E-Commerce Fraud Detection Based on Machine Learning Techniques

Authors: Ayush kakde[1], Ayush nagpure[2], Priyanshu gedam[3], Vikram ade[4], Janhavi Thak[5], Prof. Minakshi Getkar[6]

[1,2,3,4,5] Students of BTech, Computer Science, Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, INDIA

[6] Professor, Department of Computer Science, Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, INDIA

Abstract

The e-commerce industry's rapid growth, accelerated by the COVID-19 pandemic, has led to an alarming increase in digital fraud and associated losses. To establish a healthy e-commerce ecosystem, robust cyber security and anti-fraud measures are crucial. However, research on fraud detection systems has struggled to keep pace due to limited real-world datasets. Advances in artificial intelligence, Machine Learning (ML), and cloud computing have revitalized research and applications in this domain. While ML and data mining techniques are popular in fraud detection, specific reviews focusing on their application in e-commerce platforms like eBay and Facebook are lacking depth. Existing reviews provide broad overviews but fail to grasp the intricacies of ML algorithms in the e-commerce context. To bridge this gap, our study conducts a systematic literature review using the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) methodology. We aim to explore the effectiveness of these techniques in fraud detection within digital marketplaces and the broader e-commerce landscape. Understanding the current state of the literature and emerging trends is crucial given the rising fraud incidents and associated costs. Through our investigation, we identify research opportunities and provide insights to industry stakeholders on key ML and data mining techniques for combating e-commerce fraud

Keywords: *E-commerce; fraud detection; Machine Learning (ML); systematic review; organized retail fraud*

1 INTRODUCTION

According to a recent analysis by Juniper Research, losses related to online payments on e-commerce platforms are growing at a staggering rate of 18 percent annually. This highlights the critical importance of studying this area to inform fraud detection or prevention strategies to slow down the upward trend. Frequently, current strategies are unable to keep up with fraudsters, who are constantly adapting and changing their methods to exploit the platforms. What is more, low research and development efforts fueled by a lack of practical data and the need for businesses to protect their platform vulnerabilities further exacerbate the issue. For example, it makes no sense to describe fraud detection or prevention methods in the open since doing so would arm fraudsters with the knowledge they need to avoid detection.
In literature, addressing fraud of any kind can take two forms: (1) Prevention, which refers to steps taken
to avert the occurrence of the acts in the first place. This includes intricate designs, personal identity numbers, internet security for online interactions with
digital platforms, and passwords and authentication mechanisms for computers and mobile devices. Prevention techniques are not perfect; frequently, a trade-off between cost (for the business) and discomfort (for the customer) must be made. (2) On the other hand, detection entails recognizing fraudulent acts as soon as they occur. When prevention fails, detection becomes material. For example, we can prevent credit card fraud by protecting our cards insidiously, but if the card information is stolen, we must notice the fraud as soon as possible[8]. Since neither form above is perfect in reducing the risks and effects of fraud, production systems often consider a combination of the two to combat fraud. In this review, we limit our focus to detection systems.

There are two schools of thought regarding fraud detection systems. The first is in favor of statistical and computational methods, and researchers in this area include Refs. [6−8]. To identify fraud, this way of thinking applies statistical tools, including ML algorithms. Typically, labeled data are used to train classifiers to distinguish between the two classes (fraudulent and non-fraudulent). This implementation
feeds classifiers information from user profiles, including transaction values, day of the week, item category, age, gender, and geographic location. Those who argue against statistical and computational methods claim that these features are easy for sophisticated fraudsters to

fabricate[9]. Irani, Pu, and Webb[10, 11] believe that once fraudsters discover that authorities have picked up on their jargon, they can avoid keyword traps by switching to new expressions .Network analysis is advocated by the second school of thought as an alternative approach to creating fraud detection features[9, 12]. In order to derive graph-theoretical variables or scores that specifically characterize nodes of fraud, the concept makes use of the connectedness between the nodes, which are often users or items in a dataset. The theory underlying identification strategies is that abnormal users display connection patterns that are different from those of typical users. In our review, we focus on the first school of thought.

