Text Mining in Cybersecurity: A Systematic Literature Review

LUCIANO IGNACZAK, GUILHERME GOLDSCHMIDT, CRISTIANO ANDRÉ DA COSTA, and RODRIGO DA ROSA RIGHI, Laboratory of Software Innovation, Unisinos University, Brazil

The growth of data volume has changed cybersecurity activities, demanding a higher level of automation. In this new cybersecurity landscape, text mining emerged as an alternative to improve the efficiency of the activities involving unstructured data. This article proposes a **Systematic Literature Review** (**SLR**) to present the application of text mining in the cybersecurity domain. Using a systematic protocol, we identified 2,196 studies, out of which 83 were summarized. As a contribution, we propose a taxonomy to demonstrate the different activities in the cybersecurity domain supported by text mining. We also detail the strategies evaluated in the application of text mining tasks and the use of neural networks to support activities involving unstructured data. The work also discusses text classification performance aiming its application in real-world solutions. The SLR also highlights open gaps for future research, such as the analysis of non-English content and the intensification in the usage of neural networks.

CCS Concepts: • **Security and privacy**; • **Computing methodologies** → **Natural language processing**; • **General and reference** → *Surveys and overviews*;

Additional Key Words and Phrases: Cybersecurity, text mining, natural language processing, systematic literature review

ACM Reference format:
Luciano Ignaczak, Guilherme Goldschmidt, Cristiano André da Costa, and Rodrigo da Rosa Righi. 2021. Text Mining in Cybersecurity: A Systematic Literature Review. *ACM Comput. Surv.* 54, 7, Article 140 (June 2021), 36 pages.
https://doi.org/10.1145/3462477

1 INTRODUCTION

The security area is associated with the protection of different types of assets. Cybersecurity is a specialized area that aims to protect individuals, organizations, and nations that hold cyberspace presence [179]. These assets can be affected by cyberattacks that can compromise infrastructure, end devices, or information. Organizations and nations currently rely on frameworks[1] to establish

[1] An example is Cybersecurity Framework, available at https://www.nist.gov/cyberframework.

Authors' address: L. Ignaczak, G. Goldschmidt, C. Andre da Costa, and R. da Rosa Righi, Laboratory of Software Innovation, Unisinos University, São Leopoldo, Brazil; emails: lignaczak@unisinos.br, guigoldschmidt@edu.unisinos.br, {cac, rrighi}@unisinos.br.

0360-0300/2021/06-ART140 $15.00
https://doi.org/10.1145/3462477

processes to protect them against cyber threats. They can also use well-known security mechanisms to mitigate cyberattacks, such as firewalls, intrusion prevention systems, and endpoints. However, the number of incidents presented in different reports related to cybersecurity demonstrates that the current processes and mechanisms are not adequate to avoid the success of cyberattacks [77, 166].

Moreover, in many cases, cybersecurity remains focused on protecting organizations against traditional threats, but cybercrime involves personnel who are able to launch highly sophisticated attacks [93]. Nowadays, cybersecurity often needs to perform activities related to different stages of a cyberattack. These activities include collecting intelligence information to identify threats, predicting and avoiding cyberattacks, protecting assets, and investigating cybersecurity incidents. Dealing with advanced threats and performing these activities at the same time require a well-trained cybersecurity team. However, building a highly skilled team can be challenging, as $(ISC)^2$ [82] estimates a worldwide gap of approximately four million cybersecurity professionals.

Many cybersecurity activities involve the analysis of unstructured data. This type of data has increased in recent years, influenced by the web and social networks [8]. Today, most data fit in this category [114]. According to a recent report by Reinsel et al. [146], the rise of data will continue, and the volume of data produced in 2025 will reach 163 zettabytes. This report also confirms that the percentage of unstructured data will increase continuously. The growth of activities associated with cybersecurity and the rise of unstructured data impose techniques to gather useful security information from massive data. The research area addressing this challenge is text mining, defined as the process of extracting knowledge from textual data [86]. Text mining tasks, such as classification and clustering, can support many cybersecurity activities, improving the organizations' capability to deal with cyber threats.

Several secondary studies addressing cybersecurity [29, 84, 162] and text mining applications [80, 97, 127] have been published in recent years. However, to the best of our knowledge, there is no secondary study that gathers state-of-the-art research to explain cybersecurity activities covered by text mining and details how text mining tasks are applied. This study intends to fill this research gap by a consistent **Systematic Literature Review** (**SLR**) through which we selected, analyzed, and summarized 83 studies. This SLR aims to address the following research questions:

—RQ1: How would be the taxonomy representing the application of text mining in the cybersecurity domain?
—RQ2: In which contexts has cybersecurity applied text mining?
—RQ3: What strategies of text mining have been utilized?
—RQ4: How has text classification performed in the cybersecurity domain?
—RQ5: How has the cybersecurity industry been applying text mining in real-world solutions?

In order to answer the research questions and provide valuable content to the scientific community, we extracted useful information from the selected studies. We also analyzed other sources to present the text mining application in the industry. The contributions of the SLR could be summarized as follows:

(1) We propose a taxonomy to classify the use of text mining in the cybersecurity domain (in Section 3).
(2) We demonstrate the industries covered and the types of content analyzed in the studies (in Section 4).
(3) We present the strategies used to apply text mining tasks in the cybersecurity domain (in Section 5).

(4) We present the performance of text classification in different cybersecurity activities (in Section 6).
(5) We cover the application of text mining in several real-world solutions developed by the cybersecurity industry (in Section 7).
(6) We introduce open gaps to be explored in the intersection between cybersecurity and text mining research (in Section 8).

The remainder of this article is organized as follows: Section 2 introduces cybersecurity and text mining concepts. Section 3 proposes a taxonomy with the classification resulting from a systematic approach. Sections 4 and 5 explain and discuss the contexts and strategies of text mining application in the cybersecurity domain. Section 6 discusses the performance of text classification. Section 7 analyzes the use of text mining by the cybersecurity industry. Section 8 lists the open gaps that new studies can target. Section 9 summarizes the article's contribution and discusses the main findings. Finally, we present the method used in study selection, threats to the SLR's validity, and the selected studies in supplementary online material.

2 BACKGROUND

This study joins efforts to present the current status of the integration between text mining and cybersecurity. This section addresses the definitions and processes associated with both areas, aiming to present fundamental explanations to the audience.

2.1 Cybersecurity

Analysis of the technical literature can create a misunderstanding about cybersecurity since sometimes cybersecurity and information security can be presented as synonyms. ISO 27032 [79] defines cybersecurity as security in cyberspace. ISO explains cyberspace as a complex environment associated with people, software, and services on the internet, connected through devices and networks, which do not exist physically. Based on ISO definitions, it is possible to understand that cybersecurity protects against virtual threats or cyber threats. A complementary perspective is presented in Von Solms and Van Niekerk [179], distinguishing information security and cybersecurity based on asset protection. The authors emphasize that the assets defended by cybersecurity can range from a person to household appliances, to the interests of society and critical infrastructures. In contrast, information security targets the assurance of information properties.

The permanent adoption of new digital practices by users and the increase of in-digital services provided by companies, combined with the connection of new types of infrastructures to cyberspace, have significantly increased the number of threats that cybersecurity needs to manage. In addition to traditional security issues, including data breach [40], malicious software [68], and denial of service [30], cybersecurity deals with new kinds of threats. In an election scenario, cybersecurity has to protect the election system against traditional cyberattacks [151] and, additionally, address misinformation that can be spread on social networks to delegitimize a candidate[41].

As cybersecurity protects people, the information delivered to them should also be protected because misinformation and disinformation can affect social context, impacting relationships among people and organizations or governments [34, 189]. Attitudes expressed in cyberspace can also affect people. One example is cyberbullying, defined as aggressive behavior toward a victim due to some abuse of power through technological devices [155], that must be addressed by cybersecurity [51]. Other online behaviors on the cybersecurity radar are cyberstalking [156] and sextortion [136].

Cybersecurity should also monitor traditional and emerging threats in cyberspace, offering information about which threats can put the organization's infrastructure at risk. The activity

dealing with this challenge is **Cyber Threat Intelligence** (**CTI**). CTI aims to prevent cyberattacks based on the knowledge about threats [172]. It is important to differentiate threats from attacks. According to International Organization for Standardization [78], a threat is an event with the potential to cause an unwanted incident. In contrast, an attack is an attempt to destroy, expose, alter, disable, steal, or gain authorized access to an asset. So, while a threat is something with the potential to produce some harm to an organization's asset, an attack is a concrete action that can result in damage.

Cyberattacks have demonstrated a steady evolution in recent years. Today, highly skilled and well-funded cybercriminals work in a cooperative and organized model to launch cyberattacks against organizations aiming to maximize their profits [73, 145]. The cybercrime ecosystem also maintains specialized businesses offering tools and infrastructure to support other cybercriminals in performing cyberattack tactics [73]. Additionally, new technologies and models (i.e., the **Internet of Things** (**IoT**) and cloud computing) expanded the attack surface. They made organizations more susceptible to cyberattacks, enlarging the number of sectors targeted by cybercriminals. One of the consequences is cyberattacks targeting industrial cyber-physical systems [163] and critical infrastructure [167].

Cyberattacks explore existing vulnerabilities in software, network services, and humans. A vulnerability is a weakness that can be explored by a threat [78]. Kuhn et al. [95] analyzed that recent softwares are getting more complex, increasing the area for discovering vulnerabilities. The authors examined a vulnerability database and reported that about 90 percent of vulnerabilities registered in 2016 are associated with medium to high severity [95]. In addition, the adoption of social networks, mobile communication, and IoT has produced a large amount of sensitive information about people. This information has been collected and used for cybercriminals to design social engineering attacks to exploit human vulnerabilities [185].

When a cyberattack succeeds, it produces a cybersecurity incident. An incident is an event unrelated to the regular operation of a service that can cause interruption or reduce quality [81]. Although cyberattacks are a significant source of cybersecurity incidents, they can also be caused by system failures, natural phenomena, human errors, and third-party failures [60]. The main concern for organizations is the impacts produced by cybersecurity incidents. According to Couce-Vieira et al. [46], an incident can reduce the organization's income due to loss of sales, contracts, and funding. Cybersecurity incidents can also impact organization valuation, and negative information about the incident can negatively affect the organization's stock price [159].

The impacts of cybersecurity incidents demonstrate the importance of defenses, which are based on security controls. Security control is an action, device, procedure, or technique aiming to eliminate or prevent a threat, vulnerability, or attack [137]. In the current scenario, controls must extrapolate traditional security mechanisms designed to protect a logical perimeter and secure the organization's infrastructure in the cloud [65] and new technologies supporting current business models [75, 90]. Additionally, as aforementioned, cybercriminals' weaponry and skills are becoming increasingly sophisticated, so studies have evaluated the use of artificial intelligence to support different cybersecurity controls [103].

2.2 Text Mining

Text mining is a process that extracts useful knowledge from data sources, similar to data mining. Unlike data mining, the data sources used in the text mining are formed by document collections, and the process extracts knowledge from unstructured data present in the documents [55]. The unstructured data analyzed by text mining usually comprises texts written in a natural language [86]. However, the process can also be applied to extract knowledge from semi-structured data, like HTML and XML files [193]. At a high level, the core process of text mining can be divided into

three stages [115]: (i) corpus definition; (ii) preprocessing; and (iii) knowledge extraction. The first stage, corpus definition, collects relevant documents related to a specific issue.

The second stage consists of applying preprocessing techniques to transform raw text data into an intermediate format, enhancing the extraction of patterns by a task [22, 120, 173]. Preprocessing is required to improve the quality and usability of unstructured data. Usually, a unique preprocessing technique is not enough to prepare the data for the text mining tasks, so the use of multiple techniques sequentially is pretty common [55]. There are several techniques for preparing unstructured data, and the decision on which to select depends on the text mining approach used in the process. One common approach to represent a document for knowledge extraction is **Bag-of-Words** (**BoW**), a method to map the document into a fixed-length vector [197].

The BoW approach analyzes the frequency of terms in a document—often a word. The preprocessing technique that breaks up the text document into words or tokens is called tokenization [115]. An alternative to a single word frequency is N-gram, which tokenizes groups of "n" consecutive words [21]. In cases where the BoW approach calculates each word's frequency, a preprocessing technique is necessary to remove frequent words adding no value to the analysis, such as prepositions and articles. These words are named stopwords, and the technique is named stopword removal [21]. One challenge in text mining analysis is dealing with word inflection. An alternative is a stemming function that reduces variants of a word to their root form. A more complex option is lemmatization, which performs morphological analysis of each word to determine the lemma [153].

The process' last stage applies the text mining task to extract the knowledge from preprocessed documents. The text mining task can vary according to the type of knowledge targeted. According to Miner et al. [115], the knowledge extraction can be divided into the following categories: prediction, clustering, association, and trend analysis. Each category is supported for one or more text mining tasks. In this section, we explain the most used tasks in the SLR corpus.

A well-known task is text classification, a supervised learning that matches new observations according to a dataset properly categorized and labeled [113]. This dataset contains a set of documents, and each document contains a label associating it with a class value. This dataset is referred to as training data, and it is necessary to build a classification model. The model associates document features with one or more classes, and the model is used to predict the class of a new set of documents, named testing data [8]. Text classification can use shallow or deep learning. Shallow learning is based on traditional classifiers like naive Bayes, decision tree, and random forest and employs statistical models [96]. Shallow learning demands manual feature extraction to train and evaluate the classifier [17]. Deep learning explores powerful neural networks that automatically learn high-level features from a text document, dispensing manual feature extraction [48].

Text clustering is an unsupervised learning and, different from text classification, the task does not perform a training phase using labeled records from a dataset. Instead, it analyzes a document and identifies relevant terms to assign a weight. Text clustering uses term weight schemes to calculate the similarity between documents and group them into clusters [3]. The traditional text clustering approach does not consider the terms' semantic relationship in a document, so they do not deal with synonyms and polysemy. In the semantic clustering approach, algorithms can be supported by ontologies to identify each term's meaning, considering the semantic of nearby terms in clustering organization [123]. More recently, studies have evaluated integrating neural network techniques and traditional algorithms to improve text clustering [61, 191].

Topic modeling, another unsupervised learning, uses statistical methods to analyze and discover main themes pervading a collection of documents [31]. These themes are called topics and consist of word clusters that represent the ideas discussed in text data. Topic modeling uses the co-occurrence of topics to identify the subject associated with documents [83]. The most frequently

used topic modeling algorithm is **Latent Dirichlet Allocation** (**LDA**), which implements the BoW approach [175]. It is important to note that LDA instability is a known concern and should be considered by researchers [9]. Current algorithms aim to improve the topic discovery in specific analyses, such as short text and time-based topic modeling [175].

Information extraction aims to discover structured information from unstructured content [8] and is usually performed through two subtasks: **Named Entity Recognition** (**NER**) and relation extraction. These subtasks identify known entities like people, companies, and places in a text and infer relationships among them [63]. Initially, NER was developed as a rule-based system, but state-of-the-art implementations rely on statistical machine learning methods [8]. A more recent approach transformed NER into a classification problem and uses supervised learning to identify the entities. The approach is called **Named Entity Recognition and Classification** (**NERC**) [122].

