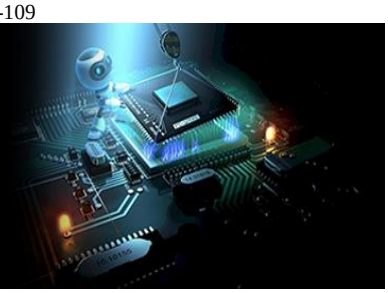
International Journal of Engineering in Computer Science



E-ISSN: 2663-3590
P-ISSN: 2663-3582
www.computersciencejournals.com/ijecs
IJECS 2024; 6(2): 106-109
Received: 21-05-2024
Accepted: 09-07-2024

Geetha Prathiba
Assistant Professor, Department of Information and Technology, Malla Reddy Engineering College for Women, Autonomous, Hyderabad, Telangana, India

Ch Shravani
Student, Department of Information and Technology, Malla Reddy Engineering College for Women, Autonomous, Hyderabad, Telangana, India

A Ashritha
Student, Department of Information and Technology, Malla Reddy Engineering College for Women, Autonomous, Hyderabad, Telangana, India

E Lavanya
Student, Department of Information and Technology, Malla Reddy Engineering College for Women, Autonomous, Hyderabad, Telangana, India

Corresponding Author:
Geetha Prathiba
Assistant Professor, Department of Information and Technology, Malla Reddy Engineering College for Women, Autonomous, Hyderabad, Telangana, India

Random forest based fraud detection method for multi-participant e-commerce transactions

Geetha Prathiba, Ch Shravani, A Ashritha and E Lavanya

DOI: https://doi.org/10.33545/26633582.2024.v6.i2b.131

Abstract
The primary goal of transactional security solutions has always been to detect and prevent fraudulent transactions on e-commerce platforms. Due to the anonymity of online transactions, it is difficult to identify attackers by just looking at past order data. Academics are busy trying to come up with fraud prevention systems, but they haven't thought about how consumers' behaviors are evolving. As a result, fraudulent behavior is not effectively detected. An innovative approach to real-time user activity monitoring for fraud detection is presented by this study, which combines process mining with algorithms grounded in machine learning. A process model with user behavior detection is first developed for the business-to-consumer online store. Secondly, we provide an anomaly-based approach to data mining that might be applied to event logs. A classification model that employs SVM (support vector machine) techniques to identify fraudulent activity is then fed the collected characteristics. The results of the studies show that our technique successfully identifies dynamic fraudulent behavior on e-commerce platforms.

Keywords: Fraud detection, e-commerce transactions, random forest

Introduction
More and more business transactions are shifting away from the old cash-based method and toward web-based methods, driven by the meteoric rise of e-commerce platforms. In spite of the recent significant detrimental effect of the COVID-19 epidemic on the entity economy, e-commerce has persisted in its constant growth. The absence of harm to the industry is largely responsible for this. Forecasts indicate that by 2023, business-to-consumer (B2C) e-commerce would have generated $6.50 trillion. Online companies have more chances than ever before because to the expansion of e-commerce made possible by contemporary technology. However, new security dangers have also surfaced in recent years. It has been reported that the annual global expenditures associated with incidents of internet fraud have increased significantly. Due to the decentralized and ever-changing character of the Internet, anti-fraud solutions are crucial for maintaining the trustworthiness of online transactions. When it comes to addressing the growing security risks, current fraud detection technologies are not up to the task. Unusual user behavior detection is the primary goal of these systems. The lack of proper process management all through the trading process is a major flaw in the current fraud detection technologies. The flawed monitoring function is a major problem that needs resolving. The detecting perspective is often insufficient in the current work due to the absence of process capture. To achieve this, we provide a process-based approach that tracks and evaluates user actions in real-time while also converting previous data into controlled knowledge. We also provide a multi-viewpoint approach to detect anomalous behaviors. In order to find outliers in data streams, this study presents a hybrid approach that uses process mining and machine learning models. It details all of the processes that make up the procedure control model. The method models and analyzes the e-commerce system's business process to identify changes in user behavior, payment methods, or issues with noncompliance. Additionally, it is capable of completely and multi-perspective identifying fraudulent transactions. Following is a synopsis of the most important findings from this research:

1. In the realm of online purchases, anomalies are identified using a conformity evaluation method based on data mining.

2. A Petri-noodle-based user behavior detection technique is suggested for comprehensive anomaly detection.
3. To autonomously define fraudulent behaviors, an SVM model is constructed by merging multi viewpoint process mining with machine learning methodologies. This is the structure that the remainder of the essay follows: Section 2 displays the matching work. In Section 3, you will get the background knowledge you need and the analytical framework. Our proposed method for detecting fraud and its theoretical foundations are laid forth in Section 4. In Section 5, we go over the experimental findings and their analysis, and in Section 6, we prove that our fraud detection system works. Our findings and proposed areas for further study are detailed in Section 7.

Related Work

The five biggest online retailers in the world are being studied in relation to the impact of the COVID-19 pandemic on the international e-commerce market

These five e-commerce giants dominate the world's market and dominate sales. The impact of COVID-19 on international e-commerce firms is the intended subject of this research. Amazon (US), Alibaba (China), Rakuten (Japan), Zalando (Germany), and ASOS (United States) all utilize "cumulative infections" and "accumulative deaths" as their daily metrics of the corona virus's prevalence. A daily return of shares of online sales enterprises to global financial markets is also computed by the dependent factor using the values of "fresh corona virus cases" and "a new corona viral-related kills" as inputs. This data demonstrates how the worldwide e-commerce industry has been impacted by the coronavirus pandemic. From March 15, 2020, until May 25, 2020, this was utilized daily. Assuming average daily returns yields good outcomes for e-commerce enterprises, according to the descriptive examination of their returns. To what degree did the aggregate model impact the return on shares of overseas electronic trading companies? The Beta Dressed Coefficients test revealed the answer. First, minute, and third place were most affected by the overall amount of cases, total number of deaths, and volume of new cases, respectively. Amazon and ASOS, both based in the United States, were "the the most significant over time cases of getting sick, based to their being to the worst adversely affected countries during the research a period," depending on the company and country in question; Zalando, based in Germany, had been the variable with the greatest impact in terms of gathering deaths, though. Alibaba and Rakuten, two Chinese companies, were the most financially advantageous "Corona virus incidents" for the Japanese stock market.

An empirical study of the internet retail business focusing on sustainable development in the e-commerce supply chain

The environmental impact of e-commerce, particularly business-to-consumer (B2C) e-commerce, has been a topic of mixed findings in previous studies. Regardless, the industry has seen explosive growth in recent years. The research initially developed two hypotheses that might provide light on the environmental impacts of e-commerce. Afterwards, a systematic questionnaire was utilized to assess the suggested models, with 303 replies from GCC nations. As part of this process, we evaluated the fundamentals and the relevance and usefulness of each component. The next step was to evaluate the expected connections between all of the concepts. The results of the first model demonstrated how green consumerism shapes consumers' positive and negative environmental views about e-commerce, and how this in turn affects their intentions to employ e-commerce strategies. Positive environmental dispositions no longer predicted behavioral intention after the addition of the perceived simplicity-for-the-sake-of-use and felt-value variables to the second model, since consumers had previously estimated the practicality and ease-of-use of e-commerce over these positive environmental traits. Despite the apparent utility and convenience of use, unfavorable environmental attitudes remained in their influence on behavioral intention. Researchers hope that policymakers and practitioners will use the study's findings to lessen the negative effects of online shopping on the environment and make the most of its beneficial aspects. The study's results are groundbreaking because they may be the first empirical effort to weigh the environmental pros and downsides of online sales and how these factors influence consumers' decisions.