1.1 Background

E-commerce platforms have intricate design architectures and multiple points of vulnerability (explored later in Section 4), which fraudsters and attackers could use against them. In Figs. 1 and 2, we
illustrate a commonly used e-commerce/marketplace

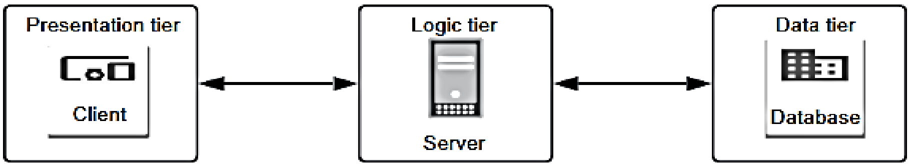

Fig. 1 High-level diagram of an e-commerce platform design architecture.

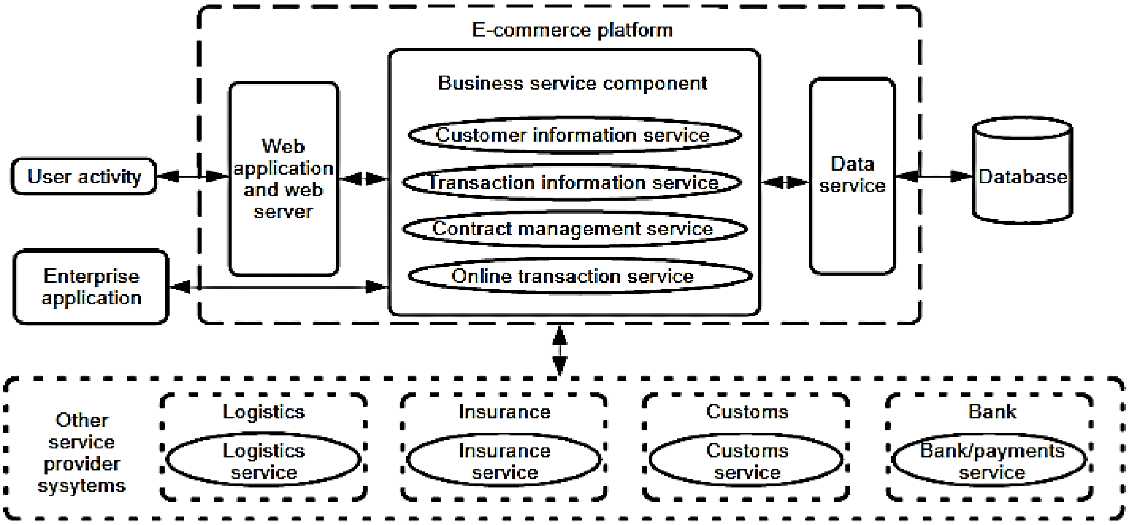
E-commerce platform
User activity
Enterprise application
Web application and web server
Business service component
Customer information service
Transaction information service
Contract management service
Online transaction service
Data service
Database
Other service provider sysytems
Logistics
Logistics service
Insurance
Insurance service
Customs
Customs service
Bank
Bank/payments service

Fig. 2 Detailed-level diagram of an e-commerce platform design architecture.

architecture in the industry to illustrate the complexity of these platforms. At a high level, an e-commerce platform comprises three layers, as shown in Fig. 1. (1) The presentation layer, which is the part that is presented to the customer. It is the user interface and communication part of the architecture, where the customer interacts with the website on the front end and the application collects data and processes requests on the back end; (2) The business layer, also known as the application or service layer, uses business logic, a specific set of business rules, to gather and process information. It can also delete, add, or change information in the data layer; (3) The data layer, which is also known as the database layer, is the final layer and is used for storing data and processing requests. In light of this complex design, we posit that the statistical and computational approach (application of ML and data mining techniques) is best suited for combating fraud on these platforms. Figure 2 not only shows the detailed connections between the tiers presented in Fig. 1, but also includes third-party connections that offer ancillary services on the e-commerce platform.

1.2 Problem Statement

Machine learning and data mining techniques have become popular in fraud detection across many domains[13], partly explained by the rapid development of artificial intelligence and the availability of affordable cloud computing technology. A review specifically concentrating on the use of these methods on e-commerce platforms like eBay and Facebook has not been published, though. What we observe is that past reviews frequently use a broad brush to describe all methodologies and domains, for example, reviews by Refs. [6, 14].

Such high-level coverage fails to produce a nuanced understanding of machine learning algorithms and their applications in the e-commerce domain.

