International Journal of Scientific Research & Engineering Trends
Volume 9, Issue 5, Sept-Oct-2023, ISSN (Online): 2395-566X

A Review of Anomaly Detection in Credit Card Fraud Using Machine Learning Techniques

M. Tech. Scholar,**Palak Shukla, HOD Vijay Shankar Mishra**
Department of Computer Science
Malwa Institute of Science and Technology,
Rajiv Gandhi Proudyogiki Vishwavidyalaya,Bhopal, India.
shuklaapalak@gmail.com

Abstract- Credit card fraud, an escalating challenge in the digital era, poses significant financial implications for businesses and compromises the security of consumers. Traditional rule-based systems, although effective to a degree, often fall short in detecting sophisticated fraud schemes. This review delves into the applicability and advantages of employing machine learning (ML) techniques, specifically anomaly detection, to mitigate the threat of credit card fraud. Anomaly detection, with its ability to identify unusual patterns in large datasets, offers a more proactive and adaptive approach to fraud prevention. As we navigate through the vast array of literature, it becomes evident that innovations in this domain are burgeoning, from the use of autoencoders, hybrid models, to cutting-edge feature selection methodologies. However, challenges persist, especially in addressing the imbalanced nature of fraud data and the dire need for real-time detection mechanisms. This review culminates in emphasizing the transformative potential of ML-driven anomaly detection, suggesting that its continuous evolution could pave the way for a more secure financial transaction environment in the imminent future.

Keywords- Machine Learning,Credit Card Fraud,Financial Implications,XGBoost, Anomaly Detection.

I. INTRODUCTION

Credit card fraud is an ever-evolving menace that poses severe threats to the financial industry and its customers. As digital transactions continue to gain traction, the ingenuity of fraudsters keeps pace, requiring innovative countermeasures. Anomaly detection, often referred to as outlier detection, is one of the advanced methods being utilized to identify patterns in data that do not conform to expected behavior. Within the ambit of credit card fraud prevention, these anomalies often signify suspicious activities.

The rapid emergence and adoption of machine learning (ML) techniques provide promising avenues in bolstering the effectiveness of fraud detection systems. Traditional rule-based systems, while valuable, often fail to adapt to the dynamic nature of fraud. In contrast, ML techniques have demonstrated their capacity to learn and predict new and unforeseen fraudulent tactics from data, ushering in a paradigm shift in how fraud detection is approached.

This review delves into the advancements in anomaly detection for credit card fraud using machine learning. From foundational techniques to state-of-the-art methodologies, we aim to provide a comprehensive overview of the current landscape, emphasizing the strengths, weaknesses, and areas of opportunity for future research.

By understanding the nuances and potential of ML-based anomaly detection, stakeholders in the financial sector can equip themselves better to tackle the multi-faceted challenges posed by credit card fraud. Furthermore, this review seeks to bridge the gap between the theoretical underpinnings of machine learning and its practical applications in the realm of fraud detection, offering insights that can guide both researchers and practitioners alike.

II. LITERATURE REVIEW

Jiang et al. (2023) introduced the **Unsupervised Attentional Anomaly Detection Network** for fraud detection. This groundbreaking work utilizes attention mechanisms within unsupervised learning frameworks to highlight salient features in the data, which assists in detecting fraudulent patterns. The model's strength lies in its ability to evolve with changing fraudulent tactics and to identify them even in the absence of labeled fraudulent data.

Leevy et al. (2023) presented a **Comparative Analysis of Binary and One-Class Classification Techniques** for credit card fraud data. Their study underscored the potential of one-class classifiers in situations where fraudulent transactions are sparse compared to legitimate ones. They opined that while binary classification remains a mainstream approach, one-class techniques offer unique advantages in specific scenarios.

© 2023 IJSRET

Aggarwal (2023) ventured into time-series anomaly detection by introducing an **LSTM-based Anomaly Detection** methodology. While this study focused on US exports and imports, the time-series nature of the data is reminiscent of credit card transactions. LSTM's ability to capture long-term dependencies can be instrumental in identifying inconsistencies over time.

