# Model Evaluation and Selection Report

## 1. Executive Summary

- **Objective**: To develop a predictive model for default risk, evaluating multiple algorithms and techniques to identify the most effective model for deployment.
- **Selected Model**: XGBoost with Under-Sampling
  - **Key Metrics**:
    - AUC: 0.98
    - Gini Coefficient: 0.97
    - KS Statistic: 86.87%
  - **Reason for Selection**: Best separation of default and non-default classes with strong interpretability using SHAP and LIME analysis.
- **Conclusion**: XGBoost (under-sampling) is recommended for deployment, supported by robust evaluation metrics and meaningful insights into feature importance.

---

## 2. Data Preparation and Feature Engineering

### 2.1 Data Preprocessing

- **Target Variable**: Default (binary classification)
- **Imbalanced Dataset**: The dataset exhibited a class imbalance with~10% defaults and~90% non-defaults.
- Addressed imbalance through:
  - **Over-sampling**: SMOTE (Synthetic Minority Oversampling Technique)
  - **Under-sampling**: Random majority class reduction

### 2.2 Selected Features

Based on VIF analysis and domain relevance, the following features were retained:

- **Numerical**: credit_utilization_ratio, age, income, dmtlm, lti, etc.
- **Categorical**: loan_purpose, residence_type, employment_status, etc.

---

# 3. Model Development

## 3.1 Models Evaluated

- Logistic Regression
- Random Forest
- XGBoost

## 3.2 Evaluation Metrics

- AUC-ROC Curve
- Precision on default class (50%+) and recall on default class (90%+)
- Gini Coefficient
- KS Statistic
- Decile Analysis (Event/Non-Event Rates)

---

# 4. Model Performance

**4.1 Comparison of Models** (All the models satisfies Precision on default class (50%+) and recall on default class (90%+)

| Model | KS Statistic (%) | AUC | Gini Coefficient | Remarks |
|---|---|---|---|---|
| XGBoost Best Model (Under-Sampling, Optuna) | **86.87** | **0.9863** | **0.9726** | Selected for deployment due to highest performance |
| XGBoost Best Model (Under-Sampling) | 86.67 | 0.9857 | 0.9715 | Excellent performance, close to the best |
| Logistic Regression (Best Over-Sampling) | 86.25 | 0.9833 | 0.9665 | Strong model, consistent results |
| Logistic Regression (Best Under-Sampling) | 86.25 | 0.9833 | 0.9666 | Similar performance to over-sampling |
| Logistic Regression (Optuna Over-Sampling) | 86.15 | 0.9833 | 0.9665 | Comparable to best logistic regression models |

| | | | | |
|---|---|---|---|---|
| Logistic Regression (Optuna Under-Sampling) | 86.25 | 0.9833 | 0.9666 | Similar to other logistic regression models |
| XGBoost (Standard Under-Sampling) | 86.25 | 0.9824 | 0.9648 | Slightly lower performance than optimized XGBoost |
| Logistic Regression (Standard Over-Sampling) | 86.25 | 0.9833 | 0.9665 | Similar performance to best logistic regression |
| Logistic Regression (Standard Under-Sampling) | 86.25 | 0.9832 | 0.9664 | Similar to other logistic regression models |

## **Key Highlights**:

- The **XGBoost (Optuna Under-Sampling)** model showed the best overall performance, with the highest KS statistic, AUC, and Gini coefficient.
- Logistic Regression models (both under-sampling and over-sampling) performed well but fell slightly short compared to XGBoost models.
- XGBoost models with Optuna tuning consistently outperformed their counterparts, making them the preferred choice for deployment.

