CREDIT CARD FRAUD DETECTION SYSTEM FOR ZESTBANK

BY MERCY KIRAGU

TABLE OF CONTENTS

1. INTRODUCTION
2. PROJECT OVERVIEW
3. PROBLEM STATEMENT
4. OBJECTIVES
5. DATA SOURCES
6. DATA PREPROCESSING
7. DATA ANALYSIS
8. CLASS IMBALANCE
9. MODELLING
10. MODEL EVALUATION
11. FEATURE IMPORTANCE
12. MODEL DEPLOYMENT
13. FUTURE ENHANCEMENT
14. CONCLUSIONS
15. RECOMMENDATIONS

INTRODUCTION

Credit card fraud involves unauthorized or fraudulent transactions made using someone else's credit card information, posing significant threats to financial security.

I am thrilled to present the culmination of my efforts in developing a state-of-the-art Credit Card Fraud Detection System called SafeSwipe tailored specifically for Zest Bank.

PROJECT OVERVIEW

- Purpose: The purpose of this project is to develop a robust Credit Card Fraud Detection System tailored specifically for Zest Bank.

- Goals: My primary goal is to enhance Zest Bank's ability to detect and prevent fraudulent transactions, thereby safeguarding the financial security of its customers and preserving trust in its services.

- Approach: I adopted a data-driven approach, leveraging advanced machine learning algorithms and data analysis techniques to build predictive models capable of identifying fraudulent activities in real-time.

- Collaboration: Throughout the project, I collaborated closely with Zest Bank's team to ensure alignment with their requirements and objectives.

- Expected Outcomes: By the end of the project, I aim to deliver a scalable and effective fraud detection solution that integrates seamlessly into Zest Bank's existing infrastructure, empowering the bank to proactively combat credit card fraud and minimize its impact on customers and stakeholders.

PROBLEM STATEMENT

- Challenges in Fraud Detection: Credit card fraud is a pervasive problem that poses significant risks to financial institutions like Zest Bank. Failure to detect fraudulent activities can result in substantial financial losses, reputational damage, and erosion of customer trust.

- Importance of Detection: Detecting credit card fraud early is essential for mitigating these risks and safeguarding the financial security of Zest Bank's customers.

- Risks of Undetected Fraud: Undetected fraudulent activities can lead to increased financial liabilities, regulatory penalties, and legal consequences for Zest Bank. Failure to address fraud promptly can result in customer dissatisfaction, attrition, and tarnished brand reputation.

- Need for Effective Solutions: Given the evolving nature of fraud tactics and the increasing volume of digital transactions, Zest Bank requires robust and adaptive fraud detection solutions to stay ahead of fraudsters and protect its customers' interests effectively.

OBJECTIVES

-The Credit Card Fraud Detection System developed for Zest Bank aims to achieve the following objectives:

1. Enhance Fraud Detection Accuracy

2. Real-time Detection Capability

3. Scalability and Adaptability

4. Integration with Existing Systems

5. Compliance with Regulatory Standards

6. Cost-effectiveness

- By achieving these objectives, the Credit Card Fraud Detection System will empower Zest Bank to effectively combat credit card fraud, protect its customers' financial assets, and maintain trust and confidence in its services.

DATA SOURCES

- Transaction Data: The transaction data used for analysis was sourced from the Kaggle dataset titled "Credit Card Fraud Detection" provided by Machine Learning Group at ULB (Université Libre de Bruxelles).
- Dataset Description: The dataset contains transaction records made by credit cards in September 2013 by European cardholders. It consists of 284,807 transactions, of which 492 are fraudulent. The data includes features such as transaction amount, time, and anonymized features obtained through Principal Component Analysis (PCA) transformation for confidentiality reasons.

By leveraging this dataset, I was able to build and evaluate the Credit Card Fraud Detection System effectively, ensuring its accuracy and reliability in identifying fraudulent transactions.

DATA PREPROCESSING

The dataset was meticulously prepared, with no missing values present.

