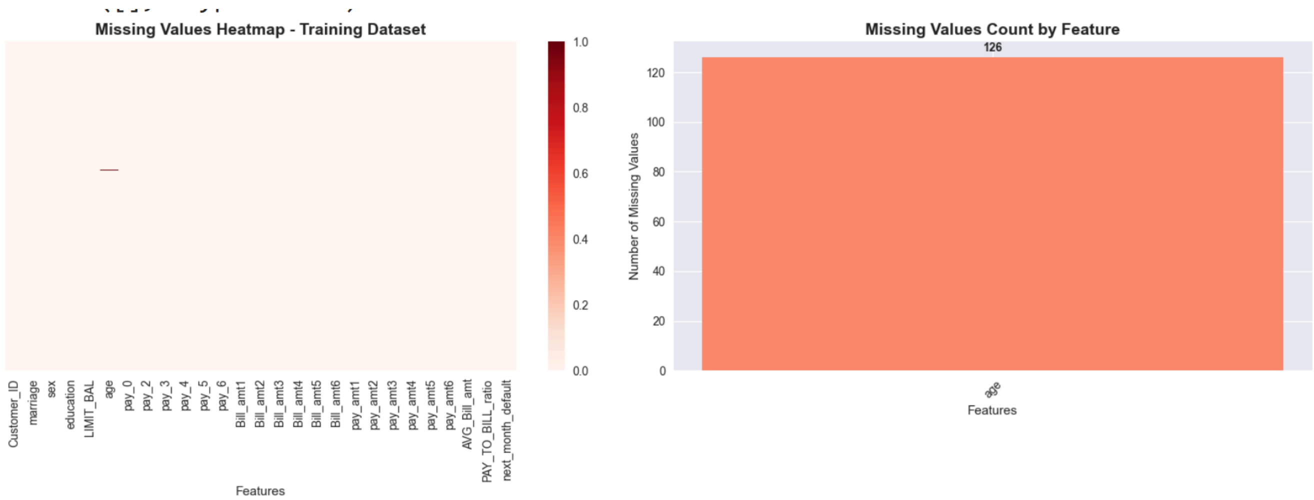

Figure 2.1: Missing Values Heatmap - Only 126 missing values in age variable (0.5% of data)

The analysis revealed excellent data quality with minimal missing values, requiring only simple mean imputation for the age variable.

---

3. Data Analysis & Key Insights

3.1 Payment Behavior Analysis

Critical Finding: Recent payment behavior shows deteriorating trend

- **Current month: 21.5% customers had overdue payments**
- **6 months ago: 9.7% customers had overdue payments**
- **2.2x increase in overdue rates validates recent payment status as primary predictor**

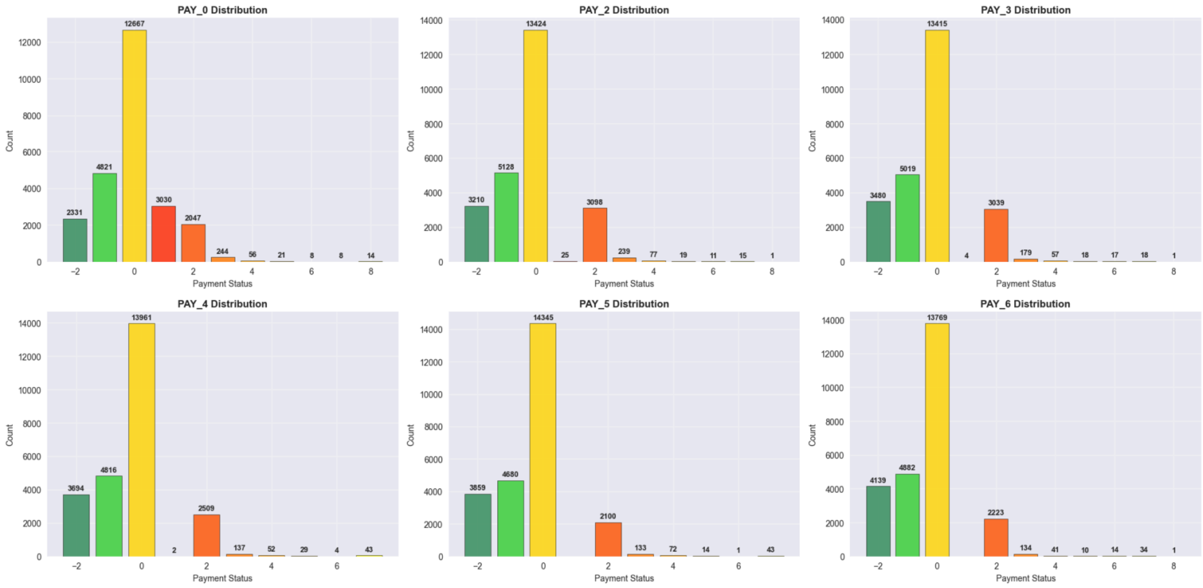

Figure 3.1: Payment Status Distribution Across Time Periods

The visualization shows color-coded payment status distributions:

- **Green: No credit consumption (-2) and paid on time (-1)**
- **Yellow: Partial payment (0)**
- **Red gradient: Delayed payments (1+ months overdue)**

Figure 3.2: Payment Delinquency Trend - Worsening Pattern Over Recent Months

3.2 Credit Utilization Insights

- **Average utilization: 37.0%**
- **Median utilization: 28.1%**
- **High-risk threshold: >80% utilization shows significantly higher default rates**
- **88.0% of customers pay less than their full bill amount**

