ENSURING ACCURACY AND FAIRNESS:

A CASE STUDY ON QUALITY ASSURANCE IN PREDICTIVE ANALYTICS FOR CREDIT RISK ASSESSMENT

Prepared by :

Kitiksha Chakrawarti

TABLE OF CONTENTS

OVERVIEW

This case study outlines a structured approach to credit risk assessment aimed at building a predictive model that is both accurate and fair. Through a comprehensive quality assurance process, the focus was on improving data quality, minimizing bias in sensitive features, and creating a scalable and transparent solution.

KEY HIGHLIGHTS:

- **Datasets Used:** Two diverse financial datasets were analyzed to validate model consistency across varying structures and customer profiles.
- **Objective:** To implement a robust quality assurance framework that enhances the accuracy, reliability, and fairness of a credit risk assessment model,
- **Focus Areas:** Ensured clean, well-prepared data and addressed bias in demographic attributes.
- **Outcome:** Developed a reliable and interpretable model supported by a reusable pipeline.

INTRODUCTION

BACKGROUND OF THE PROBLEM:

In financial institutions, predictive models play a vital role in evaluating the creditworthiness of loan applicants. However, poor data quality, overfitting, and unintended bias in these models can lead to inaccurate credit decisions. Recognizing these risks, the case-study aimed to strengthen the integrity of its credit risk model through a structured quality assurance approach. Key challenges included inconsistent data, demographic bias, and maintaining reliable model performance across datasets.

OBJECTIVE:

To develop a credit risk assessment model by implementing end-to-end quality assurance measures—ensuring data accuracy, reducing demographic bias, improving model interpretability, and supporting consistent and fair credit decision-making.

PROBLEM STATEMENT

Building reliable credit risk assessment models poses several challenges, particularly when working with large and diverse financial datasets.

Common issues include:

- Poor data quality, such as missing or inconsistent entries along with outliers.
- Overfitting of models that fail to generalize well.
- Unintended bias in predictions, especially across demographic segments.

Without a structured quality assurance framework, these problems can lead to inaccurate outcomes, unfair treatment of applicants, and significant operational and reputational risks. There is a critical need for a systematic approach that ensures clean data and fair modeling practices.

METHODOLOGY

To build a robust, accurate, and fair credit risk assessment model, a structured and multi-stage methodology was adopted. The approach combined in-depth exploratory analysis, statistical validation, advanced preprocessing, model experimentation, fairness evaluation, and automation.

1. **Exploratory Data Analysis (EDA)**
Started with a deep dive into the data to uncover patterns, gaps, and unusual values. Visual tools helped spot trends, inconsistencies, and imbalances across features.

2. **Data Cleaning and Preprocessing**

- Outliers: Treated using IQR and Z-score methods.
- Imputation: Mode imputation for categorical and Median Imputation for Numerical features
- Feature Selection: Statistical tests like VIF for multicollinearity, Chi-square and Cramér's V for categorical relevance and SelectKBest for model based selection.
- Encoding: One-Hot Encoding (low cardinality), Target Encoding (high cardinality).
- Balancing: SMOTE and class weight adjustments.
- Scaling: Standard Scaling

3. **Model Training & Evaluation**

- Trained Logistic Regression, Random Forest, XGBoost and ensemble techniques such as Voting and Stacking Classifier.
- Used Stratified K-Fold Cross Validation and hyperparameter tuning to enhance performance.
- Various metrics beyond accuracy such as precision, recall, F1 score, ROC-AUC and confusion matrix were evaluated.

4. **Fairness Assessment & Mitigation**

- Checked whether the model treated different groups fairly using fairness metrics.
- Took steps to reduce bias by adjusting model thresholds and using fairness-aware techniques.
- Created visual dashboards to keep track of fairness across sensitive features.