On the other hand, the majority of the specific fraud literature reviews, like: Refs. [15–19] only cover the financial domain, such as credit card fraud. What is more, a large number of these articles do not employ systematic literature review methodology to support replication[20]. In this work, we acknowledge these gaps and propose a systematic literature review using the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) methodology[21] to examine the use and application of machine learning and data mining techniques for fraud detection on digital marketplaces or in the e-commerce domain. This is a crucial area given the soaring trends in fraud incidents and their associated costs[22]. Understanding the current literature and trends is essential to identifying new research opportunities as well as informing the industry on the main machine learning and data mining techniques for fraud detection in this area.

2 Related Work

Reviews of general fraud detection have recently been written and published in the literature. A general review of articles on automated detection techniques (supervised, unsupervised, and hybrid) from the previous ten years is published by Unam et al.[23]. The authors of that review formalize the major fraud types and subtypes for a wide range of industries while presenting alternative information and solutions for each. Amir and Hamid[24] conducted yet another general review of articles related to fraud detection. The researchers outline five common fraud types, including credit card fraud, telecom fraud, fraud involving health insurance, fraud involving auto insurance, and fraud involving online auctions. Their work does not employ a systematic review methodology, and the review period is from 1994 to 2014. A few reviews of a particular domain are also included in the literature. Adewumi and Akinyelu[25] used the Kitchenham approach to conduct a systematic review of the financial fraud field between 2010 and 2021. Their focus is on the use of machine learning techniques in the detection of financial fraud. Ahmed et al.'s[18] review of anomaly detection methods for fraud detection is yet another review in the financial domain.

The type of fraud that receives the most reviews is credit card fraud. Reviewing credit card fraud, highlighting misuses of supervised and unsupervised techniques, and offering advice for new researchers are among Sorournejad. et al.'s[26] highlights.

Techniques for data mining are the focus of another group of reviews. For instance, Pourhabibi et al.[15] explored the interdependency between various data objects with a focus on graph-based anomaly detection. Reviewing data mining techniques with an emphasis on machine learning classification methods

Table 1 Related Work

Article	Year	Coverage	Review type	Domain
[23]	2010	2000−2010	Unknown	General fraud
[28]	2016	−	Unknown	Online fraud
[24]	2016	1994−2014	Unknown	General fraud
[18]	2016	−	Unknown	Financial fraud
[26]	2016	−	Unknown	Credit card fraud
[17]	2016	1997−2016	Unknown	Credit card fraud using nature inspired machine learning
[29]	2017	−	Systematic literature review	Credit card fraud using ML
[30]	2018	−	Unknown	General fraud using ML
[30, 31]	2018	−	Unknown	Credit card fraud in e-commerce
[14]	2020	−	Systematic literature review	General fraud with graph-based anomaly detection
[32]	2021	−	Unknown	Credit card fraud with ML
[33]	2021	−	Systematic literature review	E-commerce
[34]	2021	−	Unknown	E-commerce
[35]	2021	−	Systematic literature review	E-commerce fake reviews
[36]	2021	−	Unknown	Credit card fraud
[20]	2022	−	Systematic literature review	e-commerce (detection and prevention)
[13]	2022	−	Systematic literature review	Financial fraud (Machine learning)

In Table 1, we provide a list of the articles we consider related to our work. We develop this list by instantiating our search based on three well-known articles in this fraud domain[6, 8, 37] and snowballing to similar articles. We prioritize the list on the basis that an article covers fraud in e-commerce or a related domain.

3 Research Method

We adopt the PRISMA approach[21] to search and select articles in the scope of fraud detection in ecommerce or digital marketplaces based on machine learning or data mining techniques. The PRISMA approach generates high-quality results and supports reproducibility. It is structured in a manner that allows the identification and summarization of problems (domains), techniques, and methods used to solve the problem. The implementation of this approach follows a checklist of title, abstract, introduction, methods, results, discussion, and funding. In this structure, the title and abstract are constructed to achieve comparable objectives to any other approach, but the introduction must provide the rationale for the review and the questions to be addressed. Study characteristics, information sources, search strategy, including limits, statement process for selected studies, eligibility criteria, data collection, and data items are specified in the methods section[21]. The discussion involves a summary of the findings, a discussion of the limitations, and a general conclusion of the results and future work

3.1 Research questions

Understanding the literature on the use of machine learning and data mining techniques for fraud detection on e-commerce or digital marketplace platforms is the primary goal of this research. Our Research Question three (RQ3) ultimately encapsulates this, but in order to accomplish this successfully, we first use Research Questions one and two (RQ1 and RQ2) to establish the context. These inquiries help us understand the design architecture of e-commerce platforms and contextualize major vulnerabilities discovered therein as well as related frauds. Finding research gaps, trends, and opportunities for further research in the field is the goal of our last research question. Below, we list our research questions.