The steady increase of social media types enabled individuals to share their opinions about many topics, such as individuals, companies, and products [10]. Sentiment analysis identifies the sentiments expressed by individuals in text-written opinions, revealing if they contain positive or negative emotions [21, 111]. According to Medhat et al. [111], the task can identify the sentiment considering the whole opinion or recognize emotions expressed in distinct sentences. Similar to information extraction, sentiment analysis can be performed using a supervised learning approach [21]. In this approach, linguistic features are applied in a pre-trained machine learning classifier to identify the sentiment polarity and emotion intensity [1].

The colossal volume of unstructured data available in different digital formats resulted in a demand for new methods to retrieve useful information effectively. Information retrieval is the text mining task that deals with this challenge. Its objective is to find information based on a user's query in a document collection [117]. As information retrieval targets unstructured data, a user's query can return multiple results, so the task must compute the results' relevance and present them ordered by a relevance score. Although traditional models to rank documents by relevance are presented in Ceri et al. [35], the score computation in modern search engines is often supported by machine learning [7, 102]. Neural networks have also been applied to estimate the relevance score and rank documents in information retrieval [62, 116].

3 TAXONOMY

We identified studies evaluating the application of text mining in several cybersecurity activities, so we propose a three-level taxonomy to classify the studies and answer the first research question (RQ1). Figure 1 shows the proposed taxonomy. In Section 2.1, we presented the definitions of vulnerability, threat, attack, control, and incident, so we use them as cybersecurity domains to create the taxonomy's first level. These five cybersecurity domains cover the use of text mining tasks from the discovery of threats until the lessons learned from an incident. The next level expands each domain considering a known process or the purpose of text mining application, offering a more granular categorization. In the last level of the taxonomy, we present the cybersecurity activities or malicious behaviors associated with the upper level and link the SLR studies.

3.1 Vulnerability

Cybercriminals constantly look for vulnerabilities in their targets, so an organization should implement a vulnerability management process to identify vulnerabilities, define the severity, and decide the priorities [143]. We use the vulnerability management process and analysis types to classify the studies in this domain.

Text mining can analyze source codes to identify vulnerabilities [24, 128] or extract knowledge from vulnerability databases to discover new weaknesses in assets [88]. After discovering existing

Fig. 1. Proposed taxonomy of text mining application in the cybersecurity domain.

vulnerabilities, it is necessary to create a plan and prioritize the most critical vulnerabilities based on a severity score. Text mining supports vulnerability discovery by calculating a severity score [160, 186] or categorizing new vulnerabilities according to their characteristics [187]. Text mining can also help an organization understand the evolution of vulnerabilities to improve vulnerability management [188]. Finally, Syed et al. [165] analyzed vulnerabilities communication on Twitter and presented the most spread vulnerabilities on the platform, as well as the most likely retweeted.

3.2 Threat

Organizations need to deal with dozens of different threats in cyberspace, and many of them leave digital traces in an unstructured format. We classified studies according to the application of text mining in activities to prevent or detect a threat. The prevention targets information about threats that allows one to anticipate an event and implement countermeasures proactively. Situational awareness enables cyber threats prevention, and text mining can support this activity through online communities [11, 28, 47, 106, 108] and social networks [20, 101, 140] monitoring. Furthermore, text mining is also used to issue alerts about threats [118, 149]. Unlike prevention, detection allows an organization the knowledge about malicious behaviors that can affect it in the future. Some examples of threats detected by text mining are cyberbullying [121, 124], botnets [94], misinformation [194], and scams [177].

3.3 Attack

In this domain, studies evaluated text mining to implement cyberattacks and support defensive activities against them. An attacker can use text mining to elaborate social engineer attacks [52]

or to detect content hidden by steganography [38, 39, 190]. In both cases, the malicious actor applies text mining to attack a system or evaluate a security control. Text mining can detect attacks against a system or organization through the identification of **indicators of compromise** (**IoC**) [130], malicious behaviors indicating the initial stage of an attack [91], malicious commands in scripts [69], and the spread of fake news [12]. The prediction is also possible by detecting malicious events indicating a future attack [70, 131, 135, 141]. Lastly, text mining allows the analysis of public databases to identify useful attack patterns [5].

3.4 Control

In the control domain, we classified the studies in four categories of security controls: access control, malicious behavior detection, **data leakage prevention** (**DLP**), and content protection. Access control uses text mining to enable the automation of activities, as the creation of access control rules in a mechanism based on policies written in natural language [15, 125, 154] or the generation of access control policies based on electronic documents [19, 126].

Controlling web access is very challenging to organizations, and the most common controls use URL blacklists, which can fail due to the number of new sites created daily. Text mining can control web traffic by analyzing the site's content instead of just the URL [76]. Access control demands the identification of the entities using some authentication method, and Toor et al. [171] proposed a protocol to allow visual authentication supported by text mining.

Organizations also implement security controls to detect different types of malicious behaviors, such as phishing in websites [67, 119] or electronic messages [168]. Another type of malicious behavior is SPAM messages, which corresponded to 57.64% of the Internet's electronic mail traffic in the second quarter of 2019 [176]. Studies proposed text mining tasks to combat SPAMs in electronic messages [201], social networks [199], reviews in websites [147], SMS [87, 161], or in multiple platforms [6, 53]. Malware is also a constant concern, and security controls can use text mining to protect organizations against different malicious codes [44, 113].

The rise of data volume stored by an organization can increase data leakage risks, which can be mitigated using DLP methods [14]. Different approaches evaluated text mining to detect and prevent data leakage based on information context included in electronic documents [74, 89] or access control rules [59, 183]. Another model sanitizes the document information according to the level of access [49]. Organizations implement security classification to get visibility about the sensitivity of documents. Text mining can analyze a document's content and automatically assign the security level [13, 169, 195]. Text mining can also monitor email accounts and identify changes in user behavior to predict data leakage [110].

Security controls can apply text mining to protect content and avoid unauthorized access to information. Text mining can support steganography techniques based on word substitution because the techniques need to understand the information's context to perform substitutions [25, 36]. Content protection is also assured using watermarks [112] or anonymizing strings that can identify a user's interests [107].

3.5 Incident

Cybersecurity can deal with security events by implementing an incident response process. We used the process defined by Cichonski et al. [42] to classify the studies in the domain. Detecting an incident can be very challenging for organizations, but text mining can monitor different public data sources and issue alerts about incidents related to an organization [54, 99]. After an incident detection, an organization needs to analyze and document the details of the incident, and the application of text mining can make the triage easier [18, 26, 27]. Text mining is also used to identify the authorship of some artifact involved in the incident [4, 104, 142, 152].

Table 1. Types of Content Analyzed in the Cybersecurity Domain

Content	Studies	Total
Electronic Document	[13, 15, 25, 27, 37–39, 44, 49, 59, 74, 89, 112, 125, 126, 130, 131, 144, 154, 169, 171, 184, 190, 195]	24
Social Network	[6, 12, 45, 52, 91, 94, 99, 101, 104, 118, 124, 135, 140, 142, 149, 164, 165, 194, 199]	19
Online Forum	[11, 12, 20, 45, 47, 54, 70, 106, 108, 121, 149]	11
Web Page	[36, 67, 72, 101, 119, 141, 147, 177, 184]	9
Security Database	[5, 88, 160, 186–188]	6
Electronic Mail	[53, 110, 142, 168, 201]	5
Source Code	[4, 24, 113, 128]	4
SMS	[6, 53, 87, 161]	4
Online Community (IRC)	[28, 152]	2
Forensic Image	[18, 26]	2
Metadata	[76, 113]	2
URL	[76]	1
Behavior Profile	[183]	1
Powershell Command	[183]	1
Medical Records	[19]	1
Search String	[107]	1

The incident process establishes that an organization should learn from each incident, and the event should promote cybersecurity improvements [72]. The incident impact is also a concern to organizations, and text mining can estimate the damage to their reputation based on social media comments [45, 164] and financial losses [37]. The analysis of incidents and the extraction of useful information is also possible with text mining [144]. Finally, Wang et al. [184] studied market reactions after a security breach announcement and the data breaches' impact on company valuation.

4 CYBERSECURITY CONTEXTS APPLYING TEXT MINING

Aiming to answer the second research question (RQ2) in this section, we analyzed the following aspects associated with the application of text mining in the cybersecurity domain: (i) the type of unstructured content analyzed; (ii) the industry targeted in the study; (iii) the technology platform targeted in the study; and (iv) the dataset language.

We organized the content evaluated in Table 1. The table shows that social networks, online forums, and electronic documents as the main types of content analyzed. Among the studies based on social networks, the first choice was Twitter. The authors highlighted that the platform allows crawling tweets using a time interval or a hashtag. Another reason was the number of datasets containing tweets available on the internet. Studies also crawled different types of online forums to mine text. Although most studies crawled surface forums, some of them recovered information from forums available on Dark Web, intra-company, and online games. Lastly, we classified content as an electronic document in cases where authors used a document format (i.e., Microsoft Word) or referred to the content evaluated in a generic form (i.e., message).

Another information related to the context is the specific industry or the technology platform targeted by the studies. In both cases, we only considered works whose research was focused on applying text mining in the industry or platform. Some studies proposed general text mining approaches and only evaluated them in a context, so they were not considered. We present the results

Table 2. Industries and Technologies Targeted in the Study

Industry/Technology	Studies	Total
Digital Forensics	[4, 18, 26, 27, 106, 142, 152]	7
Mobile Devices	[53, 87, 113, 161]	4
Cloud Computing	[59, 171, 183]	3
Financial Services	[37, 119, 130]	3
Software Development	[128, 188]	2
Big Data	[74, 76]	2
Critical Infrastructure	[108]	1
E-Commerce	[67]	1
Hospitality	[147]	1
Human Resources	[177]	1
Online Game	[121]	1
Operating System	[69]	1
Healthcare	[19]	1
Cyber-Physical System	[5]	1

in Table 2. Based on the table, we understand that text mining can contribute to cybersecurity activities in many areas. We verify the interest in developing research to deal with cybersecurity challenges in industries like finance, hospitality, and health. We also identify studies driven to technologies like mobile and cloud computing.

In order to know another aspect related to the text mining applications in the cybersecurity domain, we extracted the languages of the datasets used in the selected studies. We identify that many studies did not specify the language in the section describing the methodology, only presenting details of the dataset used, such as the dataset's name or the sites in which the study crawled the data. We discovered that 70 studies focused on the English language, of which 35 we needed to infer based on the dataset content. Further, the SLR corpus only targeted six other languages: Arabic, Chinese, Italian, German, Russian, and Turkish.

Figure 2 presents the association between the taxonomy proposed and three aspects of the context discussed in this section: type of content, industry, and technology. We established that even if the content format is analyzed just by one study, it is associated with the taxonomy. The same guideline was applied to industry and technology. We diagnosed that most of the activities related to threat prevention and detection used social networks as the primary source of content, followed by internet forums. Scams were the only type of threat whose study is not based on social networks. We think this content could be targeted in future studies because we can find it on social platforms. Otherwise, vulnerability discovery studies regularly use security databases like **Common Vulnerabilities and Exposures** (**CVE**) and **National Vulnerability Database** (**NVD**), despite that many cybersecurity professionals publish vulnerability information on social networks.

5 TEXT MINING STRATEGIES IN CYBERSECURITY

This section introduces the different text mining strategies evaluated in the cybersecurity domain and answers the third research question (RQ3). We understand as strategies the three following aspects: (i) which text mining tasks were individually applied, covered in Section 5.1; (ii) which text mining tasks were combined and how they were applied together, discussed in Section 5.2; and (iii) in which cybersecurity activities text mining was supported by neural networks, covered

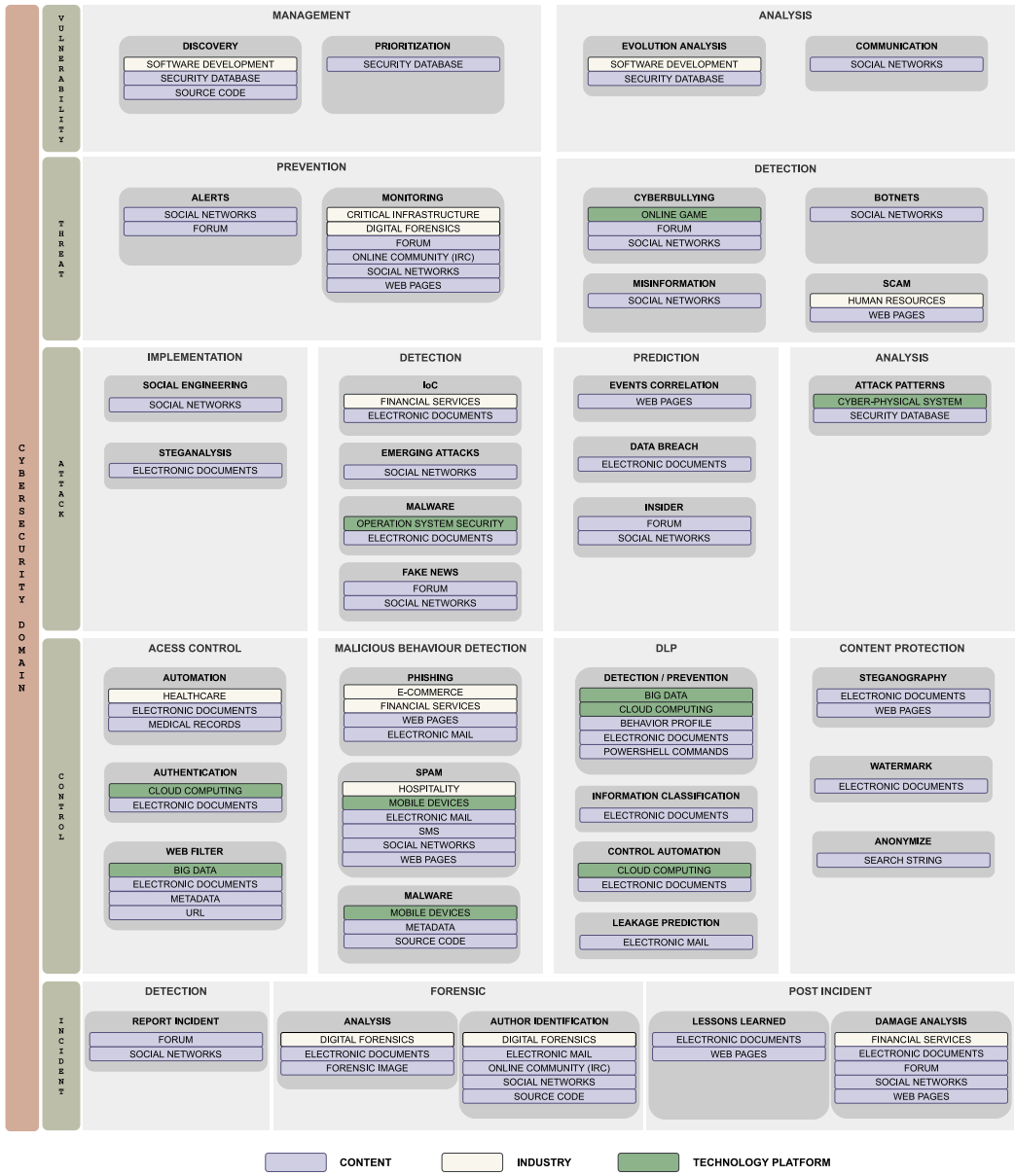

Fig. 2. Association between the proposed taxonomy and the contexts.

in Section 5.3. For each aspect, we associate the text mining application with the cybersecurity activity and briefly summarize some studies to exemplify how the tasks are applied.