5. **Model Interpretability**

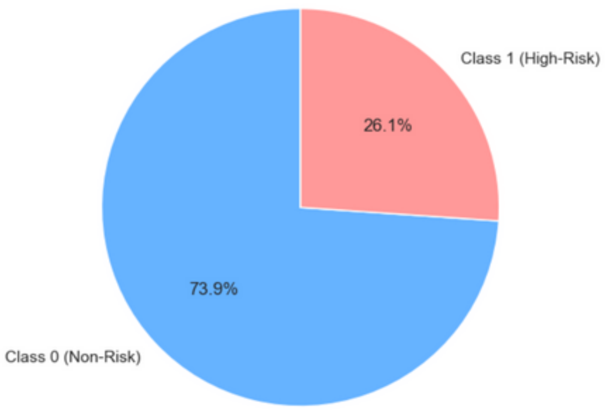
Used SHAP to explain feature influence on the model and to ensure transparency of model

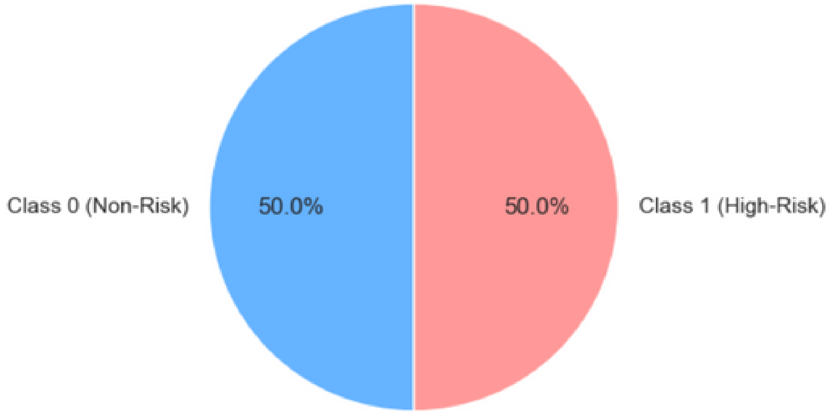
6. **Generalized QA Pipeline**

Finalized an automated preprocessing pipeline that handles cleaning, encoding, feature selection, model training, and evaluation with minimal manual effort—scalable across multiple datasets.

KEY INSIGHTS

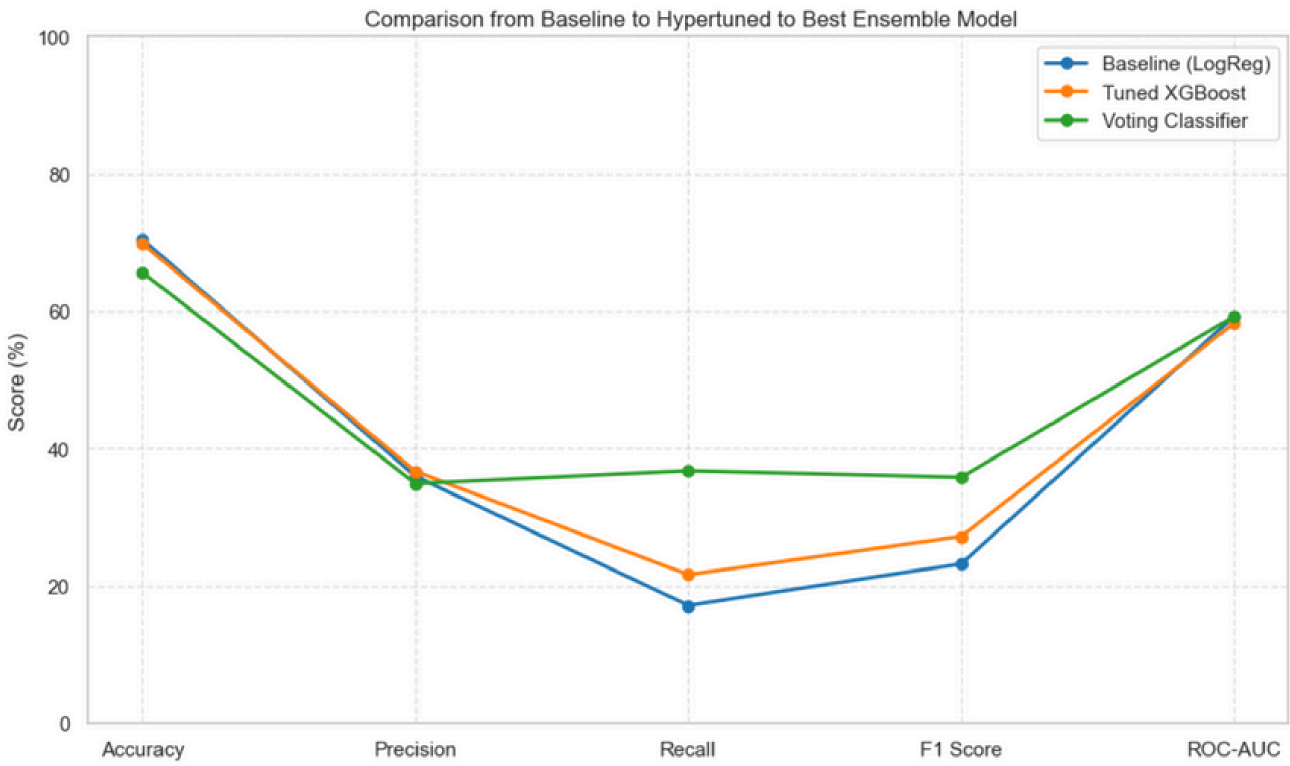
1. Class Imbalance Before and After SMOTE