### 4.2 Selected Model: XGBoost (Optuna Under-Sampling)

- **Tuned Hyperparameters**:

  {'eta': 0.0396, 'max_depth': 3, 'subsample': 0.6272, 'colsample_bytree': 0.7137, 'n_estimators': 388}

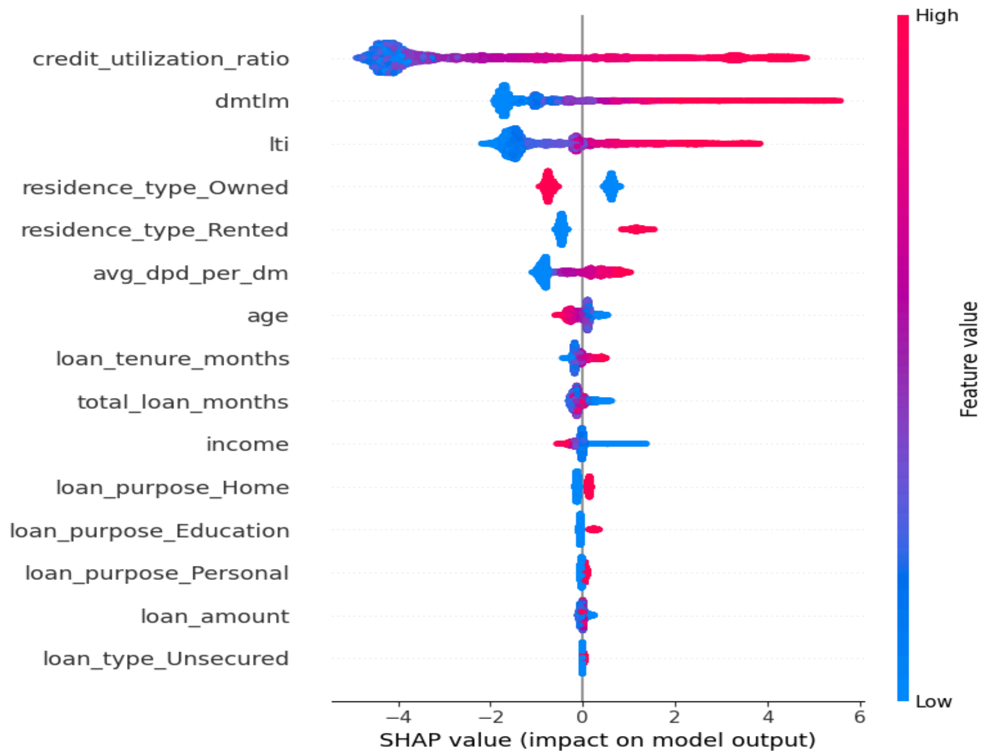
- **Key Insights**:
  - Best AUC, Gini, and KS values among all models.
  - Demonstrated superior ability to separate default and non-default classes.

