IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Realtime Fraud Detection Analysis Using Machine Learning.

A Review Article

[1]Omkar Binage, [2]Samueal D'Souza, [3]Anushka Bhagwat, [4]Sameekha Headu

[1]Student, [2]Student, [3]Student, [4]Student
[1]Department Of Computer Engineering,
[1]Vishwakarma Institute of Information Technology, Pune, India

Abstract: This collaborative review paper amalgamates insights from various studies on real-time fraud detection using machine learning in the context of online transactions. Focusing on credit card fraud, it explores innovative approaches such as clustering cardholders based on transaction amounts and employing a sliding window strategy to extract behavioral patterns. Additionally, the review addresses the challenges of network transactions, proposing fraud detection algorithms with impressive AUC values.In the realm of online banking fraud, the paper introduces models extending classical machine learning methods, incorporating economic optimization and a risk model. Real-world testing demonstrates significant reductions in expected financial losses, surpassing benchmarks. This unified perspective underscores the complexities, ongoing challenges, and potential solutions in the dynamic landscape of fraud detection. The collaborative efforts presented herein provide a holistic understanding of the intricacies involved in combating fraud systematically and economically.

IndexTerms **-Real-time fraud detection, machine learning, credit card fraud, behavioral patterns, network transactions, AUC values, online banking fraud, economic optimization, risk model, dynamic fraud landscape.**

I. INTRODUCTION

Credit cards, facilitating convenient transactions and deferred payments, are a pervasive financial tool vulnerable to fraud, particularly in the ever-expanding realm of online transactions. The ease with which fraudsters can exploit credit cards, coupled with the challenging task of detecting their illicit activities, poses a significant threat to both financial institutions and cardholders. Statistics from 2017 underscore the prevalence of credit card fraud, with 133,015 reported cases, making it the most common form among various fraud types, including employment or tax-related frauds, phone frauds, and bank frauds.

Despite advancements such as EMV cards, designed to enhance on-card payment security, credit card fraud remains a persistent issue. The shift towards card-not-present (CNP) transactions has become a focal point for fraudsters, with a notable increase in CNP fraud cases following heightened chip card security. However, the continual adaptation of fraud tactics emphasizes the need for robust fraud detection mechanisms.

Machine learning emerges as a promising solution to combat credit card fraud, leveraging techniques like clustering, sliding window strategies, and neural networks. The imbalance in credit card datasets, with significantly more legitimate transactions than fraudulent ones, necessitates innovative approaches to ensure accurate detection.

Beyond credit card fraud, the proliferation of online transactions, particularly in mobile payments, introduces new challenges. Criminals exploit the vastness of online trading platforms, posing a risk to personal property and hindering the healthy development of the network economy. Traditional fraud detection methods relying on statistical and multi-dimensional analyses are limited in capturing the nuanced patterns of fraudulent behavior within transaction data. Machine learning, with its ability to represent crucial features through extensive data, offers efficient detection methods for transaction fraud in the dynamic network environment.

This introduction further addresses the complexity of fraud detection in online and mobile payments, emphasizing the need for individualized data analysis. Existing studies underscore the importance of self-learning defense methods over static rule-based algorithms. The prevalence of imbalanced data, where fraudulent transactions are a minority, adds another layer of complexity to fraud detection algorithms.

As the financial industry grapples with rising fraud threats, the paper proposes a comprehensive fraud risk management framework. This framework encompasses an anomaly detection model, a fraud detection triage model that optimizes machine learning outputs economically, and a statistical risk model providing transparency and risk assessment.

The significance of this study lies not only in the development and validation of interconnected models but also in their real-world applicability. Utilizing real data, this research contributes a nuanced understanding of fraud detection complexities. The performance evaluation of the risk management framework demonstrates its ability to significantly reduce expected losses and prevent overestimation of fraud risk in various payment channels.

In the subsequent sections, we delve into the details of each model, presenting the methodology, validation results, and practical implications. This collaborative effort aims to provide a holistic approach to fraud detection and risk management in the ever-evolving landscape of financial transactions.

NEED OF THE STUDY.

The compelling need for the study on real-time fraud detection using machine learning is underscored by the alarming surge in cyber threats and financial fraud in the digital age. Recent statistics reveal a staggering increase in financial losses attributed to fraudulent activities, with losses in the range of $200 billion predicted between 2020 and 2024 across e-commerce, airline ticketing, money transfer, and banking services (Juniper Research, 2020). Online transactions, including credit card usage, are particularly vulnerable, experiencing a sharp rise in reported incidents.