The initial analysis focused on which text mining tasks were used in the selected studies. A summary presenting the number of studies associated with each task is demonstrated in Figure 3. The figure highlights text classification as the predominant task in the domain. Considering the 55 studies that applied text classification, 35 studies evaluated the task performance singly, whereas 20 studies combined text classification with other text mining tasks.

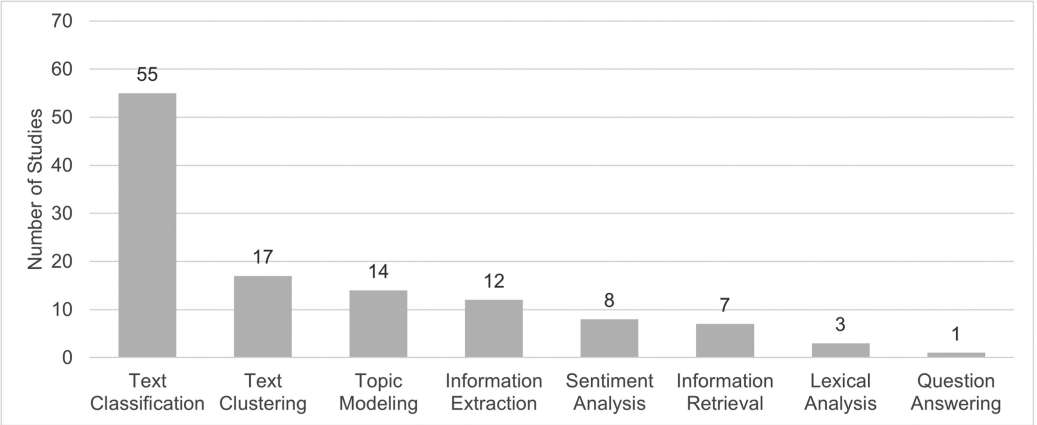

Fig. 3. Number of studies applying each text mining task.

We analyzed SLR associated with the application of text mining in other domains, and text classification is also the prevalent task in studies related to psychiatry [2], education [56], and the identification of papers for inclusion in systematic reviews [133]. However, the ratio of studies applying text classification in cybersecurity is more significant than in the other three areas. On the other hand, a study in innovation research presented text clustering as the most used task [23], and information extraction is the most common in the agricultural domain [50]. Lastly, a systematic review that assessed text mining in the financial sector did not list text classification among the used tasks [139].

In order to demonstrate the application of text mining in the cybersecurity domain, Figure 4 presents the association between the tasks and the proposed taxonomy. Our analysis identified that none of the studies applied text summarization to support cybersecurity activities. Albeit the absence of text summarization can be the difficulties cited in Gambhir and Gupta [57], we think that studies in the cybersecurity domain should consider evaluating the task. We give some suggestions about the application of text summarization in Section 8.

5.1 Text Mining Tasks

This section presents in which cybersecurity activities studies evaluated the performance of a single text mining task. As we identified that some tasks were just evaluated combined with others, they are not covered in this section.

5.1.1 Text Classification. Many cybersecurity activities must have an active posture, so the cybersecurity team should detect threats, attacks, and incidents as fast as possible. As cybersecurity deals with a massive volume of data, studies evaluated supervised learning to analyze whether, based on known patterns, it is possible to identify malicious behaviors [6, 12, 44, 54, 69, 91, 94, 121, 177, 194]. It is one reason for the predominance of text classification. Another essential attribute of cybersecurity is predicting malicious actions, and studies evaluated how text classification could improve cybersecurity effectiveness by analyzing data from distinct data sources [70, 135, 141]. Moreover, publicly available datasets contain several types of malicious behaviors that studies can use to train supervised algorithms and perform different evaluations [44, 113, 119]. We can yet consider another reason mentioned by Zhai and Massung [196], that it is easier to perform classification than other text mining tasks.

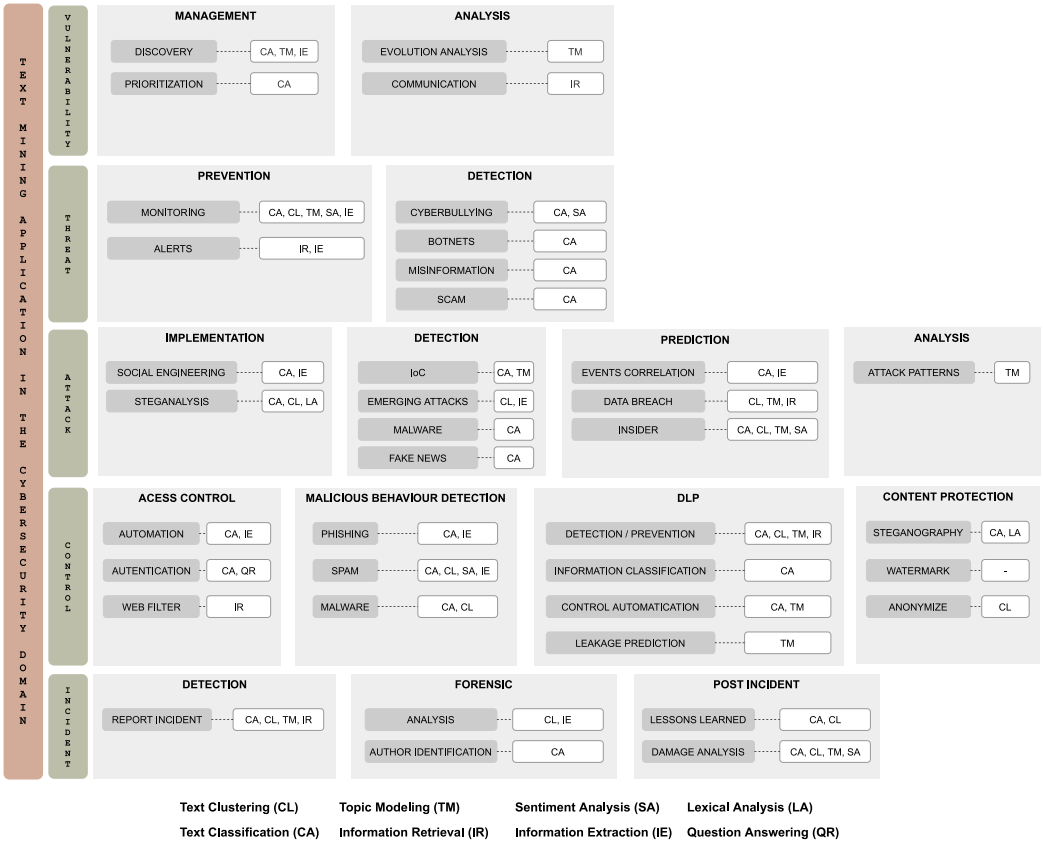

Fig. 4. Association between the proposed taxonomy and text mining tasks.

Although the examples mentioned reasons for the use of text classification in specific domains, the analysis of Figure 4 demonstrates the broader application of the task in the proposed taxonomy. In the vulnerability domain, studies applied text classification to define the priority order to patch vulnerabilities in third-party applications [160, 186, 187]. The studies used vulnerabilities databases (i.e., NVD) to extract features from textual fields and train a classifier to assess the severity [186] or categorize [187] a vulnerability.

In the threat domain, the work of Vidros et al. [177] intended to identify scams in online recruitment. The study examined malicious behavior comprising rogue job advertisements to collect information from applicants. The study evaluated two approaches using the EMSCAD, a dataset containing real-life jobs ads. The first experiment assessed the BoW approach using six traditional classifiers. The second approach transformed each dataset record into a vector of binary features.

Studies evaluated text classification performance in the control domain to identify undesired messages like spam [87] and phishing [67, 119]. Moghimi and Varjani [119] proposed two feature sets to detect phishing on internet banking websites. In order to create the first set, the authors adopted the five most frequently used features in related works. Additionally, the authors proposed two features not present in related works to create the second set. The feature sets were used to train an SVM classifier to identify phishing pages. The study pointed out that text classification was effective against phishing attacks, and highlighted the task was able to detect zero-day phishing attacks. Other studies assessed text classification to assign the security classification level to electronic documents and prevent data leakage [13, 74, 169, 195].

Table 3. Top Five Traditional Classifiers Applied in the Studies

Classifier	Studies	Total
Support Vector Machine	[6, 28, 37–39, 47, 53, 54, 67, 106, 113, 119, 135, 147, 152, 186, 187, 190, 194]	19
Naïve Bayes	[12, 19, 20, 44, 53, 54, 70, 87, 106, 113, 130, 135, 147, 154, 177]	15
Decision Tree	[19, 37, 47, 52, 104, 130, 135, 147, 160, 184]	10
Random Forest	[6, 44, 54, 72, 113, 130, 160, 161, 177]	9
K-Nearest Neighbor	[19, 47, 54, 106, 130, 140, 194, 195]	8

In the incident domain, studies evaluated text classification to identify the cyber attack authorship based on online messages [104, 152] and source code [4]. The study of Li et al. [104] extracted 233 stylometric features from Facebook's posts. Most features were based on other studies, and the authors added six novel features related to social network posts. The study evaluated the performance of multiple traditional classifiers and a two-layer neural network to identify posts' authorship. The task was also used to analyze security incidents and create a list of suggestions for cybersecurity awareness programs [72].

Since most studies applied text classification to support cybersecurity activities, we analyzed them to identify the most used classifiers. Table 3 presents the five most-used classifiers found in the corpus. We identified the **Support Vector Machine** (**SVM**) as the most used classifier, mentioned in 19 studies. When analyzing the classifiers applied to support text classification in other domains, we identified that SVM was also the most used classifier listed in two SLRs in the financial area [97, 127] and to identify relevant studies in systematic reviews [133]. Although SVM had the preference by studies in this SLR, it is interesting to point out that Hartmann et al. [64] evaluated the accuracy of text classification methods on 41 social media datasets and underlined that SVM did not outperform other classifiers.

5.1.2 Text Clustering. The SLR identified text clustering as the second most applied text mining task. As text clustering can identify similarities, studies evaluated the task to detect and prevent malicious behaviors. Some studies applied classification and clustering in a common dataset and compared the results obtained from each task [113, 147]. In these studies, text clustering played a central role in detecting cyber threats, but Milosevic et al. [113] concluded that clustering is not accurate enough to be applied in detection activities. Other works designed multi-task processes to detect malicious behaviors, and text clustering supported a stage of the process, as demonstrated in Section 5.2. In the proposed taxonomy, studies evaluated the application of text clustering singly in the control and incident domains.

In the control domain, studies evaluated the task to prevent data leakage [89] and protect user privacy when performing search queries [107]. Katz et al. [89] proposed a two-phase context model to deal with accidental and intentional data leakage. The model designed a learning phase to generate a representation of confidential content. This phase applies text clustering to identify subjects represented on confidential and non-confidential documents. For each cluster, the model discovers confidential terms and analyzes their context. The detection phase assigns a document to relevant clusters and calculates a confidentiality score. The score is based on comparing confidential and context terms present in the clusters and the document. The authors underlined that the proposed model recognized small sections of confidential content, even when they differ from examples used in the learning phase. However, they highlighted that false-positive rates could raise doubts about applying the model in real-world scenarios.

In the incident domain, studies used text clustering to enhance forensics searches [26] and identify data breaches related to personally identifiable information [144]. Beebe et al. [26] proposed a new forensic string search approach to improve the results of text searches in potential evidence. The task calculates the similarities of document vectors produced from documents obtained in a usual text search. Based on the similarity, the documents are grouped according to the thematic and ranked within each cluster. The study demonstrated that text clustering improved the effectiveness of post-retrieval searches by reducing the number of occurrences that digital forensics professionals must analyze.

5.1.3 Information Extraction. We identified two approaches in the application of information extraction singly. The first approach extracted terms and concepts from a text and used a semantic ontology to define the relationship among them [18, 118]. The second approach performed word-checking based on a handcrafted dictionary to identify malicious terms in text messages [168]. Information extraction was applied in threat prevention [118], phishing detection [168], and forensics analysis [18].

A CTI study proposed a framework to analyze Twitter and generate alerts to security analysts regarding threats and vulnerabilities [118]. The proposed framework creates a user's system profile, containing information about the operational system and installed softwares. The framework applies NER in tweets to extract terms and concepts related to security vulnerabilities and uses an ontology to identify real-world concepts and relations. The framework issues alerts to users based on their system profile and vulnerabilities information extracted from tweets. The results indicated that information extraction has the potential to be applied in CTI activities. However, the study pointed out the need for improvements to avoid unrelated alerts.

5.1.4 Topic Modeling. Although most research analyzed the combination of topic modeling with other text mining tasks, cybersecurity activities can also benefit from its exclusive application. In CTI activities, the task can identify the main topics from online communities, allowing the recognition of threads or channels associated with threats or attacks. Topic modeling can also support DLP because it can identify if the main topics of a document are related to confidential subjects.

We identified the exclusive application of the task to prevent data leakage [5] and analyze vulnerabilities [188] and attacks [5]. Two studies used time-based topic modeling. In the vulnerability domain, Williams et al. [188] analyzed the evolution of vulnerabilities and their relation to different softwares based on the NVD. The study used a temporal topical model proposed by the authors in previous work. In the attack domain, Adams et al. [5] applied topic modeling to extract information about cyberattacks from CAPEC, a database of common patterns, and linked attack patterns to a protected system. The study applied LDA to extract topics from the database and texts produced to describe the system. The approach associated cyberattacks with the system considering the distance measure between the topics learned from the system description and each attack documented in the database.

5.1.5 Sentiment Analysis. Cybersecurity needs to monitor social networks and other online communities, and sentiment analysis can identify the negative intensity of emotions and improve the discovery of threats. Two studies evaluated sentiment analysis with this goal in the threat domain [11, 108]. The research of Macdonald et al. [108] evaluated web forums to identify posts associating critical infrastructures with potential threats and used sentiment analysis to highlight the most relevant posts. The authors chose specific keywords and selected posts based on their association with hacking and critical infrastructure. A sentiment analysis tool computed the sentiment scores, and posts with negative scores should represent potential threats.