---

# 5. Interpretability

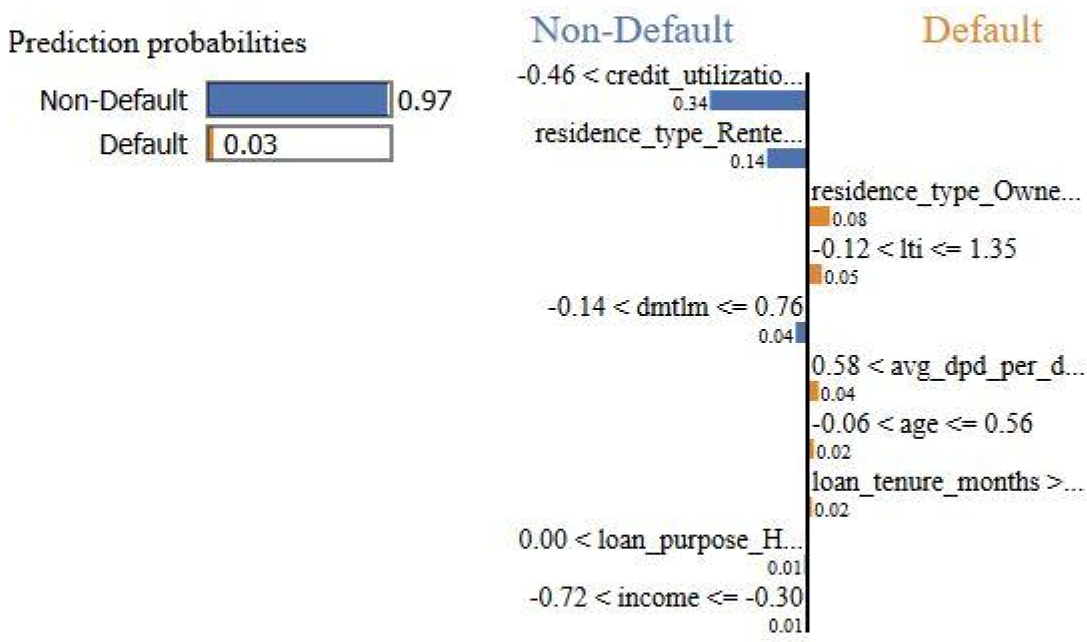
## 5.1 SHAP Analysis

- **Top Features Influencing Predictions**:
    1. credit_utilization_ratio (highly correlated with default risk)
    2. dmtlm (deliquency months to loan months ratio)
    3. lti (loan-to-income ratio)
- **Visualization** SHAP summary plot



## 5.2 LIME Analysis

- Explained individual predictions with local interpretability.
- Example (instance from first test data):
  - Predicted as non-default with 97% probability and matched with actual outcome.
  - Influencing factors: Low credit utilization ratio, no recent missed payments, and high income.




---

## 6. Decile Analysis (XGBoost Optuna Under-Sampling)

The decile analysis provides a detailed breakdown of the model's predictions, segmented into ten deciles based on predicted probabilities. Each decile represents a group of instances with similar predicted probabilities, sorted from highest to lowest risk. The table below summarizes key metrics for each decile.

| Decile | Min Probability | Max Probability | Event Count | Non-Event Count | Event Rate | Non-Event Rate | Cumulative Event Rate | Cumulativeulative Non-Event Rate | KS (%) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.00 | 0.00 | 0 | 1236 | 0.00% | 11.28% | 0.00% | 11.28% | 10.94 |
| 2 | 0.00 | 0.00 | 0 | 1236 | 0.00% | 11.28% | 0.00% | 22.57% | 21.88 |
| 3 | 0.00 | 0.00 | 0 | 1235 | 0.00% | 11.27% | 0.00% | 33.84% | 32.81 |
| 4 | 0.00 | 0.00 | 0 | 1236 | 0.00% | 11.28% | 0.00% | 45.12% | 43.75 |
| 5 | 0.00 | 0.00 | 0 | 1235 | 0.00% | 11.27% | 0.00% | 56.39% | 54.68 |
| 6 | 0.00 | 0.01 | 0 | 1236 | 0.00% | 11.28% | 0.00% | 67.67% | 65.62 |
| 7 | 0.01 | 0.05 | 0 | 1235 | 0.00% | 11.27% | 0.00% | 78.94% | 76.55 |
| **8** | **0.05** | **0.26** | **6** | **1230** | **0.49%** | **11.22%** | **0.01%** | **90.17%** | **86.87** |
| 9 | 0.26 | 0.83 | 159 | 1076 | 12.87% | 9.81% | 16.32% | 97.06% | 81.37 |
| 10 | 0.83 | 1.00 | 893 | 343 | 72.43% | 3.12% | 100.00% | 100.00% | 0.00 |

