Electronic Commerce Research and Applications 56 (2022) 101207

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/elerap

Fraud detection and prevention in e-commerce: A systematic literature review

Vinicius Facco Rodrigues [a,*], Lucas Micol Policarpo [a], Diórgenes Eugênio da Silveira [a], Rodrigo da Rosa Righi [a], Cristiano André da Costa [a], Jorge Luis Victória Barbosa [a], Rodolfo Stoffel Antunes [a], Rodrigo Scorsatto [b], Tanuj Arcot [b]

[a] Applied Computing Program, Universidade do Vale do Rio dos Sinos, São Leopoldo, RS, Brazil
[b] DELL - Eldorado do Sul, RS, Brazil

ARTICLE INFO

Keywords:
E-commerce
Fraud detection
Fraud prevention
Machine learning
Systematic literature review

ABSTRACT

The high volume of money involved in e-commerce transactions draws the attention of fraudsters, which makes fraud prevention and detection techniques of high importance. Current surveys and reviews on fraud systems focuses mainly on financial-specific domains or general areas, leaving e-commerce aside. In this context, this article presents a systematic literature review on fraud detection and prevention for e-commerce systems. Our methodology involved searching for articles published in the last six years into four different literature databases. The search of articles employs a search string composed of the following keywords: purchase, buy, transactions, fraud prevention, fraud detection, e-commerce, web commerce, online store, real-time, and stream. We apply six filtering criteria to remove irrelevant articles. The methodology resulted in 64 articles, which we carefully analyzed to answer five research questions. Our contribution appears in the updated perception of fraud types, computational methods for fraud detection and prevention, as well as the employed domains. To the best of our knowledge, this is the first survey on combining prevention and detection of e-commerce frauds, linking also architectural insights, artificial intelligence methods, and open challenges and gaps in the research area. The study main findings demonstrate that from 64 articles, only five focus on the fraud prevention problem, and credit card fraud is the most explored fraud type. In addition, current literature lacks studies that propose strategies for detecting fraudsters and automated bots in real-time.

1. Introduction

Ecommerce, also known as electronic commerce, is the buying and selling of goods or services via the internet. The digital transformation of traditional business models enabled shoppers to explore and buy products online in the comfort of their houses (Song et al., 2021). E-commerce platforms offer several advantages, including faster buying process, cost reduction, flexibility for customers, product and price comparison, faster response to buyer/market demands, and several payment modes. Besides all advantages, e-commerce is essential to the economy (Liu et al., 2021) and even more critical in pandemic times where governments impose stay-at-home and lockdown regulations to lower the circulation of people (Tran, 2021). Such situations lead to increased demand for online shopping, including products and daily necessities like food and pharmacy. Recently, a study in France showed an increase in online orders up to 35.4% during the lockdown in 2020 compared to the same month in 2019 (Guthrie et al., 2021). Additionally, according to eMarketer,[1] worldwide e-commerce sales increased 27.6% in 2020 and will increase 14.3% in 2021, reaching almost $5 trillion.

Such a high volume of money draws the attention of fraudsters which can cause tremendous loss of money. A recent study from Juniper Research[2] estimates that fraud will rise from $17.5 billion in 2020 to $20 billion by 2021. Those numbers demonstrate the importance of e-commerce and financial platforms to implement fraud detection and prevention systems to avoid financial losses. Currently, several research studies investigate different strategies to mitigate the fraud problem in

* Corresponding author.
E-mail addresses: vfrodrigues@unisinos.br (V.F. Rodrigues), lmpolicarpo@unisinos.br (L.M. Policarpo), diorgeneses@unisinos.br (D.E. da Silveira), rrrighi@unisinos.br (R. da Rosa Righi), cac@unisinos.br (C.A. da Costa), jbarbosa@unisinos.br (J.L.V. Barbosa), rsantunes@unisinos.br (R.S. Antunes), rodrigo.scorsatto@dell.com (R. Scorsatto), tanuj.arcot@dell.com (T. Arcot).

[1] https://www.emarketer.com/content/worldwide-ecommerce-will-approach-5-trillion-this-year.
[2] https://www.juniperresearch.com/press/ecommerce-losses-online-payment-fraud-exceed-20bn?ch=e-commerce-losses-online-payment-fraud-exceed-20b.

https://doi.org/10.1016/j.elerap.2022.101207
Received 27 August 2021; Received in revised form 14 April 2022; Accepted 3 October 2022
Available online 13 October 2022

e-commerce (Aziz and Ghous, 2021; Sudha and Akila, 2021; Baesens et al., 2021; Forough and Momtazi, 2021).

In general, fraud systems rely on customer data analysis, including online navigation, past activity, and behavior. Such techniques apply data engineering and data mining methods to raw events from running systems, such as log information (Baesens et al., 2021). Strategies in the field of machine learning (ML) employ data extraction and pattern analysis in classification methods aiming at predicting fraud behavior. Classification algorithms are a branch of ML that aims at identifying the category of a given observation based on a training dataset. For instance, Random Forests (Sudha and Akila, 2021), Logistic Regression (Baesens et al., 2021), and Artificial Neural Networks (Forough and Momtazi, 2021) are general classification algorithms applied for fraud detection and prevention. Those algorithms are fed with previous fraudulent observations to learn fraudulent patterns. Then, when they receive new input, such as a new customer on the website, they can classify that new client based on their similarity with the previous ones. Although there are many techniques, designing a fraud detection and prevention system requires rigorous data analysis and faster data availability. Quick availability allows algorithms to process data logs and predict fraud risks earlier in the online transaction process.

Fraud detection and prevention seem synonym, but they refer to different concepts (Abdallah et al., 2016). Fraud prevention is the process of identifying fraudulent activity and stopping the occurrence of fraud in the first place. On the other hand, fraud detection is the process of identifying a fraud only after it has already occurred. In other words, once a fraud took place, prevention is not possible anymore but only detection to take measures to mitigate its effects. Fig. 1 depicts the e-commerce online transaction flow highlighting where fraud prevention and detection can be applied. In particular, prevention can be achieved in both e-commerce and financial platforms before the banking authorization. As soon as a banking authorization is emitted, prevention is not possible anymore but only detection instead.

Moreover, customer behavior is an essential feature to consider when trying to mitigate fraudulent acts. Fraudulent behavior is a progressive set of actions that lead to a fraudulent transaction, and the reasoning behind that is hard to address. Strategies such as using Dempster–Shafer theory (Zhao et al., 2016) or ontologies (El Orche et al., 2018; El Orche and Bahaj, 2019) try to understand the thresholds that point to a customer as a fraud. The behavior is assessed by evaluating the correlation and statistical inference between the current order and previous fraudulent ones. Some of the most notable features involve the number of purchases of the user, the amount of each purchase, the number of tested credit cards, and the navigation pattern. However, those techniques have difficulties dealing with changes in the strategies of the attacks. Since the attackers also change their patterns, assessing fraud behavior is a tough task because the wrong evaluation of fraud intention may disrupt the standard user purchase experience. We summarize some of the known behaviors of a fraudster in Fig. 2. The fraud behavior starts in the e-commerce website navigation process. Normal users tend to explore the website and spend more time on each page. On the other hand, fraudsters go straight to the steps of the purchase and are fasters compared to the others. When a purchase is submitted, the payment details are crucial for identifying the user. Mainly, fraudsters have unmatching personal information in the order, temporary e-mails, and a discrepancy between billing and delivery location. Lastly, fraud behavior can be seen in the number of purchases and tryouts. Fraudsters place multiple orders with different credit cards.

Understanding the current literature and trends is essential to identify gaps and research opportunities. Survey articles about frauds in e-commerce review the literature either in financial specific domains or in a general area (Aziz and Ghous, 2021; Pourhabibi et al., 2020; Zhao et al., 2019; Chilaka et al., 2019; Adewumi and Akinyelu, 2017; Sorournejad et al., 2016; Ahmed et al., 2016; Abdallah et al., 2016). In such cases, the studies employ financial datasets, which are limited since banking institutions do not have access to all possible information

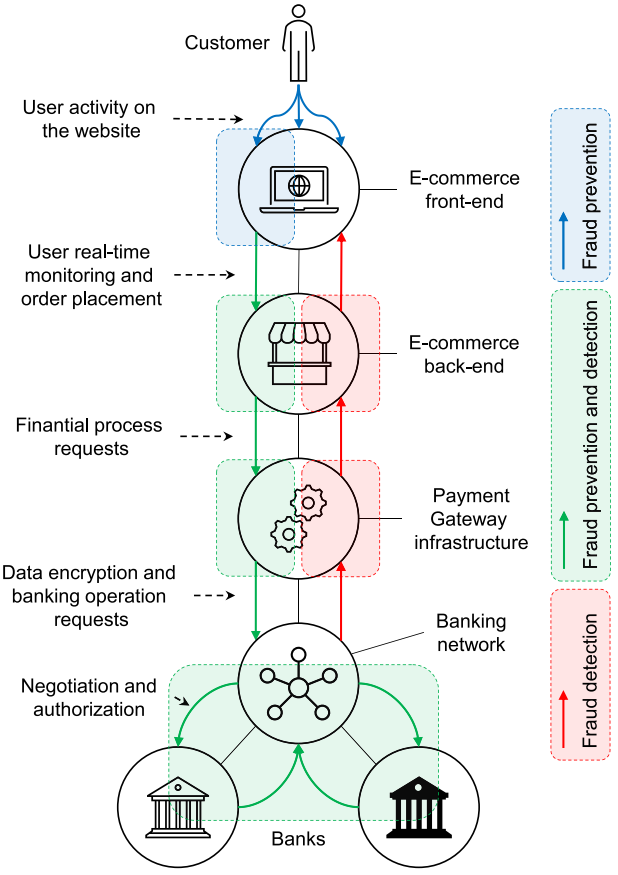

Fig. 1. E-commerce online transaction flow highlighting where fraud prevention and detection can be applied.

from the transaction. In addition, there is a lack of systematic literature reviews presenting a taxonomy of e-commerce fraud. Besides these limitations, most of the articles do not employ systematic literature reviews to provide reproducibility. Acknowledging these gaps, this article proposes a systematic literature review on fraud detection and prevention techniques for e-commerce.

This article aims at giving the reader a comprehensive vision of the main techniques, methods, datasets, and domains present in literature regarding fraud detection and prevention in e-commerce. To accomplish this, the article presents five research questions that are answered based on a literature corpus formed by 64 articles resulting from the search methodology. Our main contributions to the state-of-the-art are as follows:

(*i*) We provide an updated panorama of methods and datasets employed for fraud detection and prevention on research studies in the last six years;
(*ii*) We propose a taxonomy of e-commerce fraud classifying the fraud types, domains, and the common methods for both detection and prevention;
(*iii*) We propose an architecture to give directions on designing a fraud detection and prevention system in e-commerce.

The following sections are organized as follows. Section 2 discusses related work approaching surveys on fraud studies. Section 3 introduces the research methodology employed to gather research articles from literature. Section 4 analyzes the literature corpus based on the research questions defined in Section 3. Next, Section 5 discusses the main findings and some open challenges. Then, Section 6 presents the limitations of the current article. Finally, Section 7 concludes the document with some final remarks.

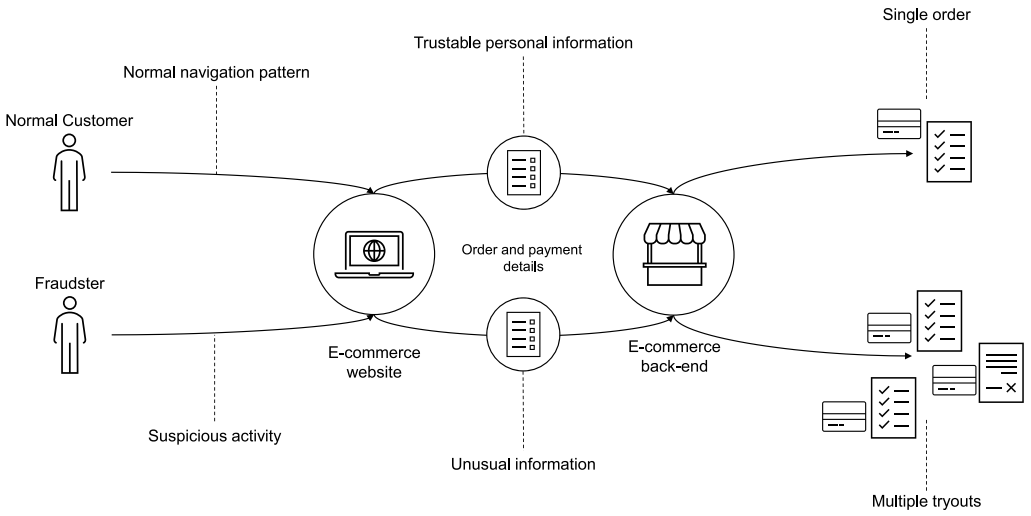

Fig. 2. Comparison of suspicious and normal behavior at the time of purchase in e-commerce.

2. Related work

In the last few years, several articles perform surveys and reviews of fraud detection and prevention strategies existent in literature. Table 1 summarizes eight articles published in the last six years. In the first article of the table, Abdallah et al. (2016) present a broad overview of types, approaches, techniques, issues, and challenges on fraud detection systems. The survey covers the time range between 1994 and 2014, not presenting a transparent methodology. The authors review fraud detection solutions in five different areas: credit card, online auction, telecommunication, healthcare insurance, and automobile insurance. However, they present just four articles that focus on online transactions in the time range 2008–2011. Focusing on a specific area, Ahmed et al. (2016) review unsupervised anomaly detection techniques in the financial domain. The article focuses on clustering techniques for fraud detection. They also discuss the availability of real datasets for modeling clustering strategies.

Credit card fraud detection is a common topic some studies address. For instance, Sorournejad et al. (2016) review credit card fraud detection techniques into two categories: supervised or unsupervised. They present a taxonomy detailing the different techniques found in the literature with a focus on the two categories. Besides, the article analyzes the datasets employed in each literature strategy. The authors present no methodology in the article. In turn, Adewumi and Akinyelu (2017) survey online credit card fraud detection employing nature-inspired and ML techniques. The authors do not detail the survey methodology that resulted in articles from 1997 to 2016. However, they summarize the leading techniques and algorithms used between 2010 and 2015. Moreover, Chilaka et al. (2019) survey credit card fraud detection techniques in the electronic finance and banking domain. The authors selected articles from 2014 to 2019 and detailed the strategies they employ. They focus on solutions aiming at fast response time and efficiency. The article does not follow a systematic literature review and does not present any taxonomy.