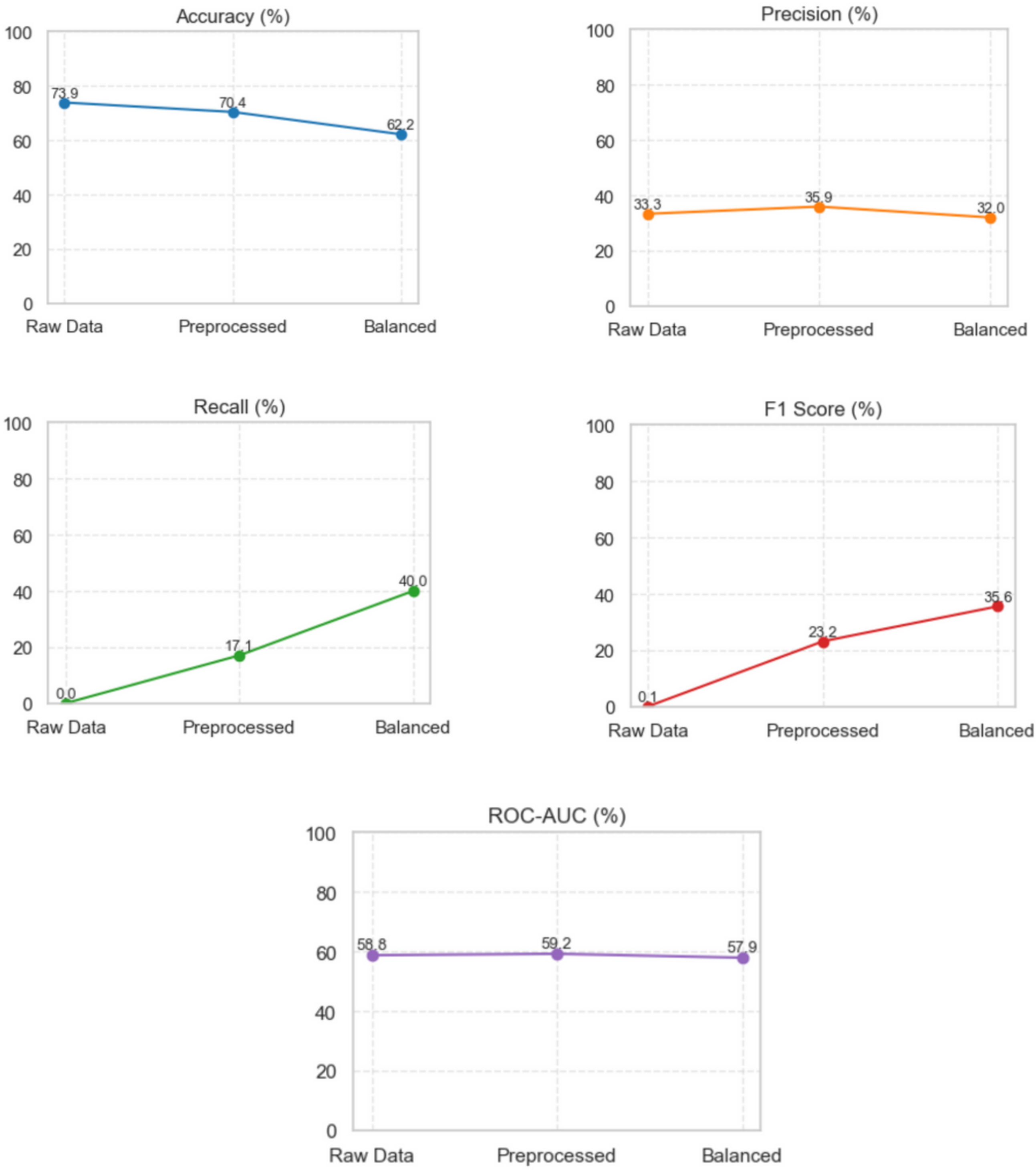
Class 1 (high-risk applicants) was underrepresented before. SMOTE balanced both classes to 50%.