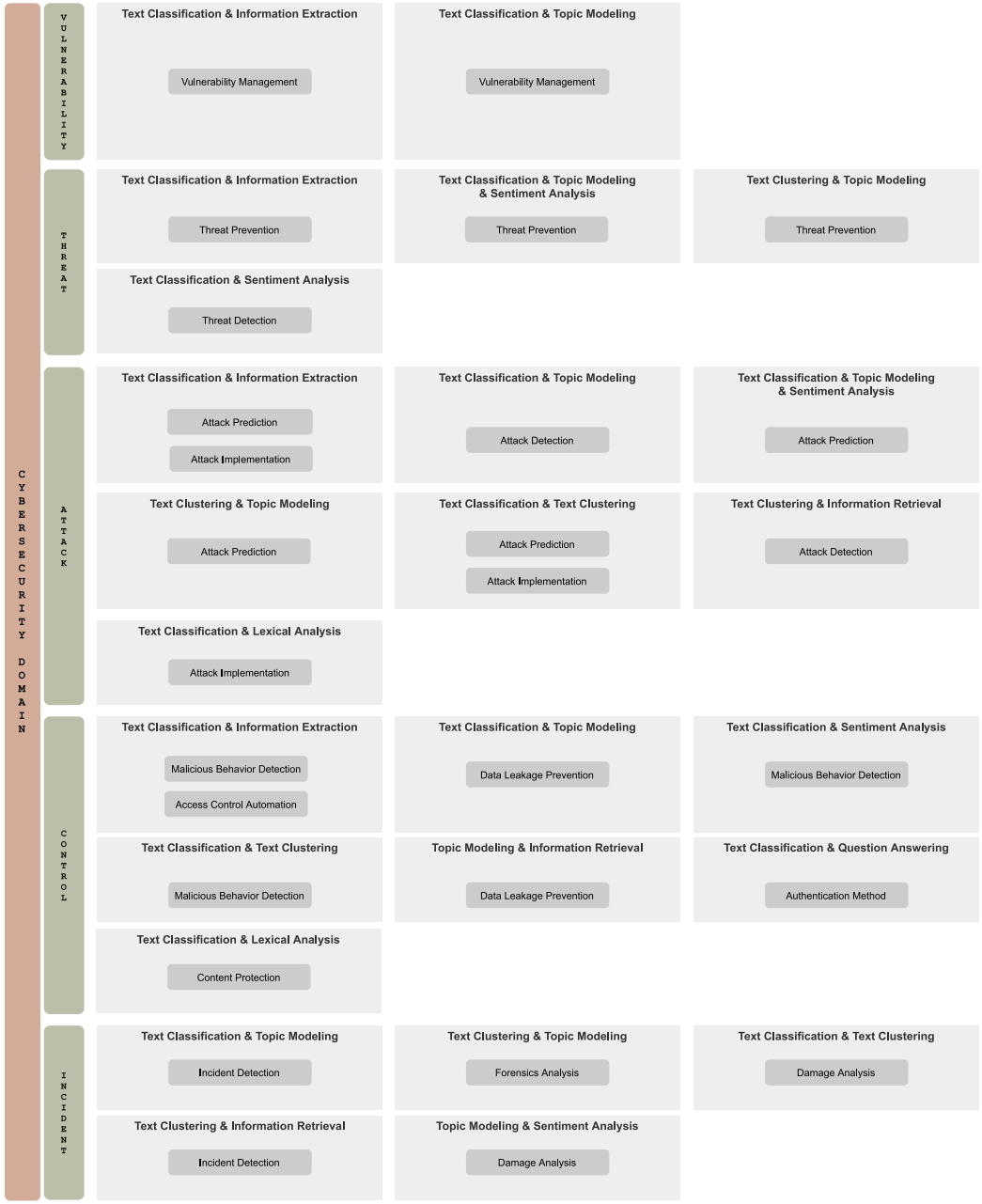

Fig. 5. Association between integrated text mining tasks and their application in the cybersecurity domain.

5.2 Integrating Text Mining Tasks

Considering the 83 studies analyzed in the SLR, 31 defined a process to apply text mining in cybersecurity comprising more than one task. Figure 5 shows the tasks applied together and associates each combination with the proposed taxonomy.

Two text mining tasks frequently used together are information extraction and text classification. Analyzing the studies, we identified the application of the tasks together in vulnerability management [88], threat prevention [20], attack prediction [141] and implementation [52], spam detection [53], and access control automation [126, 154]. We also identified three different approaches in the application of the tasks. The first used the NERC, so a classifier is trained to identify entities associated with a cybersecurity activity [88, 154]. Joshi et al. [88] applied the approach to identify entities and concepts in vulnerabilities databases and discover application vulnerabilities. The authors explained that the system met difficulties recognizing entities based on information extraction, but the results were considered promising.

In another approach, entities recognized by the application of NER are used as features to train a machine learning classifier [52, 53]. In the attack domain, Edwards et al. [52] collected information on social networks and apply NER to extract names of employees aiming to implement phishing attacks. The names allowed the identification of employees' mentions and social relations, which are features used to train the classifier. In the third approach, An and Kim [20] applied the tasks in parallel to investigate the underground economy behind cybercrime, based on a hacking community and black market forums. The study used NER to identify company names and text classification to predict messages associated with cybercrime.

Another text mining strategy is the combination of topic modeling and text classification. This combination was identified in vulnerability management [128], attack [130] and incident [54] detection, and data leakage prevention [59]. The studies of Fang et al. [54] and Noor et al. [130] applied topic modeling to obtain relevant topics from a collection of unstructured data, which were used to train machine learning classifiers. Noor et al. [130] extracted high-level IoCs from CTI documents and applied text classification to associate threat actors with cyber incidents. The results demonstrated that the text mining tasks could attribute threat actors to an incident [130].

In a distinct approach, Nembhard et al. [128] used topic modeling to categorize software projects' source code into different groups. The approach demonstrated that using both tasks was possible to detect missing terms in a source code and recommend practices to improve its security. Gonzalez-Compean et al. [59] developed a security filter acting as an intermediary to identify sensitive information in documents during sharing operations. The implementation performs a risk assessment based on the document information. The study applied topic modeling to extract topics from documents, using them to assess the risk and train a classifier to recognize sensitive information in new documents.

Two works added sentiment analysis in the combination of topic modeling and text classification, applying three text mining tasks in the process. This combination was evaluated to prevent threats [106] and predict cyberattacks [135]. Li et al. [106] applied text classification to identify publicity related to the underground economy and sentiment analysis, determining a score to measure the seller's reputation. Further, topic modeling supported the identification of the top sellers' profiles. Results outperformed benchmark methods and allowed the recognition of best-rated and least-reputable sellers of malware and stolen data.

Park et al. [135] combined the three tasks to identify malicious insiders based on tweets. Initially, the study applied sentiment analysis to determine the sentiment level in short messages. Next, the work used topic modeling to extract topics and associated distinct sentiment levels with the respective main topics. Finally, text classification was assessed to find a malicious insider based on the most frequent topics. The study also analyzed text clustering in the same stage of classification to compare the tasks. Authors underlined that text classification achieved high accuracy, while clustering was relatively inaccurate.

The combination of sentiment analysis and text classification was evaluated to detect malicious behavior [147] and cyberbullying in online multiplayer gaming [121]. Murnion et al.[121] applied

sentiment analysis to extract features to calculate a cyberbullying score. Additionally, the authors applied a custom classifier to detect message characteristics related to cyberbullying. The text classifier used to identify bullying messages was considered useful, but the authors highlighted that it did not provide a solution. The results of sentiment analysis usage in cyberbullying detection were considered quite disappointing. In both cases, the authors pointed out the message structure as one possible reason for the results. Rout et al. [147] evaluated the tasks to identify deceptive reviews associated with products in e-commerce sites. The method extracted 17 features of reviews and applied sentiment analysis to create a sentiment score, considered the 18th feature. Further, the method used text classification to indicate a review as spam or not, and the results demonstrated that sentiment score increased the classifiers' accuracy.

Text classification and text clustering were jointly applied to deal with attack prediction [70] and implementation [38], spam detection [6, 199], and incident damage analysis [184]. We identified that the predominant approach was the application of text clustering to categorize documents or extract features to train a classifier. Holton [70] applied the tasks to identify disgruntled employees through email messages analysis. Text clustering was evaluated to determine differences in emails indicating that an employee is disgruntled. As text clustering distinguished between messages associated with disgruntled or non-disgruntled employees, a classifier was trained and tested with the results. The authors described the results showed great promise to predict unhappy employees.

The combination of text clustering and topic modeling was assessed in threat prevention [101], attack prediction [131], and forensics analysis [27]. We identified two forms of combining the tasks. The first extracted the topics from the content and used the topics to create clusters. The second approach applied the tasks in parallel, so one task did not use the other's knowledge. Lee et al. [101] proposed a system to discover emerging topics based on cybersecurity experts' tweets. The system applied topic modeling to identify tweets associated with cyber threats and demonstrated that the task recognizes emerging information security threats. The study evaluated the application of text clustering to organize cybersecurity experts in communities based on security disciplines and techniques. The authors believe that clustering cybersecurity experts can improve the precision of the system in threats recognition. Noor et al. [131] employed topic modeling to extract high-level IoCs from CTI documents and associate threat actors with cyber incidents through text clustering. The results demonstrated that the tasks were able to attribute threat actors to an incident.

Text clustering was also combined with information retrieval and applied to detect cyberattacks [91] and security incidents [99]. Both studies initially used information retrieval to search on a tweet collection and posteriorly applied text clustering to detect security events. Khandpur et al. [91] proposed identifying cyberattacks in social media, determining their characteristics such as the target and the type of attack. The proposal included the use of tasks without the requirement of a training phase or labeled samples. The research used a small set of seed queries to search on a collection of tweets and automatically expanded each query based on syntactic and semantic similarities. The expanded queries were grouped in clusters to identify cybersecurity events.

Two studies applied topic modeling and sentiment analysis to analyze organizations' reputation damage after a cyber incident based on social networks' repercussions [45, 164]. Syed [164] applied topic modeling to identify topics associated with data breach incidents in Twitter. Based on the topics extracted, tweets were categorized as referring to an accidental or an intentional data breach or if the organization was a victim of the data breach. Sentiment analysis was performed to identify the users' reactions according to each category. Confente et al. [45] analyzed the relationship between data breaches and corporate reputation by examining user-generated content on social networks in an interval of days before and after the incident notification. The research applied topic modeling to identify the topics associated with corporate reputation. After, sentiment

Table 4. Neural Network Models and Methods Evaluated in the Studies

Model	Method	Studies	Total
ANN[1]	ANN[1]/NN[2]	[37, 142, 161]	**3**
	CLA_ANN[3]	[201]	**1**
	DLNN[4]	[130]	**1**
CNN[5]	CNN[5]	[15, 47, 69, 194]	**4**
	C-CNN	[4]	**1**
	S-CNN	[4]	**1**
RNN[6]	RNN[6]	[106]	**1**
	BiLSTM	[24, 171]	**2**
	Contextual LSTM	[94]	**1**
	LSTM[7]	[69, 125]	**2**

[1]Artificial Neural Network.
[2]Neural Network.
[3]Continuous Learning Approach Artificial Neural Network.
[4]Deep Learning Neural Network.
[5]Convolutional Neural Network.
[6]Recurrent Neural Network.
[7]Long Short-Term Memory.

analysis examined the tone characterizing users' posts on social networks. The authors affirmed that corporate managers could adopt the methodology to unveil company reputation.

Topic modeling and information retrieval were jointly applied to prevent data leakage in cloud computing collaboration [183]. Wang and Jin [183] proposed a system that creates user profiles based on past collaboration activities. When a user decides to share a file, the system analyzes the profile to predict the recipient's likelihood of accessing the file content. The proposed system applied information retrieval to discover information from past collaborations and create a profile for the user considering topics that represent past interactions with other users. The results demonstrated the effectiveness of the proposed system.

Two works proposed using text classification and lexical analysis, both related to linguistic steganography [36, 190]. The works initially applied lexical analysis to identify possible replacements of a word. As the studies had different objectives in linguistic steganography, the use of text classification differed between them. Chang and Clark [36] proposed a steganography system, and it used text classification to select synonyms that can substitute the original words found in the content. Xiang et al. [190] proposed a linguistic steganalysis method that applies text classification to identify frequency patterns in the use of synonymous, revealing the existence of hidden messages.

5.3 Neural Networks and Text Mining

Analyzing the corpus, we found 15 studies that applied neural networks to support text mining. We analyzed the studies to identify the neural network models and present the results in Table 4. The models were based on the work of Li et al. [105]. As the work did not mention **Artificial Neural Networks** (**ANNs**), we added this model to the table. We preserved the method's name used in the original studies, only grouping ANNs and **neural networks** (**NNs**) since we understand the authors referred to the same technique. Based on the table, we can affirm that there was no predominance of a NN model in the corpus.

Unlike conventional machine learning techniques, NNs can receive raw data as input and automatically discover the features to perform text classification [100]. However, an approach to

increase the classification performance is using an ensemble, and one method that stands out in use with NNs is word embeddings [92]. Kilimci and Akyokus [92] added that, due to the requirement of a large amount of processing, scholars usually prefer to use pre-trained word embeddings such as Word2vec, GloVe, and fastText.

We investigated the approaches applied to use unstructured data as input to NNs. The analysis identified the predominancy of word embeddings representation in the studies and the use of Word2vec and GloVe pre-trained vectors [24, 47, 94]. We also identified the adoption of customized vectors, similar to Word2vec [194]. Besides word embeddings, studies also evaluated other approaches, such as word frequencies and manual feature extraction [4, 37, 161].

We analyzed the text mining tasks and identified that all studies applied NNs for text classification, according to LeCun et al. [100] the most common form of use. We also identified the use of information extraction, sentiment analysis, question answering, and topic modeling. This section presents some examples of NN applications, ordering them by the cybersecurity domains.

In the vulnerability domain, Ban et al. [24] evaluated NNs to discover vulnerabilities in source codes. The study employed a BiLSTM to extract the features automatically from a source code representation. In order to identify the vulnerabilities, the study used functions labeled as vulnerable from NVD and CVE databases.

Studies applied NNs in the threat domain to detect misinformation [194] and botnet messages [94] on social networks, as well as to prevent threats based on online community content [47, 106]. Kudugunta and Ferrara [94] proposed an architecture using **long short-term memory** (**LSTM**) to identify whether a tweet's author is a bot or not. The authors used the pre-trained vector GloVe to transform the tweets before the LSTM processing. Additionally to the word embeddings, the architecture used the user metadata as an auxiliary input to feed LSTM. The authors highlighted the high accuracy achieved with the architecture.

The work of Li et al. [106] combined sentiment analysis and text classification to monitor cybercriminals selling credit cards on a hacker forum. The study proposed a system that used text classification to identify the threads related to card selling and sentiment analysis to determine the top-rated sellers. The system implemented an RNN to evaluate each cybercriminal's review written by customers on the forum and compute a score based on a five-degree sentiment analysis. The work also applied topic modeling to profile the top sellers, using LDA to extract seller characteristics.