Some studies focus on user behavior, anomaly detection, or data mining techniques, unlike previous articles. Zhao et al. (2019) explore studies on consistency analysis of user's behavior in e-commerce. The authors conduct an extensive overview of articles published between 1966 and 2018. They discuss behavior pattern mining strategies for online trading systems. The article analyzes 393 studies with three points of view: system behavior pattern mining, system behavior analysis, and user behavior analysis. Pourhabibi et al. (2020) review graph-based anomaly detection methods with a focus on the interdependency between different data objects. The article proposes a framework for the classification of studies according to the techniques they employ. As the authors do not target a specific area, the research returned articles focusing on different domains. Finally, Aziz and Ghous (2021) overview data mining methods for credit card fraud detection using datasets. The authors focus on ML techniques for the classification of transactions. The article does not present a review methodology and taxonomy for the concepts presented.

According to Table 1, we did not find any articles focusing specifically on e-commerce strategies for fraud detection. In summary, articles focus on reviewing the literature either in financial specific domains or in a general area. In such cases, the studies employ financial datasets, which are limited since banking institutions do not have access to all possible information from the transaction. On the e-commerce side, they can have a broader range of transaction attributes, including browsing information and personal user data. Besides these limitations in the reviews and surveys, most of them do not employ systematic literature reviews to provide reproducibility. None of the articles provide a systematic literature review presenting a taxonomy of e-commerce fraud.

3. Research methodology

This section introduces the methodology followed in this study to search and select articles in literature in the scope of fraud detection in e-commerce. This article adopts the principles of systematic literature reviews (Biolchini et al., 2005; Kitchenham and Charters, 2007) to achieve high-quality results and, mainly, reproducibility. The selected approach allows the summarization and identification of the problems, technologies, and existent methods applied to detect and prevent frauds in e-commerce platforms. Also, the review comprises the identification of primary studies, applying inclusion and exclusion criteria, and synthesizing the results. To reduce researcher bias, one of the authors developed the protocol, the others reviewed, and the authors developed a discussion, review, and iteration. Finally, we searched in the databases and reported the results that are discussed in the next section. The systematic literature review method was applied through the following procedures:

1. Research Questions: present the purposed research questions;
2. Search strategy: expose the strategy, and libraries investigated to collect data;
3. Article selection: introduce the adopted criteria for study selection;
4. Distribution of studies: explain the chronological distribution of selected articles;

Table 1
Related work.

Article	Year	Range	Systematic review?	Taxonomy?	Focus
Abdallah et al. (2016)	2016	1994–2014	–	✓	Fraud detection systems in general.
Ahmed et al. (2016)	2016	–		✓	Unsupervised anomaly detection in the financial domain (clustering).
Sorournejad et al. (2016)	2016	–	–	✓	Supervised and unsupervised credit card fraud detection techniques.
Adewumi and Akinyelu (2017)	2017	1997–2016	–	–	Credit card fraud detection using nature inspired and ML.
Chilaka et al. (2019)	2019	2014–2019	–	–	Credit card fraud detection in electronic finance and banking domain.
Zhao et al. (2019)	2019	1966–2018	–	–	Behavior analysis in online trading systems.
Pourhabibi et al. (2020)	2020	2007–2018	✓	–	Graph-based anomaly detection approaches.
Aziz and Ghous (2021)	2021	–	–	–	Data mining methods for credit card fraud detection.

5. Quality assessment: introduce the quality assessment of the studies;
6. Data extraction: apply the research questions and point out the useful information from the selected articles.

3.1. Research questions

The definition of research questions was the first step of this study. To better identify the main points in the select articles, we split the questions into two groups: (*i*) the main research questions (MQ); and (*ii*) the specific research questions (SQ). The main questions seek to identify the common domains that articles focus on when employing fraud detection techniques. Furthermore, they explore the types of fraud and methods employed in e-commerce platforms. The following statements define the two MQs this study aims to answer:

- **MQ1** *What are the common domains in the scope of e-commerce frauds?*
- **MQ2** *What are the fraud types and methods employed in e-commerce platforms to detect and prevent fraudulent activities?*

The specific questions category seeks to investigate several aspects related to fraud detection in e-commerce. First, we aim at analyzing the different datasets literature initiatives put effort into and their main characteristics. Second, real-time data analysis is an important issue to provide secure systems in e-commerce. Therefore, we focus on identifying the main requirements and benefits real-time brings to fraud detection in e-commerce. Finally, the main concerns of the e-commerce industry are not only fraudulent transactions but also automated scripts that perform identical transactions that such systems should detect. For this reason, we also look into the specifics of bot detection that literature studies employ. The following three SQs complement the two MQs listed above:

- **SQ1** *What features, datasets and information are used for detecting and preventing fraud?*
- **SQ2** *What are the requirement and benefits of real-time data analysis in this context?*
- **SQ3** *How can we differentiate human beings and bot operations on e-commerce operations and their impact on fraud concerns?*

3.2. Data sources and search strategy

Another step was to select studies to answer the research questions. Our methodology encompasses an appropriate set of databases to get extensive and broad coverage of the literature and increase the probability of finding highly relevant articles. Therefore, the search covered the following electronic literature databases:

- ACM Digital Library (https://dl.acm.org/);
- IEEE Xplore (https://ieeexplore.ieee.org/);
- ScienceDirect (https://www.sciencedirect.com/);
- SpringerLink (https://link.springer.com/).

Table 2
Detailing filtering criteria.

Identifier	Description
F1	Duplicate filtering
F2	Exclusion of all works that are surveys and reviews
F3	Title filtering using inclusion words
F4	Exclusion of all works that have less than 5 pages
F5	Title and abstract filtering using exclusion words
F6	Text filtering

Specific keywords were defined to compose a search string. This string was split into research units and combined by boolean operators. We also included acronyms, synonyms, and alternate spellings. Fig. 3 illustrates the search string we applied to the databases to search for articles published from 2015 until April 2021.

3.3. Filtering criteria and article selection

For the selection of studies, we used the data sources and the search string presented in Section 3.2. After applying the search string, we found 610 articles. We employed exclusion criteria composed of six filters applied sequentially to eliminate the irrelevant works. Table 2 describes each filter, and Fig. 4 depicts the whole filtering process showing the number of articles resulting from each phase.

Initially, F1 and F2 respectively removed all duplicate articles and all surveys and reviews. To reach articles highly related to fraud in e-commerce, F3 filtered out all articles that do not include "fraud" in the article's title. We based this decision on writing paper guidelines that suggest that the article's title should present the aim of the research (Mack, 2012). The following filter (F4) removed all articles that have less than five pages. Then, F5 removed articles that have at least one of the following words in the title or abstract: "video", "download", "reviews", "supply chain", "healthcare", "loan", "app market", "telephony", "cash-out", "social networking", "advertising", "phone call", and "auction". We performed a preliminary analysis of the 610 resulting articles to obtain this set of words. They aim at filtering out articles from different domains that do not focus on e-commerce. Finally, F6 consisted of a deep analysis of the articles' text to identify whether they present fraud detection/prevention strategies in online payment or not.

3.4. Initial analysis

Table A.1 summarizes all 64 articles resulting from the search methodology. In this subsection, we briefly analyze these articles in a broader manner showing some statistics. First, Fig. 5 shows the total number of publications over the past six years as bars. In 2018, the number of publications doubles compared to the previous years and remains at the same pace in 2019 and 2020, which demonstrates that this topic is gaining attention in the last few years. In addition, Fig. 5 presents the distribution of these studies in the publication mechanisms. The illustration reveals that the topic is not much discussed in books. However, conferences and journals have an increasing number

```
("purchase" OR "buy" OR "transactions")
AND
("fraud prevention" OR "fraud detection")
AND
("e-commerce" OR "web commerce" OR "online store")
AND
("real-time" OR "stream")
```

Fig. 3. Search string employed in research databases to collect the literature studies.

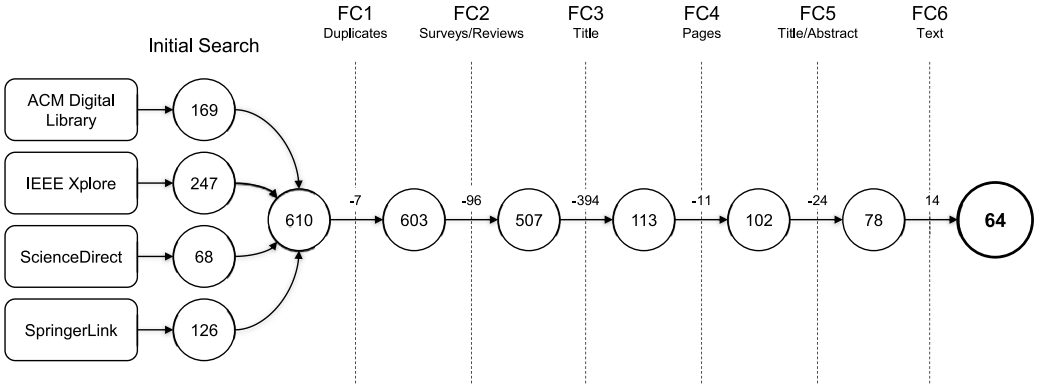

Fig. 4. Article selection process.

of studies on the subject. Another data that can be extracted from Fig. 5 is that as the publications in conferences rise, the number of publications in journals narrows down, showing greater interest of the authors in the publications in conferences.

We also performed a demographic analysis of the articles. Based on the first author's institution, to visualize the countries where the topic of fraud is most recurrent. Fig. 6 shows the heat map for the countries with the most publications. The place where the topic is most discussed is in Asia with altogether adding 27 articles, being 12 articles of China and 11 of India. In Europe, the highlight is for institutions in Italy with 5 publications, Belgium with 4, and Portugal with 3. Morocco was the only African country with selected publications, however, they add 6 more articles to the total. In America, the USA is the country with the most articles on the subject, adding up to 6 more papers. The countries: Bahrain, Canada, Croatia, France, Greece, Indonesia, Jordan, Malaysia, Pakistan, Saudi Arabia, Singapore, UAE, and the UK had one work each. Also, Brazil and Iran had two publications each.

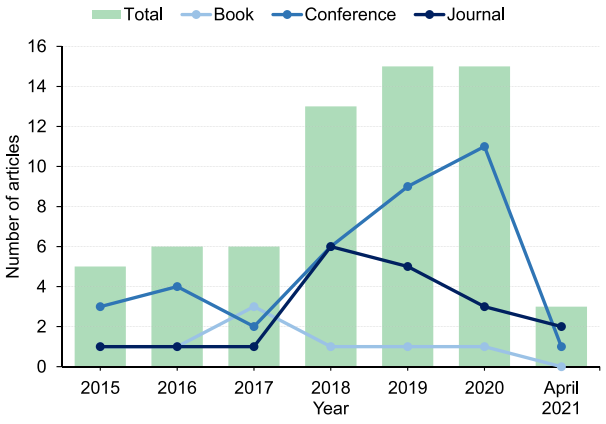

Fig. 5. Publication by year.

4. State of the art analysis

This section presents the analysis of the literature corpus to answer each one of the research questions presented in Section 3.1. To begin with, Table A.1 summarizes all selected articles with some classifications. It presents the domain and type of fraud the articles focus on in their strategies. In addition, it also shows which articles provide strategies for fraud detection or prevention. We look to the articles with different viewpoints according to our research questions. Therefore, we organized the next sections to address each question individually. Sections 4.1 and 4.2 present, respectively, the answers for the two main questions. In turn, Sections 4.3, 4.4, and 4.5 analyze the articles at the specific questions point of view.

4.1. MQ1: What are the common domains in the scope of e-commerce frauds?

Given the literature corpus resulting from the search methodology, we can identify which domain articles put effort into. E-commerce payment systems have several steps to complete online transactions involving three main domains: (*i*) the e-commerce environment itself; (*ii*) the banks; and (*iii*) a payment gateway to intermediate the process between the e-commerce and banks infrastructures. Fig. 7 illustrates the online payment process presenting its steps and all domains involved. The process always starts by a customer/attacker when first accessing the e-commerce website. For simplicity, let us refer to them as a user of the system. The users can navigate between several pages until reaching the payment page in which they inform the payment information and place a purchase order (Liu et al., 2020). At this level, the e-commerce system can track the actions the user takes to interact with the website, such as clicks and pages visited. Once the user places an order, the payment process moves on to the next domain, which is the payment gateway.

The gateway is an important element of the process because it provides security features to the financial process (Zhang and Wang, 2008). E-commerce systems request to the payment gateway a transaction between different banks using the user payment information. The gateway is in charge of encrypting the data before forwarding the request to the banking network for authorization. In turn, the financial

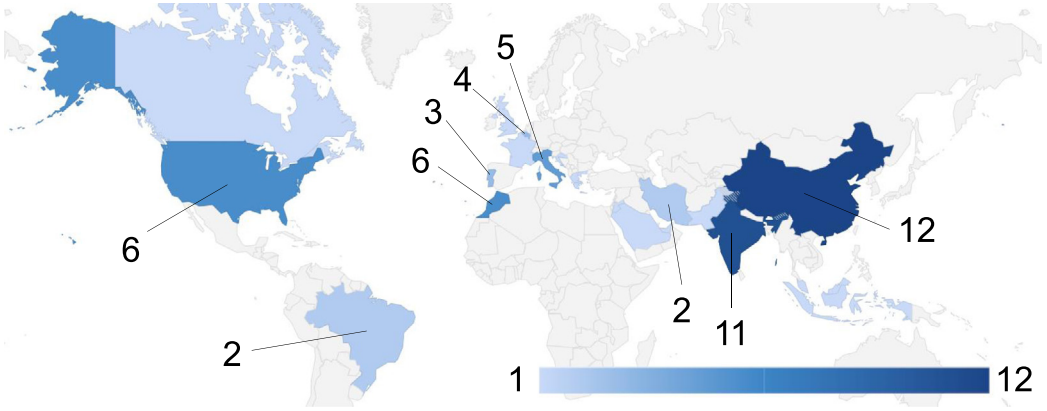

Fig. 6. Publication by countries.