Van Belle et al. (2023) unveiled **CATCHM**, a network-based fraud detection system that leverages node representation learning. The novelty of this approach lies in its utilization of graph theory, converting transaction data into nodes and edges. By understanding the relationships and patterns within this network, the system can pinpoint anomalies with high precision.

Zhu et al. (2023) tackled the perennial problem of class imbalance in fraud detection with the **NUS: Noisy-Sample-Removed Undersampling Scheme**. Their method focuses on not just undersampling the majority class but ensuring that noisy or misleading data points are removed, leading to more robust classifiers.

Yang et al. (2023) proposed a privacy-focused solution with their **Federated XGBoost** for anomaly detection. In a world increasingly concerned with data privacy, their approach ensures that data remains in its local silo while still benefiting from global model updates, striking a balance between utility and privacy.

Jayasingh& Sri (2023) introduced an **Online Transaction Anomaly Detection Model** at the ESCI conference. Their system harnesses a range of machine learning classifiers and demonstrates the importance of real-time detection in the rapidly evolving landscape of online transactions.

Strelcenia&Prakoonwit (2023) offered a comprehensive survey on **GAN Techniques for Data Augmentation**. Addressing the imbalanced data issue, their exploration into Generative Adversarial Networks reveals how synthetic data can be leveraged to enhance the robustness of fraud detection systems.

Prabhakaran &Nedunchelian (2023) ventured into optimization techniques, particularly the **Oppositional Cat Swarm Optimization**, to enhance feature selection in fraud detection. By pinpointing the most informative features, they optimize classifier performance and potentially reduce computational overhead.

Ni et al. (2023) presented a two-pronged approach, focusing on a **Fraud Feature Boosting Mechanism** coupled with a **Spiral Oversampling Balancing Technique**. Their comprehensive method tackles both feature importance and class imbalance, aiming for an optimized, well-rounded fraud detection system.

Strelcenia&Prakoonwit (2023) delved into **Data Augmentation** to enhance the classification performance in credit card fraud detection. Recognizing the challenges of imbalanced datasets, they proposed innovative augmentation strategies to bolster the minority class, improving overall model performance.

Fanai&Abbasimehr (2023) combined **Deep Autoencoder and Deep Classifiers** in a novel approach. Their framework first leverages autoencoders for dimensionality reduction and feature learning, followed by deep classifiers to identify fraudulent patterns, exhibiting promising results in their evaluations.

Pang et al. (2023) contributed to the field with **Deep Weakly-Supervised Anomaly Detection**. They posited that in many real-world scenarios, acquiring labeled data is costly; hence a weakly-supervised approach that leverages both labeled and unlabeled data can be particularly effective.

Krishna et al. (2023) provided insights on a plethora of **Anomaly Detection Techniques**. Their work serves as a foundational guide, evaluating the pros and cons of various methodologies and guiding practitioners in selecting appropriate techniques for specific scenarios
.

Eren et al. (2023) explored **Unsupervised Cyber Anomaly Detection** using non-negative tensor factorization. While this technique was generalized for cyber anomalies, its principles could be adapted for credit card fraud detection, considering both involve identifying unusual patterns in vast data streams.

Dorigo et al. (2023) introduced **RanBox**, an anomaly detection technique in the copula space. This approach, grounded in statistical theory, offers a fresh perspective, emphasizing the joint distribution of multiple variables rather than their individual distributions.

Fakiha (2023) leveraged **Deep Neural Networks** for Forensic Credit Card Fraud Detection. Recognizing the complex patterns underlying fraud, deep networks, with their hierarchical feature learning capabilities, offer a promising avenue for accurate detection.

López et al. (2023) proposed a fusion of **Anomaly Detection and False Positive Mitigation** for predictive maintenance in multivariate time series. Their methodology underscores the importance of not just detecting anomalies but also ensuring that the number of false alarms is minimized.

Du et al. (2023) combined **AutoEncoder and LightGBM** techniques. Their framework represents a blend of deep learning for feature learning and gradient boosting for classification, offering a robust solution to the fraud detection problem.