- RQ1: What are the common frauds in the marketplace or e-commerce domain?
- RQ2: What are the commonly used machine learning and data mining techniques for fraud detection on digital marketplaces or e-commerce platforms, and what does good performance of these techniques look like?
- RQ3: What are the research gaps, trends, and opportunities for future research in this area?

3.2 Data and search strategy

By extracting potential search terms from the titles, abstracts, and subject indexing of three pertinent publications[17, 23, 24], we develop an initial search strategy. We use its results to expand the list of key words and restrict it to only English-language articles in order to further hone this strategy. We then test the validity of our search strategy by checking whether it could retrieve the three known relevant studies and two more studies referenced in Ref. [17]. All the five studies are successfully identified by the strategy. A group of peer reviewers approves the final search strategy.

Using an iterative search approach, we look for publications within our search period (2010-2023) that have the following keywords in their title or abstract: e-commerce, fraud detection, machine learning, systematic review, organized retail fraud, data mining, and digital marketplace. We display the iterative approach in the workflow diagram shown in Fig. 3. To reduce the amount of noise in the results, our search strategy employs the search logics “AND”, “OR”, “LIMIT TO”, and “EXCLUDE”.

3.3 Publications Repositories

We focus our search on three international digital repositories: Scopus, Web of Science (WoS), and Google Scholar, which together hold the majority of global scientific research. The initial search query in each repository yields a wide range of publications in a multidisciplinary setting covering, among other things, computer science, engineering, decision science, mathematics, energy, physics, and astronomy. We approach our search with the knowledge that the coverage, accuracy, and access fees of these digital repositories vary. For instance, Scopus and Web of Science overlap in two out of three instances[39], with Scopus offering 20 percent more coverage than Web of Science[40]. Depending on the search terms, Google Scholar frequently provides inaccurate and inconsistent

results. Additionally, many of its articles are subpar and out of date (Falagas et al.[40]). Therefore, it helps to think of a way to minimize noise and duplicates in the combined search results. To this end, we apply the inclusion and exclusion criteria defined in Table 2 to meet that nee

3.4 PRISMA flow diagram

We use the flow diagram shown in Fig. 4 to illustrate how we apply our inclusion and exclusion criteria to narrow down the most relevant articles for our literature

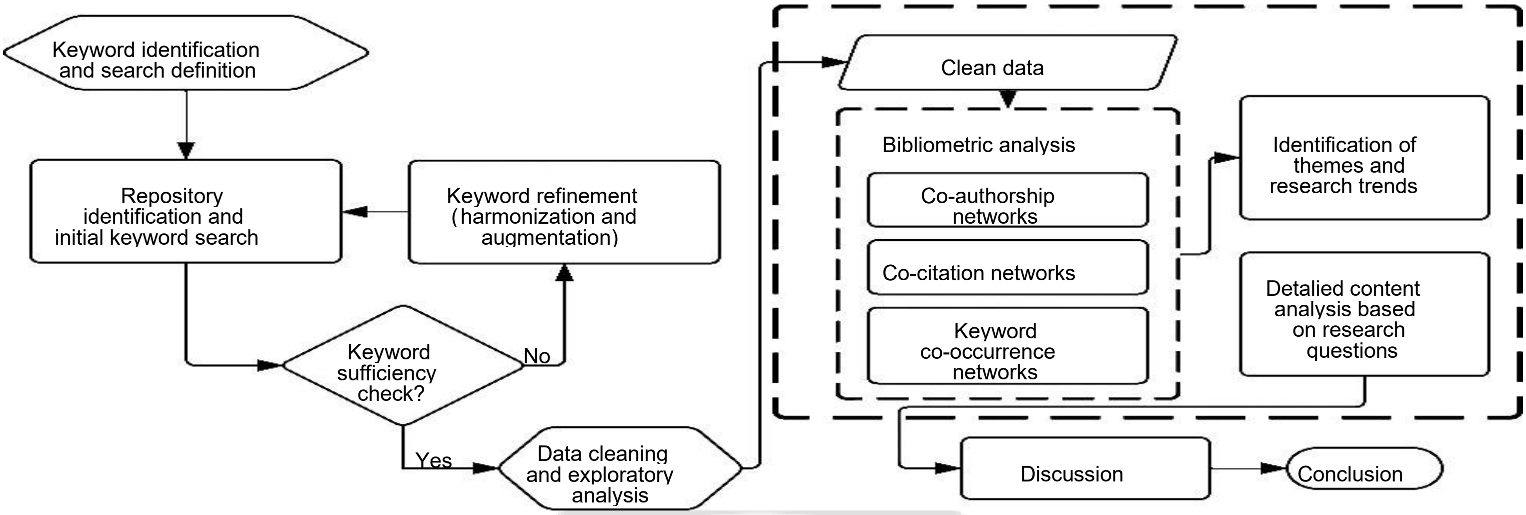
Keyword identification and search definition
Repository identification and initial keyword search
Keyword refinement (harmonization and augmentation)
Keyword sufficiency check?
No
Yes
Data cleaning and exploratory analysis
Clean data
Bibliometric analysis
Co-authorship networks
Co-citation networks
Keyword co-occurrence networks
Identification of themes and research trends
Detalied content analysis based on research questions
Discussion
Conclusion