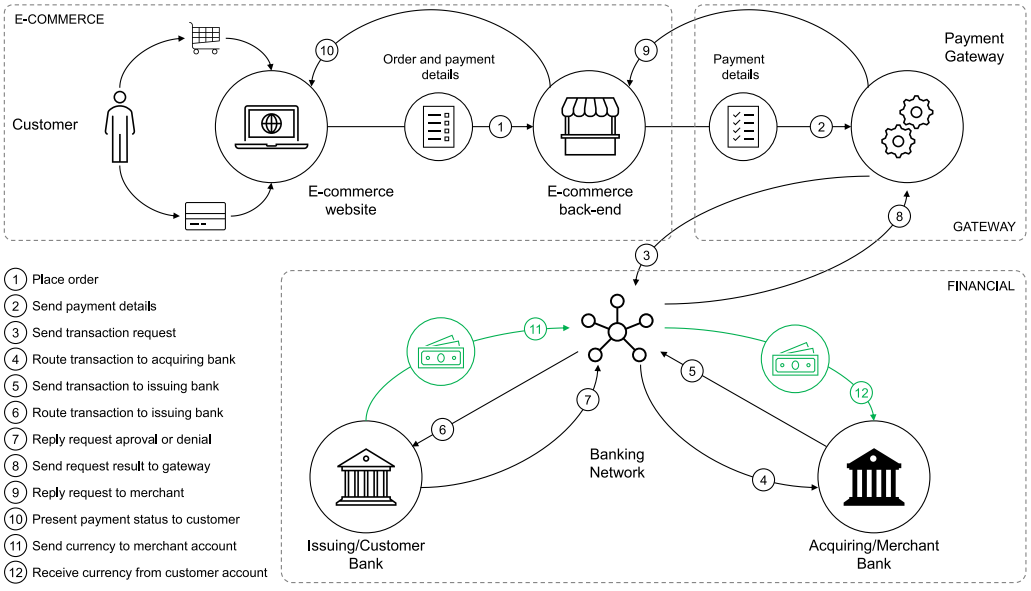

Fig. 7. Online payment process.

domain consists of a banking network and the banks involved in the requested transaction itself (Chakravorti, 2003). The authorization process includes the issuing bank, which holds the user account, and the acquiring bank holder of the destination account. The acquiring bank requests the transaction to the issuing bank that can either approve it or not. Finally, the gateway receives the authorization response and forwards it to the e-commerce platform to handle it.

The literature corpus presents articles that focus on data generated in all of the domains presented above. In particular, Fig. 8 counts the number of articles that employed fraud detection and prevention strategies in each domain. Most of the articles present strategies that focus on the financial domain. In this field, the main goal is to prevent frauds not authorizing a transaction between banks. Although that can be achieved through real-time analysis of each incoming transaction's attributes and past behavior, the articles present detection systems based on datasets. To implement and run a real-time fraud prevention system in financial institutions, the research team needs to establish a partnership with a real company. Also, financial companies willing to implement such systems in production would have to allow the publication of the research, which can be of no interest by such companies. Until now, research has been published only by employing fixed datasets available on the Internet. As a result, the articles present fraud detection strategies and comparisons of several algorithms applied to such data. They aim at providing the best options to fraud prevention systems by showing results on detecting frauds with their strategies.

In the e-commerce domain, a variety of online platforms offer goods and products in different business models. We can classify them into two categories (Kumar and Raheja, 2012): (*i*) business to consumer (B2C); and (*ii*) consumer to consumer (C2C). In the first case, nowadays, every company can deploy its e-commerce system and sell products easily. Online stores, such as DELL online shop,[3] offer the users a platform to shop products using different types of online payment. In other cases, C2C platforms provide an easy way for consumers to buy and sell products directly from other consumers. As an example, currently, there are several companies specialized in providing C2C platforms such as Alibaba[4] and Taobao.[5] What both B2C and C2C platforms have in common is that they follow the model where a customer accesses a website, navigates on it, and places an order to buy a product providing its payment information. Websites can capture and track the user behavior in real-time allowing a fine-grained analysis at the e-commerce level even before the user placing its order. That allows systems to prevent fraud at the earliest in the whole operation. Although this domain offers more possibilities regarding data availability, less than half of the literature corpus presents strategies

[3] https://www.dell.com/en-us/shop.
[4] https://www.alibaba.com/.
[5] https://world.taobao.com/.

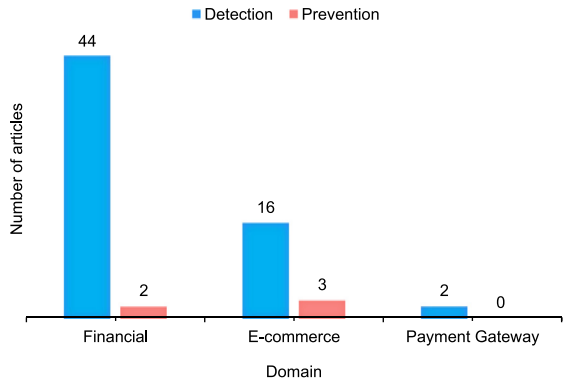

Fig. 8. Number of articles addressing each domain.

here. As for the financial field, the studies rely on fixed datasets since publishing advances require a partnership with companies and they impose restrictions regarding the publication of knowledge. For that reason, articles investigate algorithms and strategies for fraud detection instead of prevention. More specifically, strategies focus mainly on developing detection algorithms since their input data consists of a set of events instead of single real-time events.

In general, articles focus only on one domain of the online payment process. Just three articles address both e-commerce and financial domains (Taha and Malebary, 2020; Dornadula and Geetha, 2019; Behera and Panigrahi, 2015). Although integrating data from both domains could be beneficial, the articles only provide strategies that can be applied solely in datasets from each domain. In other words, they do not evaluate the entire user behavior combining the datasets. Integrating these domains can be promising since it allows a higher data availability to fraud prevention and detection systems. Each domain has access to certain portions of data that are not available to each other. For instance, in the e-commerce domain, platforms can track navigation and user history more precisely, while the financial domain can track the history of transactions of the user payment information across the whole banking network. The combination of this data can represent the best of the two worlds in which more powerful fraud prevention and detection systems can be developed.

Finally, payment gateways are almost out of the scope of research articles from the literature corpus. Only two articles address this particular domain (de Sá et al., 2018; Cao et al., 2019). Payment gateways consist of solutions to intermediate e-commerce platforms and financial institutions providing security on the payment information data traffic. Such platforms, as Alipay[6] and PayPal[7] for instance, also offer an easy way to enable online payment in e-commerce websites. The payment gateway domain is similar to the financial regarding data availability because such platforms have access mainly to the payment information. As for the other domains, fraud prevention and detection are feasible in this particular domain. Although the article sampling is short for payment gateways, it presents a particular study that proposes a system for fraud prevention in real-time (Cao et al., 2019). That was possible because the research formed a partnership with the Ant Financial Fintech,[8] which owns Alipay enabling a real deployment of the system.

4.2. MQ2: What are the fraud types and methods employed in e-commerce platforms to detect and prevent fraudulent activities?

We can answer this question in two different steps. First, we discuss the fraud types present in the literature according to the articles from our research methodology. There are several ways to successfully fraud a transaction both online and in-person. Fig. 9 presents all types of transactions approached by literature in gray boxes. Articles focus on different types and, in several cases, more than one. Therefore, we classified the fraud types into four distinct groups: (*i*) Credit Card; (*ii*) Online Transactions; (*iii*) Financial Transactions; and (*iv*) Fraud Sellers. The figure represents these groups with bold font style and dashed lines. In particular, financial transactions include both online and credit card transactions. However, some studies focus on all types of financial transactions not restricted to one of them (Yin et al., 2020; Smiles and Kamalakannan, 2020; Carminati et al., 2018). To support our decision to create the groups, we will explain each one of them to express the relationship between the articles. The order of presentation goes from the type that has more studies to the one that contains fewer.

Fig. 9. Categorization of types of fraud and transactions. Types of transactions are represented in gray boxes, and types of fraud grouped by boxes with dashed lines.

The first group and the biggest one that we classified is the credit card. The review clearly shows that this is currently the most explored field in fraudulent transactions. In total, 78.1% of the literature corpus approach this theme. The credit card has become the most common payment method employed in electronic shopping, opening a wide possibility of profit for fraudsters if the security is not approached correctly. This method of payment is also exploitable within all the entities involved in the transaction, being not a problem only for e-commerce but for the payment gateway and the financial institutions as well. The literature varies, with many approaches and algorithms being used to solve the problem. Techniques such as ML, statistical inference, data mining, ontologies, and even specific algorithms are employed in this branch. In summary, credit card fraud represents the most extensive field of exploration in e-commerce fraud literature.

Following credit card frauds, online transaction frauds is the second most topic addressed by literature. From the total corpus, 10 articles provide strategies that focus on this type of fraud. Different from articles specifically related to credit card transactions, these studies focus on purchase transactions regardless of the means of payment. The group previously presented is contained within it, but with a much more specific application. In this type, the data used has more affinity to the total purchase price and browsing history. For instance, Wang et al. (2017) present a system for transaction analysis, based on the user's browsing session and mouse behavior, and navigation between pages. For this type of strategy, the payment type does not make

[6] https://intl.alipay.com/.
[7] https://www.paypal.com/.
[8] https://www.antgroup.com/en.

Table 3
Algorithms used to prevent frauds throughout the years.

Total	Algorithm	2015	2016	2017	2018	2019	2020	2021	Articles
2	Random Forests	1	0	0	0	1	0	0	Van Vlasselaer et al. (2015) and Jhangiani et al. (2019)
2	Logistic Regression	1	0	0	0	1	0	0	Van Vlasselaer et al. (2015) and Jhangiani et al. (2019)
1	Artificial Neural Network	1	0	0	0	0	0	0	Van Vlasselaer et al. (2015)
1	Hidden Markov Model	0	1	0	0	0	0	0	dos Santos et al. (2016)
1	Adaboost	0	0	0	0	1	0	0	Jhangiani et al. (2019)
1	Gradient Boost	0	0	0	0	1	0	0	Jhangiani et al. (2019)
1	Support Vector Machine	0	0	0	0	1	0	0	Jhangiani et al. (2019)
1	Ontology	0	0	0	1	0	0	0	El Orche et al. (2018)
1	Dynamic Soft Descriptor	0	1	0	0	0	0	0	Laurens and Zou (2016)

any difference since they evaluate the online behavior of customers previous to the order placement.

Financial transactions are the third group that classify articles according to the fraud type. Three articles address this segment (Yin et al., 2020; Smiles and Kamalakannan, 2020; Carminati et al., 2018). This includes all the articles that not only view purchase transactions but in addition to purchase transfer and withdrawal transactions. For instance, Smiles and Kamalakannan (2020) propose a technique based on the balanced ensemble model that identifies fraudulent financial transactions as payment, cashout, and transfer. In this group, there are more possibilities for fraud because they have different types of transactions, and each type allows a different approach by the fraudster. Therefore, the articles in this group have a much broader scope than the previous. The last two groups are contained within financial transactions, but each with a more narrow scope.

The last group we define is the fraud seller. While the other types concentrate efforts on customer frauds, fraud sellers cover a different angle. It focuses on detecting fraudulent e-commerce platforms that try to make customers believe they are purchasing something from a trustworthy shop. This group was the minor addressed niche in the studies containing only one study in this scope (Li et al., 2019). This approach aims to predict fraudulent sellers based on fake e-commerce transactions to promote the store. Li et al. (2019) propose a solution so that search engines already identify a possible fraudulent store. Unlike the other groups, this is the one that most distances itself from the others. For this reason, it could not be considered as contained within one of the previous ones.

Now, we can proceed to the second part of the MQ2 which regards the methods employed in literature to detect and prevent frauds in e-commerce. Table 3 illustrates the studies that aim fraud prevention, while Table 4 refers to studies that focus on fraud detection. Due to the huge set of algorithms in detect fraud approach, we choose to present only the algorithms with more than three articles. The main idea in presenting these tables is to analyze the current state-of-the-art in algorithms and the evolution and adoption of the algorithms throughout the years. Several authors use more than one algorithm in their studies, therefore, the tables show the same article in several algorithms. In such cases, their main goal is to compare their performance. The tables clarify that at least nine algorithms are being used on a large scale by the studies to detect fraud.

Looking at the techniques, Random Forests is the leading algorithm for both fraud detection and prevention. Random Forests are an ensemble ML method for classification that employ multiple decision trees (see Fig. 10). The method employs a bagging technique, which consists of training several decision trees with the same input length but replacing samples with duplicates randomly for each tree. It produces low correlated trees capable of classifying different features. Thus, the algorithm can generate the final classification by either averaging the predictions of each tree or by majority voting. Random Forests are widely adopted in literature due to their performance.

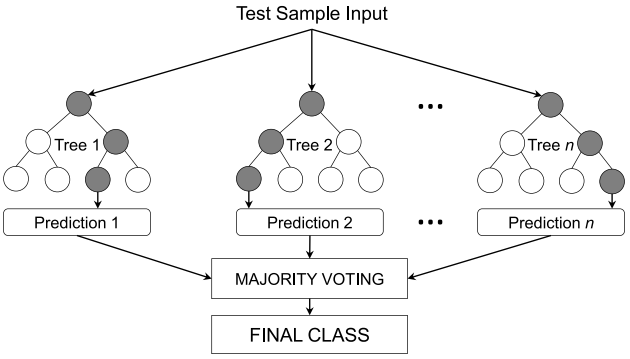

Fig. 10. Random forests classification process.

Several articles claim Random Forests to be the best choice for the fraud detection and prevention problem compared to others (Carneiro et al., 2017; Van Vlasselaer et al., 2015; Dornadula and Geetha, 2019; Puh and Brkić, 2019; Rai and Dwivedi, 2020; Armel and Zaidouni, 2019; AbdulSattar and Hammad, 2020; Patil et al., 2018). Although, there are other scenarios in which other techniques outperform, such as Artificial Neural Network (Sadineni, 2020) and Logistic Regression (Najadat et al., 2020). Even though Random Forests can be the best choice, that decision can only be taken comparing different techniques to each particular case. The literature shows that when designing an ML strategy, the model is highly sensitive to the dataset. Another important point is that there are several metrics to define each algorithm is better, and the target metric can change from problem to problem.

