Credit Card Fraud Detection System

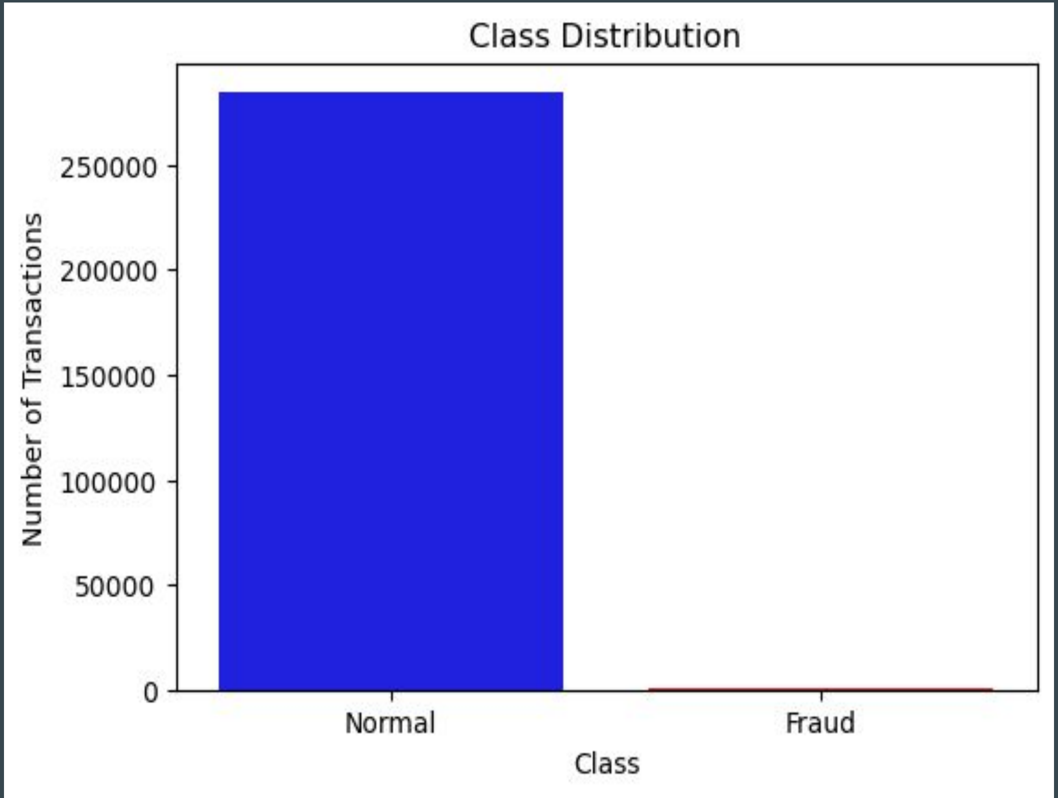
A Machine Learning–Based Predictive Approach

By
Alice Wangui Mathenge

Problem Overview

- Rapid growth in digital transactions
- Increasing fraud-related financial losses
- Manual fraud detection is inefficient and costly

Dataset Overview: Class Distribution



- Fraud is extremely rare compared to normal transactions, highlighting data imbalance.
- This imbalance is why standard accuracy isn't enough which resulted to the use of specialized approaches to detect fraud.