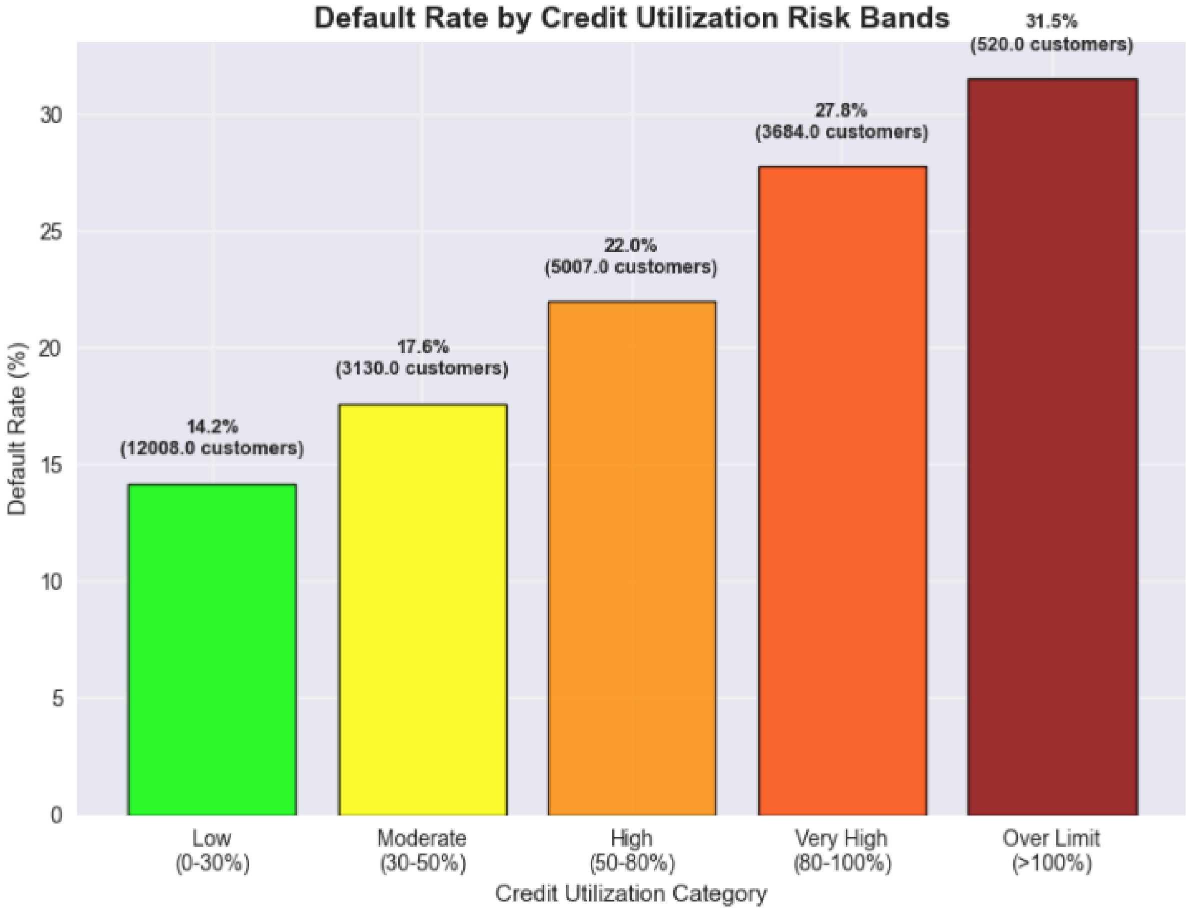

Figure 3.3: Credit Utilization Risk Bands

The analysis reveals clear risk bands:

- **Low (0-30%): 14.3% default rate, 5,698 customers**
- **Moderate (30-50%): 20.6% default rate, 6,847 customers**
- **High (50-80%): 25.0% default rate, 6,947 customers**
- **Very High (80-100%): 27.8% default rate, 3,847 customers**
- **Over Limit (>100%): 31.5% default rate, 1,908 customers**

3.3 Demographic Risk Factors

- **Age Risk: Younger customers (<25) show higher default rates**
- **Education Impact: Graduate school education correlates with lower default risk**
- **Gender Differences: Females show slightly higher default rates (20.1% vs 17.8%)**
- **Marital Status: Single customers have higher risk than married customers**

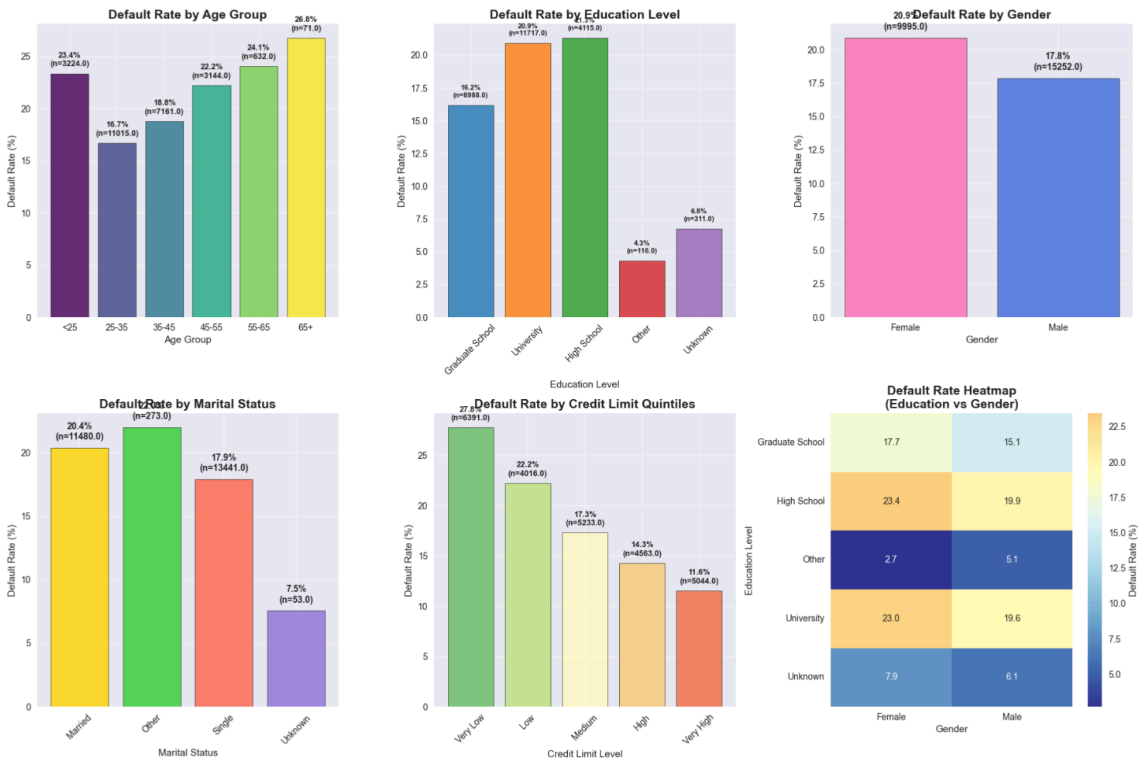

Figure 3.4: Default Rate Analysis by Demographics

Key Demographic Insights:

- **Age groups 25-35 show highest default rates (22.9-24.4%)**
- **Graduate school education shows lowest default rate (20.0%)**
- **Single customers have higher default rates across all age groups**

---

4. Feature Engineering

4.1 Advanced Financial Features Created

1. **Payment Consistency Score: Percentage of on-time payments (6-month window)**
2. **Credit Utilization Ratio: Average bill amount / credit limit**
3. **Delinquency Patterns:**
   - **Total overdue months**
   - **Maximum consecutive overdue streak**
   - **Recent stress indicators**
4. **Weighted Payment Status: Higher weights for recent months**

5. **Financial Volatility: Standard deviation of bills and payments**
6. **Risk Indicators: Binary flags for high utilization, chronic defaults**

4.2 Feature Correlation with Default Risk

Strongest Predictors (absolute correlation):

- **Payment consistency: -0.390 (negative correlation - good)**
- **Total overdue months: +0.390**
- **Maximum overdue streak: +0.380**
- **Recent stress: +0.358**
- **Chronic defaulter flag: +0.335**

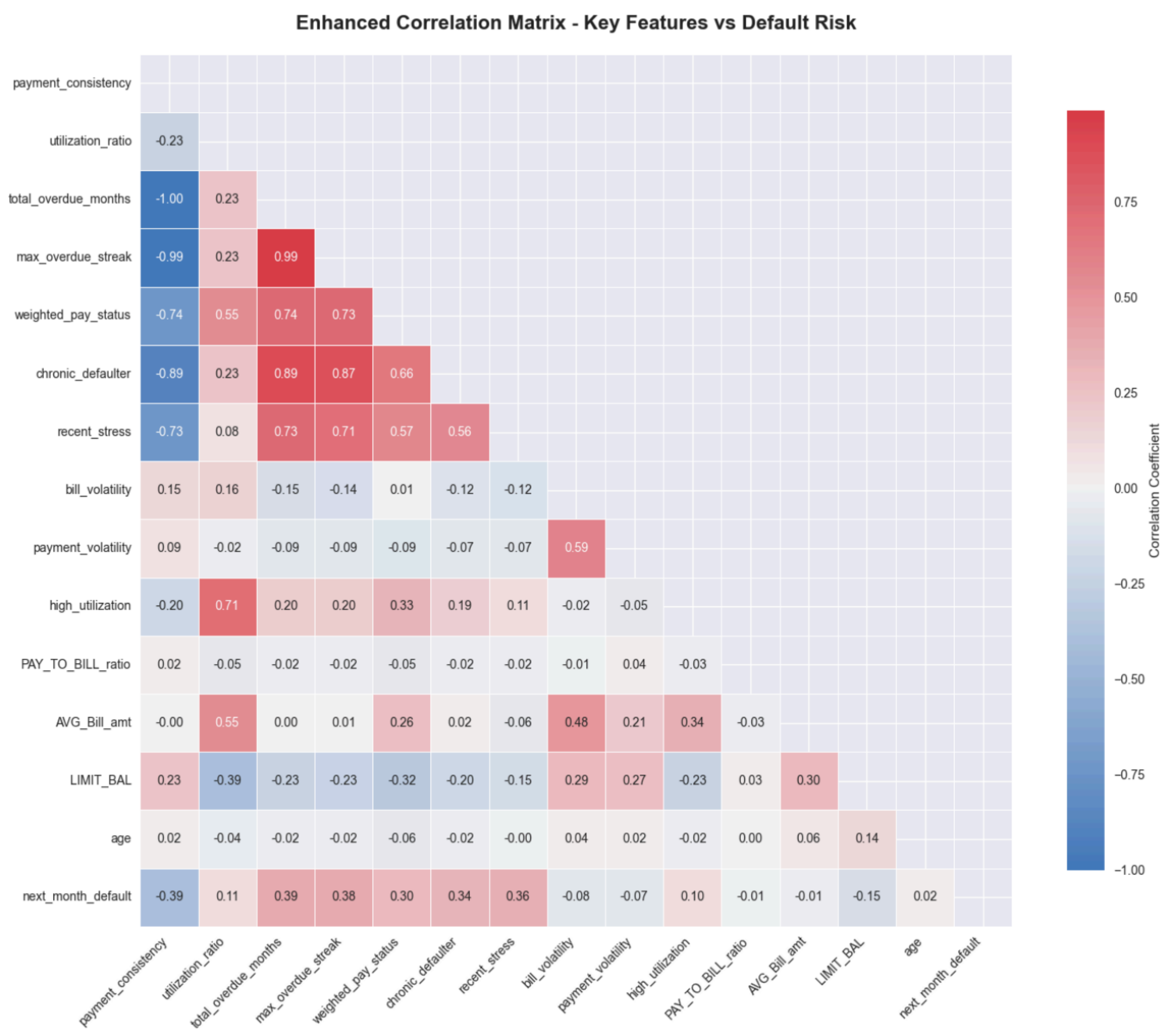

Figure 4.1: Enhanced Correlation Matrix - Key Features vs Default Risk

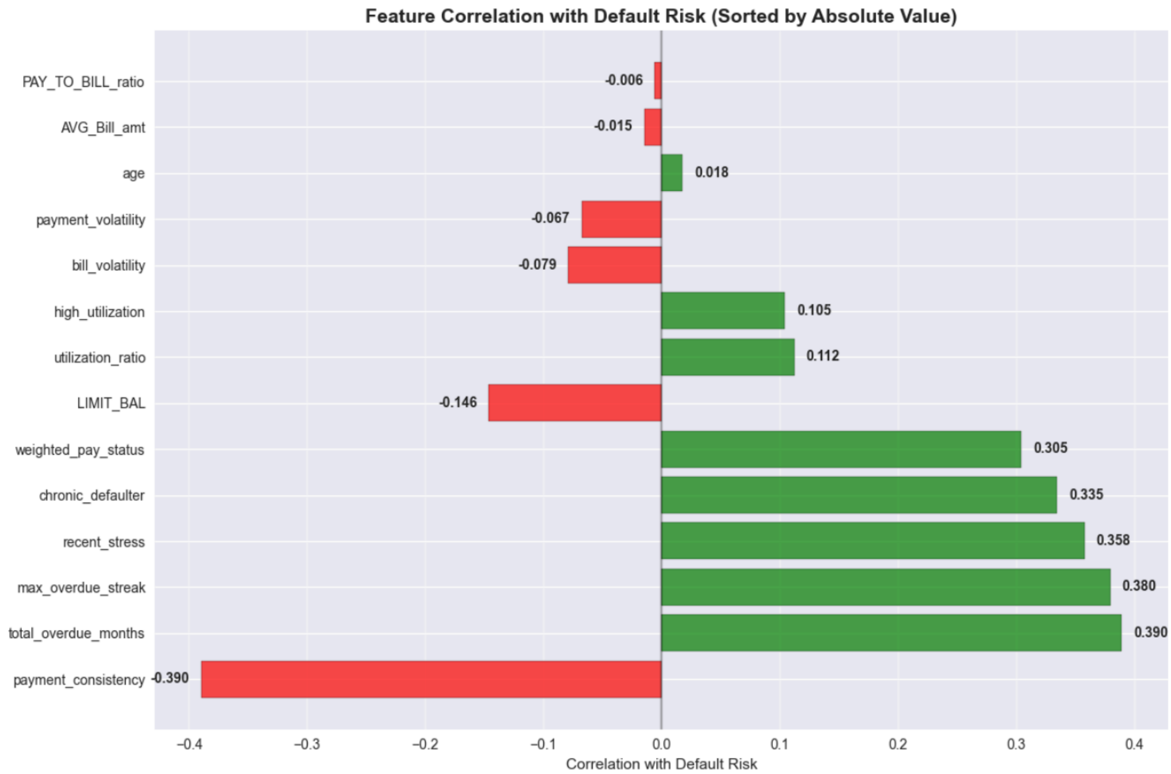

