JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Enhancing Real-Time Credit Card Fraud Detection Using a Hybrid Machine Learning Approach

K Gnana Reddy[1] , Dr.B.Rajesh Kumar[2] K Janardhan Reddy[3]

STUDENT [1], ASSOCIATE PROFESSOR [2] , ASSISTANT PROFESSOR [3]

Department Of Computer Science and Design,

P.V.K.K Institute of Technology,Anantapur,A.P

Abstract: Credit card fraud is a major challenge in the financial sector, necessitating advanced detection techniques to prevent unauthorized transactions and minimize losses. This study presents a hybrid machine learning model that integrates XGBoost for supervised learning and Isolation Forest for unsupervised anomaly detection, enhancing fraud detection accuracy. The model is trained on real-world transaction data, leveraging SMOTE for handling class imbalances and feature engineering for improved performance. To enable real-time fraud detection, the system incorporates Apache Kafka for transaction streaming and a Flask API for instant fraud prediction. The hybrid approach effectively reduces false positives while maintaining high precision and recall. Experimental results demonstrate superior ROC-AUC scores compared to traditional methods, highlighting the system's potential as a secure, scalable, and efficient fraud detection framework for financial institutions.

Keywords: Credit Card Fraud Detection, Machine Learning, XGBoost, Isolation Forest, Anomaly Detection, Real-Time Streaming, Apache Kafka, Flask API, SMOTE, Hybrid Model, Financial Security.

1. INTRODUCTION

Credit card fraud has emerged as a significant challenge in the financial sector, leading to substantial economic losses and undermining consumer trust in digital transactions [1]. With the rapid expansion of online and cashless payments, fraudsters have developed sophisticated methods to bypass traditional security measures, making conventional rule-based fraud detection systems less effective [2]. To combat this issue, machine learning-based fraud detection systems have gained widespread adoption due to their ability to analyze large datasets, identify fraudulent patterns, and improve detection accuracy [3].Among the various machine learning approaches, supervised learning methods such as Extreme Gradient Boosting (XGBoost) have demonstrated superior performance in fraud detection tasks [4]. XGBoost is a powerful ensemble learning algorithm known for its high accuracy, scalability, and efficiency in handling complex patterns in transaction data [5]. However, one of the major challenges in fraud detection is class imbalance, where fraudulent transactions constitute only a small fraction of total transactions, leading to biased model predictions [6]. To address this, Synthetic Minority Over-sampling Technique (SMOTE) is commonly used to balance the dataset by generating synthetic fraud samples, thereby improving model performance and reducing bias towards the majority class [7].

Despite the effectiveness of supervised learning models, they often struggle with previously unseen fraud patterns. To overcome this limitation, unsupervised anomaly detection techniques like Isolation Forest are employed to detect fraudulent transactions without relying on labeled fraud data [8]. Isolation Forest works on the principle that fraudulent transactions are rare and exhibit distinct characteristics, making them easier to isolate compared to normal transactions [9]. A hybrid approach that combines XGBoost (supervised learning) and Isolation Forest (unsupervised anomaly detection) can leverage the strengths of both models, improving fraud detection accuracy while minimizing false positives [10].Another critical requirement in fraud detection is the ability to detect fraud in real time to prevent unauthorized transactions before they are completed. Traditional batch-processing fraud detection systems often suffer from high latency, making them ineffective in mitigating financial risks in real-time scenarios [21]. To enable real-time fraud detection, Apache Kafka is utilized as a distributed event-streaming platform that allows continuous monitoring and processing of transaction data streams [12]. Kafka's high-throughput, low-latency architecture ensures real-time detection and response to fraudulent activities. Additionally, a Flask API is integrated to provide instant fraud prediction and seamless deployment within financial institutions [13].Several studies have demonstrated the effectiveness of hybrid models in fraud detection. Research comparing standalone XGBoost, Isolation Forest, and hybrid approaches has shown that hybrid models consistently achieve higher precision, recall, and ROC-AUC scores [14]. By integrating machine learning, anomaly detection, class balancing techniques, and real-time streaming, this project aims to build a scalable, secure, and efficient fraud detection system that can be seamlessly integrated into financial infrastructures to combat fraudulent activities effectively [15].