An explanation of how electronic payment systems use hidden codes to forestall fraud The proliferation of online marketplaces like AliExpress.com, Amazon.com, and eBay has led to a meteoric surge in the number of financial transactions conducted by electronic means in recent years. Credit cards are widely accepted as a form of payment for both online and offline transactions nowadays. When it comes to online payment systems, one of the major ethical concerns is the possibility of fraud. A person commits fraud if they knowingly and intentionally use dishonesty to benefit themselves or cause harm to another. The number of fraud litigation has also risen sharply, with the potential yearly global harm from these cases running into the billions of dollars. Consequently, it is critical and required to take measures to avoid fraud and to implement strategies that might help identify and prevent it. It is understandable that many solutions to the issue have serious limitations, given the difficulties of avoiding fraud in real-time. Encryption is the practice of encoding information or messages so that only authorized parties may decipher them. Credit card numbers and other sensitive financial data used in online purchases must be securely stored by this system. Using a secret code, the method decreases fraud. The encryption nature of this secret code makes it impossible for unauthorized individuals to utilize it. The goal of our endeavor is to clarify things.

Analyzing fraud prevention software

Most monetary transactions, including those involving debit cards, cell cell phones, and health safety measures, may now be conducted via electronic commerce platforms, thanks to the widespread use of computers and the continued expansion of businesses. Ignorant consumers and criminals alike make advantage of these technologies, which is a major setback. The networks used for online shopping have been breached by fraudsters using a variety of methods. There is a lack of adequate protection against fraud in the theft detection systems (FPSs) used by systems for electronic commerce. When it comes to protecting e-commerce systems, nevertheless, cooperation among FDSs and FPSs could be beneficial. However, FDSs aren't perfect; problems like concept drift, skewed distribution, big data

sets, real-time detection support, and more make them ineffective. These problems and obstacles hinder the operational effectiveness of FDSs, and this survey article intends to provide a thorough and methodical assessment of them. Credit card, interaction, medical security, auto insurance, and online auction systems are the five digital commerce platforms that we choose. There is an in-depth discussion of the most typical forms of fraud on those websites. In addition, only a small number of E-commerce platforms consistently use cutting-edge FDSs methods. Following that, the paper provides a concise overview of possible avenues for further study before drawing to a close.

The most popular machine learning methods for detecting fake websites and how they work

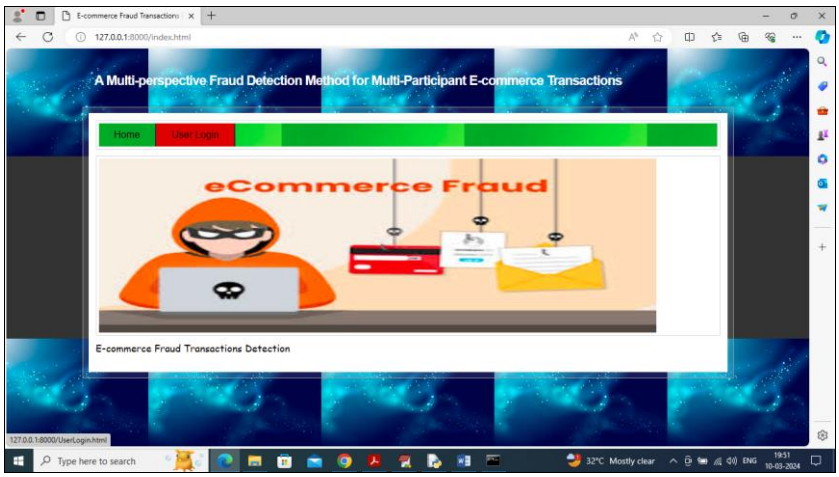
Due to their increasing complexity and worldwide reach, unlawful financial operations on the internet now cost businesses and consumers a substantial amount of money. In the digital world, there is a dearth of methods for detecting and avoiding fraud. While the end objective of all of these methods is to detect and prevent fraudulent online transactions, each of these approaches has its own set of pros and cons. The goal of these articles is to identify the currently used algorithms for fraud detection and to evaluate them according to certain criteria based on prior research in the field. A thorough statistical literature analysis was used to analyze the fraud detection investigation. Using the features of the most widely used artificial intelligence algorithms in academic publications, a classification system is established. In a novel approach, our analysis reveals the best fraud detection methods by integrating three selection criteria: coverage, accuracy, and pricing. published in the Journal of Economic Informatics.