In 2017 alone, there were 1,579 reported data breaches, comprising nearly 179 million records, and credit card fraud emerged as the most prevalent form, with 133,015 reported cases (FTC, 2017). The increasing prevalence of online payment methods has heightened the risk of fraud, necessitating advanced and adaptive fraud detection mechanisms. Moreover, the study acknowledges the dynamic nature of fraud schemes, with criminals continuously evolving their tactics to exploit vulnerabilities in online financial systems.

Machine learning stands as a promising solution, and the study focuses on developing and enhancing machine learning-based methodologies tailored for real-time fraud detection. The urgency is evident in the need to address the imbalanced nature of fraud datasets, the complexity of evolving fraud patterns, and the imperative for systems capable of processing vast transaction volumes in real-time.

This research not only aims to refine existing fraud detection methods but also endeavors to introduce innovative approaches capable of adapting to the ever-evolving tactics employed by cybercriminals. By employing statistical insights, this study seeks to provide tangible improvements, with the potential to reduce fraud-related losses. The outcomes are anticipated to have a substantial impact on financial institutions, businesses, and individual users, fortifying their defenses against fraud and mitigating financial risks.

As financial transactions continue their digital trajectory globally, this study contributes significantly to building a secure foundation for trust and reliability in the dynamic landscape of digital finance, safeguarding economic interests at both individual and institutional levels.

Literature Survey

Various supervised and semi-supervised machine learning techniques have been employed for fraud detection [8]. Challenges associated with card fraud datasets, including strong class imbalance, labeled and unlabeled samples, and the need to process a large number of transactions, have prompted innovative approaches. Supervised machine learning algorithms like Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression, and SVM are utilized for real-time fraud detection [3]. Random Forests, specifically random-tree-based and CART-based methods, are employed to train behavioral features of normal and abnormal transactions [6]. However, challenges persist, especially with imbalanced data, calling for future improvements.

While supervised learning methods show efficacy, limitations arise in certain cases. A model incorporating deep auto-encoder and restricted Boltzmann machine (RBM) is proposed to construct normal transactions and identify anomalies [2]. Additionally, a hybrid approach combining Adaboost and Majority Voting methods is developed [4]. Meta-classifiers and meta-learning approaches are explored for handling highly imbalanced credit card fraud data. The literature emphasizes the need for evolving methodologies to address the dynamic nature of fraud.

System Design

The system's design encompasses acquiring raw transaction data uploaded by users and preprocessing the data. The fraud detection algorithm, based on XGBoost, predicts the probability of a transaction being fraudulent. If the detected probability surpasses a set threshold (here, 0.85), a warning is issued immediately. The detection results are then saved in the database and returned to the user. The system flow chart outlines these steps [Figure 1].

System Realization - Fraud Detection Algorithm Based On Fully Connected Neural Network:

The algorithm involves three steps: feature selection, data preprocessing, and neural network construction [4]. Feature selection focuses on the Transaction dataset, with continuous features undergoing logarithmic transformation and Z-score standardization. Categorical features undergo One-Hot Encoding for the top 50 most common values. The Fully Connected Neural Network, built using Keras, incorporates the GELU activation function. Two cross entropy loss functions, binary cross entropy (BCE) and Focal loss (FL), are used to handle the imbalanced dataset. The model's performance is evaluated using AUC values and loss metrics.

System Realization - Fraud Detection Algorithm Based On XGBoost:
The XGBoost-based algorithm includes data preprocessing, parameter tuning, and model construction [9]. All features from both Transaction and Identity datasets are selected. LabelEncoder is used for categorical features, and PCA reduces the dimension of features from the V series. Hyperopt is employed for parameter tuning, defining an objective function and a configuration space. The best parameters for the XGBoost classifier are determined through optimization. The model's performance is evaluated using AUC values in a 7-fold cross-validation.

Algorithm Evaluation and Comparison:
Comparison between the Fully Connected Neural Network and XGBoost classifier reveals differences in AUC values, model training speed, and prediction time. XGBoost outperforms in accuracy but takes longer for prediction, making it suitable for precision-centric scenarios. The Fully Connected Neural Network, with quicker training speed, is suitable for time-sensitive situations.