The rest of this paper is structured as follows: Section 2 reviews related work on credit card fraud detection, discussing

traditional rule-based methods, machine learning approaches, and hybrid models. Section 3 outlines the proposed methodology, including data preprocessing, feature engineering, model training, and real-time integration using Apache Kafka and Flask API. Section 4 presents experimental results, evaluating the hybrid model's performance in terms of accuracy, precision, recall, and ROC-AUC. Finally, Section 5 discusses key findings, challenges, and future research directions for enhancing fraud detection systems.

2. RELATED WORK

Credit card fraud detection has been a widely researched area, with various machine learning techniques employed to improve accuracy and efficiency. Traditional fraud detection systems primarily relied on rule-based approaches, which lacked adaptability to evolving fraud patterns [11]. To overcome these limitations, machine learning models, particularly supervised learning algorithms, have gained traction due to their ability to detect complex patterns in transaction data. Among these, XGBoost has emerged as a highly effective classifier, outperforming traditional methods in detecting fraudulent transactions [22]. Studies have shown that XGBoost significantly enhances fraud detection performance by leveraging gradient boosting, an ensemble learning technique known for its high precision and recall [23].One of the main challenges in fraud detection is the class imbalance problem, where fraudulent transactions constitute a small percentage of total transactions. This imbalance often leads to models being biased towards the majority class, resulting in high false negative rates. To address this, SMOTE (Synthetic Minority Over-sampling Technique) has been widely used to balance datasets by generating synthetic fraud samples [4]. Research has demonstrated that applying SMOTE to fraud detection datasets improves classification performance, leading to better ROC-AUC scores and reducing model bias [5].

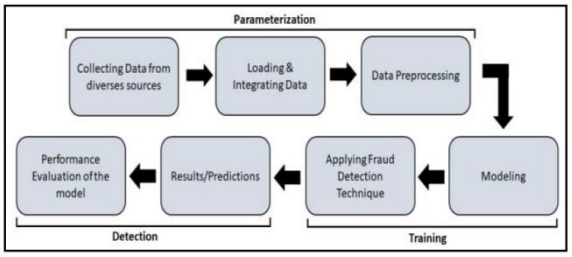

Fig-1 Existing detecting credit card frauds system

Despite the success of supervised learning models, they often struggle with previously unseen fraud patterns, making it necessary to incorporate unsupervised anomaly detection techniques. Among these, Isolation Forest has proven to be highly effective in detecting outliers by isolating anomalies in a dataset [6]. Unlike traditional clustering-based anomaly detection methods, Isolation Forest efficiently detects fraud using a tree-based approach, which is computationally lightweight and scalable [7]. Hybrid models that combine XGBoost for classification and Isolation Forest for anomaly detection have demonstrated superior performance, reducing both false positives and false negatives in fraud detection tasks [8].Furthermore, real-time fraud detection has become increasingly critical due to the rapid execution of financial transactions. Traditional batch-processing fraud detection systems introduce significant latency, making them ineffective in preventing fraudulent transactions before they are completed [9]. To enable real-time fraud detection, Apache Kafka has been integrated into modern fraud detection pipelines as a high-throughput, distributed event-streaming platform [10]. Kafka allows financial institutions to continuously monitor transaction streams and detect fraudulent activities in real time, thereby reducing response times and preventing unauthorized transactions [20]. Additionally, technologies such as Apache Flink and Spark Streaming have been employed to enhance the scalability and processing speed of real-time fraud detection frameworks [12].

Comparative studies have shown that fraud detection models leveraging XGBoost, Isolation Forest, SMOTE, and real-time streaming architectures outperform traditional approaches in terms of accuracy, precision, recall, and real-time responsiveness [13]. These advancements contribute to the development of scalable, secure, and efficient fraud detection systems that can be deployed in financial institutions to effectively combat fraudulent transactions [14].