© 2023 IJSRET

Copiaco et al. (2023) ventured into deep anomaly detection of building energy consumption using **Energy Time-Series Images**. Their innovative approach transforms time-series data into image format, leveraging convolutional networks for anomaly detection.

Koko et al. (2023) explored **Dynamic Construction of Outlier Detector Ensembles** with bisecting k-means clustering. This technique emphasizes ensemble learning's power, ensuring that the strengths of individual detectors are combined for superior performance.

Jiang et al. (2023) provided a comprehensive survey on **Weakly Supervised Anomaly Detection**. Their study accentuates the practical challenges of limited labeled data and discusses strategies to harness both labeled and unlabeled data effectively.

Habibpour et al. (2023) introduced **Uncertainty-Aware Credit Card Fraud Detection** using deep learning. By acknowledging and incorporating model uncertainty, they aim to provide more reliable and interpretable predictions.

Zhu et al. (2023) discussed **Sequential Adversarial Anomaly Detection** for one-class event data. This technique leverages adversarial training, a concept borrowed from GANs, to enhance the robustness of anomaly detectors.

Mienye& Sun (2023) proposed a **Machine Learning Method with Hybrid Feature Selection**. Recognizing the high dimensionality of fraud datasets, they emphasized the importance of effective feature selection to enhance both model performance and interpretability.

Vivek et al. (2023): This work emphasizes the utility of **Machine Learning in Credit Card Fraud Detection**. The authors' research focuses on the application of various machine learning models to detect anomalous transactions. The study provides a comprehensive evaluation of the effectiveness, accuracy, and efficiency of different algorithms in tackling the ever-evolving landscape of credit card fraud.

Alabrah (2023) :Alabrah introduces an **Improved CCF (Credit Card Fraud) Detector** specifically tailored to handle class imbalance, a prevalent issue in fraud detection datasets where fraudulent transactions are significantly outnumbered by legitimate ones. The research employs the IQR (Interquartile Range) method for outlier normalization, ensuring a robust model that is less sensitive to extreme data points, subsequently improving its generalizability and performance.

Jayanthi et al. (2023) :In a niche study, the authors address the **Detection of Credit Card Frauds in Healthcare**. Recognizing the unique challenges posed by the healthcare sector, they introduce novel machine learning strategies tailored to this context. This research stands out by merging the domains of cybersecurity, healthcare, and financial transactions, underlining the multifaceted nature of fraud detection.

Kennedy et al. (2023): Kennedy and his team delve into one of the primary challenges in fraud detection: the **Handling of Highly-Imbalanced Big Data**. Their approach, which employs unsupervised learning, emphasizes iterative cleaning and learning. By doing so, the researchers aim to enhance the quality of the data, subsequently boosting the accuracy and reliability of the fraud detection models.

Bustos-Brinez et al. (2023): In a more technical and innovative study, Bustos-Brinez and colleagues present **AD-DMKDE**, an anomaly detection mechanism that leans on Density Matrices and Fourier Features. This research taps into advanced mathematical techniques, intertwining quantum mechanics (via density matrices) and signal processing (via Fourier features) to craft a cutting-edge solution for fraud detection.

Abhaya & Patra (2023 : The study by Abhaya and Patra zooms in on **Autoencoders**, a type of neural network, for outlier detection. The authors argue for the efficiency and effectiveness of autoencoders in capturing the intrinsic data structure and subsequently identifying anomalies. Their research offers valuable insights into the application of deep learning techniques in the realm of fraud detection.

III. RESEARCH GAP

1. **Transfer Learning and Domain Adaptation**: Most of the research tends to focus on models trained and tested on the same datasets or similar datasets. However, in the real world, fraud detection mechanisms might benefit from models that can adapt to new, previously unseen data. There's a gap in research on how models trained on one dataset (or domain) can be adapted to perform well on a different yet related dataset.
2. **Real-time Anomaly Detection**: While many machine learning models boast high accuracy, there's limited research on their real-time application. Detecting fraudulent transactions in real-time is crucial for preventing potential financial losses.
3. **Evolving Frauds and Adversarial Machine Learning**: Fraudsters constantly evolve their strategies. Adversarial machine learning, where models are trained to anticipate and react to new and evolving threats, is a growing field but is yet to be deeply explored in the context of credit card fraud detection.
4. **Feature Engineering and Automatic Feature Extraction**: While deep learning models like autoencoders can automatically extract features, there's a gap in comprehensive studies on the most influential features across various datasets and techniques. There's potential in hybrid models that combine manual and automatic feature extraction.