Fig. 3 Iterative search strategy and SLR process workflow diagram.

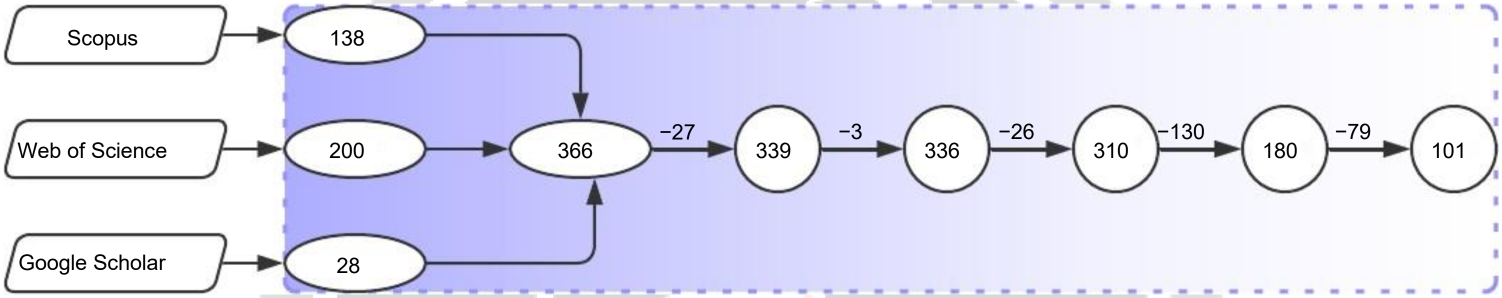

Fig. 4 PRISMA flow diagram showing detailed filtering levels from a high-level representation of publications from initial

search. Three hundred and sixty-six articles total in the combined search results are reduced to three hundred and thirty-five after duplicates are eliminated. The first step of our exclusion criteria is when EF1 eliminates three papers written in a language other than English. Our exclusion criterion, EF2, eliminates twenty-six papers in the second step that come from interdisciplinary fields like medicine. EF3 and EF4 eliminate a combined total of two hundred and nine publications, leaving us with one hundred and onepapers for our final corpus

4 . Detailed Analysis and Results from Corpus

To address each research question posed in Section 3.1, we present the findings of our analysis of the literature corpus in this section.

4.1 RQ1: What are the most common e-commerce frauds?

On e-commerce platforms, fraudsters use vulnerabilities known to them to wedge attacks and commit fraud. Once the weaknesses are clearly understood, countermeasures can be created to lessen the risk of fraud and combat its effects. In this question, we use our corpus to highlight significant ecommerce frauds and the solutions researchers have
suggested. Our corpus reveals five types of fraud that can be thwarted using machine learning and data mining techniques. These include financial or payment frauds, web application frauds, spam or phishing frauds, triangulation frauds, and bot frauds.