In the attack domain, the method was used to detect cyberattacks based on the analysis of malicious PowerShell commands [69]. The study highlighted the imbalanced dataset, which has 10 times more non-malicious commands than malicious commands. An oversampling technique was used to balance the dataset and prevent model bias. The research evaluated the use of the PowerShell commands at the character level for input data representation, as well as character level 3-gram and BoW. The work compared the performance of CNN and LSTM NNs to predict a malicious or non-malicious command. The results demonstrated that the combination of CNN and 3-gram data representation achieved the best performance.

Noor et al. [130] applied topic modeling and text classification to associate cyberattacks with cybercriminals in the attack domain. The study analyzed cyber threats intelligence reports and applied **Latent Semantic Analysis** (**LSA**) to identify concepts from attack patterns and relate them to threat actors. The use of topic modeling allows associating each threat actor with the tactics, techniques, and procedures, creating a profile. The evaluation used a deep NN trained with the profiles to predict the threat actor responsible for an unseen cyberattack.

NNs were also evaluated to support spam detection in the control domain [161, 201]. Suleiman and Al-Naymat [161] proposed a framework for spam detection and compared two conventional classifiers and ANN performance. The authors manually extracted 10 features from SMS messages

and used the most significant features to train the classifiers. In the experiment, the random forest classifier outperformed ANN.

Two studies applied NNs to support information extraction in the control domain [15, 125]. The works evaluated NNs to identify attributes like the subject and the object and automate the process of creating access rules in the ABAC model based on natural language policies. NNs also supported the construction of the relations among the identified entities. The NNs analyzed were CNN [15] and LSTM [125]. Another research combined text classification and question answering to develop an authentication method [171]. Toor et al. [171] proposed a hybrid visual question authentication protocol using a set of images and a related question. A BiLSTM NN is used to classify a single word in a question as irrelevant or the entire question as relevant. It was necessary to identify words not related to images and adjust the questions in the authentication protocol.

In the incident domain, studies assessed NNs to assist digital forensics investigators in cybercriminals identification by analyzing source codes [4] and written patterns in messages [142]. The estimation of economic damages caused by a phishing attack on an organization was also evaluated [37]. Phan and Zincir-Heywood [142] proposed a forensic system to identify the users based on their writing styles and employed ANN. The system focused on short messages like email and tweets and applied a two-stage feature extraction. The first stage converted the message to a word embedding vector representation, extracting low-level features. The second stage applied the BoW approach to extract high-level features from short messages and train the ANN.

We also investigated NNs application in different domains based on other SLRs. Kumar and Ravi [97] mapped fewer studies using NNs in the financial domain than we identified in our corpus. However, it is relevant to highlight that the SLR was published in 2016, so, though it is not explicit in the SLR, we can infer that the domain started to apply NNs to extract knowledge from unstructured content before the cybersecurity domain. Drury and Roche [50] also mentioned the use of NNs to support NERC and sentiment classification in the agricultural domain. The SLR did not present the number of studies that applied the approach but highlighted the use of **Recurrent Neural Networks (RNNs)**.

We performed a broader analysis, and works have presented positive results in applying NNs in text mining tasks [178, 182]. Previous studies compared the performance of NNs to traditional classifiers and "shallow architectures" [98, 178], and experiments demonstrated that NNs reached results equal or superior in text classification. In this SLR, four studies compared the performance of traditional classifiers and NNs in their experiments. One study suggested that NNs performed better than traditional classifiers [130]. The others reported that traditional classifiers reached similar [47] or better [37, 161] performance than the NN evaluated.

6 TEXT CLASSIFICATION PERFORMANCE IN THE CYBERSECURITY DOMAIN

In the previous sections, we presented several characteristics concerning the application of text mining in cybersecurity, so we believe it is also relevant to show the results obtained in the studies. This section answers the fourth research question (RQ4). It is important to highlight that we do not intend to compare the results among the studies listed in this section. Instead, we aim to present that text classification has reached significant results in different activities in the cybersecurity domain. In order to organize the discussion addressing text classification performance, we divided this section into two parts. Section 6.1 presents a table including four metrics to represent the performance achieved by text classification in the studies. The metrics are related to the cybersecurity activities represented in the second level of the proposed taxonomy. In Section 6.2, we discuss the relevance of the metrics achieved in each cybersecurity activity, considering if the results indicate that the activity can be supported by text classification in the real world.

6.1 Summary of the Metrics

Table 5 lists the activities introduced in the proposed taxonomy and presents some text classification performance measures. We took care to include the experiment dataset to reinforce the use of different data and that the results cannot be compared. Moreover, the table lists algorithms in alphabetical order in each activity, not considering the best performance. The metrics presented in the table are accuracy, precision, recall, and F1-Score. We decided to show multiple metrics because text classification studies can use different performance measures to evaluate an experiment [157]. Distinct studies presented the results with slight differences, such as the use of decimal values or percentages. When it was possible, we uniformed the results in percentage, considering one decimal place.

In cases where a study evaluated multiple classification algorithms, we present the classifier that achieved the best precision because this metric and recall provide the best perspective for performance on text classification [157]. Some studies evaluated more than one dataset. When it happened, we decided to show the first or the primary dataset results. Several studies also evaluated distinct configurations associated with the classifier, so we decided to present the best configuration result. If a study did not calculate all metrics, we filled the table only with the measures presented. A few experiments fragmented the evaluation into smaller parts and did not present the overall metrics. In such cases, we decided not to show the results in the table. Finally, if we did not fully understand a study experiment's results, we preferred not to include them.

As presented in Table 5, most studies evaluated a customized dataset. In these cases, the authors collected unstructured information from a data source (i.e., a social network) to evaluate how well a proposed method performs. We consider that creating a customized dataset has a positive side because the authors collected real data representing a landscape associated with their research problem. So it is important to highlight the authors' efforts to assess their proposals simulating a real cybersecurity scenario.

On the other hand, it is worth emphasizing that using a customized dataset can positively or negatively impact the results obtained in text classification since the task has known difficulties in dealing with an imbalanced dataset [66]. The effect of an imbalanced dataset can also impact the use of NNs [71]. According to Pawlicki et al. [138], the cybersecurity domain usually deals with imbalanced datasets since most data are related to benign activities. Therefore, in many cases, studies should consider the imbalanced dataset problem in the experiment design.

Moreover, the information collected to create a dataset reflects a time interval associated with a specific data source. The patterns created for the same type of malicious activity in different moments of time or data sources can present significant variation, so the text classifier's performance can be significantly impacted. Based on the datasets used in the studies, we understand that there are opportunities to create new datasets associated with different cybersecurity activities.

6.2 Discussion by Cybersecurity Activity

Table 5 demonstrated variations in text classification performance according to the cybersecurity activity. Although the metrics represent a specific evaluation of the text classification in the cybersecurity domain, we present details and discuss the experiments' results.

The study of Wen et al. [187] applied text classification to categorize vulnerabilities according to information extracted from public databases. The accuracy demonstrated the importance of evaluating new strategies in this vulnerability management activity because a significant percentage of misclassified vulnerabilities could offer wrong orientation to the cybersecurity personnel. In another study, Wen et al. [186] evaluated text classification to calculate the vulnerability severity. Although the proposal has excellent value in the real world, its application depends on better results to prevent minor vulnerabilities from being prioritized.

Table 5. The Table Presents the Results from the Application of Text Classification in the Cybersecurity Domain

				Classification				
		Study	Dataset	Algorithm	Accuracy	Precision	Recall	F1-Score
Vulnerability	Management	[88]	Custom	CRF		83%	76%	80%
		[187]	Custom	SVM	82.5%			
	Analysis	[186]	Custom	SVM	82.5%			
Threat	Prevention	[106]	Custom	SVM		100%	62.9%	77.2%
		[140]	Custom	LR[1]	77.7%			
		[20]	Custom	NB[2]	76.7%	76.7%	76.7%	
		[47]	Custom	SVM	98.6%	98.4%	98.1%	98.2%
	Detection	[94]	Custom	AdaBoost	99.8%	100%	100%	100%
		[194]	Weibo	CNN	94.8%	95.6%	94.4%	95%
		[124]	Custom	Custom		87%	98%	91%
		[177]	EMSCAD	RF[3]		90.6%	90.6%	90.6%
Attack	Implementation	[39]	Custom	SVM	98.6%	97.7%	99.5%	
		[38]	Custom	SVM	98.8%	99%	98.6%	
	Detection	[12]	Custom	LR[1]		99.4%	99.3%	99.3%
		[130]	CTA Profile	RF	88%	92%	89%	89%
	Prediction	[135]	Sentiment140	NB[2]	96.2%	100%	96.2%	98%
		[70]	Custom	NB[2]		90%	90%	
	Analysis	[52]	Custom	DT		58%	90%	65%
Control	Access Control	[126]	Custom	Adaboost		73%	71%	71%
		[154]	Custom	K-NN		83%	96%	89%
		[171]	Custom	LSTM	73.2%	78.1%	68.5%	73%
	Malicious Behavior Detection	[201]	SpamAssassin	CLA_ANN		99.1%	68.3%	
		[53]	Spambase	Custom	98.3%	96.4%	99.6%	97.9%
		[147]	Custom	DT	92.1%			
		[113]	M0Droid	Multi-C[4]		95.8%	95.7%	95.6%
		[161]	SMS Collection V1	RF[3]	97%	95%	85%	89%
		[6]	SMS Collection V1	RF[3]		99.2%	99.2%	99.1%
		[119]	Custom	SVM			99.1%	99%
	DLP	[13]	Custom	Custom		91.4%	90.6%	
		[74]	Reuters	LPA	100%		100%	
		[195]	Custom	TsF-kNN	68%	47%	71%	55%
Incident	Detection	[54]	Custom	NB[2]	98%			
	Forensic	[142]	Enron	ANN	82.7%			
		[104]	Custom	DT	78%			
		[4]	Custom	S-CNN	99.5%			
		[152]	Custom	SVM	100%			
	Post Incident	[184]	Custom	DT	77%			
		[72]	Custom	RF[3]		98%	81%	

The table's goal is to present the results of different studies in cybersecurity areas, but the metrics must not be compared.

[1]Logistic Regression.
[2]Naive Bayes.
[3]Random Forest.
[4]Multi-Classifiers.

In the threat domain, studies evaluating text mining to detect misinformation [94, 194] and online frauds [177] achieved precision and recall metrics higher than 90%. Online monitoring is an arduous task to organizations and even to nations that need to protect their citizens and critical infrastructures. The results demonstrated the actual value of text classification to identify cyber threats. Although the use of text mining in this activity results in missing some cyber threats, the detection rate will offer a significant number of threats that cannot be discovered manually or using traditional techniques. The threat detection automation can also offer extra time to cybersecurity personnel to manually mine the information gathered to find missed cyber threats. In contrast, the

use of text classification to identify cyberbullying did not reach a precision as significant as other threats [124].

The application of text classification to identify cyber threats and anticipate cyberattacks achieved relevant results in some studies. Deliu et al. [47] obtained more than 98% in all metrics applying SVM to identify posts on online forums. The study achieved the results applying only text classification. Another study obtained a 100% precision metric in recognizing stolen data advertisements on online forums [106] through the combined use of topic modeling and text classification. Although the other two studies obtained metric results below 80%, we believe that the overall results demonstrated that text mining could be a valuable tool to monitor threat-related content for the cybersecurity team analysis.

Another way to anticipate cyberattacks is through their prediction, and studies demonstrated that text mining could also offer appropriate guidance to cybersecurity personnel about imminent attacks. Like threat detection, attack prediction reached precision and recall metrics equal to or greater than 90%. Both studies listed in Table 5 predicted internal attacks based on posts on online forums and social networks but applied different strategies. While Holton [70] combined text clustering and text classification to identify disgruntled employees, Park et al. [135] applied topic modeling, sentiment analysis, and text classification to recognize malicious insiders. Considering the results of both strategies, we believe that text mining can offer reliable information based on users' behavior monitoring and determine signals of misbehavior that indicate a possible internal attack.

In the area defined as attack implementation, two studies associated with steganalysis reached a satisfactory performance. Although steganography is not commonly used for organizations, cyberterrorists have used this technique [200]. Thus, results demonstrated that security forces could rely on text classification to detect content hidden by linguistic steganography.

Text mining has also achieved satisfactory results to detect distinct types of cyberattacks. In the real world, organizations have faced difficulties in cyberattack detection. Frequently, organizations take several weeks to discover that they were hit by a cyberattack [33]. The study of Noor et al. [130] achieved a precision of 92% in the use of high-level IoCs to identify an attacker by combining topic modeling and text classification tasks. Discovering who is behind a cyberattack is very relevant to cybersecurity personnel to understand the authors' tactics, techniques, and procedures.

Aldwairi and Alwahedi [12] also showed that text mining could effectively detect misinformation campaigns on social media, achieving 99.4% precision. The proposal applied only text classification to identify misinformation and did not use other text mining tasks. Vosoughi et al. [180] analyzed more than 100,000 Twitter stories and pointed out that misinformation spreads farther, faster, deeper, and more broadly than true stories on social media. Based on the results, we believe that text mining is a valuable tool to identify and censor misinformation campaigns.

Following the trend identified in threat and attack detection, traditional classifiers also obtained competent results identifying malicious behavior in the control domain. Five studies focused on spam message identification, and all achieved precision more significant than 95% [6, 53, 147, 161, 201]. Zitar and Hamdan [201] and Suleiman and Al-Naymat [161] applied only text classification, but the works did not obtain a recall metric as good as the precision. Consequently, many legitimate messages could be identified as spam, and possibly they will not be delivered into the recipient's mailbox.

The remaining three studies that focused on spam detection proposed strategies combining text classification and another text mining task: Rout et al. [147] used sentiment analysis; Adewole et al. [6] applied text clustering; El-Alfy and AlHasan [53] used information extraction. We considered good results for spam detection proposals reaching precision and recall metrics higher than 99% based on the problem related to legitimate messages misclassification. We also highlight that four studies that analyzed spam messages used datasets available on the internet. The usage of public

datasets is positive because it allows comparing distinct strategies, as Suleiman and Al-Naymat [161] performed in their study.

Another study related to malicious behavior detection analyzed the application of multiple classifiers to detect Android malware and achieved precision and recall results greater than 95% [113]. Different from other cybersecurity areas, we think these results are not appropriate for antimalware software. However, the results were obtained evaluating app permissions, and this type of text mining use could be considered by the antimalware industry integrating with traditional analysis. Phishing is a known cybersecurity challenge, and Moghimi and Varjani [119] evaluated a distinct approach to detect this malicious behavior, achieving results greater than 99%. Traditionally, anti-phishing software focuses on identifying messages associated with phishing campaigns; however, the authors' approach recognized phishing based on the site content and URL. We think the results demonstrated that anti-phishing software could rely on text mining in phishing identification. We also understand that studies should integrate traditional approaches with the authors' proposal to evaluate the performance in identifying messages and websites associated with phishing campaigns.

Two studies applied text classification to automate the creation of access control rules based on natural language documents, and both employed information extraction to identify access control entities [126, 154]. Access control is vital to any organization's security, and any failure in its rules can result in a severe incident. We understand that studies should evaluate new strategies since the current metrics do not ensure the activity's reliability.

