CREDIT CARD FRAUD DETECTION PROJECT

By – Kumar Kanishka

Introduction

The Credit Card Fraud Detection project is designed to enhance the financial ecosystem's security by applying advanced machine learning methodologies and data analysis techniques. This initiative focuses on building a predictive model capable of discerning fraudulent credit card transactions from legitimate ones with high accuracy and efficiency.

Fraudulent activities are increasingly sophisticated, often outpacing traditional rule-based detection systems. To address this, the project leverages both supervised learning and anomaly detection techniques to capture the dynamic and evolving nature of fraudulent patterns. These approaches enable the identification of fraudulent behavior that may not align with pre-defined rules or historical patterns.

Project Objectives

The core objective of this project is to leverage machine learning techniques to predict fraudulent credit card transactions, enabling financial institutions to take proactive measures against fraud. Key steps include:

- **Data Preprocessing:** Cleaning and structuring the dataset, handling missing values, imbalanced classes, and outliers.
- **Dataset Splitting:** Ensuring balanced class distribution using techniques like Stratified Sampling.
- **Model Training:** Training multiple algorithms including logistic regression, decision trees, random forests, and SVMs.
- **Performance Evaluation:** Using metrics such as Precision, Recall, F1 Score, and ROC-AUC to assess model performance.
- **Hyperparameter Tuning:** Optimizing models using Grid Search, Random Search, or advanced techniques like Bayesian Optimization.
- **Advanced Techniques:** Employing ensemble methods like XGBoost and exploring deep learning architectures.

Problem Understanding

Digital transactions in India have experienced rapid growth, with a 151% increase during 2020-23. However, this has also led to a rise in fraudulent activities, with approximately 93,544 cases of credit and debit card fraud reported in FY 2023. Detecting fraudulent transactions promptly is crucial to protect consumers and financial institutions from substantial losses.

Machine learning provides a powerful solution by analyzing patterns and detecting fraudulent activities more effectively than traditional methods. This project explores model selection, class imbalance handling, and advanced fraud detection techniques to deliver a robust solution.

Summary Results

The project strategically selected the XGBoost model as the optimal solution for detecting fraudulent credit card transactions. Key outcomes include:

- **Accuracy:** 99.94%, reflecting near-perfect classification of transactions.
- **ROC-AUC Score:** 0.982, indicating an excellent balance between sensitivity and specificity.
- **Optimal Decision Threshold:** 0.0044, maximizing the trade-off between precision and recall.

Additionally, Logistic Regression was explored as a competitive alternative for scenarios requiring interpretability and computational simplicity.

Conclusion

The XGBoost model, combined with Random Oversampling and StratifiedKFold Cross-Validation, emerged as the most effective solution for credit card fraud detection. Its superior accuracy and robust ROC-AUC score make it an ideal choice for addressing evolving fraud challenges.

Hyperparameter tuning for XGBoost is recommended to further refine its capabilities. Logistic Regression also demonstrated strong performance, highlighting its potential as a simpler, interpretable alternative for specific use cases.

This project positions financial institutions to enhance fraud detection measures, safeguarding user trust and mitigating risks in the financial ecosystem.