Web Development Technology:
The transaction fraud detection system is implemented as a web platform based on the Django framework. Users can upload CSV files containing transaction and identity information. The system preprocesses the data and employs the trained XGBoost model for fraud prediction. Detection results are returned to users for download, and the system maintains a database of detection history. This web-based approach enhances accessibility and usability for end-users.

RESEARCH METHODOLOGY

Data Collection:
The research utilizes diverse datasets, including European credit card fraud data, online banking transaction records, and synthetic data for algorithm validation. Real-world data is prioritized to ensure the applicability of the proposed methodologies in practical scenarios.

Algorithm Development:
Two main fraud detection algorithms are explored:

Fully Connected Neural Network (FCNN): The algorithm involves feature selection, data preprocessing, and neural network construction. Various loss functions, including binary cross-entropy and focal loss, are considered for optimal model performance. XGBoost Classifier: This algorithm undergoes data preprocessing, parameter tuning, and model construction. Hyperopt is employed for efficient parameter tuning, ensuring robust model performance.

Algorithm Evaluation:
The developed algorithms are rigorously evaluated based on key criteria, including AUC values, training speed, and prediction time. Comparative analysis between FCNN and XGBoost provides insights into their respective strengths and weaknesses.

System Design and Realization:
The study involves the design and implementation of a real-time fraud detection system based on the XGBoost model. The system acquires raw transaction data, preprocesses it, and utilizes the trained model to predict the probability of fraud. The results are saved, and the detection history is stored in a database.

Statistical Analysis:
Statistical analysis is integral to the research methodology, with real-world figures on financial losses, reported fraud cases, and the effectiveness of existing models informing the development and evaluation processes.

Web Development Technology:
The final stage involves deploying the fraud detection system as a web platform. The system, built on Django framework, ensures user-friendly interaction. Users can upload transaction data, receive real-time fraud predictions, and access historical detection records.

Validation and Optimization:
The developed algorithms undergo extensive validation using real and synthetic data to ensure their effectiveness in diverse scenarios. Optimization strategies are explored to enhance model performance and address challenges such as imbalanced datasets.

Ethical Considerations:
Ethical guidelines, including data privacy and confidentiality, are strictly adhered to throughout the research. The use of real-world data is handled with sensitivity, and all procedures align with ethical standards in research.

Expected Outcomes:
The research anticipates providing novel insights into real-time fraud detection, with tangible outcomes including enhanced algorithms, optimized models, and a user-friendly web platform. The expected outcomes aim to contribute significantly to the field of cybersecurity and financial risk management.

IV. RESULTS AND DISCUSSION

Model Comparison Before SMOTE:

Table 3 presents the results of various fraud detection models on the original dataset before applying Synthetic Minority Over-sampling Technique (SMOTE). Notably, the Support Vector Machine (SVM), Logistic Regression, Decision Tree, and Random Forest models achieved high accuracy, precision, and Matthews Correlation Coefficient (MCC) values. The SVM demonstrated remarkable precision, capturing a substantial portion of fraud cases. However, the Local Outlier Factor exhibited lower precision, indicating a higher false-positive rate.

Model Comparison After SMOTE:

Table 4 displays the model performance after applying SMOTE to address class imbalance. The Logistic Regression, Decision Tree, and Random Forest models exhibit exceptional accuracy, precision, and MCC values after SMOTE. Particularly, the Random Forest model achieves near-perfect accuracy and MCC, showcasing its effectiveness in handling imbalanced datasets.

Conclusions:

1. **Model Performance Enhancement with SMOTE:**
   - SMOTE significantly improved the performance of various models, particularly Local Outlier Factor and Isolation Forest, which exhibited substantial increases in accuracy, precision, and MCC after addressing class imbalance.

2**. Model Suitability:**
   - Logistic Regression, Decision Tree, and Random Forest models, post-SMOTE, showcased outstanding results, making them suitable candidates for fraud detection in the given dataset.

3. **One-Class SVM Consideration:**
   - The One-Class SVM, despite being designed for binary class datasets, displayed lower accuracy and precision than other models. Consideration should be given to its relevance and efficacy in the context of the specific dataset characteristics.

4**. Web Transaction Fraud Detection System:**
   - The proposed online fraud detection system, based on the XGBoost classifier, demonstrated accurate prediction of fraud probabilities. Integration with online transactions provides proactive fraud interception before user payments.

Defense Mechanisms and Risk Management:

The study delves into the complexity of online banking fraud detection, emphasizing the need for a multi-faceted approach. The ensemble of models, including Local Outlier Factor, Isolation Forest, Support Vector Machine, Logistic Regression, Decision Tree, and Random Forest, contributes to a comprehensive defense mechanism against sophisticated fraud attacks.