As with access control, DLP is also crucial because confidential information should not leave organization domains without proper authorization. Alneyadi et al. [13] and Zardari and Jung [195] applied text classification to identify sensitive information and assign security classification labels. Although security classification is not mandatory for leakage prevention, it is highly recommended because assigning a label to associate the information with its sensitivity allows visibility to the DLP tool. However, studies should achieve better results to apply text mining in supporting security classification in the real world.

In a different DLP approach, Huang et al. [74] consider a document's context to assign a score representing its sensitivity. The results showed that text classification was highly accurate in the evaluation. However, it is important to highlight that the study used as a dataset distinct categories of news articles, assigning to a category the sensitive status. We understand that the proposal should be evaluated with a dataset consisting of sensitive and non-sensitive documents to verify if the approach is successful.

Forensics studies applied text classification to recognize the authorship of cybercriminals involved in an incident analyzing IRC posts [152] and source codes [4]. The studies applied only text classification and reached an accuracy greater than 99%. Today, cybercriminals use several tools to attack an organization, and the recognition of coding style can offer tips to optimize forensics investigation. Studies also evaluated other artifacts to help forensics investigators search for the culprits. However, text classification did not perform well in evaluating authorship identification based on social network posts [104] and short messages [142].

7 TEXT MINING IN THE CYBERSECURITY INDUSTRY

Text mining results have also attracted the cybersecurity industry's attention, which has adopted the technology in cybersecurity solutions. This section presents examples of text mining applications in the real world and answers the fifth research question (RQ5). CTI is one activity that benefited from technology. According to Papadopoulos [134], the "cyber blindness" produced by the relevant volume of structured and unstructured data makes many cyberattacks go undetected, so IBM has applied machine learning techniques in its security analytics platform. Another cyber

threat intelligence company, Recorded Future, has applied text mining to deal with cyber threats in multiple languages and search for clues in massive volumes of text [170]. The technology allows the company to identify cyber threat actors, targets of an attack, and attack methods. Microsoft has also implemented the technology to automate CTI extraction from unstructured text [158]. The company evaluated the application of information extraction and text classification to identify threat actors, malware families, attack techniques, and relationships between entities.

DLP is another cybersecurity area benefited by text mining application. According to Sambamoorthy [148], co-founder of Armorblox, the lack of understanding of general methods used in the solutions results in imprecise detection and excessive noise. The DLP solution developed by Armorblox applied deep learning and natural language models to understand human language and classify confidential data. The large amount of data managed by Google made the company develop a text mining–based tool to identify and redact sensitive information [129]. The DLP solution developed by Google became a cloud service, and it is used by Google applications and by the company's customers. Amazon has also launched the platform Amazon Macie, a DLP platform designed for AWS cloud that applies text mining tasks to classify data stored in the company's cloud [181].

The daily worldwide email traffic in 2019 was estimated at 293 billion messages [43]. Email messages can carry cyber threats, and solutions designed to protect this platform have applied text mining. Arguing that traditional email protection solutions produce many noise and false positives, Armorblox developed a solution that uses text mining [32]. According to the company, text mining enables the solution to detect cyber extortion and scam in email messages that do not contain links or attachments. Another cyber threat associated with email messages is phishing. An email security solution developed by Vade Secure applied text mining to scan for patterns and behaviors expected in spear phishing, a specialized type of phishing targeting a specific individual [58].

Another approach to fight against phishing is the identification of the DNS address associated with a cyberattack. Using its DNS infrastructure, Cisco applied text mining–based predictive models to recognize and block malicious DNS addresses related to an APT group [132]. The incident response team can also benefit from the use of text mining. IBM integrated the technology in its solution, so when the incident response team registers a new incident, the solution can discover similar incidents already stored in the database [150]. The access to past events enables the incident response team to understand what actions were taken to remediate a similar incident.

The industry's interest in using text mining demonstrates that the technology reached a maturity level to support critical applications in the cybersecurity domain. One example is text mining application in data leakage prevention, a vital security control that can significantly compromise a company's reputation if a failure occurs. We believe we will see a rise in the number of cybersecurity companies integrating text mining in their solutions and services in the following years. Moreover, we think that the results obtained in cybersecurity research in areas like CTI, information classification, and forensics will stimulate new services aimed at cyber threats recognition, multicloud classification, and incident authorship identification. Lastly, the application of text mining to deal with misinformation has been evaluated by non-academic organizations, as demonstrated by Rand Corporation in a study contracted by the United Kingdom government [109]. Based on the study, we can assume that new cybersecurity challenges, like misinformation, can foster new text mining–based solutions.

8 OPEN CHALLENGES AND FUTURE DIRECTIONS

Cybersecurity is an emerging research topic in computer science that has received increasing attention from scientists. In parallel, the rise of unstructured data and cyberattacks demands new

studies to discover valuable patterns to support daily cybersecurity activities. Moreover, the future outlook requires the continuous improvement of today's cybersecurity activities and automation of new ones. Therefore, we focus on challenges arising in this field of research that the research community can address.

(1) **Non-English Texts:** In the SLR corpus, only eight studies analyzed contents not written in English. The Symantec Report [166] demonstrated that non-English speaking countries are both top sources and top targets of cyberattacks. So, we think that researchers could perform studies using datasets based on non-English content to evaluate text mining performance in other languages. Moreover, most pre-processing tools and libraries support only the English language, and new studies can develop tools targeting non-English languages. A few studies [47, 52, 106, 108] have already cited this as a direction to evolve their research.

(2) **Regional Analysis:** The implementation of text mining in non-English content makes some analysis related to regional context possible, as performed in Huang et al. [72]. New studies can analyze reports and posts on social networks from featured companies and professionals to extract information about the cybersecurity status in a country or world region. For example, studies can compare cybersecurity threats among world regions [28] or analyze the impact of incidents in different countries.

(3) **Text Summarization:** Although many activities involve analyzing large volumes of textual data, none of the studies summarized content related to cybersecurity. The increase of unstructured data available in organizations and on the internet makes monitoring activities hard to perform, meaning cybersecurity teams can miss valuable data [172]. The application of text summarization in these activities can narrow the data analysis, and cybersecurity personnel can focus on scrutinizing the suspect content. CTI activities could take advantage of text summarization in the analysis of information collected from multiple digital reports, e-mail threads, and online forums. The incident domain is also a candidate in the use of text summarization. In several situations, the time between incident detection and response can be longer than expected, so studies can analyze if the task improves the performance based on summarizing multiple similar incidents already remediated.

(4) **New or Broader Datasets:** We noticed that many studies applied text mining tasks in the same dataset. For example, different studies analyzed e-mail messages using the Enron dataset [53, 104, 110, 131]. Although the Enron database stores real e-mail messages and is good to verify the application of an experiment in the real world, more recent and different datasets are preferable to validate studies. The use of new or broader datasets is also mentioned in studies because an experiment is based on just one source [27, 108, 118], in a small dataset [20, 47], or a specific type of message [28, 44, 59]. The evaluation of specific-context studies on multiple datasets allows the analysis in different scenarios.

(5) **Algorithm or NNs Comparison:** Some studies' experiments are based on evaluating just one traditional classifier or NN. However, the cybersecurity team should rely on comparative tests to decide which classifier must be implemented to support a specific activity. Thus, as suggested by Deliu et al. [47] and Phan and Zincir-Heywood [142], new studies comparing the results of the current and novel algorithms or NNs are relevant to provide enough information to support the cybersecurity community to choose the best option.

(6) **Application of NNs in New Cybersecurity Areas:** We identified that just 15 studies evaluated the use of NNs in the cybersecurity domain. Additionally, NNs have presented promising results in different applications related to text mining [178, 182]. Based on this, the application of NNs could expand to new cybersecurity activities, and different techniques need to be compared with those already applied. For example, in the threat domain, NNs can

support information extraction to identify cyber threats based on online community analysis. In the incident domain, studies can evaluate the application of NNs to identify artifacts associated with known threat actors. Furthermore, new NN techniques have been developed, and their results in specific text mining tasks are encouraging [174], so studies can evaluate the techniques to support cybersecurity activities.

(7) **Data Leakage Discovery:** The growth of data leakages has caused losses to both organizations and consumers in recent years. This topic needs to receive considerably more attention from cybersecurity researchers. We believe the application of text mining tasks can identify data leakages in a big data scenario, so new studies associated with the surface or the dark web should be considered.

(8) **Adversarial Text Mining:** Cyberattacks have become more sophisticated recently, and malicious actors can use new techniques to automate their actions. The proposed taxonomy demonstrated that text mining could leverage cyberattack implementation [38, 39, 52]. At the same time, organizations are using text mining–based mechanisms to protect their infrastructure and systems against cyberattacks. Current research has analyzed the application of text mining tasks to improve adversarial learning [192]. Thus, new studies can analyze adversarial text mining to evaluate the security of text mining–based mechanisms [16, 198]. In a similar approach, works can also evaluate new uses of text mining to support the discovery of vulnerabilities in running protocols or applications [85] and weaknesses in traditional security mechanisms.

9 CONCLUSION

Text mining tasks are used to extract knowledge from unstructured content and improve the processes in several domains like politics, financial, publishing, and media. The increase of essential services supported by digital platforms makes cybersecurity more relevant, and researchers started to evaluate the application of text mining to improve cybersecurity activities. Considering the recent attention received by cybersecurity and the role that text mining can perform to automate activities, this study proposes an SLR to understand the current application of text mining in the cybersecurity domain. The review initially identified 2,196 studies by combining different search strings, and 83 primary studies were in-depth analyzed after filtering out based on well-defined exclusion and inclusion criteria. This study presents the text mining tasks applied in the domain and which cybersecurity activities have taken advantage of the technology. The analysis allowed us to answer the proposed research questions.

The relevance of using text mining in the cybersecurity domain is evident, mainly in monitoring activities since the data volume makes human analysis almost impossible. Text mining can also offer great support to organizations by automating the detection of threats, attacks, and incidents, delivering the results to the cybersecurity team that will be able to act promptly. It is worth highlighting that text mining plays an ancillary role in these cases since the action is performed by cybersecurity personnel. On the other hand, in some activities, text mining can automate cybersecurity actions without human intervention, playing the primary role. For example, systems based on text mining can block phishing sent to an organization. In data leakage systems, text mining can detect an attempt to transmit classified information to external networks and avoid a data breach. So, text mining plays different roles in the cybersecurity domain. However, we believe that, regardless of the role played, text mining is crucial to face the massive volume of unstructured data that cybersecurity personnel need to handle.

Although we think this SLR presents valuable information about the intersection between cybersecurity and text mining, we also identify improvements that can result in future works. This SLR analyzed the result of search queries and did not perform a snowballing to expand the

corpus. Therefore, we suggest new studies adding backward and forward snowballing methods. Due to the time spent analyzing the SLR corpus to get reliable results, more recent publications are not included in this work, so mainly the forward snowballing can bring relevant enhancements. Another future work is the expansion of the search string, using other terms related to natural language processing, and the inclusion of new scientific databases.

REFERENCES

[1] Ahmed Abbasi, Stephen France, Zhu Zhang, and Hsinchun Chen. 2010. Selecting attributes for sentiment classification using feature relation networks. *IEEE Transactions on Knowledge and Data Engineering* 23, 3 (2010), 447–462.

[2] Adeline Abbe, Cyril Grouin, Pierre Zweigenbaum, and Bruno Falissard. 2016. Text mining applications in psychiatry: A systematic literature review. *International Journal of Methods in Psychiatric Research* 25, 2 (2016), 86–100.

[3] Laith Mohammad Abualigah, Ahamad Tajudin Khader, and Essam Said Hanandeh. 2018. A novel weighting scheme applied to improve the text document clustering techniques. *Innovative Computing, Optimization and its Applications*. Springer, 305–320. https://doi.org/10.1007/978-3-319-66984-7_18

[4] Mohammed Abuhamad, Ji su Rhim, Tamer AbuHmed, Sana Ullah, Sanggil Kang, and DaeHun Nyang. 2019. Code authorship identification using convolutional neural networks. *Future Generation Computer Systems* 95 (2019), 104–115.

[5] Stephen Adams, Bryan Carter, Cody Fleming, and Peter A. Beling. 2018. Selecting system specific cybersecurity attack patterns using topic modeling. In *2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)*. Institute of Electrical and Electronics Engineers Inc., 490–497.

[6] Kayode Sakariyah Adewole, Nor Badrul Anuar, Amirrudin Kamsin, and Arun Kumar Sangaiah. 2017. SMSAD: A framework for spam message and spam account detection. *Multimedia Tools and Applications* 78, 4 (July 2017), 3925–3960.

[7] Charu C. Aggarwal. 2018. *Machine Learning for Text*. Springer.

[8] Charu C. Aggarwal and ChengXiang Zhai. 2012. *Mining Text Data*. Springer Science & Business Media.

[9] Amritanshu Agrawal, Wei Fu, and Tim Menzies. 2018. What is wrong with topic modeling? And how to fix it using search-based software engineering. *Information and Software Technology* 98 (2018), 74–88.

[10] Tareq Al-Moslmi, Nazlia Omar, Salwani Abdullah, and Mohammed Albared. 2017. Approaches to cross-domain sentiment analysis: A systematic literature review. *IEEE Access* 5 (2017), 16173–16192.

[11] Khalid Al-Rowaily, Muhammad Abulaish, Nur Al-Hasan Haldar, and Majed Al-Rubaian. 2015. BiSAL—A bilingual sentiment analysis lexicon to analyze Dark Web forums for cyber security. *Digital Investigation* 14 (2015), 53–62.

[12] Monther Aldwairi and Ali Alwahedi. 2018. Detecting fake news in social media networks. *Procedia Computer Science* 141 (2018), 215–222.

[13] Sultan Alneyadi, Elankayer Sithirasenan, and Vallipuram Muthukkumarasamy. 2013. Adaptable N-gram classification model for data leakage prevention. In *Proceedings of the 7th International Conference on Signal Processing and Communication Systems (ICSPCS')* https://doi.org/10.1109/ICSPCS.2013.6723919

[14] Sultan Alneyadi, Elankayer Sithirasenan, and Vallipuram Muthukkumarasamy. 2016. A survey on data leakage prevention systems. *Journal of Network and Computer Applications* 62 (2016), 137–152.

[15] Manar Alohaly, Hassan Takabi, and Eduardo Blanco. 2018. A deep learning approach for extracting attributes of ABAC policies. In *Proceedings of the 23rd ACM Symposium on Access Control Models and Technologies (SACMAT'18)*. Association for Computing Machinery, New York, NY, 137–148.

[16] Basemah Alshemali and Jugal Kalita. 2020. Improving the reliability of deep neural networks in NLP: A review. *Knowledge-Based Systems* 191 (2020), 105210.

[17] Berna Altınel and Murat Can Ganiz. 2018. Semantic text classification: A survey of past and recent advances. *Information Processing & Management* 54, 6 (2018), 1129–1153.

[18] Flora Amato, Giovanni Cozzolino, Vincenzo Moscato, and Francesco Moscato. 2019. Analyse digital forensic evidences through a semantic-based methodology and NLP techniques. *Future Generation Computer Systems* 98 (2019), 297–307.