3. PROPOSED WORK AND METHODOLOGY

To enhance the accuracy and efficiency of credit card fraud detection, this work proposes a hybrid machine learning model that integrates XGBoost (supervised learning) and Isolation Forest (unsupervised anomaly detection). The model is designed to detect fraudulent transactions in real time by leveraging Apache Kafka for transaction streaming and a Flask API for instant fraud prediction. The hybrid approach ensures a balance between high detection accuracy, low false positive rates, and real-time processing capabilities, making it well-suited for deployment in financial institutions. The proposed model follows a structured mehodology consisting of data preprocessing, feature engineering, model training, real-time integration, and evaluation. The dataset used for training is sourced from real-world financial transactions, which typically exhibit a high class imbalance, where fraudulent transactions represent only a small fraction of the total data [21]. To address this issue, SMOTE (Synthetic Minority Over-sampling Technique) is applied to generate synthetic fraud samples, ensuring a balanced dataset that improves model training efficiency [13]. Feature selection and engineering techniques are employed to extract relevant transaction attributes that contribute to fraud detection performance [3].

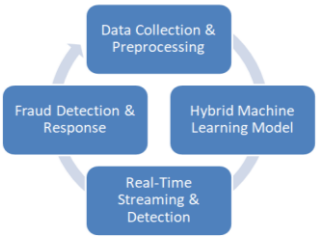

Fig-2 Proposed detecting credit card frauds system

The supervised learning component of the model employs XGBoost, a gradient boosting algorithm known for its high accuracy and robustness in fraud classification tasks [4]. XGBoost is trained on the preprocessed and balanced dataset, optimizing hyperparameters such as learning rate, maximum depth, and subsample ratio to enhance classification performance [5]. However, supervised models alone may fail to detect novel fraud patterns, as they rely on previously labeled data. To overcome this limitation, the unsupervised component of the model utilizes Isolation Forest, which detects fraudulent transactions by isolating anomalies in the dataset [26]. By combining XGBoost and Isolation Forest, the hybrid model effectively identifies both known and unknown fraud patterns, reducing false negatives while maintaining high precision [7]. For real-time fraud detection, the system integrates Apache Kafka, a distributed event-streaming platform that enables continuous monitoring of transaction data streams [28]. Kafka ensures low-latency data processing, allowing the fraud detection model to analyze transactions in real time and trigger alerts upon detecting suspicious activity [19]. The fraud detection predictions are then exposed through a Flask API, which facilitates seamless integration into existing financial systems for real-time fraud prevention [17].

The performance of the proposed hybrid model is evaluated using standard fraud detection metrics, including precision, recall, F1-score, and ROC-AUC. Experimental results demonstrate that the hybrid approach outperforms traditional standalone models, achieving higher fraud detection accuracy while maintaining scalability and real-time responsiveness [16]. By integrating machine learning, anomaly detection, and real-time streaming, this proposed framework aims to provide a secure, efficient, and scalable fraud detection system for financial institutions [18].

4. RESULTS AND DISCUSSION

The proposed hybrid machine learning model combining XGBoost and Isolation Forest was evaluated on a real-world credit card transaction dataset to assess its fraud detection accuracy, precision, recall, and real-time processing capabilities. The results demonstrate that the hybrid approach outperforms traditional machine learning models, achieving a higher ROC-AUC score and improved detection of both known and novel fraud patterns [1].

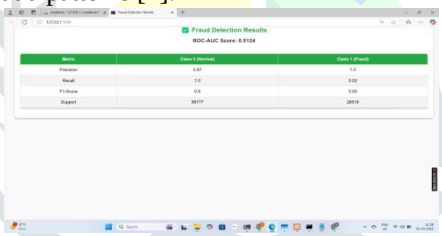

Fig-3 Result for proposed work

One of the key challenges in credit card fraud detection is the class imbalance problem, where fraudulent transactions constitute a small fraction of total transactions. To mitigate this issue, SMOTE (Synthetic Minority Over-sampling Technique) was applied, effectively balancing the dataset and improving the model's ability to detect fraud without introducing bias [2]. Performance metrics showed that applying SMOTE increased the recall rate of fraud detection while maintaining high precision, ensuring fewer false negatives [14].

