**Enhancing Credit Card Fraud Detection Using Machine Learning**

**By**

**Sai Tulasi Kolapudi  
Meghana Patibandla  
Rakesh Sarma Karra**

**GROUP 10 - FINANCE**

**Course: ADTA 5340 Section 002 - Discovery and Learning with Big Data (Fall 2023 1)**

**Main Problem :** The primary challenge tackled in this project is the detection of fraudulent transactions in credit card data. Credit card fraud represents a significant and growing concern in the financial sector, impacting both consumers and financial institutions. Traditional fraud detection systems are often plagued by high rates of false positives and an inability to adapt to the continuously evolving patterns of fraud. The core problem this project aims to address is enhancing the accuracy and efficiency of fraud detection mechanisms. By reducing false positives and negatives, the project seeks to establish a secure and reliable financial environment for users, thus bolstering consumer trust and financial security.

**Approach :** The project's approach is comprehensive and multi-dimensional, encompassing several critical steps:

1. **In-Depth Data Analysis and Preprocessing:** The project began with a thorough analysis of the credit card dataset. This step was followed by essential data preprocessing techniques, including the standardization of key features and the meticulous handling of any missing values.
2. **Critical Outlier Analysis:** A pivotal part of the approach was the focused analysis and inclusion of outliers within the dataset. Given that fraudulent transactions frequently manifest as outliers, this strategy was instrumental in bringing to light fraudulent patterns and behaviors.
3. **Innovative Data Augmentation with GANs:** To combat the challenge of imbalanced data, the project employed Generative Adversarial Networks (GANs) to generate synthetic data samples. This augmentation was key in creating a balanced dataset, which is crucial for effective model training.
4. **Implementation and Evaluation of Diverse Models:** The project saw the implementation of a variety of machine learning models, including but not limited to Logistic Regression, Random Forest, XGBoost, KNN, Naive Bayes, SVM, and Decision Trees. Each model underwent rigorous evaluation based on various metrics such as accuracy, precision, recall, and AUPRC.
5. **Focus on Feature Importance and Model Explainability:** An integral part of the project was dedicated to understanding which features were most significant in detecting fraud and enhancing the transparency and explainability of the predictive models.

**Conclusions :** The project concluded with several key insights:

1. **Remarkable Efficacy of Models:** The majority of the models, particularly Random Forest, XGBoost, KNN, and Decision Trees, demonstrated exceptionally high efficacy, with some achieving 100% accuracy. Logistic Regression, meanwhile, offered a more balanced and realistic perspective on fraud detection.
2. **The Crucial Role of Outliers:** The analysis and inclusion of outliers were found to be significantly influential in detecting fraudulent transactions, underscoring their crucial role in predictive modeling.
3. **Effective Data Augmentation:** The use of GANs for data augmentation successfully balanced the dataset, leading to enhanced performance of the models.
4. **Future Directions:** Moving forward, the project will focus on further validation of the models using different datasets and cross-validation techniques. Continuous updating with new data, exploration of feature importance, model explainability, and real-time integration into fraud detection systems are also on the horizon. A steadfast commitment to ethical considerations and data privacy will remain a cornerstone of future developments.

**Contributions:**

**Meghana Patibandla**

* Managed data analysis, preprocessing, and standardization.
* Implemented and optimized Logistic Regression and SVM models.
* Wrote significant portions of the project report, focusing on model results and data insights.

**Sai Tulasi Kolapudi**

* Led the outlier analysis and data augmentation using GANs.
* Oversaw the development and tuning of Random Forest and Naive Bayes models.
* Created the PowerPoint presentation, focusing on methodology and data augmentation.

**3. Rakesh Sarma Karra**

* Focused on model evaluation for XGBoost, KNN, and Decision Trees.
* Drafted the conclusions, future work, and ethical considerations sections.
* Managed project documentation and ensured compliance with data privacy standards.

**Collaboration:**

* Regular team meetings for progress updates and decision-making.
* Peer review sessions for quality assurance and cohesive output.