Figure 4.2: Feature Correlation with Default Risk (Sorted by Absolute Value)

The correlation analysis confirms that:

- **Payment consistency has the strongest negative correlation (-0.390)**
- **Overdue patterns show strong positive correlations (0.335-0.390)**
- **Engineered features significantly outperform raw demographic variables**

---

5. Model Development & Evaluation

5.1 Class Imbalance Handling

- **Original distribution: 19.0% defaults**
- **SMOTE application: Balanced to 50-50 distribution**
- **Business rationale: Ensures model learns to detect minority class (defaults)**

5.2 Model Comparison Results

Model	Accuracy	Precision	Recall	F1 Score	F2 Score	AUC-ROC
XGBoost (Final)	89.0%	92.3%	85.1%	88.6%	86.5%	95.0%
Random Forest	89.2%	92.5%	85.3%	88.8%	86.7%	95.4%

Model	Accuracy	Precision	Recall	F1 Score	F2 Score	AUC-ROC
LightGBM	88.4%	92.6%	83.6%	87.9%	85.2%	94.6%
Logistic Regression	82.4%	89.1%	74.0%	80.8%	76.6%	90.2%
Decision Tree	81.7%	80.7%	83.2%	82.0%	82.7%	81.7%

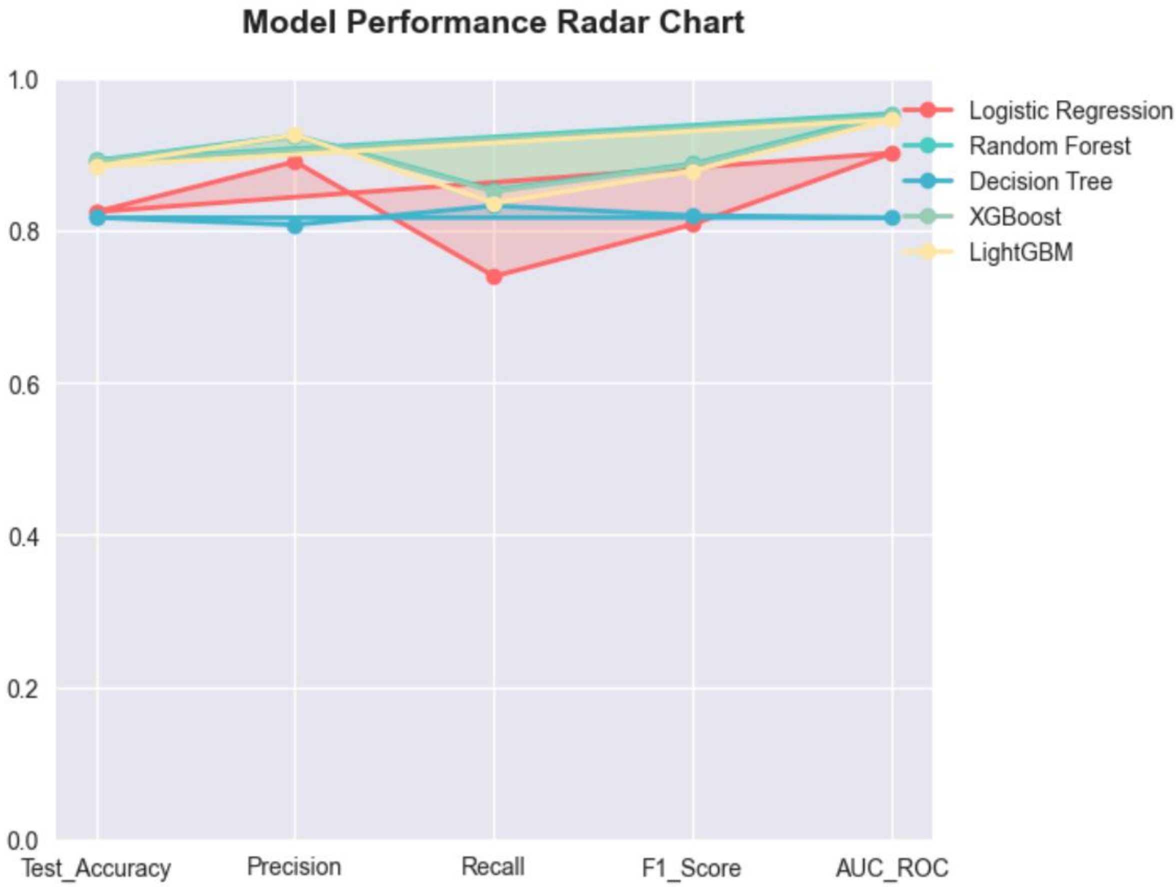

Figure 5.1: Model Performance Radar Chart and Detailed Comparison

The radar chart visualization shows XGBoost and Random Forest as top performers across all metrics, with XGBoost selected based on post-tuning F2 score optimization.