2. Model Progression: From Baseline to Final



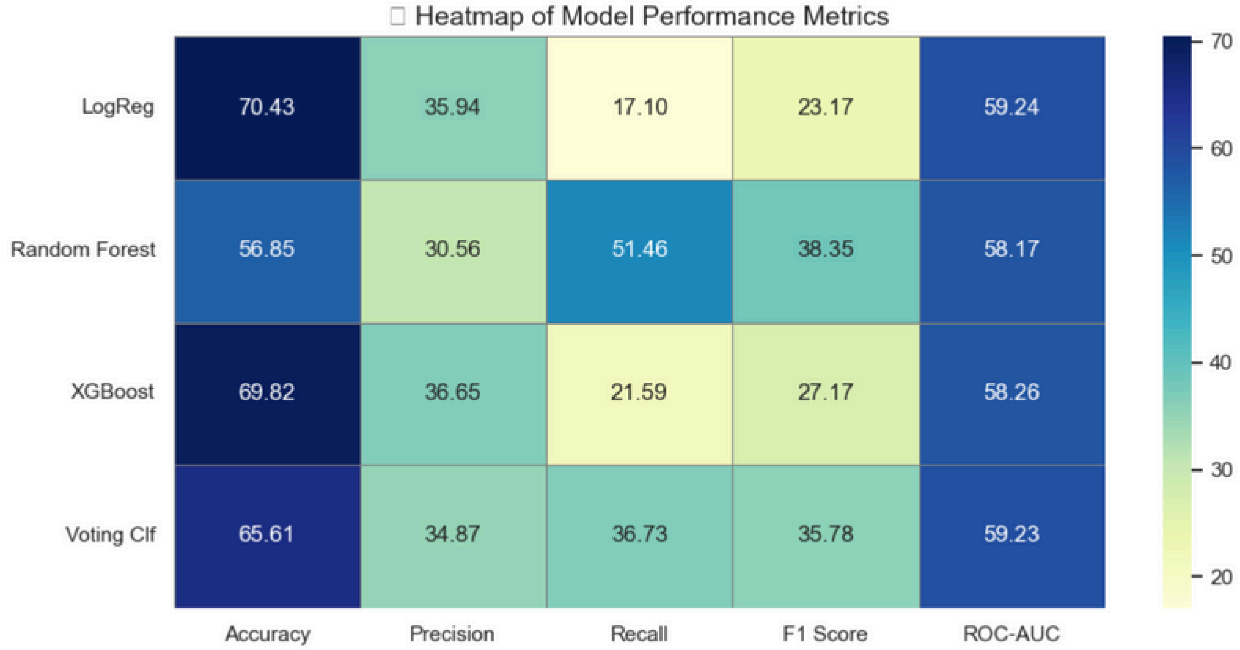
Significant gains in recall and F1 score were achieved from baseline logistic regression to ensemble model.