Future Directions:

1. **Reinforcement Learning Integration:**
   - Future research directions may explore the integration of reinforcement learning methods to establish a feedback loop, creating a learning system for continuous improvement in fraud detection.

2. **Optimization of Fraud Detection:**
   - Further optimization of fraud detection, considering transaction-dependent loss risks and customer segmentation, could enhance the model's adaptability to evolving fraud patterns.

3**. Ethical Considerations:**
   - Ongoing emphasis on ethical considerations, including data privacy and confidentiality, ensures responsible and transparent use of real-world data in fraud detection research.

The research not only contributes valuable insights into fraud detection methodologies but also provides a foundation for advancing risk management practices in the dynamic landscape of online financial transactions.

II. ACKNOWLEDGMENT

REFERENCES

[1] Jiang, Changjun et al. “Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism.” IEEE Internet of Things Journal 5 (2018): 3637-3647.
[2]Pumsirirat, A. and Yan, L. (2018). Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. International Journal of Advanced Computer Science and Applications, 9(1).

[3]Mohammed, Emad, and Behrouz Far. "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study." IEEE Annals of the History of Computing, IEEE, 1 July 2018, doi.ieeecomputersociety.org/10.1109/IRI.2018.00025.
[4]Randhawa, Kuldeep, et al. "Credit Card Fraud Detection Using AdaBoost and Majority Voting." IEEE Access, vol. 6, 2018, pp. 14277–14284., doi:10.1109/access.2018.2806420.
[5]Roy, Abhimanyu, et al. "Deep Learning Detecting Fraud in Credit Card Transactions." 2018 Systems and Information Engineering Design Symposium (SIEDS), 2018, doi:10.1109/sieds.2018.8374722.
[6]Xuan, Shiyang, et al. "Random Forest for Credit Card Fraud Detection." 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), 2018, doi:10.1109/icnsc.2018.8361343.
[7]Awoyemi, John O., et al. "Credit Card Fraud Detection Using Machine Learning Techniques: A Comparative Analysis." 2017 International Conference on Computing Networking and Informatics (ICCNI), 2017, doi:10.1109/iccni.2017.8123782.
[8]Melo-Acosta, German E., et al. "Fraud Detection in Big Data Using Supervised and Semi-Supervised Learning Techniques." 2017 IEEE Colombian Conference on Communications and Computing (COLCOM), 2017, doi:10.1109/colcomcon.2017.8088206.
[9] http://www.rbi.org.in/Circular/CreditCard
[10] https://www.ftc.gov/news-events/press-releases/2019/02/imposter-scams-top-complaints-made-ftc-2018
[11] https://www.kaggle.com/mlg-ulb/creditcardfraud
[12] https://www.kaggle.com/uciml/default-of-credit-card-clients-dataset
[13] https://www.kaggle.com/ntnu-testimon/paysim1/home

[14] E. Kurshan, H. Shen: Graph Computing for Financial Crime and Fraud Detection: Trends, Challenges and Outlook. 2020 Second International Conference on Transdisciplinary AI (TransAI), 2020.
[15] Dheepa V, Dhanapal R: Analysis of credit card fraud detection methods. International journal of recent trends in engineering, 2009, 2(3):126.
[16] Zhang Z, Zhou X, Zhang X, et al: A model Based on Convolutional Neural Network for Online Transaction Fraud Detection. Security and Communication Networks, 2018, 10(2):1-9.
[17] Jain V: Perspective analysis of telecommunication fraud detection using data stream analytics and neural network classification based data mining. International Journal of Information Technology, 2017, 9(3):303-310.
[17] Hendrycks D, Gimpel K: Gaussian Error Linear Units (GELUs), 2016.
[18] Glorot X, Bordes A, Bengio Y: Deep Sparse Rectifier Neural Networks. Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS), 2011, 15:315- 323.
[19] Lin T Y, Goyal P, Girshick R, et al: Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 42:2999-3007.
[20] Prusti D, Rath S K: Fraudulent Transaction Detection in Credit Card by Applying Ensemble Machine Learning techniques. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2019.
[21] Lei, Shimin, et al: An Xgboost based system for financial fraud detection. E3S Web of Conferences, 2020, 214(2).
[22] Bergstra J, et al: Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms. Computational Science & Discovery, 2015, 8(1).