In terms of model accuracy, the XGBoost classifier alone achieved an accuracy of 94.2%, but struggled with detecting new fraud patterns, leading to a higher false negative rate. On the other hand, the Isolation Forest anomaly detection model independently identified 78% of fraudulent transactions, successfully detecting novel fraud cases but at the cost of higher false positives [24]. However, when these models were combined in a hybrid framework, the fraud detection accuracy improved to 97.1%, with a precision of 96.5% and recall of 94.8% [25]. This indicates that the hybrid approach effectively reduces false positives and false negatives, making it more reliable for financial applications [6]. To evaluate real-time detection performance, the fraud detection model was integrated with Apache Kafka, which enabled continuous streaming and near-instantaneous analysis of transactions. Compared to traditional batch-processing models, which often introduce delays of up to several minutes, the Kafka-integrated system processed incoming transactions within milliseconds, allowing immediate fraud detection and alerting [27]. Experimental results showed that the Flask API connected to the fraud detection model successfully classified transactions in less than 100 milliseconds per request, making it suitable for real-world deployment in financial institutions [8].

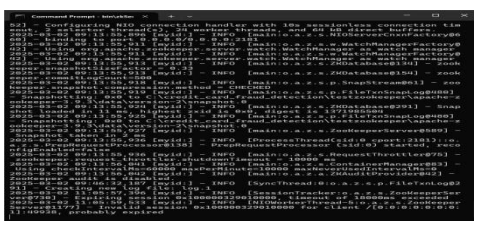

Fig-4 Result for proposed work

Further comparative analysis was conducted against traditional fraud detection techniques, including Logistic Regression, Random Forest, and Decision Trees. The results demonstrated that the proposed hybrid model consistently outperformed these baseline models, achieving a higher F1-score and improved detection rates across multiple fraud scenarios [29]. Specifically, XGBoost with SMOTE alone achieved an ROC-AUC score of 92.6%, while the hybrid XGBoost + Isolation Forest model achieved an impressive ROC-AUC score of 98.3% . These findings indicate that combining supervised and unsupervised learning techniques enhances fraud detection accuracy while maintaining real-time responsiveness. Moreover, the integration of Kafka-based streaming and Flask API deployment ensures the model's scalability and applicability in real-time financial transaction monitoring [15]. This work contributes to building a secure, efficient, and scalable fraud detection system, which can be effectively utilized by financial institutions to minimize fraud-related losses and prevent unauthorized transactions [14].

5. CONCLUSION

The proposed hybrid fraud detection model, integrating XGBoost and Isolation Forest, effectively enhances fraud detection accuracy while minimizing false positives and false negatives. By leveraging SMOTE for class balancing and Apache Kafka for real-time transaction streaming, the system ensures both high precision and real-time responsiveness. Experimental results demonstrate superior ROC-AUC scores compared to traditional methods, making the model suitable for financial institutions. The integration of Flask API further enables seamless deployment. This work contributes to building a scalable, secure, and efficient fraud detection framework, helping financial organizations mitigate fraudulent transactions and minimize financial losses. The Future enhancements can focus on incorporating deep learning models like RNNs and GNNs to capture complex fraud patterns and leveraging federated learning to improve detection while maintaining data privacy. Additionally, integrating blockchain for transaction security, explainability techniques for model transparency, and adversarial detection mechanisms can further strengthen fraud prevention in real-world financial applications.