3. Metric-Wise Improvement Across Data Processing Stages

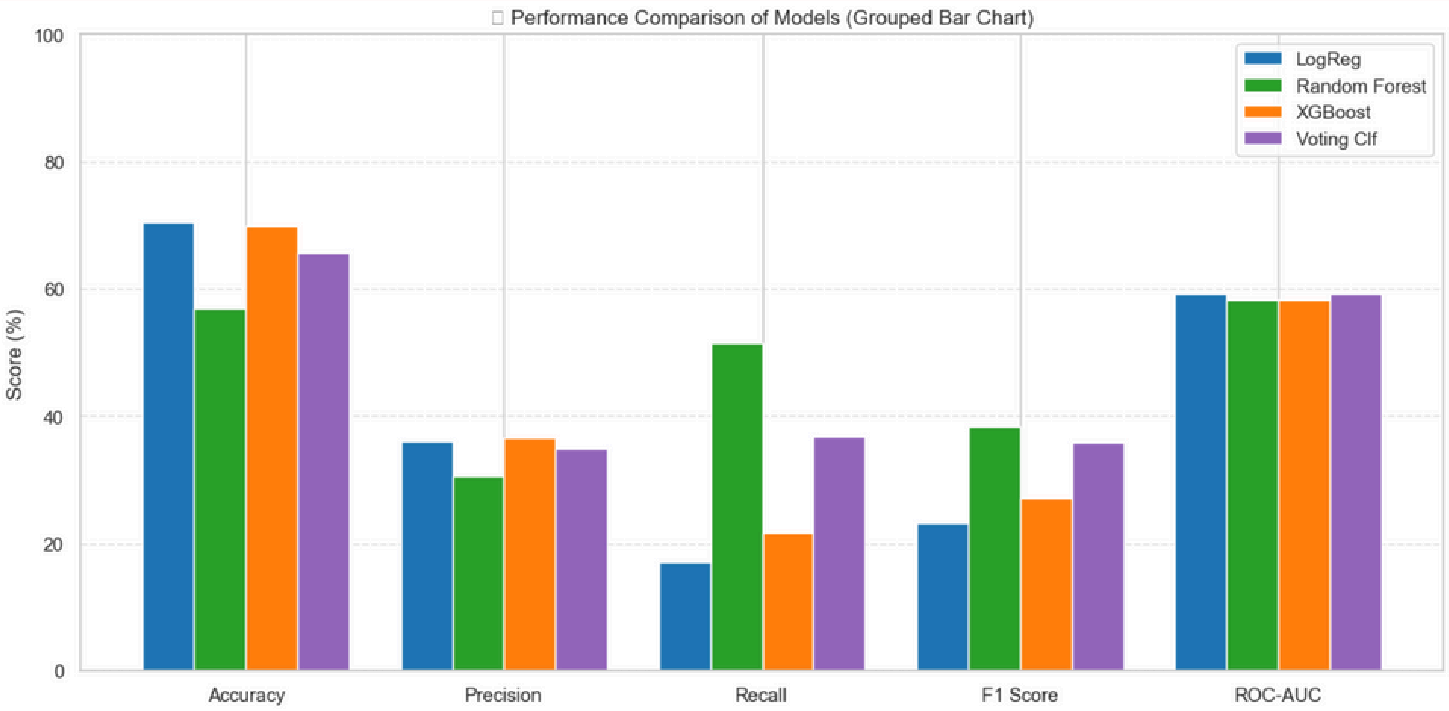


Metrics like recall and F1 score saw major improvements, with recall increasing by 40% and F1 score by over 35%, while only incurring a small 8% drop in accuracy and maintaining stable ROC-AUC performance.