Methodology

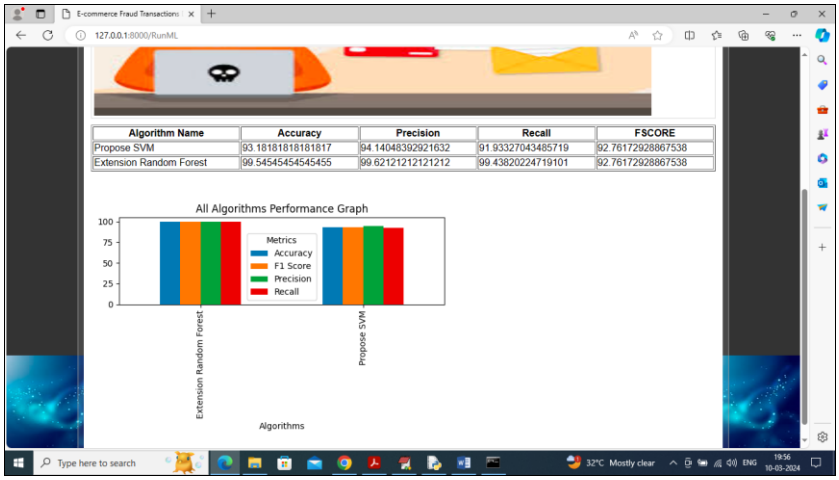
We have created the following modules in order to carry out this project.

1. The system's login credentials are "admin."
   The second step is to load and process the dataset.
2. After logging in, the individual may upload a dataset and then utilize various processing techniques including normalization, shifting, and missing value removal. Finally, they can convert any variables that are not numeric to numeric ones.
3. Process Mining: An algorithm will be used to determine normal and aberrant user behavior from a processed collection. The outcome is a graph that represents both types of transactions.
4. Execute ML techniques: Following the data's partitioning into learning and testing sets, the former is used to train a model using Svc and random forests techniques. The latter is then applied to the former to ascertain the prediction accuracy.
5. Fraud Detection: An extension algorithm will be used to categorize samples as Normal or Fraud based on the findings of behavior extraction via process mining and the uploaded test data.

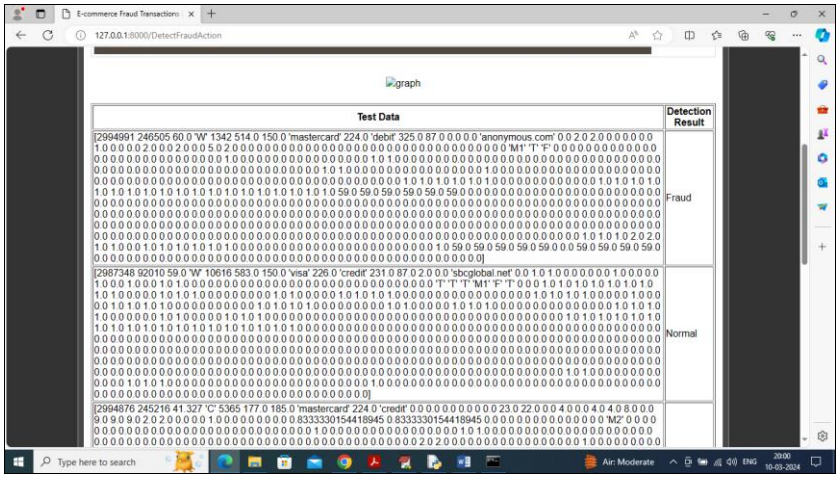
Results and Discussion



Click the "User Login" option in the upper screen to access the page below.