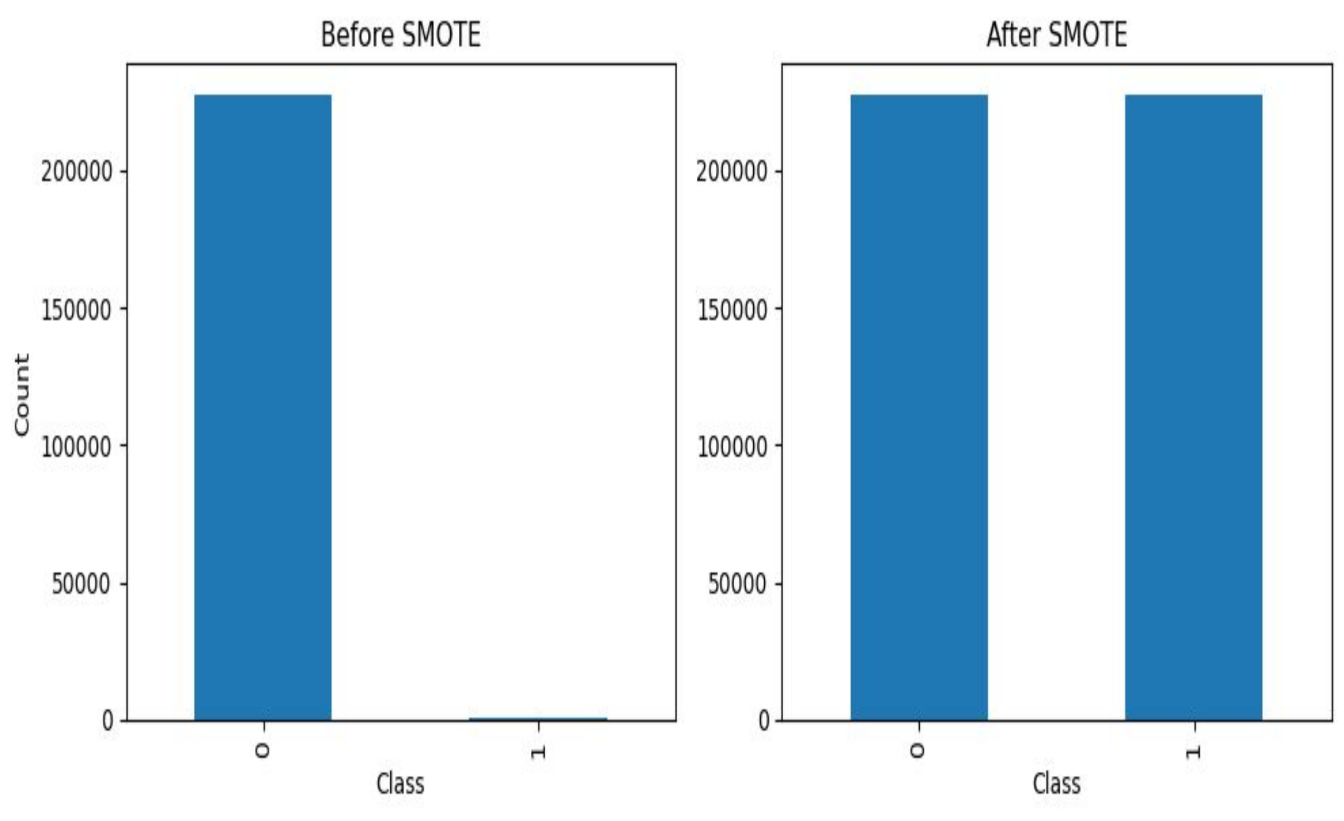
The Imbalance Problem

- Fraud cases are extremely rare
- Models may ignore the minority class
- Accuracy alone is misleading

Feature Preparation

- Time and Amount features were scaled
- PCA features (V1–V28) already standardized
- Scaling ensures fair model learning

Handling Class Imbalance with SMOTE

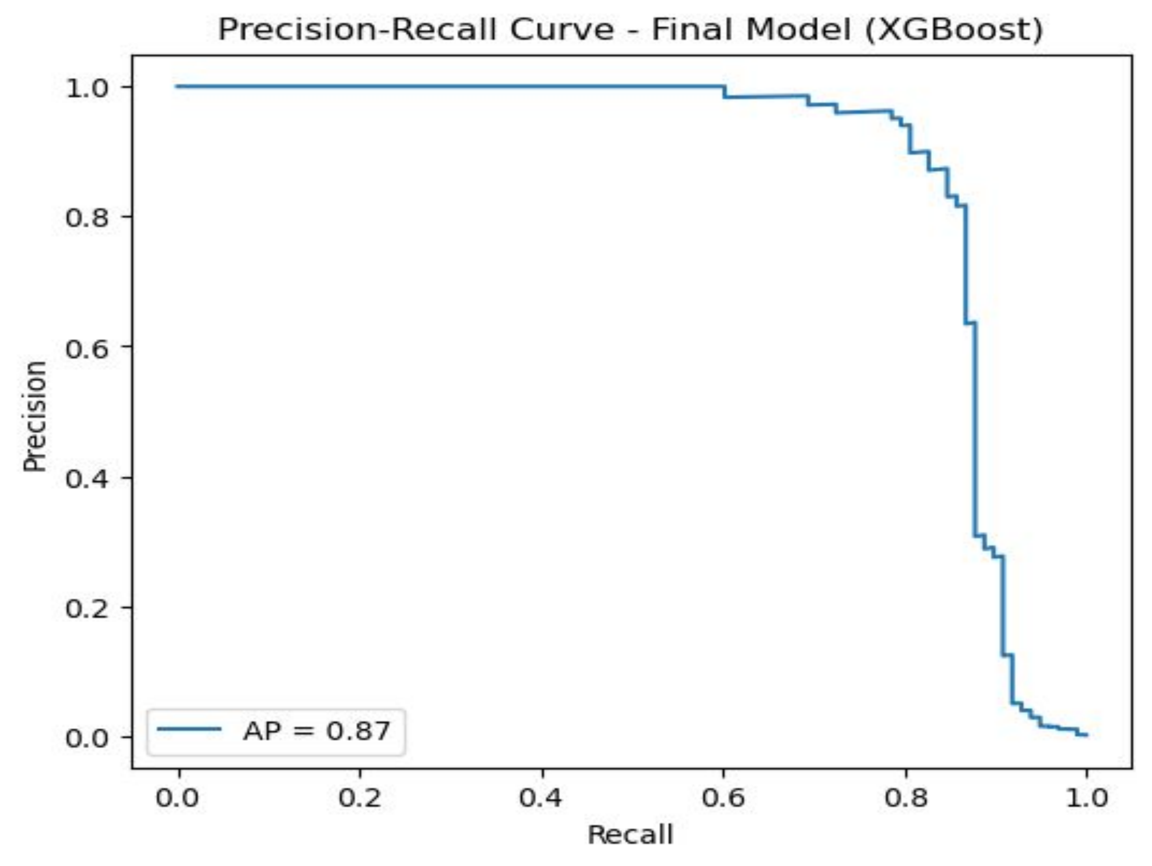


- Dataset is heavily imbalanced before SMOTE.
- SMOTE generates synthetic fraud examples to balance classes.
- Balancing helps the model learn fraud patterns more effectively.

Models Evaluated

- Logistic Regression (baseline)
- LightGBM
- XGBoost

Final Model Performance: XGBoost



The graph shows XGBoost performance:

- high precision and recall means effective fraud detection with few false alerts.
- The high area under this curve shows the model is effective at detecting fraud while minimizing false alarms.

Model Evaluation Focus

- Accuracy is not sufficient for imbalanced data
- Precision and Recall are prioritized
- Precision–Recall curve used for evaluation

Conclusion & Impact

- Machine learning improves fraud detection efficiency.
- XGBoost(Final Model) + SMOTE handles class imbalance effectively.
- Solution is scalable for real-world deployment.