[19] Flora Amato, Giuseppe De Pietro, Massimo Esposito, and Nicola Mazzocca. 2015. An integrated framework for securing semi-structured health records. *Knowledge-Based Systems* 79 (2015), 99–117.

[20] Jungkook An and Hee-Woong Kim. 2018. A data analytics approach to the cybercrime underground economy. *IEEE Access* 6 (2018), 26636–26652.

[21] Murugan Anandarajan, Chelsey Hill, and Thomas Nolan. 2019. Text preprocessing. *Practical Text Analytics*. Springer, 45–59.

[22] Giulio Angiani, Laura Ferrari, Tomaso Fontanini, Paolo Fornacciari, Eleonora Iotti, Federico Magliani, and Stefano Manicardi. 2016. A comparison between preprocessing techniques for sentiment analysis in twitter. In *KDWeb*.
[23] David Antons, Eduard Grünwald, Patrick Cichy, and Torsten Oliver Salge. 2020. The application of text mining methods in innovation research: Current state, evolution patterns, and development priorities. *R&D Management* 50, 3 (2020), 329–351.
[24] Xinbo Ban, Shigang Liu, Chao Chen, and Caslon Chua. 2018. A performance evaluation of deep-learnt features for software vulnerability detection. *Concurrency Computation* 31, 19 (2018).
[25] Barnali Gupta Banik and Samir Kumar Bandyopadhyay. 2018. Novel text steganography using natural language processing and part-of-speech tagging. *IETE Journal of Research* 66, 3 (2018), 1–12. https://doi.org/10.1080/03772063.2018.1491807
[26] Nicole Lang Beebe, Jan Guynes Clark, Glenn B. Dietrich, Myung S. Ko, and Daijin Ko. 2011. Post-retrieval search hit clustering to improve information retrieval effectiveness: Two digital forensics case studies. *Decision Support Systems* 51, 4 (2011), 732–744.
[27] Nicole L. Beebe and Lishu Liu. 2014. Clustering digital forensic string search output. *Digital Investigation* 11, 4 (2014), 314–322.
[28] Victor Benjamin, Bin Zhang, Jay F Nunamaker, Jr., and Hsinchun Chen. 2016. Examining hacker participation length in cybercriminal internet-relay-chat communities. *Journal of Management Information Systems* 33, 2 (2016), 482–510.
[29] Daniel S. Berman, Anna L. Buczak, Jeffrey S. Chavis, and Cherita L. Corbett. 2019. A survey of deep learning methods for cyber security. *Information* 10, 4 (2019), 122.
[30] Sajal Bhatia, Sunny Behal, and Irfan Ahmed. 2018. Distributed denial of service attacks and defense mechanisms: Current landscape and future directions. *Versatile Cybersecurity*. Springer, 55–97.
[31] David M. Blei. 2012. Probabilistic topic models. *Communications of the ACM* 55, 4 (2012), 77–84.
[32] Tony Bradley. 2019. New Cybersecurity Company Focuses on Addressing the Weakest Link in the Security Chain. Retrieved on April 2021 from https://www.forbes.com/sites/tonybradley/2019/02/20/new-cybersecurity-company-focuses-on-addressing-the-weakest-link-in-the-security-chain/?sh=3dcd66753745.
[33] Matt Bromiley. 2019. *SANS 2019 Incident Response (IR) Survey: It's Time for a Change*. Technical Report. SANS Institute.
[34] Kevin Matthe Caramancion. 2020. An exploration of disinformation as a cybersecurity threat. In *Proceedings of the 3rd International Conference on Information and Computer Technologies (ICICT'20)*. IEEE, 440–444.
[35] Stefano Ceri, Alessandro Bozzon, Marco Brambilla, Emanuele Della Valle, Piero Fraternali, and Silvia Quarteroni. 2013. An introduction to information retrieval. *Web Information Retrieval*. Springer, 3–11.
[36] Ching-Yun Chang and Stephen Clark. 2014. Practical linguistic steganography using contextual synonym substitution and a novel vertex coding method. *Computational Linguistics* 40, 2 (June 2014), 403–448.
[37] Xi Chen, Indranil Bose, Alvin Chung Man Leung, and Chenhui Guo. 2011. Assessing the severity of phishing attacks: A hybrid data mining approach. *Decision Support Systems* 50, 4 (2011), 662–672.
[38] Zhili Chen, Liusheng Huang, Haibo Miao, Wei Yang, and Peng Meng. 2011. Steganalysis against substitution-based linguistic steganography based on context clusters. *Computers & Electrical Engineering* 37, 6 (2011), 1071–1081.
[39] Zhili Chen, Liusheng Huang, and Wei Yang. 2011. Detection of substitution-based linguistic steganography by relative frequency analysis. *Digital Investigation* 8, 1 (2011), 68–77.
[40] Long Cheng, Fang Liu, and Danfeng Yao. 2017. Enterprise data breach: Causes, challenges, prevention, and future directions. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 7, 5 (2017), e1211.
[41] Qi Cheng, Conor Cunningham, Fabian Gacayan, Anni Gu, Alex Hall, Olivia Lee, Ashley Sawyer, Safy Sayoud, Vriti Wadhwa, and Jion Yi. 2018. Hacking Democracy: Cybersecurity and global election interference. http://hdl.handle.net/1773/43754.
[42] Paul Cichonski, Tom Millar, Tim Grance, and Karen Scarfone. 2012. *Computer Security Incident Handling Guide*. Technical Report 800-61. National Institute of Standards and Technology (NIST).
[43] Jessica Clement. 2020. Number of sent and received e-mails per day worldwide from 2017 to 2024. Retrieved from https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/.
[44] Aviad Cohen, Nir Nissim, Lior Rokach, and Yuval Elovici. 2016. SFEM: Structural feature extraction methodology for the detection of malicious office documents using machine learning methods. *Expert Systems with Applications* 63 (2016), 324–343.
[45] Ilenia Confente, Giorgia Giusi Siciliano, Barbara Gaudenzi, and Matthias Eickhoff. 2019. Effects of data breaches from user-generated content: A corporate reputation analysis. *European Management Journal* 37, 4 (2019), 492–504.
[46] Aitor Couce-Vieira, David Rios Insua, and Alex Kosgodagan. 2020. Assessing and forecasting cybersecurity impacts. *Decision Analysis* 17, 4 (2020), 356–374.
[47] Isuf Deliu, Carl Leichter, and Katrin Franke. 2018. Extracting cyber threat intelligence from hacker forums: Support vector machines versus convolutional neural networks, In *Proceedings of the 2017 IEEE International Conference on Big Data (Big Data'17)* , 3648–3656.

[48] Li Deng and Yang Liu. 2018. *Deep Learning in Natural Language Processing*. Springer.
[49] Prasad M. Deshpande, Salil Joshi, Prateek Dewan, Karin Murthy, Mukesh Mohania, and Sheshnarayan Agrawal. 2014. The Mask of ZoRRo: Preventing information leakage from documents. *Knowledge and Information Systems* 45, 3 (Dec. 2014), 705–730.
[50] Brett Drury and Mathieu Roche. 2019. A survey of the applications of text mining for agriculture. *Computers and Electronics in Agriculture* 163 (2019), 104864.
[51] April Edwards, David Demoll, and Lynne Edwards. 2020. Detecting cyberbullying activity across platforms. In *Proceedings of the 17th International Conference on Information Technology–New Generations (ITNG'20)*. Springer, 45–50.
[52] Matthew Edwards, Robert Larson, Benjamin Green, Awais Rashid, and Alistair Baron. 2017. Panning for gold: Automatically analysing online social engineering attack surfaces. *Computers & Security* 69 (2017), 18–34.
[53] El-Sayed M. El-Alfy and Ali A. AlHasan. 2016. Spam filtering framework for multimodal mobile communication based on dendritic cell algorithm. *Future Generation Computer Systems* 64 (2016), 98–107.
[54] Yong Fang, Yusong Guo, Cheng Huang, and Liang Liu. 2019. Analyzing and identifying data breaches in underground forums. *IEEE Access* 7 (2019), 1–1.
[55] Ronen Feldman, James Sanger, et al. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
[56] Rafael Ferreira-Mello, Máverick André, Anderson Pinheiro, Evandro Costa, and Cristobal Romero. 2019. Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 9, 6 (2019), e1332.
[57] Mahak Gambhir and Vishal Gupta. 2017. Recent automatic text summarization techniques: A survey. *Artificial Intelligence Review* 47, 1 (2017), 1–66.
[58] Adrien Gendre. 2019. Vade Secure Expands AI-Based Threat Detection with New Computer Vision Engine. Retrieved from https://www.vadesecure.com/en/vade-secure-expands-ai-based-threat-detection-with-new-computer-vision-engine/.
[59] J. L. Gonzalez-Compean, Oscar Telles, Ivan Lopez-Arevalo, Miguel Morales-Sandoval, Victor J. Sosa-Sosa, and Jesus Carretero. 2019. A policy-based containerized filter for secure information sharing in organizational environments. *Future Generation Computer Systems* 95 (2019), 430–444.
[60] NIS Cooperation Group. 2018. *Cybersecurity Incident Taxonomy*. Technical Report. European Union Agency for Network and Information Security.
[61] Renchu Guan, Hao Zhang, Yanchun Liang, Fausto Giunchiglia, Lan Huang, and Xiaoyue Feng. 2020. Deep feature-based text clustering and its explanation. *IEEE Transactions on Knowledge and Data Engineering* (2020).
[62] Jiafeng Guo, Yixing Fan, Liang Pang, Liu Yang, Qingyao Ai, Hamed Zamani, Chen Wu, W. Bruce Croft, and Xueqi Cheng. 2020. A deep look into neural ranking models for information retrieval. *Information Processing & Management* 57, 6 (2020), 102067.
[63] Vishal Gupta, Gurpreet S. Lehal, et al. 2009. A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence* 1, 1 (2009), 60–76.
[64] Jochen Hartmann, Juliana Huppertz, Christina Schamp, and Mark Heitmann. 2019. Comparing automated text classification methods. *International Journal of Research in Marketing* 36, 1 (2019), 20–38.
[65] Keiko Hashizume, David G. Rosado, Eduardo Fernández-Medina, and Eduardo B. Fernandez. 2013. An analysis of security issues for cloud computing. *Journal of Internet Services and Applications* 4, 1 (2013), 1–13.
[66] Haibo He and Edwardo A. Garcia. 2009. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering* 21, 9 (2009), 1263–1284.
[67] Mingxing He, Shi-Jinn Horng, Pingzhi Fan, Muhammad Khurram Khan, Ray-Shine Run, Jui-Lin Lai, Rong-Jian Chen, and Adi Sutanto. 2011. An efficient phishing webpage detector. *Expert Systems with Applications* 38, 10 (2011), 12018–12027.
[68] Wu He, Ivan Ash, Mohd Anwar, Ling Li, Xiaohong Yuan, Li Xu, and Xin Tian. 2019. Improving employees' intellectual capacity for cybersecurity through evidence-based malware training. *Journal of Intellectual Capital* (2019), 203–213.
[69] Danny Hendler, Shay Kels, and Amir Rubin. 2018. Detecting malicious powershell commands using deep neural networks. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security (ASIACCS'18)*. Association for Computing Machinery, New York, NY, 187–197.
[70] Carolyn Holton. 2009. Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. *Decision Support Systems* 46, 4 (2009), 853–864.
[71] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. 2016. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5375–5384.
[72] Cheng Huang, JiaYong Liu, Yong Fang, and Zheng Zuo. 2016. A study on Web security incidents in China by analyzing vulnerability disclosure platforms. *Computers & Security* 58 (2016), 47–62.
[73] Keman Huang, Michael Siegel, and Stuart Madnick. 2018. Systematically understanding the cyber attack business: A survey. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 1–36.

[74] Xiaohong Huang, Yunlong Lu, Dandan Li, and Maode Ma. 2018. A novel mechanism for fast detection of transformed data leakage. *IEEE Access* 6 (2018), 35926–35936.
[75] Abdulmalik Humayed, Jingqiang Lin, Fengjun Li, and Bo Luo. 2017. Cyber-physical systems security—A survey. *IEEE Internet of Things Journal* 4, 6 (2017), 1802–1831.
[76] Mubashar Hussain, Mansoor Ahmed, Hasan Ali Khattak, Muhammad Imran, Abid Khan, Sadia Din, Awais Ahmad, Gwanggil Jeon, and Alavalapati Goutham Reddy. 2018. Towards ontology-based multilingual URL filtering: A big data problem. *The Journal of Supercomputing* 74, 10 (Apr. 2018), 5003–5021.
[77] Breach Level Index. 2018. *2018: Data Privacy and New Regulations Take Center Stage*. Technical Report. Breach Level Index.
[78] International Organization for Standardization 2018. *ISO 27000:2018(E)*. International Organization for Standardization, Geneva, Switzerland.
[79] International Electrotechnical Commission International Organization for Standardization. 2012. ISO/IEC 27032: 2012—Information technology—Security techniques—Guidelines for cybersecurity.
[80] Rizwana Irfan, Christine K. King, Daniel Grages, Sam Ewen, Samee U. Khan, Sajjad A. Madani, Joanna Kolodziej, Lizhe Wang, Dan Chen, Ammar Rayes, et al. 2015. A survey on text mining in social networks. *The Knowledge Engineering Review* 30, 2 (2015), 157–170. https://doi.org/10.1017/S0269888914000277
[81] ISACA. 2019. Glossary of Terms. Retrieved December 3, 2019 from https://csrc.nist.gov/publications/detail/nistir/7298/rev-3/final.
[82] $(ISC)^2$. 2019. Cybersecurity Workforce Study 2019. Retrieved from https://www.isc2.org/Research/2019-Cybersecurity-Workforce-Study.
[83] Carina Jacobi, Wouter Van Atteveldt, and Kasper Welbers. 2016. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism* 4, 1 (2016), 89–106.
[84] Julian Jang-Jaccard and Surya Nepal. 2014. A survey of emerging threats in cybersecurity. *Journal of Comput.er and System Sciences* 80, 5 (2014), 973–993. https://doi.org/10.1016/j.jcss.2014.02.005
[85] Samuel Jero, Maria Leonor Pacheco, Dan Goldwasser, and Cristina Nita-Rotaru. 2019. Leveraging textual specifications for grammar-based fuzzing of network protocols. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 9478–9483.
[86] Taeho Jo. 2018. *Text Mining: Concepts, Implementation, and Big Data Challenge*. Vol. 45. Springer.
[87] Jae Woong Joo, Seo Yeon Moon, Saurabh Singh, and Jong Hyuk Park. 2017. S-Detector: An enhanced security model for detecting Smishing attack for mobile computing. *Telecommunication Systems* 66, 1 (Jan. 2017), 29–38.
[88] Arnav Joshi, Ravendar Lal, Tim Finin, and Anupam Joshi. 2013. Extracting cybersecurity related linked data from text. In *Proceedings of the 2013 IEEE 7th International Conference on Semantic Computing*. IEEE Computer Society, 252–259.
[89] Gilad Katz, Yuval Elovici, and Bracha Shapira. 2014. CoBAn: A context based model for data leakage prevention. *Information Sciences* 262 (2014), 137–158.
[90] Saad Khan, Simon Parkinson, and Yongrui Qin. 2017. Fog computing security: A review of current applications and security solutions. *Journal of Cloud Computing* 6, 1 (2017), 1–22.
[91] Rupinder Paul Khandpur, Taoran Ji, Steve Jan, Gang Wang, Chang-Tien Lu, and Naren Ramakrishnan. 2017. Crowdsourcing cybersecurity: Cyber attack detection using social media. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM'17)*, Vol. Part F131841. Association for Computing Machinery, 1049–1057.
[92] Zeynep H. Kilimci and Selim Akyokus. 2018. Deep learning-and word embedding-based heterogeneous classifier ensembles for text classification. *Complexity* 2018 (2018).
[93] Nir Kshetri. 2006. The simple economics of cybercrimes. *IEEE Security & Privacy* 4, 1 (2006), 33–39.
[94] Sneha Kudugunta and Emilio Ferrara. 2018. Deep neural networks for bot detection. *Information Sciences* 467 (2018), 312–322.
[95] Rick Kuhn, Mohammad Raunak, and Raghu Kacker. 2017. It doesn't have to be like this: Cybersecurity vulnerability trends. *IT Professional* 19, 6 (2017), 66–70.
[96] Akshi Kumar, Vaibhav Singh, Tuba Ali, Saurabh Pal, and Jeevanjot Singh. 2020. Empirical evaluation of shallow and deep classifiers for rumor detection. In *Advances in Computing and Intelligent Systems*. Springer, 239–252.
[97] B. Shravan Kumar and Vadlamani Ravi. 2016. A survey of the applications of text mining in financial domain. *Knowledge-Based Systems* 114 (2016), 128–147. https://doi.org/10.1016/j.knosys.2016.10.003
[98] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*.
[99] Quentin Le Sceller, ElMouatez Billah Karbab, Mourad Debbabi, and Farkhund Iqbal. 2017. SONAR: Automatic detection of cyber security events over the twitter stream. In *Proceedings of the 12th International Conference on Availability, Reliability and Security*. Association for Computing Machinery, 1–11.