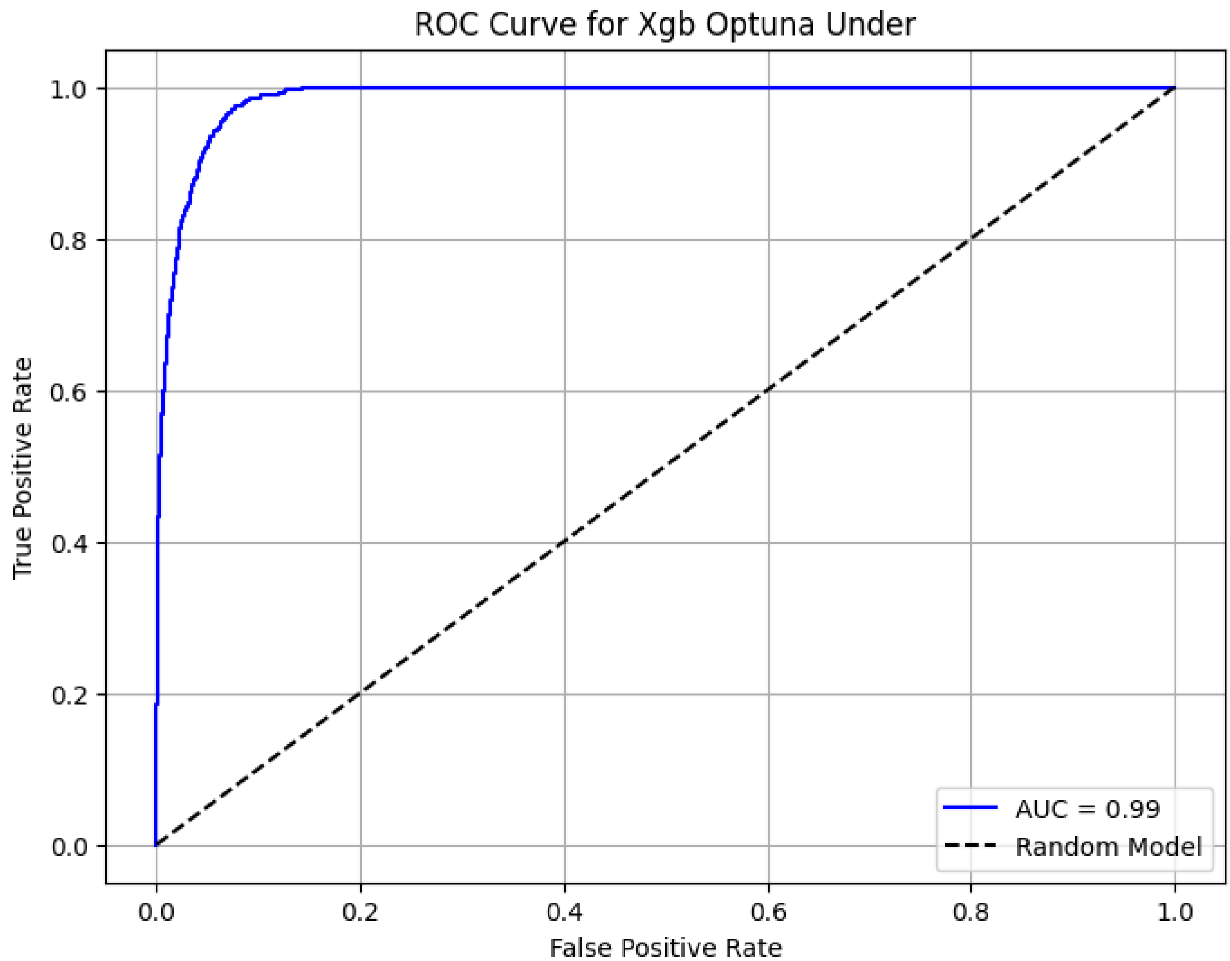
## Key Observations:

1. **Event Distribution**: The majority of events (defaults) are concentrated in deciles 9 and 10, confirming the model's ability to rank high-risk instances accurately.
2. **KS Statistic**: The **KS statistic** reaches its peak at **86.87% in decile 8**, indicating excellent discriminatory power between events and non-events.
3. **Non-Event Concentration**: Deciles 1–7 contain only non-events, aligning with their lower predicted probabilities.
4. **Cumulative Analysis**: The cumulative event rate and non-event rate trends highlight the model's consistent separation between high- and low-risk groups.

This decile analysis further supports the **XGBoost Optuna Under-Sampling** model as a strong candidate for deployment due to its robust performance in identifying and separating events and non-events effectively.

# 7. Visualizations

1. **ROC Curve**:




- **AUC: 0.99 (Excellent classification performance)**

---

# 8. Deployment Considerations

## 8.1 Justification for Deployment

- XGBoost (under-sampling) shows consistent performance across key metrics and decile analysis.
- Interpretability tools (SHAP, LIME) make the model explainable and suitable for regulatory and business use.

## 8.2 Risks and Mitigations

- **Risk**: Under-sampling may lose information from the majority class.
- **Mitigation**: Regular retraining, monitoring, and exploration of hybrid sampling methods.

---

# 9. Conclusion

- XGBoost (under-sampling) is the recommended model for deployment based on its superior discriminatory power and robust evaluation.
- Next Steps:
    1. Develop a monitoring framework after deployment.
    2. Perform stress testing on newer data for validation.
    3. Ensure interpretability reports accompany the deployed model.