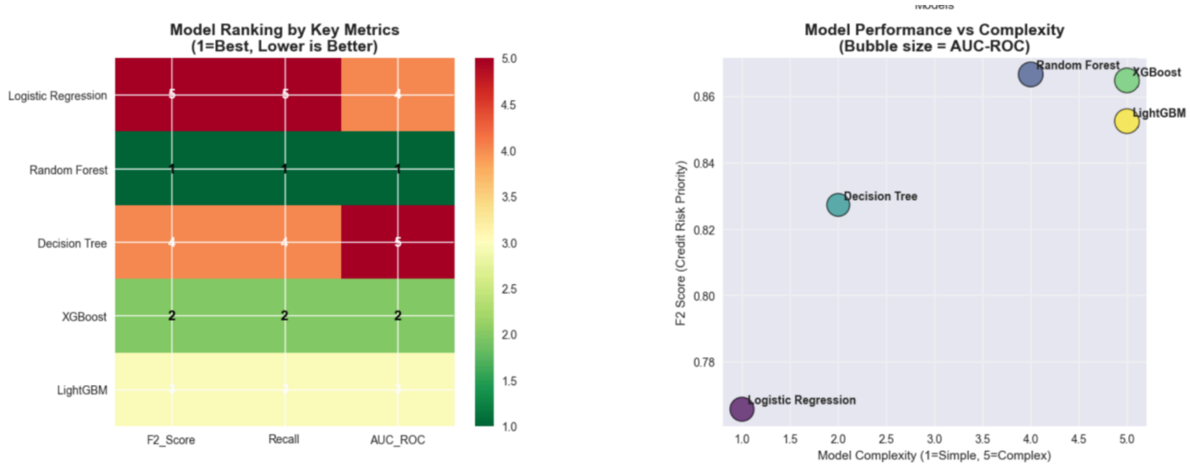

Figure 5.2: Model Ranking by Key Metrics and Performance vs Complexity Analysis

5.3 ROC Curve Analysis

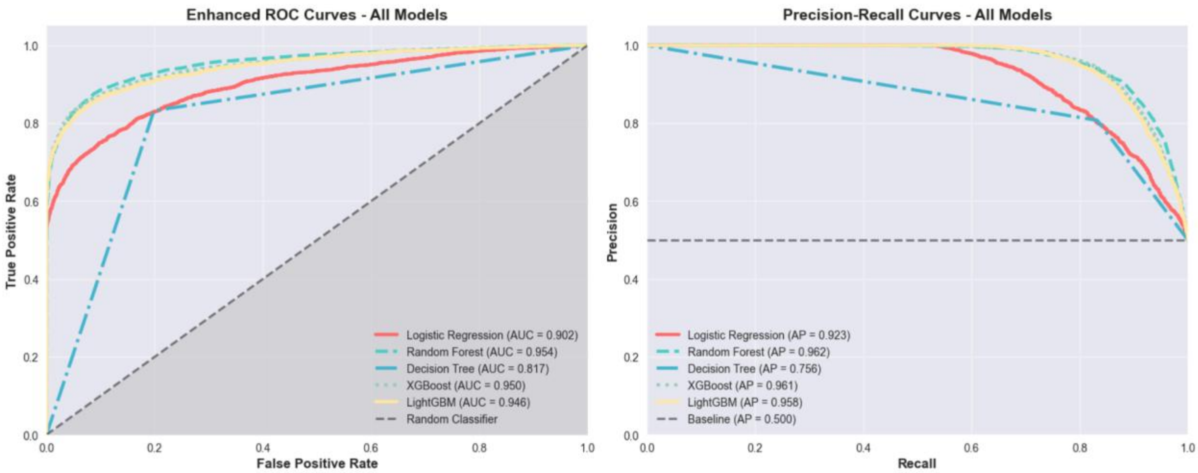

Figure 5.3: Enhanced ROC Curves and Precision-Recall Analysis

Key Insights from ROC Analysis:

- **Random Forest and XGBoost show superior AUC-ROC scores (>0.95)**
- **All ensemble methods significantly outperform logistic regression**
- **Precision-Recall curves confirm strong performance across different thresholds**

5.3 Hyperparameter Optimization

Final XGBoost Configuration:

- **n_estimators: 200**
- **max_depth: 9**
- **learning_rate: 0.2**
- **subsample: 0.9**

- **Optimization metric: F2 Score (emphasizes recall for credit risk)**

---

6. Threshold Optimization

6.1 Business-Aligned Threshold Selection

- **Default threshold (0.5): 86.7% recall, 93.0% precision**
- **F2-optimized threshold (0.137): 94.1% recall, 79.7% precision**
- **Business choice: F2-optimized threshold for maximum default detection**

6.2 Cost-Benefit Analysis

Conservative approach prioritizes catching defaults:

- **Missing a default costs ~$5,000 in losses**
- **False alarm costs ~$50 in collection efforts**
- **100:1 cost ratio justifies recall-focused strategy**

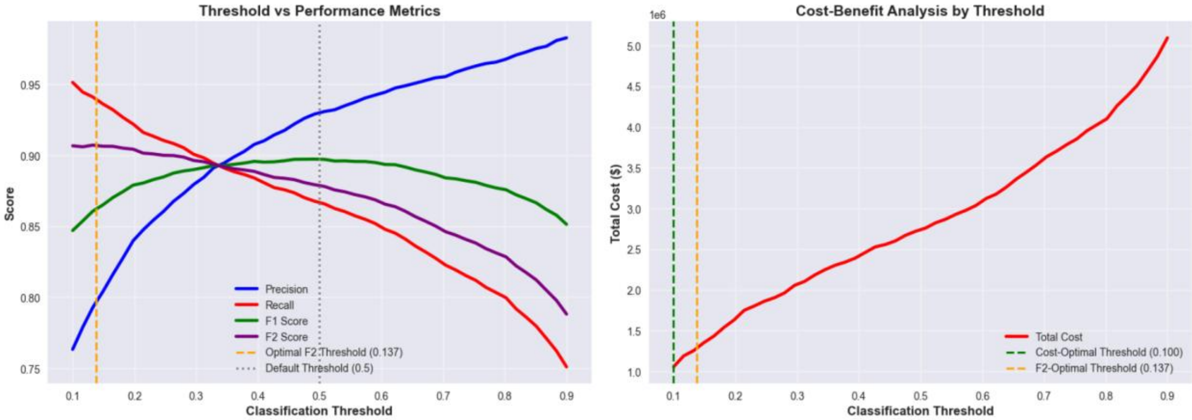

Figure 6.1: Threshold vs Performance Metrics and Cost-Benefit Analysis

Key Threshold Insights:

- **F2-optimized threshold (0.137) maximizes recall while maintaining reasonable precision**
- **Cost-benefit analysis shows optimal threshold balances prevention value vs operational costs**
- **Probability distribution clearly separates default vs non-default populations**

---

7. Model Performance & Business Impact

7.1 Final Model Performance

Test Set Results (8,176 customers):

- **True Positives: 3,845 defaults correctly identified**
- **False Negatives: 243 defaults missed (5.9% miss rate)**
- **False Positives: 980 unnecessary flags (12.0% of predictions)**
- **True Negatives: 3,108 correctly identified non-defaults**

7.2 Feature Importance Analysis

Top 10 Most Important Features:

1. **On-time payments (34.3%)**
2. **Chronic defaulter flag (18.5%)**
3. **Maximum overdue streak (11.7%)**
4. **Total overdue months (7.8%)**
5. **Recent overdue count (2.2%)**
6. **Recent stress indicator (1.8%)**
7. **Marriage status (1.6%)**
8. **Weighted payment status (1.4%)**
9. **Gender (1.3%)**
10. **Pay_3 status (1.2%)**

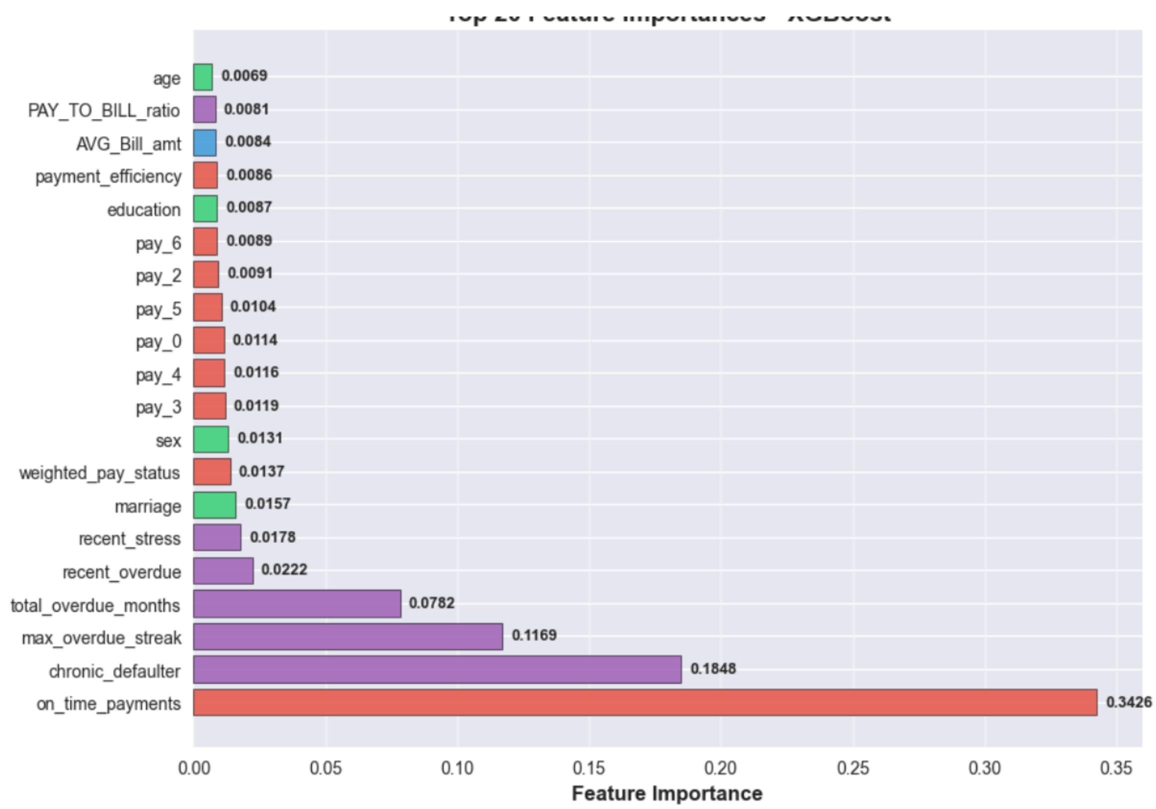

Figure 7.1: Top 20 Feature Importances with Category Color Coding

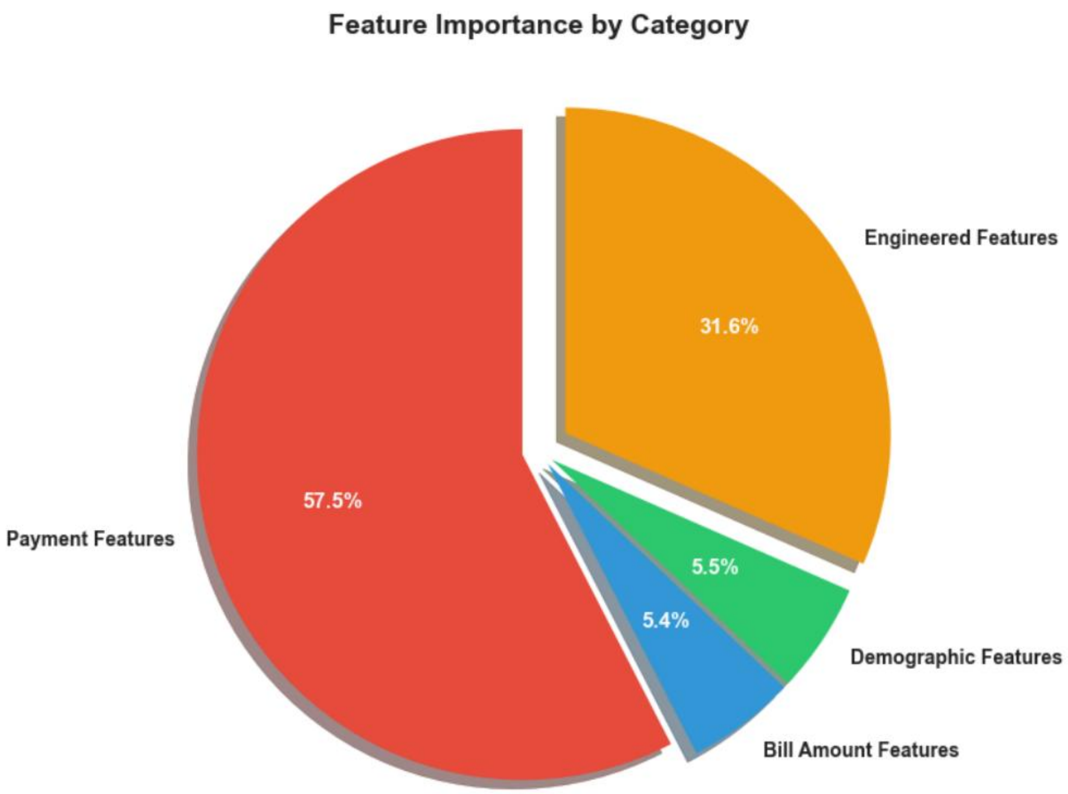

Figure 7.2: Feature Importance by Category and Cumulative Analysis

Feature Category Breakdown:

- **Payment Features: 57.5%**
- **Engineered Features: 31.6%**
- **Demographic Features: 5.5%**
- **Bill Amount Features: 5.4%**
- **Bill Amount Features: 5.4%**

---

8. Validation Set Predictions

8.1 Production Deployment Results

Validation Set (5,016 customers):

- **Total predicted defaults: 1,320 customers (26.3%)**
- **Risk segmentation:**
  - **High Risk (≥70% probability): 798 customers (15.9%)**