4.1.1 Financial frauds or payment frauds

This type of fraud is the most prevalent on e-commerce platforms and has existed since the beginning of businesses' shift from physical to online locations. Using financial or payment information obtained through the exploitation of the aforementioned vulnerabilities, fraudsters frequently carry out unauthorized transactions. In our work, we do not address the architecture of the online payment process or the classification of the sub-fraud types under financial frauds

Using financial or payment information obtained through the exploitation of the aforementioned vulnerabilities, fraudsters frequently carry out unauthorized transactions. In our work, we do not address the architecture of the online payment process or the classification of the sub-fraud types under financial frauds

4.1.2 Spam/Phishing fraud

Phishing is a type of fraud to gain access to a user's credentials to defraud the user or connected services[59], for example, e-commerce platforms and online merchants. There are numerous tools used by phishers to lure users into traps (unsecure sites) where their sensitive information, like e-commerce account login credentials, payment passwords, home addresses, and birthdays, among others, is exposed. Emails and fake websites are good examples of such tools. Emails are a key marketing channel for e-commerce platforms, and fraudsters can exploit them to obtain customer details. Spam emails with links to fake e-commerce platforms and products are commonplace. We find seven articles in our literature corpus focused on spam or phishing detection, making it the third highest ranked category after financial and web application frauds. These articles are almost evenly split, with three of them looking at data mining techniques and the rest applying machine learning methods.

4.1.3 Triangulation fraud

It is an emerging fraud type on e-commerce platforms that occurs when a customer makes a genuine purchase on an e-commerce platform, but the seller (fake) fraudulently purchases the product from another merchant. First, the fraudster sets up operations as a third-party seller on the marketplace site, for example, eBay. The criminal then lists products for sale at unusually low prices. When a cardholder makes a purchase, the fraudster then turns around and buys the goods from a legitimate seller using stolen card information. The fraudster sets the shipping address to match that of the customer, and therefore the legitimate merchant ships the product to the buyer. The fraudster pockets money from the original sale, while the legitimate merchant gets paid with a stolen payment card. Eventually, the buyer requests a chargeback to their card when they notice an unauthorized transaction, leaving the legitimate merchant defrauded and with legal ramifications.

Total	Algorithm	Article
14	Logistic regression	[45, 51, 61, 67−75]
17	Decision tree	[44−46, 61, 69, 70 , 74 , 76−84]
21	Random forest	[45, 63, 69, 81, 85−88]
10	Naïve Bayes	[45, 60, 62, 77, 81−83, 89, 90]
13	SVM	[50, 51, 60, 68, 69, 74 , 79, 82, 86, 91−94]
30	ANN	[44, 45, 48, 54, 59, 60, 68, 79, 86, 95−112]
3	K-nearest neighbor	[69, 76]
12	Boosting algorithm	[22, 53, 65, 85, 103, 113−119]
16	Others	[43, 62, 64, 66, 89, 120−127]

Table 2 Methods used to detect e-commerce frauds through the years.

4.2 RQ2: What are the commonly used machine learning and data mining techniques for fraud detection on digital marketplaces or e-commerce platforms, and what does good performance of these techniques look like?

This is the most important question for our review, and we use the previous questions to set the scene for it. In this context, we focus on machine learning and data mining applications for tackling fraud detection in the e-commerce domain. This implies that we do not look at other methods like statistical inference techniques, ontologies, or even bespoke algorithms that could be relevant in the domain. One more thing to note is that we only focus on detection methods.

There are many algorithms applied in these articles, and therefore we only consider those that are used in more than two articles. In Table below, we show the evolution and frequency of use of the algorithms from the corpus over the years and The results also show that the use of ANNs gained more traction in e-commerce fraud detection around 2019.

The second largest category of algorithms is the Random Forests algorithm, which features in about 21 articles in our literature corpus. Many articles conduct performance tests in experimental settings where, for each data set, several algorithms are jointly tested. In these cases, the authors report the Random Forest and the ANNs as the best performers[44, 60, 133]. Decision trees and logistic regression are some of the other notable algorithms.

Our search strategy exposes a couple of data mining strategies for e-commerce fraud detection in addition to common machine learning methods. The algorithms discussed in these strategies do not appear to be clustered, so we collectively refer to them as the "other" category. There are roughly 16 articles in this category, and they primarily cover three types of fraud: web application fraud, credit card fraud, and phishing.

4.3 RQ3: What are the research gaps, trends, and opportunities for future research in this area?

In this question, we show research gaps to inform future research directions. We synthesize all the articles in the final corpus to understand how the articles apply machine learning and data mining techniques for e-commerce fraud detection and to surface gaps in their usage. We cover bespoke gaps in the following subsections.

4.3.1 Class asymmetry

The issue of imbalanced classes between fraudulent and legitimate transactions is rife in fraud data. It occurs when there is an asymmetric distribution between classes in the data. In the machine learning domain, most algorithms do not perform well on imbalanced data, as the minority class contributes less to the learning objective[172].