Given the analysis provided to answer MQ1 and MQ2, we build a taxonomy to describe the E-commerce Fraud landscape on the lens of domain, types, and methods. Fig. 11 depicts the taxonomy showing the three main domains and four types of fraud in e-commerce. More importantly, the taxonomy shows all the main methods employed in literature to prevent and detect fraud. According to the taxonomy, five methods appear for both fraud prevention and detection: (*i*) Random Forests; (*ii*) Logistic Regression; (*iii*) Decision Tree; (*iv*) Support Vector Machines; and (*v*) Hidden Markov Model (HMM). Other than these methods, there are several algorithms employed either for detection or prevention. On the one hand, Artificial Neural Network, Naive Bayes, Multi-Layer Perceptron, and K-Nearest Neighbors cover fraud detection systems. On the other hand, Adaboost, Gradient Boost, Ontology, and Dynamic Descriptor are the ones applied specifically for fraud prevention.

4.3. SQ1: What features, datasets and information are used for detecting and preventing fraud?

We identified 41 datasets with different characteristics employed by studies from the literature corpus. Table A.2 summarizes their main

Table 4
Algorithms used to detect frauds throughout the years.

Total	Algorithm	2015	2016	2017	2018	2019	2020	2021	Articles
22	Random Forests	1	0	1	5	7	7	1	Saia et al. (2015), Carneiro et al. (2017), Banerjee et al. (2018), Patil et al. (2018), Xuan et al. (2018), Xuan et al. (2018), Carcillo et al. (2018), Dornadula and Geetha (2019), Puh and Brkić (2019), Saia and Carta (2019), Carta et al. (2019), Armel and Zaidouni (2019), Raghavan and Gayar (2019), Mittal and Tyagi (2019), Najadat et al. (2020), Smiles and Kamalakannan (2020), Sadineni (2020), Mrozek et al. (2020), Rai and Dwivedi (2020), AbdulSattar and Hammad (2020), Lucas et al. (2020) and Sudha and Akila (2021)
17	Logistic Regression	1	1	1	2	5	6	1	Yang et al. (2015), Dai et al. (2016), Carneiro et al. (2017), Banerjee et al. (2018), Patil et al. (2018), Dornadula and Geetha (2019), Puh and Brkić (2019), Saia and Carta (2019), Mittal and Tyagi (2019), Cao et al. (2019), Najadat et al. (2020), Sadineni (2020), Sahu et al. (2020), Mrozek et al. (2020), Rai and Dwivedi (2020), Misra et al. (2020) and Baesens et al. (2021)
11	Decision Tree	0	1	0	1	5	3	1	Dai et al. (2016), Patil et al. (2018), Dornadula and Geetha (2019), Saia and Carta (2019), Armel and Zaidouni (2019), Cao et al. (2019), Sadgali et al. (2019b), Najadat et al. (2020), Sadineni (2020), AbdulSattar and Hammad (2020) and Baesens et al. (2021)
9	Artificial Neural Network	1	1	1	0	3	2	1	Behera and Panigrahi (2015), Kamaruddin and Ravi (2016), Artikis et al. (2017), Mittal and Tyagi (2019), Sadgali et al. (2019a,b), Sadineni (2020), Ali et al. (2020) and Forough and Momtazi (2021)
9	Support Vector Machines	0	0	1	1	5	1	1	Carneiro et al. (2017), Banerjee et al. (2018), Dornadula and Geetha (2019), Puh and Brkić (2019), Raghavan and Gayar (2019), Mittal and Tyagi (2019), Sadgali et al. (2019b), Sadineni (2020) and Sudha and Akila (2021)
9	Naive Bayes	0	1	0	1	5	2	0	Dai et al. (2016), Banerjee et al. (2018), Saia and Carta (2019), Carta et al. (2019), Armel and Zaidouni (2019), Mittal and Tyagi (2019), Sadgali et al. (2019b), Rai and Dwivedi (2020) and Ali et al. (2020)
7	Multi-Layer Perceptron	0	0	0	0	4	3	0	Saia and Carta (2019), Carta et al. (2019), Sadgali et al. (2019a,b), Sahu et al. (2020), Misra et al. (2020) and Anowar and Sadaoui (2020)
5	K-Nearest Neighbors	0	0	0	0	4	1	0	Saia and Carta (2019), Raghavan and Gayar (2019), Mittal and Tyagi (2019), Sadgali et al. (2019b) and Misra et al. (2020)
4	Hidden Markov Model	0	2	0	0	1	1	0	Dai et al. (2016), Rajeshwari and Babu (2016), Sadgali et al. (2019b) and Lucas et al. (2020)

information giving them an ID to be referenced in the text. Describing datasets is not the primary goal of the studies, therefore, several datasets lack some kind of information. In such cases, if they are not publicly available or described by other authors, we could not fill some information in the table. By analyzing the table, we can see that most of the datasets are composed of transactions from real companies and banks from different platforms. Only DS03, DS13, DS25, and DS37 are composed of synthetic data generated to test the authors' models. In particular, DS25 comprises all articles that use simulation to generate data without having historic data from model training.

For fraud detection and prevention strategies, the dataset must provide historical data to be used to design the algorithms. More importantly, the dataset should be well balanced between legitimate and fraudulent transactions. However, that is not the case in most of the datasets employed in the literature. Most of the datasets are unbalanced having much more legitimate transactions than fraudulent

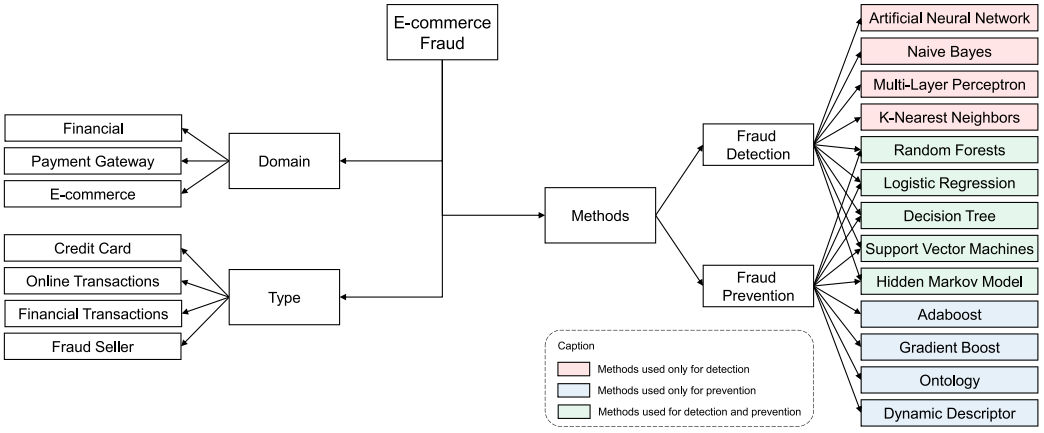

Fig. 11. Proposal of a taxonomy to define domain, types, and methods in e-commerce fraud.

ones. Unbalanced datasets create a challenge when modeling algorithms because the low sampling of fraudulent transactions can lower the system accuracy. With exception of DS07, DS14, and DS16, the average rate of fraudulent transactions in all datasets is 1.91%, which is pretty low. Although not being heavily unbalanced, DS07 and DS14 have a maximum of 1000 transactions, which is low compared to the others. In turn, DS16 presents the ideal case since it has half legitimate and half fraudulent transactions. To form DS16, Guo et al. (2019) have access to data from Taobao. They were able to sample Taobao for the data they needed to form a balanced dataset.

Fig. 12 depicts the number of articles using each dataset. It shows the variety of datasets employed a single time. These cases comprise studies in which authors form a partnership with a real company or bank to provide access to data. From the 41 datasets, only three are explored by more than one article, and one (DS25) comprises simulations. DS02 is largely employed by many articles (17 in total) and is publicly available on Kaggle.[9] The datasets contain transactions made by credit cards in September 2013 by European cardholders. Both DS07 and DS19 appear in three articles because they were publicly available at the time of the researches. However, currently only DS07[10] is still available. DS19 was a dataset made available by the University of California San Diego (UCSD) in the 2009 Data Mining contest sponsored by Fair Isaac Corporation (FICO).

Fig. 13 depicts the number of articles by data type. This figure shows that there is a high variety of data types used in literature to detect and prevent fraud. Some are more common, like credit card and online payments, while others appear just once. In particular, credit card fraud is the most commonly explored problem in the literature. It represents 21 from 41 datasets including DS02, the most commonly used dataset. The remaining 20 datasets are well distributed between the other eight data types. Online payments comprise datasets that contain payment transactions without defining if the payment is being made through credit card, debit card, or other payment modalities. Navigation datasets contain behavioral data of the users including the pages they visit before placing an order. In turn, financial transactions include not online payments but also banking transactions such as transfers between accounts or ATM withdraws.

Four other datasets contain a unique type of information. Yang et al. (2015) employ a dataset from Dangdang[11] that contains order details from customers. Smiles and Kamalakannan (2020) employ a dataset from Kaggle focusing specifically on mobile payment transactions. Molloy et al. (2017) use a banking dataset containing several transactions

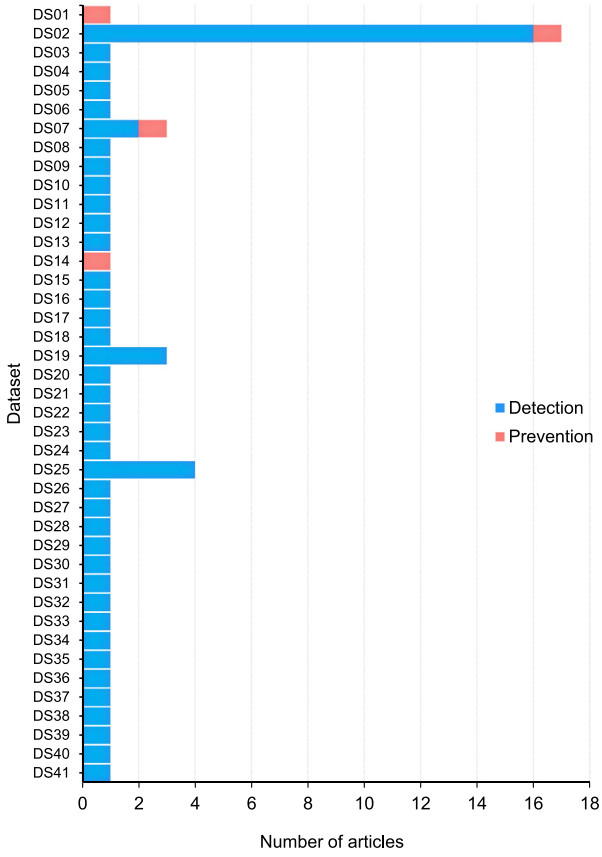

Fig. 12. Number of times each dataset is used in the literature corpus.

from online banking, point of sale, ATM, mobile banking, and person-to-person payments. Finally, Laurens et al. (2019) use Domain Name System (DNS) information from buyers at the time of the transaction.

4.4. SQ2: What are the requirement and benefits of real-time data analysis in this context?

One of the most important factors in fraud prevention and detection is the time to process and analyze data. Fraud must be identified as soon as possible to protect the victim and ensure reliability for e-commerce and their customers. Furthermore, we bring in this section a discussion about the studies that present possible real-time applications of fraud

[9] https://www.kaggle.com/mlg-ulb/creditcardfraud.
[10] http://weka.8497.n7.nabble.com/file/n23121/credit_fruad.arff.
[11] http://www.dangdang.com.

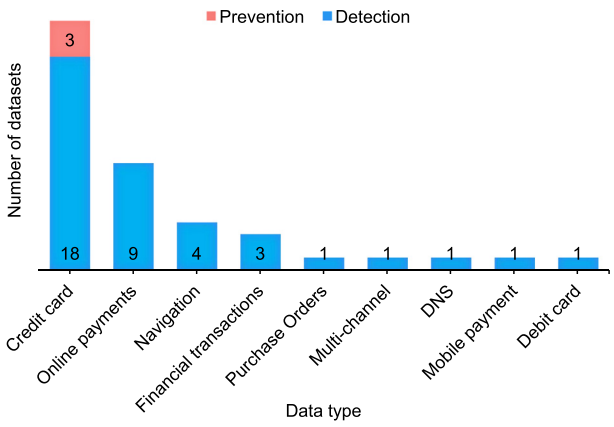

Fig. 13. Main data types from the datasets used in the literature corpus.

analysis proposals, as well as the most used tools and, architectural similarity.

We emphasize that studies such as Artikis et al. (2017), dos Santos et al. (2016), Dornadula and Geetha (2019), Jing et al. (2018), Dal Pozzolo et al. (2018), Sadineni (2020), Correia et al. (2015), Jing et al. (2019), Patil et al. (2018) and Molloy et al. (2017) can execute in real-time due to the prototype execution time, implemented process management or, real-time simulation. However, in this section, we will focus on applied or tested work in production with data streaming, to analyze the similarities of applications to the industry. From the analyzed articles, we obtained 12 studies with a focus on real-time execution. Of these, only three do not focus on detecting credit card fraud, presenting unique and yet effective proposals for operating in production.

That said, Li et al. (2019) propose an impression regulation system using reinforced learning. The strategy aims at avoiding the impression of fraudulent sellers in e-commerce search engines, being able to avoid possible fraud search results at the time of the search. Wang et al. (2017) present a system for transaction analysis based on the user's browsing session. Testing the system in real production over eight months and presenting a scalable architecture of the system able to perform real-time prediction. Cao et al. (2019) propose a complete system for detecting fraudulent transactions in real-time. It contemplates an architecture of execution and processing of data, defining a network of transactions, and also combining different ML models for analysis and classification of transactions. The model was tested at Ant Financial using real data. Moreover, the architecture presented by the study contemplates the use of several machines to execute the system, making it extremely necessary to use multiple instances for the system to be agile.

Within the scope of credit card fraud, Carneiro et al. (2017) use a real dataset from an e-tail/e-commerce to which they extract 70 features, including one computed by functions (past behavior, field comparison). In addition, the strategy employs a weekly script that updates the model with orders older than four months. Abakarim et al. (2018) propose a 6-layer deep learning auto-encoder to detect credit card fraud in real-time. They employ Kafka, MemSQL pipeline, and TensorFlow models in the architecture. Van Vlasselaer et al. (2015) present a model that focuses on real-time feedback for transactions on the credit card issuer side. Their model relies on two strategies for feature extraction: recency-frequency-monetary framework, and data exploration using network relations between merchants and cardholders.