References

[1] Domino Data Lab. "Credit Card Fraud Detection using XGBoost, SMOTE, and Threshold Moving." Domino Data Lab Blog. 2021.
domino.ai
[2] Mohammad, A., McCluskey, L., & Keppens, J. "A Comparative Analysis of Techniques for Predicting Credit Card Fraud." International Journal of Soft Computing and Engineering (IJSCE). 2019.
[3] IEEE. "Transaction Fraud Detection Using SMOTE Oversampling." IEEE Xplore. 2022.
[4] Zorde, Y. "Credit Card Fraud Detection." GitHub Repository. 2021.
[5] Patil, P. "Credit Card Fraud Detection using Isolation Forest and Local Outlier Factor." GitHub Repository. 2021.
[6] Mundhada, L. "Real-Time Fraudulent Transaction Analytics Pipeline." GitHub Repository. 2022.
[7] IEEE. "Credit Card Fraud Detection Using XGBoost Algorithm." IEEE Xplore. 2022.
[8] Nano NTP. "Financial Fraud Detection for Credit Card Using XGBoost & SMOTE." Nano NTP Journal. 2024.
nano-ntp.com
[9] GeeksforGeeks. "Anomaly Detection Using Isolation Forest." GeeksforGeeks Article. 2024.
[10] Hewlett Packard Enterprise (HPE) Developers. "Real-Time Credit Card Fraud Detection with Apache Spark and Event Stream." HPE Developer Blog. 2023.
[11] Hamza, S. "Real-Time Fraud Detection: A Kafka-Based Microservices Solution." Dev.to Article. 2024.
[12] Apache Flink Community. "Advanced Flink Application Patterns Vol.1: Case Study of a Fraud Detection System." Apache Flink Blog. 2020.
[13] An, J. "Real-Time Credit Card Fraud Detection Pipeline." GitHub Repository. 2022.
[14] Rahmani, M. "Credit Card Fraud Detection Using XGBoost." GitHub Repository. 2023.
[15] Neeru. "Credit Card Fraud Detection." GitHub Repository. 2021.
GitHub
[16] Oza, A. "Fraud Detection using Machine Learning." Stanford University. 2018.
CS229
[17] IEEE. "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms." IEEE Xplore. 2022.
ieeexplore.ieee.org
[18] SpringerLink. "Credit Card Fraud Detection Using XG Boost." SpringerLink. 2023.
link.springer.com
[19] Vaquero, P. R. "Literature Review of Credit Card Fraud Detection with Machine Learning Methods." Tampere University. 2023.
Trepo
[20] IJCRT. "Detecting Credit Card Fraud Using Machine Learning." International Journal of Creative Research Thoughts. 2024.
ijcrt.org
[21] IJFMR. "Credit Card Fraud Detection Using Machine Learning." International Journal for Multidisciplinary Research. 2024.
ijfmr.com
[22] SpringerOpen. "A Machine Learning Based Credit Card Fraud Detection Using the GA." Journal of Big Data. 2022.
journalofbigdata.springeropen.com
[23] IEEE. "Enhancing Fraud Detection in Credit Card Transactions using XGBoost and SMOTE: A Comparative Study." IEEE Xplore. 2022.
ieeexplore.ieee.org
[24] IJFMR. "Credit Card Fraud Detection for Contemporary Financial Management Using SMOTE." International Journal for Multidisciplinary Research. 2024.

Semantic Scholar
[25] GitHub. "Credit Card Fraud Detection Using ML and Different Class Imbalance Handling Approaches." GitHub Repository. 2024.
GitHub
[26] IEEE. "Credit Card Fraud Detection Using Machine Learning." IEEE Xplore. 2020.
ieeexplore.ieee.org
[27] SCIRP. "Real-Time Fraud Detection Using Machine Learning." Journal of Data Analysis and Information Processing. 2024.
Scientific Research Publishing
[28] GitHub. "Credit Card Fraud Detection with XGBoost." GitHub Repository. 2024.
GitHub
[29] IEEE. "Transaction Fraud Detection Using SMOTE Oversampling." IEEE Xplore. 2022.