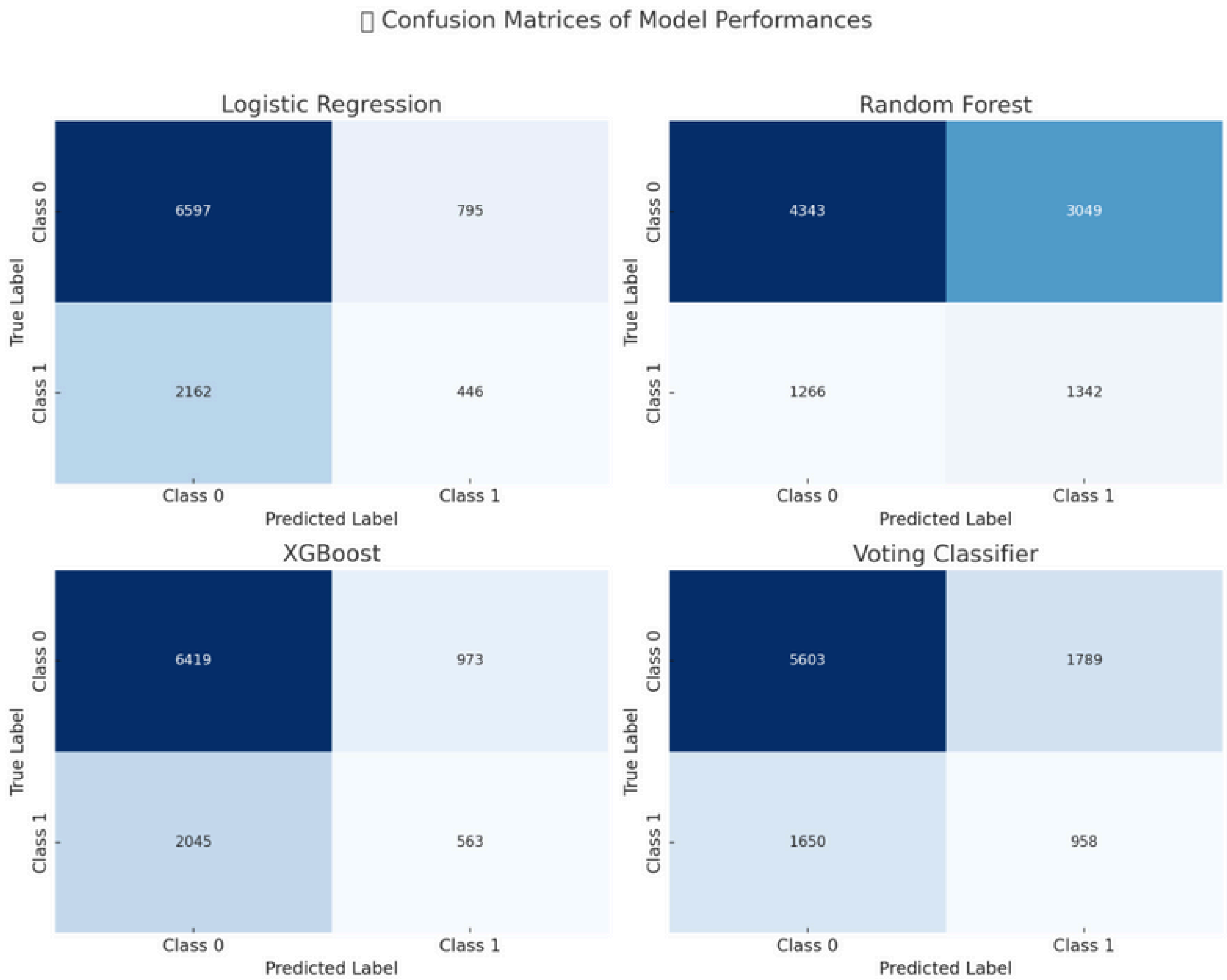
4. Model Comparison



The Voting Classifier (bottom row), which is an ensemble of Logistic Regression, Random Forest, and XGBoost, demonstrates the most balanced performance across all key metrics.