© 2023 IJSRET

5. **Explainability and Interpretability**: With the increased complexity of models, there's a dire need for them to be explainable. While a model might offer high accuracy, it's crucial for financial institutions to understand why a particular transaction was flagged.
6. **Handling Imbalanced Data**: While techniques like oversampling, undersampling, and synthetic data generation have been explored, there's a gap in newer techniques or hybrid techniques that might offer better performance, especially when considering very large-scale datasets.
7. **Integration with Other Data Sources**: Most research focuses on the transaction data alone. However, integrating other sources of data (like user behavior analytics, device fingerprinting, etc.) might offer a richer context and better detection capabilities.
8. **Ethical and Privacy Concerns**: With the use of more data and more complex models, there's a potential risk to user privacy. Research into methods that ensure user privacy (like differential privacy) in the context of fraud detection is an emerging area.
9. **Scalability and Deployment**: While many models might work well in a controlled, experimental setup, there's a gap in research focused on the scalability of these models in real-world scenarios, handling millions of transactions daily.
10. **Benchmarking and Evaluation Metrics**: There's a need for a standardized benchmarking dataset and evaluation metrics to truly compare the performance of various techniques. While accuracy, precision, recall, and F1-score are commonly used, considering the business impact and cost implications might offer a more holistic evaluation.

IV. ADVANTAGE OF ANOMALY DETECTION IN CREDIT CARD FRAUD

1. **Early Detection**: Anomaly detection algorithms can identify suspicious patterns in real-time or near-real-time. This quick identification can lead to faster reactions, potentially stopping a fraudulent transaction before it's completed.
2. **High Accuracy**: With the right training data and fine-tuning, anomaly detection models can achieve high levels of accuracy, leading to reduced false positives and false negatives.
3. **Adaptable to New Threats**: Unlike traditional rule-based systems that rely on predefined conditions to detect fraud, anomaly detection systems can identify previously unseen types of fraud. This is particularly useful as fraudsters continually evolve their tactics.
4. **Efficiency**: Automating the fraud detection process with anomaly detection techniques can considerably reduce the workload on human analysts, allowing them to focus on more complex investigation tasks.
5. **Scalability**: Anomaly detection systems can process vast amounts of transaction data in relatively short periods, making them suitable for large-scale operations of major financial institutions.
6. **Reduced Financial Loss**: By identifying and stopping fraudulent activities earlier, financial institutions can significantly reduce the associated monetary losses.
7. **Enhanced Customer Trust**: A robust fraud detection system can enhance customer trust and confidence. Customers appreciate knowing their financial institutions are taking proactive measures to safeguard their funds.
8. **Holistic View of Transactions**: Anomaly detection doesn't just consider a single transaction in isolation; it looks at patterns over time, providing a more holistic view of a user's behavior. This can be crucial in identifying subtle, sophisticated fraud schemes.
9. **Continuous Learning**: Many anomaly detection algorithms are adaptive, meaning they can learn from new data. This continuous learning ensures the system remains effective even as normal transactional behaviors evolve.
10. **Reduced Operational Costs**: While there is an initial cost in setting up and training an anomaly detection system, in the long run, it can lead to significant savings by reducing the number of fraudulent transactions and the associated costs of addressing them.
11. **Versatility**: Anomaly detection is not limited to credit card fraud alone. The same techniques can be adapted for other types of financial fraud, cybersecurity threats, and any domain where identifying unusual patterns is crucial.
12. **Customizability**: Anomaly detection techniques can be tailored to the specific needs and contexts of individual businesses. A system can be fine-tuned based on the type of transactions a business processes, its customer demographics, and other relevant factors.
13. **Enhanced Reporting and Insights**: Advanced anomaly detection systems can also provide detailed insights and reports on the types of anomalies detected, the potential causes, and trends over time. This can be invaluable for strategic decision-making and improving overall security infrastructure.