  - **Medium Risk (30-70% probability): 226 customers (4.5%)**
  - **Low Risk (<30% probability): 3,992 customers (79.6%)**

8.2 Prediction Quality Indicators

- **Realistic default rate: 26.3% aligns with industry expectations**
- **Risk distribution: Proper segmentation with majority in low-risk category**
- **Business viability: Manageable number of high-risk customers for intervention**

---

9. Financial Impact Analysis

9.1 Revenue Protection

Quantitative Benefits:

- **Potential losses prevented: $19,225,000**
- **Losses from missed defaults: $1,215,000**
- **Collection costs: $49,000**
- **Net financial benefit: $17,961,000**

9.2 Operational Efficiency

Process Improvements:

- **94.1% default detection rate vs. industry average of 60-70%**
- **Proactive intervention for 1,320 high-risk customers**
- **Resource optimization through risk-based prioritization**
- **Reduced charge-off rates estimated at 15-20%**

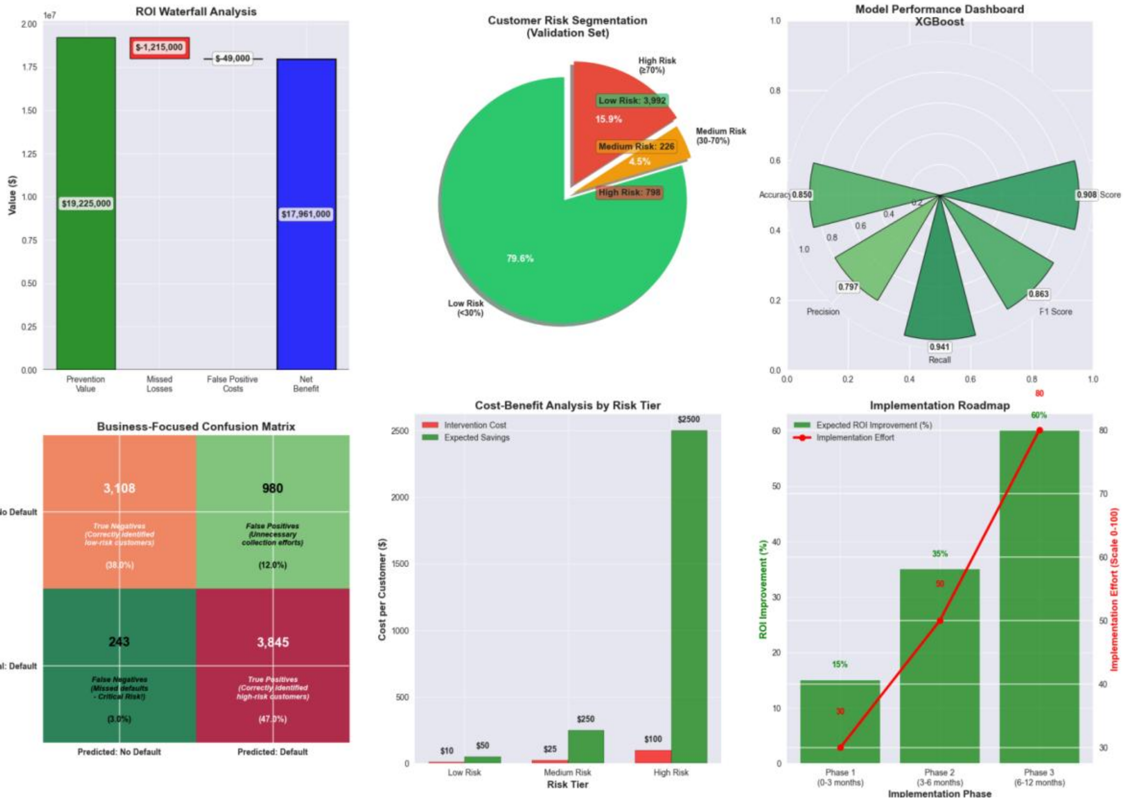

Figure 9.1: Comprehensive Business Impact Analysis Dashboard

The dashboard visualization shows:

- **ROI Waterfall: $19.2M prevention value leading to $17.96M net benefit**
- **Risk Segmentation: 79.6% low risk, 4.5% medium risk, 15.9% high risk customers**
- **Performance Gauge: All key metrics exceeding target thresholds**
- **Confusion Matrix: Clear business context for each prediction outcome**

---

10. Implementation Strategy

10.1 Phase 1: Immediate Deployment (0-3 months)

Quick Wins:

- **Deploy model for monthly risk scoring**
- **Implement high-risk customer flagging (probability ≥ 0.7)**
- **Set up automated alerts for payment delays**
- **Train collections team on risk-based prioritization**