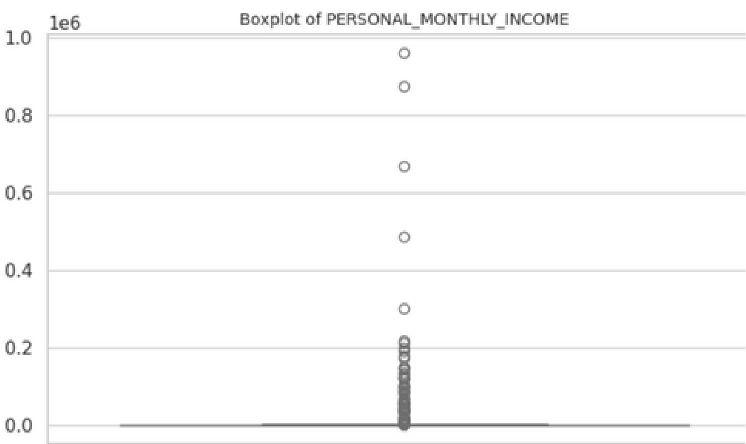


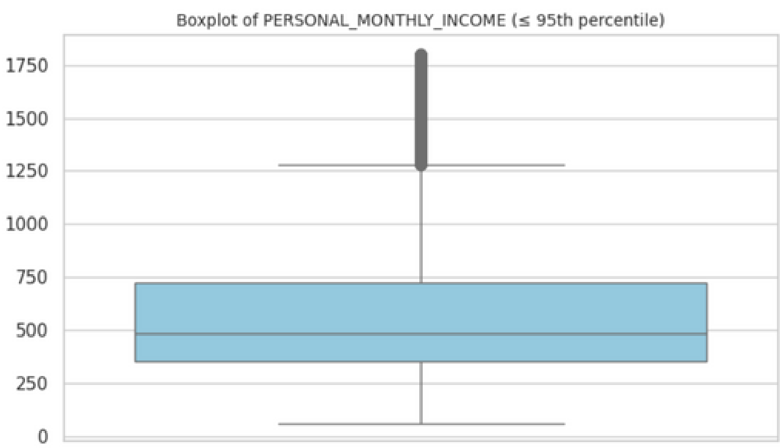
5. Confusion Matrices of All Models



Ensemble model (Voting Classifier) reduced false negatives significantly vs baseline model (Logistic Regression).

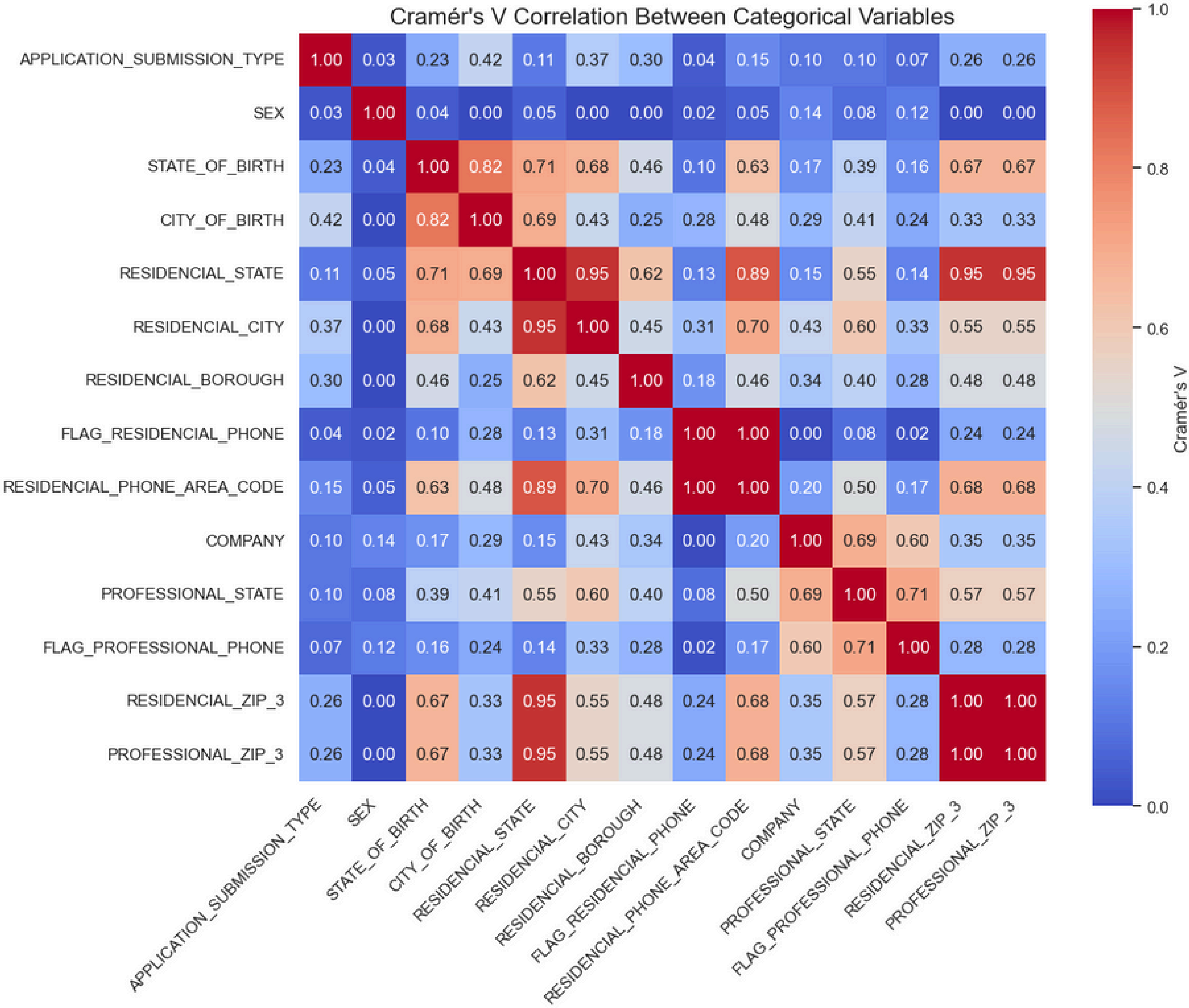
6. Boxplot Comparison: Raw vs Treated Outliers