The proposed and extension techniques are shown within table and graph style on the screen above. Both algorithms show that the Random Forest extension gives excellent accuracy. Click the "Detect Fraud" link to access the page below.



The first column of the screen above shows test data from user behaviour; the second column shows the transaction status as either normal or fraudulent.

Conclusion

Using a combination of formal process models and dynamic user behaviors, this article provided a hybrid method to fraud event capturing. The e-commerce purchase process was examined from five main angles: data opinions, command flow, thing, time, and user behavior patterns. This study built a support vector machine (SVM) model to predict suspicious user activities and identify fraudulent purchases by building on top of high-level Petri nets for process modeling. Our comprehensive testing proved that the suggested strategy effectively identifies fraudulent transactions and patterns. In terms of overall index, our suggested multi-perspective detection approach was superior to the single-perspective method. In order to make the suggested framework more accurate, we want to include model validation and relevant deep learning techniques. Future studies will focus on incorporating more temporal elements into the behavioral patterns in order to improve the accuracy of risk detection. Also, by bridging the computational models, we'll look at making a shared library of fraud scenarios and extend the suggested approach to other harmful areas of behavior.

References

1. Kuscu RA, Circumcise Y, Bookly U. Electronic payment systems in electronic commerce. Turkey: IGI Global; c2020. p. 114-139.
2. Abdulrahim M, Elsayed A. The effect of COVID-19 spread on the e-commerce market: The case of the 5 largest e-commerce companies in the world. Available at SSRN 3621166; c2020. DOI: 10.2139/ssrn.3621166.
3. Rao P, et al. The e-commerce supply chain and environmental sustainability: An empirical investigation on the online retail sector. Cogent Bus Manag. 2021;8(1):1938377.
4. Dhobi SD, Tighare KK, Dake SS. A review on prevention of fraud in electronic payment gateway using secret code. Int J Res Eng Sci Manag. 2020;3(1):602-606.
5. Abdallah A, Marof MA, Zainal A. Fraud detection system: A survey. J Newt Compute Appl. 2016;68:90-113.
6. Ministering EA, Manita G. An analysis of the most used machine learning algorithms for online fraud detection. Info Econ. 2019;23(1).
7. Niu X, Wang L, Yang X. A comparison study of credit card fraud detection: Supervised versus unsupervised. arXiv preprint arXiv:1904.10604; c2019. DOI: 10.48550/arXiv.1904.10604.
8. Zheng L, et al. Transaction fraud detection based on total order relation and behaviour diversity. IEEE Trans Comput Soc Syst. 2018;5(3):796-806.
9. Li Z, Liu G, Jiang C. Deep representation learning with full censor loss for credit card fraud detection. IEEE Trans Comput Soc Syst. 2020;7(2):569-579.
10. Mary IM, Priyadarshini M. Online transaction fraud detection system. In: 2021 Int Conf Adv C Innov Tech Engr (ICACITE); c2021. p. 14-16.
11. Choi D, Lee K. Machine learning based approach to financial fraud detection process in mobile payment system. IT Conv P (INPRA). 2017;5(4):12-24.
12. Sarno R, et al. Hybrid association rule learning and process mining for fraud detection. IAENG Int J Comput Sci. 2015;42(2).
13. Stoop JJ. Process mining and fraud detection—A case study on the theoretical and practical value of using process mining for the detection of fraudulent behaviour in the procurement process. M.S. thesis. Netherlands: University of Twente; c2012.
14. Jans M, et al. A business process mining application for internal transaction fraud mitigation. Expert Syst Appl. 2011;38(10):13351-13359.
15. Rinner C, et al. Process mining and conformance checking of long running processes in the context of melanoma surveillance. Int J Env Res Public Health. 2018;15(12):2809.
16. Asare E, Wang L, Fang X. Conformance checking: Workflow of hospitals and workflow of open-source EMRs. IEEE Access. 2020;8:139546-139566.