In training an imbalanced data set with a standard classification method, the minority class contributes less towards the minimization of the objective function[173], leading to lower classification accuracy for the minority class and poor performance of the classifier as a whole. For example, a binary classifier that achieves 99 percent training accuracy on imbalanced data with 1 percent minority samples would be irrelevant for predictions on out-of-sample data. In this case, the classifier is only accurate at predicting the majority, while its performance on the minority class is poor (often, all the instances of the minority class are misclassified as instances of the majority class). This is a costly decision because, in most practical applications, classifying the minority instances correctly is more important[173]. Therefore, it is of paramount importance to improve a classifier's ability to recognize the minority class in these settings

4.3.2 Training data

One criticism of machine learning and data mining for fraud detection is the lack of good practical data to use for training algorithms[23]. Real fraud data often carry sensitive information about consumers, and as such, companies are constrained by data protection laws from sharing such data. Additionally, it is counterintuitive to openly share data and fraud detection strategies, as fraudsters can use that information to escape detection systems. These challenges make it hard to advance fraud detection research in general. We observe minimal use of realworld fraud data in our corpus, and in those few cases, the actual details of features used for training the detection algorithms are hardly mentioned.

Some of the articles using real-world data include, which uses a real dataset from one of Egypt's top e-payment gateways,, which tests their fraud detection system with real-world data from European banks' day-to-day transaction data. The implication of the lack of real-world data is that the majority of the research articles in this domain are experimental and likely will not result in real-world fraud detection systems. Future work could look for sandbox environments that can allow fraud researchers to work with real-world fraud data to advance the field.

4.3.3 Detection algorithms

The use of ANNs to create fraud detection systems in the e-commerce fraud domain is a clear trend in our data. More than 30 percent of all articles use ANNs as their primary learning technique. ANNs use data and information processing techniques inspired by biological neural network behavior, and they are powerful when used on big data[154]. This explains their popularity in credit card fraud detection, where they can be trained using massive amounts of high-velocity transaction data. This trend is reflected in our data set, in which nearly 60 percent of all articles using ANNs are geared towards credit card fraud detection. Despite being widely used and achieving good discriminatory performance, these techniques lack interpretability, making it difficult for researchers and practitioners to comprehend the signals that lead to fraud. As a result, their use necessitates a conscious decision to optimize performance as opposed to deciphering the underlying indicators associated with fraudulent instances. Future research and applications can put their lack of interpretability into design considerations. We also observe that the random forest is highly featured in our data. It achieves high performance and is interpretable. It is possible to tease out the importance of features' contributions towards the minimization of the objective function. As such, researchers and practitioners can glean from features highly associated with fraudulent instances. In summary, ANNs and the Random Forest algorithm provide a healthy trade-off between performance

Conclusion

In this article, we employed a combined PRISMA and content synthesis approach to identify and analyze relevant articles focusing on fraud detection in the ecommerce domain using machine learning and data mining techniques.

To structure our analysis, we formulated four research questions, with the first two providing context for our main question. Among the machine learning algorithms utilized, ANNs emerged as the most frequently employed, closely followed by random forest. Notably, the majority of articles centered around the detection of credit card fraud, showcasing its prevalence in the field. However, we found a dearth of detailed research addressing reseller fraud, also known as product flipping or scalping, within our corpus, highlighting a potential avenue for future investigation

given its significance in the e-commerce domain and potential impact on the economy and households. Further exploration of various techniques, including machine learning, to combat reseller fraud could be a fruitful area for future work.

Our review also shed light on emerging fraud types in e-commerce, namely triangulation and bot fraud, which have received limited attention in the realm of machine learning and data mining techniques. This observation underscores the need for further research to address these novel fraud types effectively.

Furthermore, our analysis revealed a growing demand for the application of imbalanced learning techniques to enhance future fraud detection systems. This indicates an opportunity for the concerted use of such techniques to tackle the challenge posed by imbalanced datasets in fraud detection.

The findings of our work have practical implications for practitioners in the e-commerce industry. They can replicate the approaches discussed in our corpus and implement them to proactively identify and eliminate malicious actors from their platforms, thereby reducing losses and safeguarding their brand reputations. Additionally, our survey contributes to the existing body of knowledge and literature on fraud detection in the e-commerce domain, providing valuable insights for future research endeavors.