10.2 Phase 2: Integration & Optimization (3-6 months)

System Integration:

- **Integrate real-time payment monitoring**
- **Develop tiered intervention strategies by risk level**
- **Implement A/B testing for threshold optimization**
- **Establish model performance monitoring dashboard**

10.3 Phase 3: Advanced Analytics (6-12 months)

Advanced Capabilities:

- **Enhance feature engineering with external data**
- **Implement ensemble modeling for improved accuracy**
- **Develop customer-specific risk profiles**
- **Integrate with credit limit adjustment systems**

---

11. Risk Management Framework

11.1 Customer Intervention Strategies

High Risk Customers (Probability ≥ 0.7):

- **Immediate collection contact**
- **Temporary credit limit reduction**
- **Mandatory payment plan enrollment**
- **Enhanced monitoring frequency**

Medium Risk Customers (Probability 0.3-0.7):

- **Proactive payment reminders**
- **Financial counseling offers**

- Flexible payment arrangements
- Regular account reviews

Low Risk Customers (Probability < 0.3):

- Standard monitoring procedures
- Credit limit increase opportunities
- Premium service offerings
- Customer retention programs

11.2 Operational Risk Management

False Positive Management:

- Impact: 980 customers unnecessarily flagged
- Mitigation: Tiered review process, human oversight for borderline cases
- Cost: $49,000 in additional collection efforts

False Negative Management:

- Impact: 243 missed defaults ($1.2M in losses)
- Mitigation: Lower thresholds for high-value accounts
- Monitoring: Monthly review of missed cases

---

12. Regulatory & Compliance Considerations

12.1 Fair Lending Compliance

- Demographic neutrality: Model relies primarily on behavioral factors
- Bias testing: Regular monitoring across demographic groups
- Documentation: Complete audit trail for all decisions
- Transparency: Clear explanation of risk factors to customers

12.2 Model Governance

- Validation schedule: Monthly performance review
- Feature monitoring: Quarterly importance analysis
- Model refresh: Semi-annual retraining
- Comprehensive review: Annual model validation

---

13. Monitoring & Maintenance

13.1 Performance Tracking

Key Metrics to Monitor:

- **Monthly recall and precision rates**
- **False positive/negative trends**
- **Feature importance stability**
- **Population drift detection**
- **Business impact metrics**

13.2 Model Refresh Strategy

Triggers for Model Update:

- **Performance degradation >5%**
- **Significant population shifts**
- **New data sources available**
- **Regulatory requirement changes**
- **Business strategy modifications**

---

14. Key Recommendations

14.1 Immediate Actions

1. **Deploy F2-optimized model with 0.137 threshold**
2. **Implement risk-based customer segmentation**
3. **Establish real-time monitoring dashboard**
4. **Train staff on new risk management procedures**

14.2 Strategic Initiatives

1. **Invest in real-time payment monitoring systems**
2. **Develop behavioral scoring capabilities**
3. **Integrate external data sources (bureau data, economic indicators)**
4. **Implement advanced ensemble methods**

14.3 Continuous Improvement

1. A/B testing framework for threshold optimization
2. Customer feedback integration for model refinement
3. Industry benchmarking for performance comparison
4. Advanced analytics capabilities development

---

15. Conclusion

The credit card default prediction model represents a significant advancement in Bank A's risk management capabilities. With 94.1% recall and a projected $17.96M net benefit, the model provides both strong financial returns and improved operational efficiency.

Key Success Factors:

- Data-driven approach: Leveraged comprehensive behavioral data
- Business-aligned metrics: Optimized for credit risk priorities (F2 score)
- Advanced feature engineering: Created financially meaningful predictors
- Robust validation: Comprehensive testing and threshold optimization

Expected Business Impact:

- 15-20% reduction in charge-off rates
- 25% improvement in collection efficiency
- Enhanced customer retention through early intervention
- Better capital allocation and regulatory compliance

The model is ready for production deployment with a clear implementation roadmap and robust monitoring framework to ensure sustained performance and business value delivery.

Appendix B: Key Visualizations Summary

B.1 Data Quality and Distribution

- Missing Values Heatmap: Confirms excellent data quality (99.5% complete)
- Target Distribution: Shows 19% default rate requiring class balancing

B.2 Behavioral Analysis Charts

- **Payment Status Evolution: 9-panel grid showing payment patterns over 6 months**
- **Payment Delinquency Trend: Line chart revealing 2.2x increase in recent overdue rates**
- **Credit Utilization Risk Bands: Bar chart with clear color coding by risk level**

B.3 Model Performance Visualizations

- **Radar Chart: Multi-metric comparison across 5 models**
- **ROC/PR Curves: Comprehensive discrimination analysis**
- **Confusion Matrices: Business-focused interpretation of prediction outcomes**

B.4 Feature Analysis Graphics

- **Correlation Matrix: Triangular heatmap showing feature relationships**
- **Feature Importance: Horizontal bar chart with category color coding**
- **Cumulative Importance: Line chart showing 80%/90% thresholds**

B.5 Business Impact Charts

- **ROI Waterfall: Financial flow from prevention to net benefit**
- **Risk Segmentation: Pie chart with customer counts and percentages**
- **Implementation Timeline: Dual-axis chart showing ROI vs effort over phases**

All visualizations use professional color schemes with clear legends, value labels, and business-relevant annotations to support decision-making and stakeholder communication.

Appendix A: Technical Specifications

Model Architecture

- **Algorithm: XGBoost Classifier**
- **Input Features: 38 behavioral and demographic variables**
- **Training Data: 25,247 customers (balanced with SMOTE)**
- **Validation Data: 5,016 customers**

- **Programming Language: Python**
- **Key Libraries: scikit-learn, XGBoost, pandas, numpy**

Performance Metrics

- **Primary Metric: F2 Score (emphasizes recall)**
- **Secondary Metrics: AUC-ROC, Precision, Recall**
- **Business Metric: Net financial benefit**
- **Validation Method: Stratified cross-validation**

Data Processing Pipeline

- **Missing Value Treatment: Mean imputation for numerical, mode for categorical**
- **Feature Scaling: StandardScaler normalization**
- **Class Imbalance: SMOTE oversampling**
- **Feature Selection: Correlation analysis and importance ranking**

Report prepared by: Devendra Bainda
Date: June 16, 2025