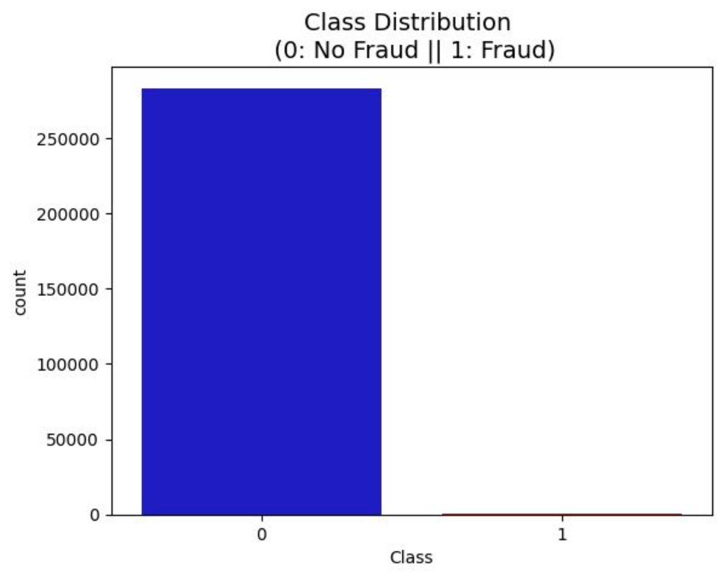
Additionally, I conducted preprocessing steps such as removing duplicate entries and scaling the transaction amount column to ensure uniformity and enhance the performance of our models.

I also dealt with skewness of the features,by addressing skewness, I aimed to ensure that our Credit Card Fraud Detection System is trained on data that follows a more Gaussian /normal
distribution, thereby enhancing the accuracy and reliability of our predictive models.

EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) involved summarizing data statistics, visualizing feature distributions, analyzing transaction amounts and times, examining correlations, identifying feature importance, , and assessing class imbalance. It provided insights crucial for subsequent model development and evaluation.

I observed a class imbalance between fraudulent and non-fraudulent transactions. This imbalance necessitated careful consideration and implementation of strategies to address it effectively during model training.



CLASS IMBALANCE

To address the class imbalance observed in the dataset, I employed various strategies during model training. These included:

1. **Resampling Techniques**: We utilized resampling techniques such as oversampling and undersampling to balance the class distribution. Oversampling involves increasing the number of minority class samples, while undersampling involves reducing the number of majority class samples.

By implementing these strategies, we aimed to mitigate the effects of class imbalance and improve the overall performance of our Credit Card Fraud Detection System.

MODELLING

For our Credit Card Fraud Detection System, we utilized a range of machine learning models tailored to handle the challenges posed by imbalanced data. Here's a summary:

Machine Learning Models:

1. Logistic Regression

2. Gaussian Naive Bayes

3. Decision Trees

4. XGBoost

5. AdaBoost

Model Training Process:

- Hyperparameter Tuning: Performed grid search for some of the models to find optimal hyperparameters for each model.

By leveraging these models and following a systematic training process, our goal was to develop an accurate and reliable Credit Card Fraud Detection System capable of effectively identifying fraudulent transactions while minimizing false positives and negatives.

MODEL EVALUATION

Model evaluation for the Credit Card Fraud Detection System involved assessing the performance of various machine learning models using key metrics such as accuracy, precision, recall, F1-score, and ROC AUC score. Evaluation techniques included analyzing confusion matrices, ROC curves and conducting k-fold cross-validation for ROC_AUC mean and standard deviation. The outcome of the evaluation process provided insights into the effectiveness of the models in accurately detecting fraudulent transactions and guided decisions regarding model deployment and refinement

I found the Xgboost model with SMOTE oversampled dataset to be the best performer across all metrics.

METRICS DEFINITIONS

1. **Accuracy:**

   - Accuracy tells us how often the model is correct. It's the ratio of correctly predicted instances to the total instances.

2. **ROC AUC Score (Receiver Operating Characteristic Area Under the Curve):**

   - ROC AUC Score shows how well the model can distinguish between the classes. Higher scores mean better separation.

3. **F1 Score:**

   - F1 Score is a balance between precision and recall. It's the harmonic mean of precision and recall, providing a single value that summarizes both.

4. **Precision:**

   - Precision is the proportion of true positive predictions among all positive predictions made by the model. It tells us how many of the predicted positive instances are actually positive.