Kamaruddin and Ravi (2016) propose a hybrid architecture involving particle swarm optimization (PSO) and auto-associative neural network (AANN) to get a solution for one-class classification (OCC) in a big data paradigm in a Spark cluster. Branco et al. (2020) focus on streaming data, which means the model should process events arriving in an unexpected order. In addition, they focus on mission-critical low latency systems to quickly provide feedback. Dai et al. (2016) propose a four-layer fraud detection system, using different ML tools and methods to achieve good performance in real-time classifications. The system makes use of software such as Hadoop, Spark, Storm, and Hbase, for data storage, processing, and analysis.

Rajeshwari and Babu (2016) propose the use of streaming analytic to analyze the sequence of credit card transactions. The solution models an HMM-based on the previous cardholder transactions. With that, the transaction streaming sequences are analyzed and managed with spark. Carcillo et al. (2018) propose the integration of Kafka, Spark, and Cassandra tools for data handling and management of a large-scale fraud detection system. SCARFF's focus is the implementation of an architecture capable of collecting, processing data, and training ML models for fraud detection. In contrast, Laurens and Zou (2016) use a unique identifier during the transaction to certify that the user of the credit card has access to the referring bank. This identifier is a code that appears on the card statement and can be used as an authenticator.

The analysis of the architecture of studies focused on real-time shows that certain tools are essential for real-time applications. Apache Spark is the tool most used by the papers, providing a resource for cluster management and parallelism. Also from Apache, Kafka is widely used for data streaming management, providing high capacity and low latency for real-time data processing. In addition, big data management tools like Hadoop, Hbase, and MemSQL are also widely used. For ML algorithms, the most used framework in the real-time analysis is the tensor-flow, this occurs due to the easy integration and packages available for the implementation of Kafka along with the framework's features.

Summarizing the architectural ideas presented by the mentioned articles, Fig. 14 explains the architecture pattern and data flow that is addressed by the studies. The flow starts in "Data Source" at the top-left of the figure. This is the platform where the detection or prevention will be applied and the data collected. The system forwards its collected logs in real-time for two different directions simultaneously. We represent these two procedures by the continuous line and the dashed line. The first one represents the off-line procedures or pre-deploy procedures. This is the storage of the data in a data management system and the ML modeling of the architecture. The ML modeling is in charge of calibrating the ML algorithm to be applied later in production. Usual ML procedures of train/test split, evaluation, and tunning. These procedures aim to generate a final model based on the historical database with the best prediction results. Following, once ML Modeling is finished, the result is a trained model capable of performing a fraud analysis. And so it can be deployed in production to carry out its functionalities, and this is the part that must be replicable to generate scalability in the system.

The second part of the architecture is the online modules. Starting at the top of Fig. 14, the data streaming manager is responsible for caching data and making the requests to the ML module to assess possible frauds. The cache stores recent data which is also stored in the historical database. However, it has a higher reading velocity for faster requisition and a smaller storage capacity. Finally, every time a new fraud score request comes into the system, the previously trained model selects the desired features from the cache and evaluates the user based on recent events. With architectures similar to this presented approach, the evaluated articles presented convincing results for real-time fraud detection and prevention.

4.5. SQ3: How can we differentiate human beings and bot operations on e-commerce operations and their impact on fraud concerns?

Our last research question focuses on how the studies that relate fraud with e-commerce and online payment methods deal with possible automation and bot usage. Automating credit card testing through online shopping is a high-speed method of testing whether a stolen

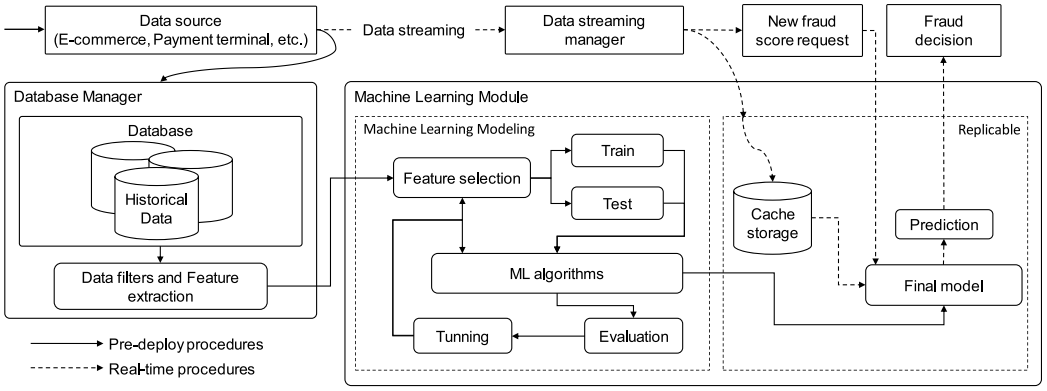

Fig. 14. Common architecture of the studies that addressed the scalability and real-time execution of the fraud detection and prevention systems.

card works. Fraudsters commonly use this type of practice to carry out repeated attacks on a card to use or withdraw all available money (Jing et al., 2018). On the website's side, bots are also increasingly used to liquidate inventories for re-sellers to retain products and resell them at prices higher than the official market value. Automation is also present on websites with the intuition of creating a reliable browsing behavior for e-commerce, thus not being stopped immediately when making a purchase (Wang et al., 2017). This section aims to discuss papers that present ways to differentiate human behavior from automation to prevent fraud. More specifically, following Fig. 1, our point of interest is the studies that present techniques to be used in "Fraud prevention" to deny the purchases before their processing. Thus, only three articles among those analyzed discussed the topic of bot detection.

Laurens et al. (2019) propose using disposable domain names to collect data from the buyer's DNS at the time of the purchase. Using this technique, the customer is dependent on the seller for name resolution, so the seller collects DNS data that allows the identification of network behavior. Therefore, the seller can allow or not a transaction based on the DNS analysis. The study also presents the possible behaviors identified according to the domain resolution. For example, suppose the client's IP address is identical to the DNS address. In that case, the user is probably using a resolver stub, which is not typical of a conventional user but of a server that also runs a DNS service. Another atypical behavior is when there are sharing subnets between IP address and DNS server. In this case, the client's IP is on the same subnet as a DNS server farm, which is not typical for a client and can represent traffic coming from a server. Even though the paper is not focused on identifying bots, the proposed method helps identify traffic coming from servers, which points to the possible use of automation by the fraudster.

Wang et al. (2017) present a system for online purchase analysis, based on the user's browsing session and mouse behavior, and navigation between pages. Also presented is the architecture of the system and how it can be scalable for large amounts of data and still perform real-time prediction. The authors discuss in the text the behavior of fraudulent users and their browsing patterns. Highlighting the uniform pattern, where the user accesses the e-commerce, selects an item, and proceeds to purchase directly, and randomly where the user accesses the site and randomly browses the site until the purchase is complete. This study aims to evaluate the user's last sequence of clicks in a recurrent neural network (RNN) composed of a long–short term memory (LSTM). The window is based on the last 50 clicks that generate the user's traffic data. This data is then used as an input to the RNN. Of the selected studies, this was the only one that focused on detecting possible fraud in section-time, so analyzing user behavior up to the time of purchase.

The last article that discussed the topic of bot detection is (Guo et al., 2019). The authors use adversarial training to elucidate the security problems of deep learning algorithms for detecting fraudulent transactions in real scenarios. Two disturbance methods are proposed. The first is based on the fast gradient sign method (FGSM) and the second on value of perturbation (VoP). One of the methods of disturbance addressed navigation automation. The authors present an example script that bypasses the fraud detector generating similar behavior as humans, such as dragging the cursor across the screen, browsing categories, etc. The study focuses on adversarial training. So, this type of disturbance is later considered by the robust optimization algorithm. The perturbations presented by the authors made the accuracy of the current algorithm used by Taobao decrease from approximately 90% to 20%. Finally, the authors propose adverse training to improve the fraud detection model reaching 85.9% accuracy in detecting fraudulent transactions even with disruption methods.

5. Discussion and open challenges

This section presents the main findings on studying the current state-of-the-art in fraud prevention and detection systems in e-commerce. We provide thoughts on the main problems as well as on the main techniques. Looking at the types of fraud studies focus on, as presented in question MQ2, we have identified four groups. The credit card transactions group was addressed by most studies. The direct correlation that we identified in the fraud process would be the ease of purchasing with someone else's card details. Studies have been looking for ways to link different information so that even if the card data is valid, the system can identify whether the purchase transaction with this card is fraudulent or not. The articles address several solutions, each with a different approach from the other. Another point to have so much work focusing on this niche is the complexity of the problem. Alternatives to using a credit card for transactions appear worldwide, mainly to remedy the security flaws that the credit card allows. There is an excellent example of this in Brazil, which is named PIX. In this solution the money is transferred from one account to another with a unique key, guaranteeing authenticity and carrying out the operation instantly.

In addition to the variety of approaches to detect or prevent fraud, too many algorithms can be applied to proposals. The algorithms that were listed in this study have some points of discussion. The first notable factor is that both studies that proposed to detect or prevent frauds follow the same trend. However, the Decision Tree algorithm applied by eleven articles that detect fraud does not appear in any prevention article. This fact can be interpreted as the application of this algorithm for this purpose is not ideal.

The second interesting factor is that the majority of the articles had a significant distribution between the years. But two algorithms started to gain attention only from 2019, and even with recent appearances, they are already among the most used in researches. They are Multi-Layer Perceptron and K-Nearest Neighbors. In addition to the algorithms already consolidated and used on a large scale by the studies, these two tend to be used more in this area.

5.1. Open gaps and common problems

The analysis of literature through the lens of our research questions produced a fair content of information regarding fraud detection and prevention systems in e-commerce. This enables us to see the current trends and weaknesses of strategies employed to this end. In summary, we can list the following open issues in literature:

- Lack of data integration;
- Few systems are evaluated in production;
- Bot detection is not commonly approached;
- Explainable AI is not approached;
- Companies do not allow publications of strategies.

One of the main issues of the literature regards the lack of data integration from many domains. In general, studies employ datasets from only a single source domain. Each domain in the online payment process can provide different attributes from users because they have access to different databases. For instance, the financial domain has access to historical data regarding the specific payment details of a transaction, which is not available to the e-commerce domain. On the other hand, the e-commerce domain can track user activity by logging the pages the access and, likewise the contrary, this data is not available to the financial domain. Integrating several data sources in the fraud system allows a wider analysis of the behavior of users. Although a more complete dataset can impose more complexity on the data analysis, it offers more possibilities for fraud system architects to process these data. Looking at the literature, most strategies fall in the lack of information problem since they have access to fixed data that is already preprocessed in many cases due to privacy issues. Therefore, data integration from different domains can open possibilities for new algorithms and models.

Another common issue we identified is the lack of studies deploying their strategies in production platforms. Most of the articles present algorithms, methods, and ML models applied to static datasets that are either available on the Internet or provided privately by companies. Studies design their strategies with these datasets and report their results applying their algorithms to these same datasets. A fraud system needs to be evaluated in production in a log-term run to analyze its performance faced to the real behavior of fraudsters. A common issue that fraud systems face is the concept drift problem (Webb et al., 2016). While ML models are static once they are trained, the real world is always changing and shoppers follow the same behavior. Likewise, fraudsters are also constantly changing their behavior and tactics to bypass fraud systems. This poses a challenging environment in which the fraud systems must constantly adapt to the new trends. The only way to test if a system is resilient to such changes is to test it in production in a long period. Unfortunately, this is not widely covered in literature, which represents interesting research and technical problem.

Furthermore, one of the least discussed problems is the detection of bots on the e-commerce platform. Bots make it possible to automate purchases and test stolen cards. In the case of stolen cards, the fraud is clear and usually dealt with in the part of financial fraud. However, this remains a problem for the e-commerce domain since automation can liquidate an entire stock without a normal user having the opportunity to buy a product. One of the major problems in automation detection is sensitivity to disturbances from the detectors (Guo et al., 2019). Since the bot detection algorithm can test every user, it is crucial not to intervene with the non-bot user. Bot detectors algorithms tend to compromise accuracy to ensure that the average user is not affected by false positives. In general, they are less effective in the overall detection. These adjustments in bot detection methods still need improvement, and few are works addressing the topic.

This research shows that the primary fraud detection and prevention methods are ML algorithms. One issue that is not addressed is the explainability of the decision of the algorithms. The traceability of what triggered the occurrence of fraud is fundamental for improving the e-commerces security and validating the algorithm's decision. The field of explainable AI (XAI) aims at making AI decisions transparent and understandable for the user of the system. However, no work in the ML scope addresses XAI, and few are those provide a fairer perception of the cause of fraud.

We identified that fraud detection and prevention is an industrial problem that involves sensitive data. On one hand, data from customers must be protected and not made available publicly, which difficult access to data for research studies. On the other hand, e-commerce platforms from the same field compete with each other. Therefore, providing a better user experience and a more secure platform can result in better performance and consequently in more profit for companies ahead technologically. This is a major issue for companies that do not want to share their solutions publicly. As a consequence, there are not many articles published in the literature presenting those strategies in production in e-commerce platforms.

5.2. Proposal and directions on intelligent anti-fraud systems

While Section 5.1 detailed open gaps and problems on several initiatives, this subsection reveals our proposal towards providing an intelligent anti-fraud system to bypass the aforementioned opportunities. From the e-commerce companies' point of view, having a fraud prevention system is ideal to avoid loss of money due to fraudsters attacks. However, defining the data workflow to achieve real-time performance able to stop attackers before the banking authorization is difficult. Taking into consideration the articles reviewed on this study and the authors' experience, we propose in Fig. 15 an architecture to provide directions on how to design a real-time fraud prevention system. We put together our experience with fraud prevention systems and the knowledge present in the literature to build an architecture showing what would be an ideal system for real-time fraud prevention on the e-commerce platform.