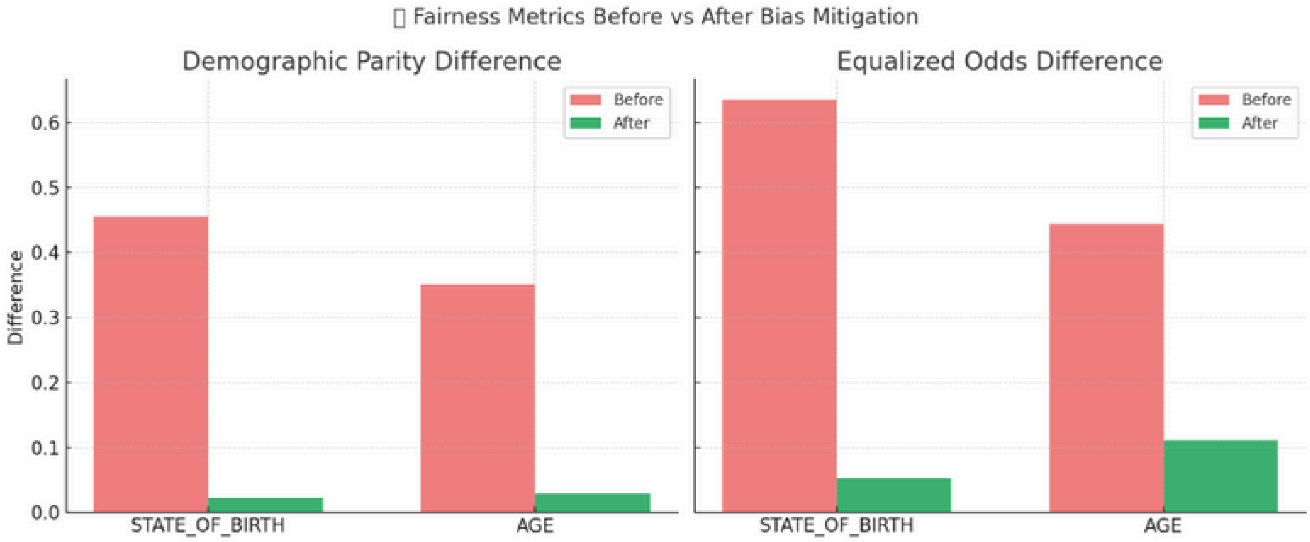
This helps improve model robustness and prevents extreme values from distorting learning.

7. Cramer’s V Heatmap for Visualizing Association between Categorical features



Cramér’s V heatmap helps visually quantify the association between categorical features, especially when traditional correlation metrics (like Pearson) are not applicable.
It reveals hidden dependencies and multicollinearity between categorical variables, guiding better feature selection and reducing redundancy in modeling

8. Fairness Metrics Before vs After Bias Mitigation



Bias against sensitive demographic groups dropped significantly after mitigation

---

CONCLUSION

This case study demonstrated the development of a robust, fair, and high-performing credit risk assessment pipeline through a comprehensive quality assurance process. By addressing data quality, handling class imbalance, and applying bias mitigation techniques, the model's predictive power and fairness improved significantly.

The final ensemble model achieved a strong balance across all evaluation metrics while ensuring ethical treatment of sensitive demographic groups—making it a reliable tool for informed and responsible decision-making in financial risk management.

Falcon Informatics

Transforming Businesses with ERP Systems, Data Science, and Intelligence

Delivering data-driven Insights

Contact :

www.falconinformatics.com