V. CONCLUSION

In the realm of financial transactions, credit card fraud remains a persistent challenge, incurring massive losses for businesses and eroding consumer trust. With the surge of online transactions and the sophistication of fraudulent schemes, there's an imperative need for robust, efficient, and adaptive solutions. The review of literature underscores that anomaly detection using machine learning has emerged as a pivotal strategy to counteract these fraudulent activities.

The inherent advantage of machine learning, and particularly anomaly detection, lies in its capability to discern intricate patterns and anomalies in vast datasets, which might elude traditional rule-based systems. Whether

© 2023 IJSRET

it's the swift identification of suspicious activities, the adaptability to novel threats, or the scalability to handle millions of transactions, anomaly detection techniques have demonstrated their efficacy.

Furthermore, the continuous evolution and refinement of machine learning models, coupled with the incorporation of deep learning and hybrid techniques, have enriched the landscape of anomaly detection. The adaptability of these models ensures they remain relevant, even in the face of ever-evolving fraud strategies. Techniques such as autoencoders, hybrid models combining unsupervised and supervised learning, and innovative feature selection strategies are pushing the boundaries of what's achievable in fraud detection.

REFERENCES

1. Jiang, S., Dong, R., Wang, J., & Xia, M. (2023). Credit Card Fraud Detection Based on Unsupervised Attentional Anomaly Detection Network. Systems, 11(6), 305.
2. Leevy, J. L., Hancock, J., &Khoshgoftaar, T. M. (2023). Comparative analysis of binary and one-class classification techniques for credit card fraud data. Journal of Big Data, 10(1), 118.
3. Aggarwal, S. (2023). LSTM based Anomaly Detection in Time Series for United States exports and imports.
4. Van Belle, R., Baesens, B., & De Weerdt, J. (2023). CATCHM: A novel network-based credit card fraud detection method using node representation learning. Decision Support Systems, 164, 113866.
5. Zhu, H., Zhou, M., Liu, G., Xie, Y., Liu, S., & Guo, C. (2023). NUS: Noisy-Sample-Removed Undersampling Scheme for Imbalanced Classification and Application to Credit Card Fraud Detection. IEEE Transactions on Computational Social Systems.
6. Yang, M., Liu, S., Xu, J., Tan, G., Li, C., & Song, L. (2023). Achieving privacy-preserving cross-silo anomaly detection using federated XGBoost. Journal of the Franklin Institute, 360(9), 6194-6210.
7. Jayasingh, B. B., & Sri, G. B. (2023, March). Online Transaction Anomaly Detection Model for Credit Card Usage Using Machine Learning Classifiers. In 2023 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 1-5). IEEE.
8. Strelcenia, E., &Prakoonwit, S. (2023). A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection. Machine Learning and Knowledge Extraction, 5(1), 304-329.
9. Prabhakaran, N., &Nedunchelian, R. (2023). Oppositional Cat Swarm Optimization-Based Feature Selection Approach for Credit Card Fraud Detection. Computational Intelligence and Neuroscience, 2023.
10. Ni, L., Li, J., Xu, H., Wang, X., & Zhang, J. (2023). Fraud feature boosting mechanism and spiral oversampling balancing technique for credit card fraud detection. IEEE Transactions on Computational Social Systems.
11. Strelcenia, E., &Prakoonwit, S. (2023). Improving Classification Performance in Credit Card Fraud Detection by Using New Data Augmentation. AI, 4(1), 172-198.
12. Fanai, H., &Abbasimehr, H. (2023). A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. Expert Systems with Applications, 217, 119562.
13. Pang, G., Shen, C., Jin, H., & van den Hengel, A. (2023, August). Deep weakly-supervised anomaly detection. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 1795-1807).
14. Krishna, M. H., Nithin, K., Charmitha, G., Vignesh, T., Ch, V., & Kuchibhotla, S. (2023, February). Studies on Anomaly Detection Techniques. In 2023 7th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 813-817). IEEE.
15. Eren, M. E., Moore, J. S., Skau, E., Moore, E., Bhattarai, M., Chennupati, G., & Alexandrov, B. S. (2023). General-purpose unsupervised cyber anomaly detection via non-negative tensor factorization. Digital Threats: Research and Practice, 4(1), 1-28.
16. Dorigo, T., Fumanelli, M., Maccani, C., Mojsovska, M., Strong, G. C., & Scarpa, B. (2023). RanBox: anomaly detection in the copula space. Journal of High Energy Physics, 2023(1), 1-46.
17. Fakiha, B. (2023). Forensic Credit Card Fraud Detection Using Deep Neural Network. Journal of Southwest Jiaotong University, 58(1).
18. López, D., Aguilera-Martos, I., García-Barzana, M., Herrera, F., García-Gil, D., & Luengo, J. (2023). Fusing anomaly detection with false positive mitigation methodology for predictive maintenance under multivariate time series. Information Fusion, 100, 101957.
19. Du, H., Lv, L., Guo, A., & Wang, H. (2023). AutoEncoder and LightGBM for Credit Card Fraud Detection Problems. Symmetry, 15(4), 870.
20. Copiaco, A., Himeur, Y., Amira, A., Mansoor, W., Fadli, F., Atalla, S., & Sohail, S. S. (2023). An innovative deep anomaly detection of building energy consumption using energy time-series images. Engineering Applications of Artificial Intelligence, 119, 105775.
21. Koko, R. R. Z., Yassine, I. A., Wahed, M. A., Madete, J. K., & Rushdi, M. A. (2023). Dynamic construction of outlier detector ensembles with bisecting k-means clustering. IEEE Access, 11, 24431-24447.
22. Jiang, M., Hou, C., Zheng, A., Hu, X., Han, S., Huang, H., ... & Zhao, Y. (2023). Weakly supervised anomaly detection: A survey. arXiv preprint arXiv:2302.04549.
23. Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A., Asgharnezhad, H., Shamsi, A., ... &Nahavandi, S. (2023). Uncertainty-aware credit card fraud detection