Conflict of interest

The authors declare no potential conflict of interest with respect to research, authorship and/or publication of this article.

The authors confirm that the materials included in this chapter do not violate copyright laws. Where relevant, appropriate permissions have been obtained from the original copyright holder(s), and all original sources have been appropriately acknowledged or referenced.

References

[1] S. Monteith, M. Bauer, M. Alda, J. Geddes, P. C. Whybrow, and T. Glenn, Increasing cybercrime since the pandemic: Concerns for psychiatry, Curr. Psychiatry Rep., vol. 23, no. 4, p. 18, 2021.

[2] R. Samani and G. Davis, McAfee mobile threat report, https://www.mcafee.com/enterprise/en-us/assets/reports/rp-mobile-threat-report-2019

[3] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011.

[4] Sam Smith and Juniper Research, Online payment fraud: Market forecasts, emerging threats & segment analysis 2022-2027, https://www.juniperresearch.com

[5] S. Sorournejad, Z. Zojaji, R. E. Atani, and A. H. Monadjemi, A survey of credit card fraud detection techniques: Data and technique oriented perspective, arXiv preprint arXiv: 1611.06439, 2016.

[6] R. J. Bolton and D. J. Hand, Statistical fraud detection: A review, *Statistical Science*, vol. 17, no. 3, pp. 235–255, 2002.

[7] C. Phua, V. Lee, K. Smith, and R. Gayler, A comprehensive survey of data mining-based fraud detection research, arXiv preprint arXiv: 1009.6119, 2010.

[8] L. Akoglu, H. Tong, and D. Koutra, Graph based anomaly detection and description: A survey, *Data Min. Knowl. Discov.*, vol. 29, no. 3, pp. 626–688, 2015,

[9] D. Irani, S. Webb, and C. Pu, Study of static classification of social spam profiles in MySpace, *Proc. Int. AAAI Conf. Web Soc. Med.*, vol. 4, no. 1, pp. 82–89, 2010.

[10] A. Bhowmick and S. M. Hazarika, Machine learning for E-mail spam filtering: Review, techniques and trends, arXiv preprint arXiv: 1606.01042, 2016.

[11] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, Anomaly detection in online social networks, *Soc. Netw.*, vol. 39, pp. 62–70, 2014.

[12] A. Ali, S. Abd Razak, S. H. Othman, T. A. E. Eisa, A. Al-Dhaqm, M. Nasser, T. Elhassan, H. Elshafie, and A. Saif, Financial fraud detection based on machine learning: A systematic literature review, *Appl. Sci.*, vol. 12, no. 19, p. 9637, 2022.

[13] T. Pourhabibi, K. L. Ong, B. H. Kam, and Y. L. Boo, Fraud detection: A systematic literature review of graphbased anomaly detection approaches, *Decis. Support Syst.*, vol. 133, p. 113303, 2020.

[14] R. Banerjee, G. Bourla, S. Chen, M. Kashyap, and S. Purohit, Comparative analysis of machine learning algorithms through credit card fraud detection, in *Proc. IEEE MIT Undergraduate Research Technology Conf.*, Cambridge, MA, USA, 2018, pp. 1–4.

[15] N. Carneiro, G. Figueira, and M. Costa, A data mining based system for credit-card fraud detection in e-tail, *Decis. Support Syst.*, vol. 95, pp. 91–101, 2017.

[16] A. O. Adewumi and A. A. Akinyelu, A survey of machine-learning and nature-inspired based credit card fraud detection techniques, *Int. J. Syst. Assur. Eng. Manag.*, vol. 8, no. S2, pp. 937–953, 2017.

[17] M. Ahmed, A. N. Mahmood, and M. R. Islam, A survey of anomaly detection techniques in financial domain, *Future Gener. Comput. Syst.*, vol. 55, pp. 278–288, 2016.

[18] V. Rodrigues, L. Policarpo, and D. E. da Silveira, Fraud detection and prevention in e-commerce: A systematic literature review, https://www.sciencedirect.com/science/article/pii/S1567422322000904?casa_token=UOjgVT_FXuwAAAAA:YgIpy5PUX5dEdF_dJ2Nd1Hz-664Vr32oHJPDq_ ZbevxtOazQ38tP_I-PVDtKsCBFXXu_6-Ri6Q, 2022.

[19] J. West and M. Bhattacharya, Intelligent financial fraud detection: A comprehensive review, *Comput. Secur.*, vol. 57, pp. 47–66, 2016