PACE UNIVERSITY

FALL 2023

Analytics Capstone Project (By Prof. Krystyn Gutu)

"DeepGuard": High-Accuracy Credit Card Fraud Detection

Literature Review

Submitted by :- Piyush Gupta

CRN: - 74233; CS 668 FALL 2023

Introduction

With the relentless evolution of the digital transaction landscape, the urgency for effective credit card fraud detection has substantially amplified due to a spike in online transactions and the sophistication of fraudulent tactics. The current literature delineates various methodologies, challenges, and significant technological advancements, providing a foundational understanding for navigating ongoing discussions and projecting the future trajectory of fraud detection mechanisms.

Technological Advancements in Fraud Detection

The past decade has witnessed pivotal technological advancements with Machine Learning (ML) algorithms emerging at the forefront of fraud detection. Algorithms such as Logistic Regression, Naïve Bayes, Random Forest, K-Neighbor, and Support Vector Machines (SVM) are extensively utilized due to their proficiency in discerning transaction patterns, effectively differentiating between legitimate and fraudulent activities. However, these traditional ML algorithms exhibit limitations, mainly when interpreting complex fraud patterns. Given the escalating sophistication of fraudulent activities, there is a pressing need to explore and integrate more advanced algorithms, including those involving fuzzy logic and neural networks, promising to enhance the accuracy and efficiency of fraud detection systems significantly.

Data Imbalance in Fraud Detection

Data imbalance poses a significant challenge due to the disproportionate representation of legitimate and fraudulent transactions within datasets. Techniques such as under-sampling and over-sampling have emerged as practical solutions. Under-sampling techniques aim to reduce the number of legitimate transactions in datasets, balancing it against the number of fraudulent transactions. In contrast, over-sampling techniques may involve synthesizing fraudulent transaction data points to achieve dataset equilibrium, facilitating practical model training and performance enhancement.

Deep Learning Applications

The application of deep learning techniques for credit card fraud detection, with Convolutional Neural Networks (CNN) receiving particular attention for their efficacy. CNNs are lauded for their superior pattern recognition and classification capabilities, autonomously learning hierarchical features. When trained with extensive datasets, CNNs significantly enhance detection accuracy and effectively minimize false positives. However, further research and exploration are requisite to optimize their performance fully.

Deployment Challenges and Strategies

Deploying fraud detection models presents a unique set of challenges, concerning the prerequisites of real-time detection. The development of efficient and accurate models is paramount. In response to this need, solutions such as cloud platforms AWS EC2 and automated deployment workflows through GitHub have been identified as promising strategies to bolster deployed models' speed, performance, and reliability.

Methodology

1. **Data Collection and Preprocessing:** "DeepGuard" initiates the collection and preprocessing of data, targeting a dataset comprising at least 200,000 credit card transactions.

    1.1 Data Cleaning and Transformation: The raw data will undergo a rigorous cleaning and transformation process through PCA, ensuring confidentiality while preserving essential transaction features.

    1.2 Under-sampling and Over-sampling Techniques Assessment: This stage will rigorously assess the efficacy of under-sampling and over-sampling techniques, addressing data imbalance issues.

2. **Implementation and Evaluation of ML Algorithms:** This phase entails implementing and meticulously evaluating various ML algorithms, using a suite of performance metrics to gauge their effectiveness in fraud detection.
3. **CNN Model Development:** "DeepGuard" will seamlessly integrate a TensorFlow-developed CNN to enhance the project's fraud detection capabilities substantially.
4. **Model Deployment:** The project foresees deployment through GitHub Actions and AWS EC2 instances, ensuring a smooth transition to practical use.

Conclusion

As "DeepGuard" approaches its final stages, the model's performance will be meticulously evaluated using key metrics such as Accuracy, AUC Score, Precision, Recall, F1 Score, and the Confusion Matrix. The AUC Score is remarkably esteemed due to its widespread utilization across various performance metrics, serving as a concise graphical representation delineating each model's capabilities. Furthermore, integrating a TensorFlow-developed CNN within the "DeepGuard" framework is set to significantly enhance the system's fraud detection prowess, marking a substantial advancement in the reliability and accuracy of credit card fraud detection initiatives.

References

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.
- Jagtap, A., & Shirbahadurkar, S. (2020). Detection of Credit Card Fraud Using Machine Learning Algorithms: A Survey. Procedia Computer Science, 167, 191–201.
- Chawla, N.V., Bowyer, K.W., Hall, L.O., & Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321-357.
- Hassan, M.M., Huda, S., Uddin, M.Z., Almogren, A., & Alamri, A. (2019). Secure and Robust Cloud-Based Large-Scale IoT Infrastructure for Smart Cities. IEEE Access, 7, 50724-50734.