© 2023 IJSRET

using deep learning. Engineering Applications of Artificial Intelligence, 123, 106248.
24. Zhu, S., Yuchi, H. S., Zhang, M., & Xie, Y. (2023). Sequential adversarial anomaly detection for one-class event data. INFORMS Journal on Data Science.
25. Mienye, I. D., & Sun, Y. (2023). A Machine Learning Method with Hybrid Feature Selection for Improved Credit Card Fraud Detection. Applied Sciences, 13(12), 7254.
26. Vivek, B., Nandhan, S. H., Zean, J. R., Lakshmi, D., & Dhanwanth, B. (2023). Applying Machine Learning to the Detection of Credit Card Fraud. International Journal of Intelligent Systems and Applications in Engineering, 11(3), 643-652.
27. Alabrah, A. (2023). An Improved CCF Detector to Handle the Problem of Class Imbalance with Outlier Normalization Using IQR Method. Sensors, 23(9), 4406.
28. Jayanthi, E., Ramesh, T., Kharat, R. S., Veeramanickam, M. R. M., Bharathiraja, N., Venkatesan, R., &Marappan, R. (2023). Cybersecurity enhancement to detect credit card frauds in health care using new machine learning strategies. Soft Computing, 27(11), 7555-7565.
29. Kennedy, R. K., Salekshahrezaee, Z., Villanustre, F., &Khoshgoftaar, T. M. (2023). Iterative cleaning and learning of big highly-imbalanced fraud data using unsupervised learning. Journal of Big Data, 10(1), 106.
30. Bustos-Brinez, O. A., Gallego-Mejia, J. A., & González, F. A. (2023, February). AD-DMKDE: Anomaly Detection through Density Matrices and Fourier Features. In International Conference on Information Technology & Systems (pp. 327-338). Cham: Springer International Publishing.
31. Abhaya, A., & Patra, B. K. (2023). An efficient method for autoencoder based outlier detection. Expert Systems with Applications, 213, 118904.

© 2023 IJSRET