Given the presented literature, there are multiple choices when selecting the best fraud prevention algorithm. From statistical inference to deep learning, all the algorithms can be adapted to a fraud mechanism depending only on the data modeling of the system and the main concerns to be addressed. For example, if the reasoning behind a fraudulent detection is a crucial point, the use of ontologies or statistical inference can be the way to go. If one is dealing with large sets of data with multiple attributes and complex data correlation, neural networks are a good choice, for example, the multi-layer perceptron. On the other hand, random forest is the most used method for being easy to adopt compared to more robust and complex methods and still delivering excellent results. In the literature, there are widely used algorithms that outperform, in most cases, others of the same scope. This is mainly because the algorithm is an improved version of its predecessor. Random forest, for instance, is an improved version of decision trees. The same is valid for multi-layer perceptrons, which are a variety of artificial neural networks. However, in our review, it is clear that the essential part of fraud prevention is data handling. Feature discovery, exploration, and modeling are crucial and much more critical in fraud prevention modeling when compared with the algorithm choice. There are even works that we reviewed that address this by doing extensive data modeling into a dataset and testing it with multiple algorithms. In summary, it is more important to make an extensive data exploration and feature selection into your dataset than to select the best performing algorithm within state of the art.

The fraud prevention mechanism can take place in three different moments throughout the purchase process. First, the e-commerce front-end is a key real-time data source since it can log specific information from the customer side. At each interaction of the customer with the website, a real-time data collection process can catch information including IP address, location, browser version, among others. Capturing and forwarding this information to a database as soon as possible

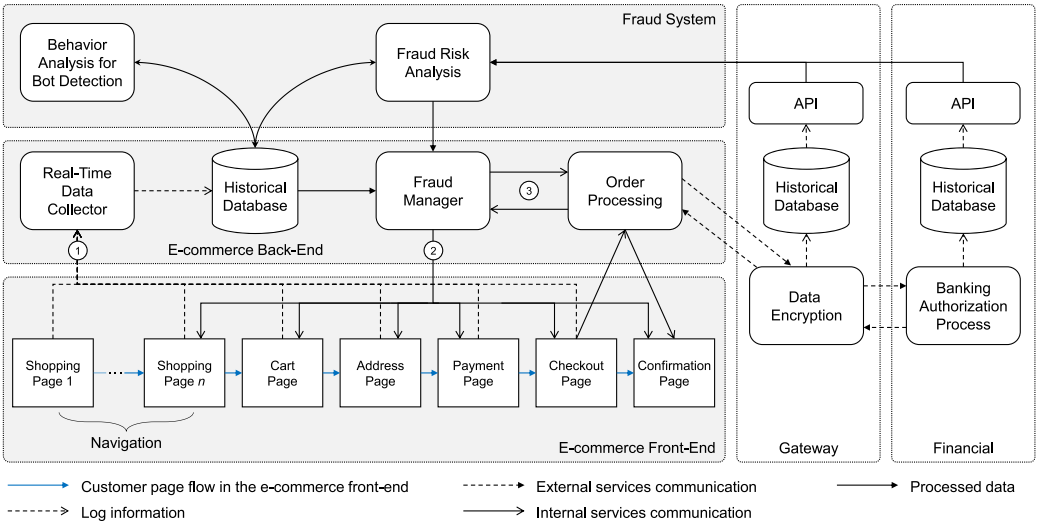

Fig. 15. Directions on a real-time anti-fraud system. Interactions with gateway and financial are not mandatory but would improve the fraud analysis.

is essential. Internal processing depends on this so that the fraud system can provide bot detection and risk analysis before the end of the transaction. Second, real-time analysis brings a lot of advantages mainly for behavior analysis and bot detection. The Behavior Analysis for Bot Detection module analyzes the real-time data stream to produce a probability of a given customer session be an automated script, for instance. The Fraud Manager can use this information to block the customer page flow or to provide additional checking strategies to filter bots, as captchas for instance.

A third moment regards the order processing itself. When a customer places an order, the normal workflow would be to reach the payment gateway to request the banking authorization. However, the Fraud Risk Analysis module provides a probability of this particular order is a fraud attempt. In this phase, payment and shipping information is available to this module that employs fraud detection techniques based on historical data. In addition, a key factor that would increase the accuracy of the system is the availability of data from both payment gateways and financial domains. These domains can provide historical information on the payment data provided and, thus, allow a wider analysis of behavior. For instance, a credit card used in different countries and different e-commerce platforms would suggest a higher probability of fraud. E-commerce generally can only trace an order profile with data available locally. Integration of data from other sources would provide a better profile and, thus, better results. However, this integration is not required for the functioning of the system.

Here, based on our experience in data analysis from e-commerce, we propose a heuristic to define the probability of fraud using user behavior, payment information, and shipment data. A set of rules should be applied to verify the following statements: (R1) billing data is different from shipping data; (R2) IP address from a different country than shipping address; (R3) the customer has no browsing history; (R4) the customer used a minimum of different credit card numbers; (R5) the customer used a minimum of different IP addresses; and (R6) the credit card was used by a different customer. The rules R4 and R5 require a minimum value as a threshold to trigger the rule. The fraud probability can be achieved by setting a weight for each rule. If the rule is triggered, its weight is considered to compute the final probability. Mathematically speaking, Eq. (1) defines the final probability. In the equation, Rn is 1 if rule n triggers or 0 otherwise. Rules compose three different groups: (*i*) consistency (R1 and R2); (*ii*) browsing (R3); and (*iii*) frequency (R4, R5, and R6). Weights should be distributed so each group composes 1/3 of the final weight. Thus, weights should respect the following: (*i*) $x = w1 + w2$; (*ii*) $x = w3$; and (*iii*) $x = w4 + w5 + w6$.

$$\frac{R1 \times w1 + R2 \times w2 + R3 \times w3 + R4 \times w4 + R5 \times w5 + R6 \times w6}{w1 + w2 + w3 + w4 + w5 + w6} \quad (1)$$

Today, e-commerce platforms can already capture data in several phases of an online transaction. Designing a fraud system should consider proposing different data collection, varying from consumer metrics, navigation logs, known patterns on each part of the site to different event correlation data sources. In this way, the idea is to collect data systematically along with the navigation in the e-commerce website. At each capturing procedure, data analysis is a crucial step where different algorithms should be tested to discover the best solution for optimizing the fraud detection system. In particular, we can gather data on each navigation step, so inserting them on algorithms such as Bot Behavior Detection and Fraud Risk Analysis. Logically, these algorithms can be adapted in accordance with the target company's context. A particular behavior index of a suspicious client for company A could represent the typical navigation process in company B, for instance. Also, the same company can change its goals along with the time: for instance, in a Black Friday week, we can change the algorithms policies and parameters. Also, we live in an era with constant changes in fraudsters' tactics, so imposing modifications in the fraud detection algorithms to catch a malicious action. Concluding, it is essential to note that the same algorithm will not be the best option forever.

6. Limitations

This section presents limitations regarding the review methodology employed throughout the research. Our idea here is to clarify to readers what could be explored in future research by continuing the current survey. We considered only research initiatives and did not look at commercial solutions. Unfortunately, a systematic literature review relies on search strings to research databases that do not reach commercial solutions. There are several solutions available in the market that are outside the scope of this research, such as Clearsale,[12] Signifyd,[13] and Kount.[14] An additional problem is that large e-commerce companies, such as Alibaba, Amazon, DELL, and Walmart, do not open their anti-fraud policies publicly. The limitations they impose prevent us to deeply review such strategies in a research study.

Besides, we approached in the review only fraud in online transactions excluding reviews, phishing, login, and telecom. There is also a rising online market of cryptocurrency that we did not address in this study since it was not the main topic of this research. Future

[12] https://www.clear.sale/.
[13] https://www.signifyd.com/.
[14] https://kount.com/.

Table A.1
Summary of articles resulting from the search methodology. Caption: (A) article, (B) year, (D) detection, and (P) prevention.

A	Y	Publication	Domain	Type	D	P
Van Vlasselaer et al. (2015)	2015	Journal	Financial	Credit card		✓
Yang et al. (2015)		Book	E-commerce	Online transaction	✓	
Behera and Panigrahi (2015)		Conference	Financial and e-commerce	Credit card	✓	
Saia et al. (2015)		Conference	Financial	Credit card	✓	
Correia et al. (2015)		Conference	Financial	Credit card	✓	
dos Santos et al. (2016)	2016	Book	E-commerce	Credit card		✓
Kamaruddin and Ravi (2016)		Conference	Financial	Credit card	✓	
Zhao et al. (2016)		Journal	E-commerce	Online transaction	✓	
Dai et al. (2016)		Conference	Financial	Credit card	✓	
Rajeshwari and Babu (2016)		Conference	Financial	Credit card	✓	
Laurens and Zou (2016)		Conference	E-commerce	Credit card		✓
Carneiro et al. (2017)	2017	Journal	E-commerce	Credit card	✓	
Saia (2017)		Book	Financial	Credit card	✓	
Artikis et al. (2017)		Conference	Financial	Credit card	✓	
Molloy et al. (2017)		Book	Financial	Credit card	✓	
Laurens et al. (2017)		Conference	E-commerce	Credit card	✓	
Wang et al. (2017)		Book	E-commerce	Online transaction	✓	
de Sá et al. (2018)	2018	Journal	Payment gateway	Credit card	✓	
Abakarim et al. (2018)		Conference	Financial	Credit card	✓	
Banerjee et al. (2018)		Conference	E-commerce	Credit card	✓	
Randhawa et al. (2018)		Journal	Financial	Credit card	✓	
Zamini and Montazer (2018)		Conference	E-commerce	Credit card	✓	
Dal Pozzolo et al. (2018)		Journal	E-commerce	Credit card	✓	
El Orche et al. (2018)		Conference	Financial	Online transaction		✓
Patil et al. (2018)		Journal	Financial	Credit card	✓	
Xuan et al. (2018)		Conference	E-commerce	Credit card	✓	
Xuan et al. (2018)		Book	Financial	Credit card	✓	
Jing et al. (2018)		Conference	Financial	Online transaction	✓	
Carcillo et al. (2018)		Journal	Financial	Credit card	✓	
Carminati et al. (2018)		Journal	Financial	Financial transactions	✓	
El Orche and Bahaj (2019)	2019	Conference	Financial	Credit card	✓	
Dornadula and Geetha (2019)		Journal	Financial and e-commerce	Credit card	✓	
Puh and Brkić (2019)		Conference	Financial	Credit card	✓	
Saia and Carta (2019)		Journal	Financial	Credit card	✓	
Li et al. (2019)		Conference	E-commerce	Fraud sellers	✓	
Carta et al. (2019)		Journal	Financial	Credit card	✓	
Sadgali et al. (2019a)		Conference	Financial	Credit card	✓	
Armel and Zaidouni (2019)		Conference	Financial	Credit card	✓	
Raghavan and Gayar (2019)		Conference	Financial	Credit card	✓	
Jhangiani et al. (2019)		Conference	E-commerce	Credit card		✓
Mittal and Tyagi (2019)		Conference	Financial	Credit card	✓	
Sadgali et al. (2019b)		Journal	Financial	Online transaction	✓	
Guo et al. (2019)		Conference	E-commerce	Online transaction	✓	
Jing et al. (2019)		Book	Financial	Online transaction	✓	
Cao et al. (2019)		Journal	Payment gateway	Online transaction	✓	
Laurens et al. (2019)		Conference	E-commerce	Credit card	✓	
Sahu et al. (2020)	2020	Conference	Financial	Credit card	✓	
Yin et al. (2020)		Conference	Financial	Financial transactions	✓	
Misra et al. (2020)		Journal	Financial	Credit card	✓	
Ali et al. (2020)		Conference	Financial	Credit card	✓	
Taha and Malebary (2020)		Journal	Financial and e-commerce	Credit card	✓	
Najadat et al. (2020)		Conference	Financial	Credit card	✓	
Smiles and Kamalakannan (2020)		Conference	Financial	Financial transactions	✓	
Sadineni (2020)		Conference	Financial	Credit card	✓	
Mrozek et al. (2020)		Conference	Financial	Credit card	✓	
Rai and Dwivedi (2020)		Book	Financial	Credit card	✓	
Liu et al. (2020)		Conference	E-commerce	Credit card	✓	
AbdulSattar and Hammad (2020)		Conference	Financial	Credit card	✓	
Anowar and Sadaoui (2020)		Conference	Financial	Credit card	✓	
Branco et al. (2020)		Conference	Financial	Credit card	✓	
Lucas et al. (2020)		Journal	Financial	Credit card	✓	
Sudha and Akila (2021)	2021	Conference	Financial	Credit card	✓	
Baesens et al. (2021)		Journal	Financial	Online transaction	✓	
Forough and Momtazi (2021)		Journal	Financial	Credit card	✓	

research should consider looking at the different types of fraud and also commercial solutions.

Finally, the analysis of algorithms employed in the literature does not dive deeply into the technical details of each technique. We focus on bringing light to the main techniques in use to show the reader some directions and the current state-of-the-art. In future research, it would be possible to explore each algorithm showing some examples of use and how to use them.

Table A.2
Description of each dataset employed in the literature corpus.