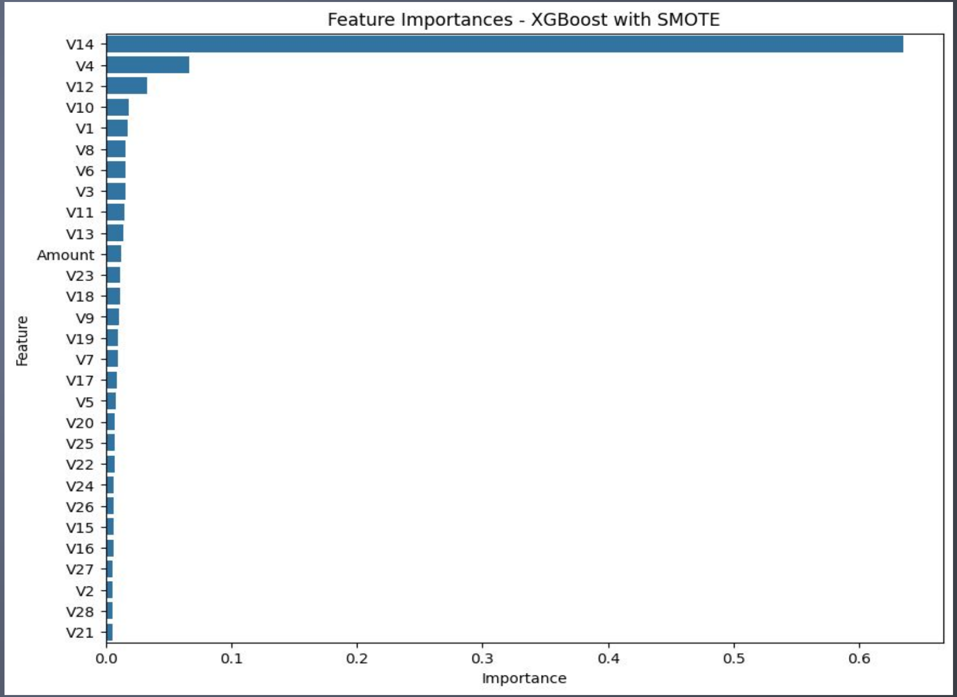
5. **Recall (Sensitivity or True Positive Rate):**

   - Recall is the proportion of true positive predictions among all actual positive instances. It tells us how many of the actual positive instances were captured by the model.

FEATURE IMPORTANCE

feature importance analysis is essential for understanding the factors influencing the prediction of fraudulent transactions in our Credit Card Fraud Detection System.

V14 is the most impotant feature for credit card fraud detection



MODEL DEPLOYMENT

I plan to deploy the XGBOOST model on the smote data as part of a real-time transaction monitoring system, where it will analyze incoming transactions and flag suspicious activities for further review.

CONCLUSION

Our Credit Card Fraud Detection System ensures Zest Bank's security through innovative machine learning models and robust deployment strategies. Continuous improvement and future enhancements promise ongoing protection against evolving fraud tactics, reinforcing our commitment to financial security and innovation.

RECOMMENDATIONS

1. Regular Evaluation:

2. Collaboration: Foster collaboration between data scientists, cybersecurity experts, and banking professionals to leverage diverse expertise and enhance fraud detection capabilities.

3. Customer Awareness: Educate Zest Bank's customers about common fraud schemes and best practices for protecting their accounts, reducing the risk of falling victim to fraud.

4. Technology Adoption: Stay informed about advancements in fraud detection technology and consider adopting new tools and techniques to strengthen the security of Zest Bank's systems.

5. Regulatory Compliance: Ensure adherence to regulatory requirements governing fraud detection and prevention in the banking industry to mitigate legal and financial risks associated with non-compliance.

NEXT STEPS

- Continuous Improvement:

  - Regular monitoring and evaluation of the fraud detection system.

  - Updates to the xgboost model with new data and insights to maintain effectiveness against evolving fraud tactics.

- Areas for Research:

  - Integration of advanced machine learning techniques like deep learning and reinforcement learning.

  - Exploration of anomaly detection algorithms, behavioral biometrics, and blockchain technology for enhanced fraud detection capabilities and system security.

THANK YOU!

You're welcome! If you have any further questions or need assistance, feel free to ask. Additionally, if you'd like to contact me for further discussions or inquiries, you can reach out to me via email at [mercykiragu75@gmail.com] or I'm here to help!