ID	Articles	Source	Data type	Type	Transactions	Frauds	Rate
DS01	Van Vlasselaer et al. (2015)	Belgium Credit Card	Credit card	Real	3.3M	48,000	1.45%
DS02	Saia (2017), Abakarim et al. (2018), Zamini and Montazer (2018), Dornadula and Geetha (2019), Puh and Brkić (2019), Saia and Carta (2019), Carta et al. (2019), Raghavan and Gayar (2019), Jhangiani et al. (2019), Sahu et al. (2020), Misra et al. (2020), Ali et al. (2020), Taha and Malebary (2020), Mrozek et al. (2020), Rai and Dwivedi (2020), Anowar and Sadaoui (2020) and Forough and Momtazi (2021)	European Cardholders	Credit card	Real	284,807	492	0.17%
DS03	Artikis et al. (2017)	Feedzai	Credit card	Synthetic	10M	20,000	0.20%
DS04	Laurens et al. (2017)	MaximusCards	Navigation	Real	129,116	0	0.00%
DS05	de Sá et al. (2018)	PagSeguro	Credit card	Real	903,801	16,639	1.84%
DS06	Randhawa et al. (2018)	Malasyan Financial Institution	Credit card	Real	284,224	102	0.04%
DS07	Patil et al. (2018), Saia and Carta (2019) and Jhangiani et al. (2019)	German Credit Data	Credit card	Real	1000	300	30.00%
DS08	Xuan et al. (2018)	Chinese E-commerce	Credit card	Real	31,755,151	82,931	0.26%
DS09	Xuan et al. (2018)	Chinese Financial Company 1	Credit card	Real	5M	150,000	3.00%
DS10	Jing et al. (2018)	Asian Bank Part 1	Online transactions	Real	4,690,000	38,601	0.82%
DS11	Jing et al. (2018)	Asian Bank Part 2	Online transactions	Real	13.3M	33,248	0.25%
DS12	Carcillo et al. (2018)	Industrial Company	Credit card	Real	8,356,811	33,427	0.40%
DS13	Sadgali et al. (2019a)	Generated Synthetically	Financial transactions	Synthetic	60,000	168	0.28%
DS14	Jhangiani et al. (2019)	Australian Credit Approval	Credit card	Real	690	307	44.49%
DS15	Carminati et al. (2018)	Italian Bank	Financial transactions	Real	890,997	–	–
DS16	Guo et al. (2019)	TaoBao 1	Online transactions	Real	3M	1.5M	50.00%
DS17	Jing et al. (2019)	Private Bank	Financial transactions	Real	3,502,048	65,291	1.86%
DS18	Yin et al. (2020)	Chinese Financial Company 2	Online transactions	Real	3.5M	–	–
DS19	Banerjee et al. (2018), AbdulSattar and Hammad (2020) and Najadat et al. (2020)	UCSD-FICO Data Mining Contest 2009	Online transactions	Real	97,346	2094	2.15%
DS20	Sadineni (2020)	Financial From Kaggel	Credit card	–	150,000	–	–
DS21	Liu et al. (2020)	Alibaba	Online transactions	Real	2,904,611	22,497	0.77%
DS22	Sudha and Akila (2021)	Best Pay	Online payment	Real	–	–	–
DS23	Forough and Momtazi (2021)	Brazilian Bank	Credit card	Real	360,792	14,031	3.89%
DS24	Yang et al. (2015)	DangDang E-commerce	Purchase orders	Real	14,235	2075	14.58%
DS25	Behera and Panigrahi (2015), Dai et al. (2016), Rajeshwari and Babu (2016) and Armel and Zaidouni (2019)	Simulation	Credit card	Synthetic	–	–	–
DS26	Saia et al. (2015)	Private Company	Credit card	Real	204	–	–
DS27	Kamaruddin and Ravi (2016)	ccFraud	Credit card	Real	9403,986	596,014	6.34%
DS28	Zhao et al. (2016)	TaoBao 2	Navigation	Real	8885	561	6.31%
DS29	Carneiro et al. (2017)	Private Retailer	Credit card	Real	347,572	6500	1.87%
DS30	Molloy et al. (2017)	ABN AMRO Bank	Multi-channel	Real	>200M	–	–
DS31	Wang et al. (2017)	JD.com E-commerce	Navigation	Real	–	–	–
DS32	Dal Pozzolo et al. (2018)	European E-commerce 1	Credit card	Real	21,830,330	41,000	0.19%
DS33	Dal Pozzolo et al. (2018)	European E-commerce 2	Credit card	Real	54,764,384	130,000	0.24%
DS34	Cao et al. (2019)	Alipay	Online payment	Real	–	–	–
DS35	Laurens et al. (2019)	Private E-commerce	DNS	Real	18,974	–	–
DS36	Najadat et al. (2020)	IEEE-CIS Fraud Detection (VESTA)	Online payment	Real	569,875	20,663	3.63%
DS37	Smiles and Kamalakannan (2020)	Synthetic From Kaggle	Mobile payment	Synthetic	6,354,407	8213	0.13%
DS38	Branco et al. (2020)	European Financial Institution A	Credit card	Real	1B	5M	0.50%
DS39	Lucas et al. (2020)	European Financial Institution B	Credit card	Real	4B	571,429	0.01%
DS40	Lucas et al. (2020)	Belgium Company	Credit card	Real	16.5B	96,730	0.59%
DS41	Baesens et al. (2021)	European Bank	Debit card	Real	31,763	506	1.59%

7. Final remarks

This article presented a survey on the scope of fraud detection and prevention. In particular, we have explained the current domains of these issues in e-commerce systems and the most used algorithms and used datasets. In conclusion, first, among the 64 analyzed articles, only 5 of them focused on prevention. This means that most of the initiatives work after an eventual problem is detected. Second, we observed that most articles are from China and India. This makes sense since these countries present the most effective e-commerce systems worldwide. Third, we observed that the fraud keyword is more evident when discussing credit card usage and online payments. In other words, statistical-based methods and artificial intelligence algorithms are mainly applied to these data contexts.

As one of the article's main findings, it is clear for the authors that we have research gaps toward automatically identifying what user intent is and what is done a bot action. Also, we envisage as an opportunity the fact of combining multiple data sources to compute a

scalable Artificial Intelligence-based heuristic to define a probability of fraud for each purchase individually. The challenges here are threefold: define the amount of historical data that could be used, use efficient methods to correlate data, and compile appropriate warnings based on the obtained fraud probability.

CRediT authorship contribution statement

Vinicius Facco Rodrigues: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Lucas Micol Policarpo:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Diórgenes Eugênio da Silveira:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Rodrigo da Rosa Righi:** Conceptualization, Methodology, Writing – review & editing, Visualization. **Cristiano André da Costa:** Supervision. **Jorge Luis Victória Barbosa:** Supervision. **Rodolfo Stoffel Antunes:** Supervision. **Rodrigo Scorsatto:** Supervision. **Tanuj Arcot:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was support by Dell Inc. via the 7th Amendment to the Technical and Scientific Cooperation Agreement No. 01/2017 – Information Technology Innovation Support Law – Brazilian Government. The authors would like to thank Dell Inc. for financing this research project.

Appendix. Tables

See Tables A.1 and A.2

References

Abakarim, Y., Lahby, M., Attioui, A., 2018. An efficient real time model for credit card fraud detection based on deep learning. In: Proceedings of the 12th International Conference on Intelligent Systems: Theories and Applications. SITA '18, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3289402.3289530.

Abdallah, A., Maarof, M.A., Zainal, A., 2016. Fraud detection system: A survey. J. Netw. Comput. Appl. 68, 90–113. http://dx.doi.org/10.1016/j.jnca.2016.04.007, URL https://www.sciencedirect.com/science/article/pii/S1084804516300571.

AbdulSattar, K., Hammad, M., 2020. Fraudulent transaction detection in FinTech using machine learning algorithms. In: 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT). pp. 1–6. http://dx.doi.org/10.1109/3ICT51146.2020.9312025.

Adewumi, A.O., Akinyelu, A.A., 2017. A survey of machine-learning and nature-inspired based credit card fraud detection techniques. Int. J. Syst. Assur. Eng. Manag. 8 (2), 937–953.

Ahmed, M., Mahmood, A.N., Islam, M.R., 2016. A survey of anomaly detection techniques in financial domain. Future Gener. Comput. Syst. 55, 278–288. http://dx.doi.org/10.1016/j.future.2015.01.001, URL https://www.sciencedirect.com/science/article/pii/S0167739X15000023.

Ali, I., Aurangzeb, K., Awais, M., ul Hussen Khan, R.J., Aslam, S., 2020. An efficient credit card fraud detection system using deep-learning based approaches. In: 2020 IEEE 23rd International Multitopic Conference (INMIC). pp. 1–6. http://dx.doi.org/10.1109/INMIC50486.2020.9318202.

Anowar, F., Sadaoui, S., 2020. Incremental neural-network learning for big fraud data. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 3551–3557. http://dx.doi.org/10.1109/SMC42975.2020.9283136.

Armel, A., Zaidouni, D., 2019. Fraud detection using apache spark. In: 2019 5th International Conference on Optimization and Applications (ICOA). pp. 1–6. http://dx.doi.org/10.1109/ICOA.2019.8727610.

Artikis, A., Katzouris, N., Correia, I., Baber, C., Morar, N., Skarbovsky, I., Fournier, F., Paliouras, G., 2017. A prototype for credit card fraud management: Industry paper. In: Proceedings of the 11th ACM International Conference on Distributed and Event-Based Systems. DEBS '17, Association for Computing Machinery, New York, NY, USA, pp. 249–260. http://dx.doi.org/10.1145/3093742.3093912.

Aziz, A., Ghous, H., 2021. Fraudulent transactions detection in credit card by using data mining methods: A review. Int. J. Sci. Prog. Res. 79 (179), 31–48.

Baesens, B., Höppner, S., Verdonck, T., 2021. Data engineering for fraud detection. Decis. Support Syst. 113492. http://dx.doi.org/10.1016/j.dss.2021.113492, URL https://www.sciencedirect.com/science/article/pii/S0167923621000026.

Banerjee, R., Bourla, G., Chen, S., Kashyap, M., Purohit, S., 2018. Comparative analysis of machine learning algorithms through credit card fraud detection. In: 2018 IEEE MIT Undergraduate Research Technology Conference (URTC). pp. 1–4. http://dx.doi.org/10.1109/URTC45901.2018.9244782.

Behera, T.K., Panigrahi, S., 2015. Credit card fraud detection: A hybrid approach using fuzzy clustering neural network. In: 2015 Second International Conference on Advances in Computing and Communication Engineering. pp. 494–499. http://dx.doi.org/10.1109/ICACCE.2015.33.

Biolchini, J., Mian, P.G., Natali, A.C.C., Travassos, G.H., 2005. Systematic Review in Software Engineering. Technical Report, System Engineering and Computer Science Department COPPE/UFRJ.

Branco, B., Abreu, P., Gomes, A.S., Almeida, M.S.C., Ascensão, J.T., Bizarro, P., 2020. Interleaved sequence RNNs for fraud detection. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '20, Association for Computing Machinery, New York, NY, USA, pp. 3101–3109. http://dx.doi.org/10.1145/3394486.3403361.

Cao, S., Yang, X., Chen, C., Zhou, J., Li, X., Qi, Y., 2019. TitAnt: Online real-time transaction fraud detection in ant financial. Proc. VLDB Endow. 12 (12), 2082–2093. http://dx.doi.org/10.14778/3352063.3352126.

Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.-A., Caelen, O., Mazzer, Y., Bontempi, G., 2018. SCARFF: A scalable framework for streaming credit card fraud detection with spark. Inf. Fusion 41, 182–194. http://dx.doi.org/10.1016/j.inffus.2017.09.005, URL https://www.sciencedirect.com/science/article/pii/S1566253517305444.

Carminati, M., Polino, M., Continella, A., Lanzi, A., Maggi, F., Zanero, S., 2018. Security evaluation of a banking fraud analysis system. ACM Trans. Priv. Secur. 21 (3), http://dx.doi.org/10.1145/3178370.

Carneiro, N., Figueira, G., Costa, M., 2017. A data mining based system for credit-card fraud detection in e-tail. Decis. Support Syst. 95, 91–101. http://dx.doi.org/10.1016/j.dss.2017.01.002, URL https://www.sciencedirect.com/science/article/pii/S0167923617300027.

Carta, S., Fenu, G., Reforgiato Recupero, D., Saia, R., 2019. Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model. J. Inf. Secur. Appl. 46, 13–22. http://dx.doi.org/10.1016/j.jisa.2019.02.007, URL https://www.sciencedirect.com/science/article/pii/S2214212618304216.

Chakravorti, S., 2003. Theory of credit card networks: A survey of the literature. Rev. Netw. Econ. 2 (2).

Chilaka, U.L., Chukwudebe, G.A., Bashiru, A., 2019. A review of credit card fraud detection techniques in electronic finance and banking. Conic Res. Eng. J. 3 (2), 456–467.

Correia, I., Fournier, F., Skarbovsky, I., 2015. The uncertain case of credit card fraud detection. In: Proceedings of the 9th ACM International Conference on Distributed Event-Based Systems. DEBS '15, Association for Computing Machinery, New York, NY, USA, pp. 181–192. http://dx.doi.org/10.1145/2675743.2771877.

Dai, Y., Yan, J., Tang, X., Zhao, H., Guo, M., 2016. Online credit card fraud detection: A hybrid framework with big data technologies. In: 2016 IEEE Trustcom/BigDataSE/ISPA. pp. 1644–1651. http://dx.doi.org/10.1109/TrustCom.2016.0253.

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., Bontempi, G., 2018. Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Trans. Neural Netw. Learn. Syst. 29 (8), 3784–3797. http://dx.doi.org/10.1109/TNNLS.2017.2736643.

de Sá, A.G., Pereira, A.C., Pappa, G.L., 2018. A customized classification algorithm for credit card fraud detection. Eng. Appl. Artif. Intell. 72, 21–29. http://dx.doi.org/10.1016/j.engappai.2018.03.011, URL https://www.sciencedirect.com/science/article/pii/S0952197618300605.

Dornadula, V.N., Geetha, S., 2019. Credit card fraud detection using machine learning algorithms. Procedia Comput. Sci. 165, 631–641. http://dx.doi.org/10.1016/j.procs.2020.01.057, URL https://www.sciencedirect.com/science/article/pii/S187705092030065X 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION, 2019 November 11-12, 2019.

dos Santos, M.V.M., da Silva, P.D.B., Otero, A.G.L., Wisnieski, R.T., Goncalves, G.S., Maria, R.E., Dias, L.A.V., da Cunha, A.M., 2016. Applying scrum in an interdisciplinary project for fraud detection in credit card transactions. In: Latifi, S. (Ed.), Information Technology: New Generations. Springer International Publishing, Cham, pp. 461–471.

El Orche, A., Bahaj, M., 2019. Approach to use ontology based on electronic payment system and machine learning to prevent fraud. In: Proceedings of the 2nd International Conference on Networking, Information Systems & Security. In: NISS19, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3320326.3320369.

El Orche, A., Bahaj, M., Alhayat, S.A., 2018. Ontology based on electronic payment fraud prevention. In: 2018 IEEE 5th International Congress on Information Science and Technology (CiSt). pp. 143–148. http://dx.doi.org/10.1109/CIST.2018.8596486.

Forough, J., Momtazi, S., 2021. Ensemble of deep sequential models for credit card fraud detection. Appl. Soft Comput. 99, 106883. http://dx.doi.org/10.1016/j.asoc.2020.106883, URL https://www.sciencedirect.com/science/article/pii/S1568946620308218.

Guo, Q., Li, Z., An, B., Hui, P., Huang, J., Zhang, L., Zhao, M., 2019. Securing the deep fraud detector in large-scale E-commerce platform via adversarial machine learning approach. In: The World Wide Web Conference. WWW '19, Association for Computing Machinery, New York, NY, USA, pp. 616–626. http://dx.doi.org/10.1145/3308558.3313533.

Guthrie, C., Fosso-Wamba, S., Arnaud, J.B., 2021. Online consumer resilience during a pandemic: An exploratory study of e-commerce behavior before, during and after a COVID-19 lockdown. J. Retail. Consum. Serv. 61, 102570. http://dx.doi.org/10.1016/j.jretconser.2021.102570, URL https://www.sciencedirect.com/science/article/pii/S0969698921001363.

Jhangiani, R., Bein, D., Verma, A., 2019. Machine learning pipeline for fraud detection and prevention in E-commerce transactions. In: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON). pp. 0135–0140. http://dx.doi.org/10.1109/UEMCON47517.2019.8992993.

Jing, C., Wang, C., Yan, C., 2018. Replay attack: A prevalent pattern of fraudulent online transactions. In: 2018 5th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2018 4th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom). pp. 75–82. http://dx.doi.org/10.1109/CSCloud/EdgeCom.2018.00022.

Jing, C., Wang, C., Yan, C., 2019. Thinking like a fraudster: Detecting fraudulent transactions via statistical sequential features. In: Goldberg, I., Moore, T. (Eds.), Financial Cryptography and Data Security. Springer International Publishing, Cham, pp. 588–604.

Kamaruddin, S., Ravi, V., 2016. Credit card fraud detection using big data analytics: Use of PSOAANN based one-class classification. In: Proceedings of the International Conference on Informatics and Analytics. In: ICIA-16, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/2980258.2980319.

Kitchenham, B., Charters, S., 2007. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical Report, Keele University.

Kumar, V., Raheja, E.G., 2012. Business to business (b2b) and business to consumer (b2c) management. Int. J. Comput. Technol. 3 (3b), 447–451.

Laurens, R., Jusak, J., Zou, C.C., 2017. Invariant diversity as a proactive fraud detection mechanism for online merchants. In: GLOBECOM 2017 - 2017 IEEE Global Communications Conference. pp. 1–6. http://dx.doi.org/10.1109/GLOCOM.2017.8254499.

Laurens, R., Rezaeighaleh, H., Zou, C.C., Jusak, J., 2019. Using disposable domain names to detect online card transaction fraud. In: ICC 2019 - 2019 IEEE International Conference on Communications (ICC). pp. 1–7. http://dx.doi.org/10.1109/ICC.2019.8761144.

Laurens, R., Zou, C.C., 2016. Using credit/debit card dynamic soft descriptor as fraud prevention system for merchant. In: 2016 IEEE Global Communications Conference (GLOBECOM). pp. 1–7. http://dx.doi.org/10.1109/GLOCOM.2016.7842369.

Li, Z., Song, J., Hu, S., Ruan, S., Zhang, L., Hu, Z., Gao, J., 2019. FAIR: Fraud aware impression regulation system in large-scale real-time E-commerce search platform. In: 2019 IEEE 35th International Conference on Data Engineering (ICDE). pp. 1898–1903. http://dx.doi.org/10.1109/ICDE.2019.00205.

Liu, Y., Gao, W., Hua, R., Chen, H., 2021. Decomposition and measurement of economic effects of E-commerce based on static feder model and improved dynamic feder model. In: 2021 2nd International Conference on E-Commerce and Internet Technology (ECIT). pp. 213–217. http://dx.doi.org/10.1109/ECIT52743.2021.00054.

Liu, C., Zhong, Q., Ao, X., Sun, L., Lin, W., Feng, J., He, Q., Tang, J., 2020. Fraud transactions detection via behavior tree with local intention calibration. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '20, Association for Computing Machinery, New York, NY, USA, pp. 3035–3043. http://dx.doi.org/10.1145/3394486.3403354.

Lucas, Y., Portier, P.-E., Laporte, L., He-Guelton, L., Caelen, O., Granitzer, M., Calabretto, S., 2020. Towards automated feature engineering for credit card fraud detection using multi-perspective HMMs. Future Gener. Comput. Syst. 102, 393–402. http://dx.doi.org/10.1016/j.future.2019.08.029, URL https://www.sciencedirect.com/science/article/pii/S0167739X19300664.

Mack, C.A., 2012. How to write a good scientific paper: title, abstract, and keywords. J. Micro/Nanolithography, Mems, Moems 11 (2), 020101.

Misra, S., Thakur, S., Ghosh, M., Saha, S.K., 2020. An autoencoder based model for detecting fraudulent credit card transaction. Procedia Comput. Sci. 167, 254–262. http://dx.doi.org/10.1016/j.procs.2020.03.219, URL https://www.sciencedirect.com/science/article/pii/S1877050920306840 International Conference on Computational Intelligence and Data Science.

Mittal, S., Tyagi, S., 2019. Performance evaluation of machine learning algorithms for credit card fraud detection. In: 2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence). pp. 320–324. http://dx.doi.org/10.1109/CONFLUENCE.2019.8776925.

Molloy, I., Chari, S., Finkler, U., Wiggerman, M., Jonker, C., Habeck, T., Park, Y., Jordens, F., van Schaik, R., 2017. Graph analytics for real-time scoring of cross-channel transactional fraud. In: Grossklags, J., Preneel, B. (Eds.), Financial Cryptography and Data Security. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 22–40.

Mrozek, P., Panneerselvam, J., Bagdasar, O., 2020. Efficient resampling for fraud detection during anonymised credit card transactions with unbalanced datasets. In: 2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC). pp. 426–433. http://dx.doi.org/10.1109/UCC48980.2020.00067.

Najadat, H., Altiti, O., Aqouleh, A.A., Younes, M., 2020. Credit card fraud detection based on machine and deep learning. In: 2020 11th International Conference on Information and Communication Systems (ICICS). pp. 204–208. http://dx.doi.org/10.1109/ICICS49469.2020.239524.

Patil, S., Nemade, V., Soni, P.K., 2018. Predictive modelling for credit card fraud detection using data analytics. Procedia Comput. Sci. 132, 385–395. http://dx.doi.org/10.1016/j.procs.2018.05.199, URL https://www.sciencedirect.com/science/article/pii/S1877050918309347 International Conference on Computational Intelligence and Data Science.

Pourhabibi, T., Ong, K.-L., Kam, B.H., Boo, Y.L., 2020. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. Decis. Support Syst. 133, 113303. http://dx.doi.org/10.1016/j.dss.2020.113303, URL https://www.sciencedirect.com/science/article/pii/S0167923620300580.

Puh, M., Brkić, L., 2019. Detecting credit card fraud using selected machine learning algorithms. In: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). pp. 1250–1255. http://dx.doi.org/10.23919/MIPRO.2019.8757212.

Raghavan, P., Gayar, N.E., 2019. Fraud detection using machine learning and deep learning. In: 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). pp. 334–339. http://dx.doi.org/10.1109/ICCIKE47802.2019.9004231.

Rai, A.K., Dwivedi, R.K., 2020. Fraud detection in credit card data using machine learning techniques. In: Bhattacharjee, A., Borgohain, S.K., Soni, B., Verma, G., Gao, X.-Z. (Eds.), Machine Learning, Image Processing, Network Security and Data Sciences. Springer Singapore, Singapore, pp. 369–382.

Rajeshwari, U., Babu, B.S., 2016. Real-time credit card fraud detection using Streaming Analytics. In: 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (ICATccT). pp. 439–444. http://dx.doi.org/10.1109/ICATCCT.2016.7912039.

Randhawa, K., Loo, C.K., Seera, M., Lim, C.P., Nandi, A.K., 2018. Credit card fraud detection using AdaBoost and majority voting. IEEE Access 6, 14277–14284. http://dx.doi.org/10.1109/ACCESS.2018.2806420.

Sadgali, I., Sael, N., Benabbou, F., 2019a. Fraud detection in credit card transaction using neural networks. In: Proceedings of the 4th International Conference on Smart City Applications. SCA '19, Association for Computing Machinery, New York, NY, USA, http://dx.doi.org/10.1145/3368756.3369082.

Sadgali, I., Sael, N., Benabbou, F., 2019b. Performance of machine learning techniques in the detection of financial frauds. Procedia Comput. Sci. 148, 45–54. http://dx.doi.org/10.1016/j.procs.2019.01.007, URL https://www.sciencedirect.com/science/article/pii/S1877050919300079 The second international conference on intelligent computing in data sciences, ICDS2018.

Sadineni, P.K., 2020. Detection of fraudulent transactions in credit card using machine learning algorithms. In: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). pp. 659–660. http://dx.doi.org/10.1109/I-SMAC49090.2020.9243545.

Sahu, A., GM, H., Gourisaria, M.K., 2020. A dual approach for credit card fraud detection using neural network and data mining techniques. In: 2020 IEEE 17th India Council International Conference (INDICON). pp. 1–7. http://dx.doi.org/10.1109/INDICON49873.2020.9342462.

Saia, R., 2017. A discrete wavelet transform approach to fraud detection. In: Yan, Z., Molva, R., Mazurczyk, W., Kantola, R. (Eds.), Network and System Security. Springer International Publishing, Cham, pp. 464–474.

Saia, R., Boratto, L., Carta, S., 2015. Multiple behavioral models: A Divide and Conquer strategy to fraud detection in financial data streams. In: 2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K), Vol. 01. pp. 496–503.

Saia, R., Carta, S., 2019. Evaluating the benefits of using proactive transformed-domain-based techniques in fraud detection tasks. Future Gener. Comput. Syst. 93, 18–32. http://dx.doi.org/10.1016/j.future.2018.10.016, URL https://www.sciencedirect.com/science/article/pii/S0167739X18306423.

Smiles, J.A., Kamalakannan, T., 2020. Data mining based hybrid latent representation induced ensemble model towards fraud prediction. In: 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). pp. 376–382. http://dx.doi.org/10.1109/ICISS49785.2020.9316080.

Song, Y., Escobar, O., Arzubiaga, U., De Massis, A., 2021. The digital transformation of a traditional market into an entrepreneurial ecosystem. Rev. Manage. Sci. 1–24.

Sorournejad, S., Zojaji, Z., Atani, R.E., Monadjemi, A.H., 2016. A survey of credit card fraud detection techniques: Data and technique oriented perspective. CoRR abs/1611.06439 arXiv:1611.06439 URL http://arxiv.org/abs/1611.06439.

Sudha, C., Akila, D., 2021. Credit card fraud detection system based on operational transaction features using SVM and random forest classifiers. In: 2021 2nd International Conference on Computation, Automation and Knowledge Management (ICCAKM). pp. 133–138. http://dx.doi.org/10.1109/ICCAKM50778.2021.9357709.

Taha, A.A., Malebary, S.J., 2020. An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. IEEE Access 8, 25579–25587. http://dx.doi.org/10.1109/ACCESS.2020.2971354.

Tran, L.T.T., 2021. Managing the effectiveness of e-commerce platforms in a pandemic. J. Retail. Consum. Serv. 58, 102287. http://dx.doi.org/10.1016/j.jretconser.2020.102287, URL https://www.sciencedirect.com/science/article/pii/S0969698920312959.

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., Baesens, B., 2015. APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. Decis. Support Syst. 75, 38–48. http://dx.doi.org/10.1016/j.dss.2015.04.013, URL https://www.sciencedirect.com/science/article/pii/S0167923615000846.

Wang, S., Liu, C., Gao, X., Qu, H., Xu, W., 2017. Session-based fraud detection in online E-commerce transactions using recurrent neural networks. In: Altun, Y., Das, K., Mielikäinen, T., Malerba, D., Stefanowski, J., Read, J., Žitnik, M., Ceci, M., Džeroski, S. (Eds.), Machine Learning and Knowledge Discovery in Databases. Springer International Publishing, Cham, pp. 241–252.

Webb, G.I., Hyde, R., Cao, H., Nguyen, H.L., Petitjean, F., 2016. Characterizing concept drift. Data Min. Knowl. Discov. 30 (4), 964–994.

Xuan, S., Liu, G., Li, Z., 2018. Refined weighted random forest and its application to credit card fraud detection. In: Chen, X., Sen, A., Li, W.W., Thai, M.T. (Eds.), Computational Data and Social Networks. Springer International Publishing, Cham, pp. 343–355.

Xuan, S., Liu, G., Li, Z., Zheng, L., Wang, S., Jiang, C., 2018. Random forest for credit card fraud detection. In: 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC). pp. 1–6. http://dx.doi.org/10.1109/ICNSC.2018.8361343.

Yang, Q., Hu, X., Cheng, Z., Miao, K., Zheng, X., 2015. Based big data analysis of fraud detection for online transaction orders. In: Leung, V.C., Lai, R.X., Chen, M., Wan, J. (Eds.), Cloud Computing. Springer International Publishing, Cham, pp. 98–106.

Yin, S., Liu, G., Li, Z., Yan, C., Jiang, C., 2020. An accuracy-and-diversity-based ensemble method for concept drift and its application in fraud detection. In: 2020 International Conference on Data Mining Workshops (ICDMW). pp. 875–882. http://dx.doi.org/10.1109/ICDMW51313.2020.00125.

Zamini, M., Montazer, G., 2018. Credit Card Fraud Detection using autoencoder based clustering. In: 2018 9th International Symposium on Telecommunications (IST). pp. 486–491. http://dx.doi.org/10.1109/ISTEL.2018.8661129.

Zhang, X., Wang, L., 2008. Key technologies for security enhancing of payment gateway. In: 2008 International Symposium on Electronic Commerce and Security. pp. 743–748. http://dx.doi.org/10.1109/ISECS.2008.37.

Zhao, P., Ding, Z., Wang, M., Cao, R., 2019. Behavior analysis for electronic commerce trading systems: A survey. IEEE Access 7, 108703–108728. http://dx.doi.org/10.1109/ACCESS.2019.2933247.

Zhao, J., Lau, R.Y., Zhang, W., Zhang, K., Chen, X., Tang, D., 2016. Extracting and reasoning about implicit behavioral evidences for detecting fraudulent online transactions in e-Commerce. Decis. Support Syst. 86, 109–121. http://dx.doi.org/10.1016/j.dss.2016.04.003, URL https://www.sciencedirect.com/science/article/pii/S0167923616300562.