[100] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436–444.
[101] Kuo-Chan Lee, Chih-Hung Hsieh, Li-Jia Wei, Ching-Hao Mao, Jyun-Han Dai, and Yu-Ting Kuang. 2016. Sec-Buzzer: Cyber security emerging topic mining with open threat intelligence retrieval and timeline event annotation. *Soft Computing* 21, 11 (July 2016), 2883–2896. https://doi.org/10.1007/s00500-016-2265-0
[102] Hang Li. 2014. Learning to rank for information retrieval and natural language processing. *Synthesis Lectures on Human Language Technologies* 7, 3 (2014), 1–121.
[103] Jian-hua Li. 2018. Cyber security meets artificial intelligence: A survey. *Frontiers of Information Technology & Electronic Engineering* 19, 12 (2018), 1462–1474.
[104] Jenny S. Li, Li-Chiou Chen, John V. Monaco, Pranjal Singh, and Charles C. Tappert. 2017. A comparison of classifiers and features for authorship authentication of social networking messages. *Concurrency Computation* 29, 14 (2017).
[105] Qian Li, Hao Peng, Jianxin Li, Congyin Xia, Renyu Yang, Lichao Sun, Philip S. Yu, and Lifang He. 2020. A survey on text classification: From shallow to deep learning. Retreived on April 2021 from https://arxiv.org/abs/2008.00364.
[106] Weifeng Li, Hsinchun Chen, and Jay F. Nunamaker, Jr. 2016. Identifying and profiling key sellers in cyber carding community: AZSecure text mining system. *Journal of Management Information Systems* 33, 4 (2016), 1059–1086.
[107] Junqiang Liu and Ke Wang. 2012. Anonymizing bag-valued sparse data by semantic similarity-based clustering. *Knowledge and Information Systems* 35, 2 (June 2012), 435–461.
[108] Mitch Macdonald, Richard Frank, Joseph Mei, and Bryan Monk. 2015. Identifying digital threats in a hacker web forum. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 (ASONAM'15)*. Association for Computing Machinery, New York, NY, 926–933.
[109] William Marcellino, Kate Cox, Katerina Galai, Linda Slapakova, Amber Jaycocks, and Ruth Harris. 2020. *Human-Machine Detection of Online-Based Malign Information*. Technical Report. RAND Corporation.
[110] I. V. Mashechkin, M. I. Petrovskiy, D. S. Popov, and Dmitry V. Tsarev. 2015. Applying text mining methods for data loss prevention. *Programming and Computer Software* 41, 1 (Jan. 2015), 23–30. https://doi.org/10.1134/S0361768815010041
[111] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal* 5, 4 (2014), 1093–1113.
[112] Hasan Mesut Meral, Bülent Sankur, A. Sumru Özsoy, Tunga Güngör, and Emre Sevinç. 2009. Natural language watermarking via morphosyntactic alterations. *Computer Speech & Language* 23, 1 (2009), 107–125.
[113] Nikola Milosevic, Ali Dehghantanha, and Kim-Kwang Raymond Choo. 2017. Machine learning aided Android malware classification. *Computers & Electrical Engineering* 61 (2017), 266–274.
[114] Natalia Miloslavskaya and Alexander Tolstoy. 2016. Big data, fast data and data lake concepts. *Procedia Computer Science* 88 (2016), 300–305.
[115] Gary Miner, John Elder IV, Andrew Fast, Thomas Hill, Robert Nisbet, and Dursun Delen. 2012. *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*. Academic Press.
[116] Bhaskar Mitra and Nick Craswell. 2017. Neural models for information retrieval. Retreived on April 2021 from https://arxiv.org/abs/1705.01509.
[117] Mandar Mitra and B. B. Chaudhuri. 2000. Information retrieval from documents: A survey. *Information Retrieval* 2, 2 (2000), 141–163.
[118] Sudip Mittal, Prajit Kumar Das, Varish Mulwad, Anupam Joshi, and Tim Finin. 2016. CyberTwitter: Using Twitter to generate alerts for cybersecurity threats and vulnerabilities. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'16)*. Institute of Electrical and Electronics Engineers Inc., 860–867.
[119] Mahmood Moghimi and Ali Yazdian Varjani. 2016. New rule-based phishing detection method. *Expert Systems with Applications* 53 (2016), 231–242.
[120] Daša Munková, Michal Munk, and Martin Vozár. 2013. Data pre-processing evaluation for text mining: Transaction/sequence model. *Procedia Computer Science* 18 (2013), 1198–1207.
[121] Shane Murnion, William J. Buchanan, Adrian Smales, and Gordon Russell. 2018. Machine learning and semantic analysis of in-game chat for cyberbullying. *Computers & Security* 76 (2018), 197–213.
[122] David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes* 30, 1 (2007), 3–26.
[123] Maitri P. Naik, Harshadkumar B. Prajapati, and Vipul K. Dabhi. 2015. A survey on semantic document clustering. In *Proceedings of the 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT'15)*. IEEE, 1–10.
[124] B. Sri Nandhini and J. I. Sheeba. 2015. Online social network bullying detection using intelligence techniques. *Procedia Computer Science* 45 (2015), 485–492.
[125] Masoud Narouei, Hamed Khanpour, Hassan Takabi, Natalie Parde, and Rodney Nielsen. 2017. Towards a top-down policy engineering framework for attribute-based access control. In *Proceedings of the 22nd ACM Symposium on Access Control Models and Technologies*. 103–114.

[126] Masoud Narouei, Hassan Takabi, and Rodney Nielsen. 2018. Automatic extraction of access control policies from natural language documents. *IEEE Transactions on Dependable and Secure Computing* (2018), 1–1.
[127] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, and David Chek Ling Ngo. 2014. Text mining for market prediction: A systematic review. *Expert Systems with Applications* 41, 16 (2014), 7653–7670.
[128] Fitzroy Nembhard, Marco Carvalho, and Thomas Eskridge. 2018. A hybrid approach to improving program security. In *Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (SSCI'18)*, Vol. 2018-January. Institute of Electrical and Electronics Engineers Inc., 1–8.
[129] Lily Hay Newman. 2019. Google's Making It Easier to Safeguard Sensitive Data Troves. Retrieved from https://www.wired.com/story/google-data-loss-prevention-interface/.
[130] Umara Noor, Zahid Anwar, Tehmina Amjad, and Kim-Kwang Raymond Choo. 2019. A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise. *Future Generation Computer Systems* 96 (2019), 227–242.
[131] Umara Noor, Zahid Anwar, Asad Waqar Malik, Sharifullah Khan, and Shahzad Saleem. 2019. A machine learning framework for investigating data breaches based on semantic analysis of adversary's attack patterns in threat intelligence repositories. *Future Generation Computer Systems* 95 (2019), 467–487.
[132] Jeremiah O'Connor. 2020. Utilizing NLP To Detect APT in DNS. Retrieved from https://umbrella.cisco.com/blog/nlp-apt-dns.
[133] Alison O'Mara-Eves, James Thomas, John McNaught, Makoto Miwa, and Sophia Ananiadou. 2015. Using text mining for study identification in systematic reviews: A systematic review of current approaches. *Systematic Reviews* 4, 1 (2015), 5.
[134] Lecia Papadopoulos. 2017. How Watson AI is helping companies stay ahead of hackers and cybersecurity attacks. Retrieved from https://www.ibm.com/blogs/watson/2017/08/how-watson-ai-is-helping-companies-stay-ahead-of-cybersecurity-attacks/.
[135] Won Park, Youngin You, and Kyungho Lee. 2018. Detecting potential insider threat: Analyzing insiders' sentiment exposed in social media. *Security and Communication Networks* 2018 (2018). https://doi.org/10.1155/2018/7243296
[136] Justin W. Patchin and Sameer Hinduja. 2020. Sextortion among adolescents: Results from a national survey of US youth. *Sexual Abuse* 32, 1 (2020), 30–54.
[137] Celia Paulsen and Robert Byers. 2019. *Glossary of Key Information Security Terms*. Technical Report NISTIR 7298. National Institute of Standards and Technology (NIST).
[138] Marek Pawlicki, Michał Choraś, Rafał Kozik, and Witold Hołubowicz. 2020. On the Impact of network data balancing in cybersecurity applications. *International Conference on Computational Science*. Springer, 196–210.
[139] Mirjana Pejić Bach, Živko Krstić, Sanja Seljan, and Lejla Turulja. 2019. Text mining for big data analysis in financial sector: A literature review. *Sustainability* 11, 5 (2019), 1277.
[140] Hector Pellet, Stavros Shiaeles, and Stavros Stavrou. 2019. Localising social network users and profiling their movement. *Computers & Security* 81 (2019), 49–57.
[141] Ian Perera, Jena Hwang, Kevin Bayas, Bonnie Dorr, and Yorick Wilks. 2019. Cyberattack prediction through public text analysis and mini-theories. In *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data'18)*. Institute of Electrical and Electronics Engineers Inc., 3001–3010.
[142] Tien D. Phan and Nur Zincir-Heywood. 2018. User identification via neural network based language models. *International Journal of Network Management* (2018). https://doi.org/10.1002/nem.2049
[143] Raymond Pompon. 2016. Vulnerability management. *IT Security Risk Control Management*. Springer, 165–174.
[144] Clay Posey, Uzma Raja, Robert E. Crossler, and A. J. Burns. 2017. Taking stock of organisations' protection of privacy: Categorising and assessing threats to personally identifiable information in the USA. *European Journal of Information Systems* 26, 6 (Nov. 2017), 585–604.
[145] Santiago Quintero-Bonilla and Angel Martín del Rey. 2020. A new proposal on the advanced persistent threat: A survey. *Applied Sciences* 10, 11 (2020), 3874.
[146] David Reinsel, John Gantz, and John Rydning. 2017. *Data Age 2025: The Evolution of Data to Life-Critical*. Technical Report. IDC, Framingham.
[147] Jitendra Kumar Rout, Smriti Singh, Sanjay Kumar Jena, and Sambit Bakshi. 2016. Deceptive review detection using labeled and unlabeled data. *Multimedia Tools and Applications* 76, 3 (Aug. 2016), 3187–3211.
[148] Arjun Sambamoorthy. 2019. Applying Human Language Understanding to DLP's Biggest Challenges. Retrieved from https://www.armorblox.com/blog/reinventing-dlp-with-natural-language-understanding/.
[149] Anna Sapienza, Alessandro Bessi, Saranya Damodaran, Paulo Shakarian, Kristina Lerman, and Emilio Ferrara. 2017. Early warnings of cyber threats in online discussions. *Proceedings of the IEEE International Conference on Data Mining Workshops (ICDMW'17)*, 667–674.
[150] Mark Scherfling. 2020. Machine Learning with Natural Language Processing. Retrieved from https://community.ibm.com/community/user/security/blogs/mark-scherfling1/2020/05/04/nlp-for-resilient.

[151] Marian K. Schneider. 2020. Election security: Increasing election integrity by improving cybersecurity. *The Future of Election Administration*. Springer, 243–259.
[152] Sicong Shao, Cihan Tunc, Amany Al-Shawi, and Salim Hariri. 2019. Autonomic author identification in internet relay chat (IRC). In *Proceedings of the IEEE/ACS International Conference on Computer Systems and Applications (AICCSA'18)*. https://doi.org/10.1109/AICCSA.2018.8612780
[153] Jasmeet Singh and Vishal Gupta. 2017. A systematic review of text stemming techniques. *Artificial Intelligence Review* 48, 2 (2017), 157–217.
[154] John Slankas, Xusheng Xiao, Laurie Williams, and Tao Xie. 2014. Relation extraction for inferring access control rules from natural language artifacts. In *Proceedings of the 30th Annual Computer Security Applications Conference. ACM International Conference Proceeding Series*, 366–375.
[155] Robert Slonje and Peter K. Smith. 2008. Cyberbullying: Another main type of bullying? *Scandinavian Journal of Psychology* 49, 2 (2008), 147–154.
[156] Melissa Smoker and Evita March. 2017. Predicting perpetration of intimate partner cyberstalking: Gender and the dark tetrad. *Computers in Human Behavior* 72 (2017), 390–396.
[157] Marina Sokolova and Guy Lapalme. 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management* 45, 4 (2009), 427–437.
[158] Bhavna Soman. 2019. From unstructured data to actionable intelligence: Using machine learning for threat intelligence. Retrieved from https://www.microsoft.com/security/blog/2019/08/08/from-unstructured-data-to-actionable-intelligence-using-machine-learning-for-threat-intelligence/.
[159] Georgios Spanos and Lefteris Angelis. 2016. The impact of information security events to the stock market: A systematic literature review. *Computers & Security* 58 (2016), 216–229.
[160] Georgios Spanos and Lefteris Angelis. 2018. A multi-target approach to estimate software vulnerability characteristics and severity scores. *Journal of Systems and Software* 146 (2018), 152–166.
[161] Dima Suleiman and Ghazi Al-Naymat. 2017. SMS spam detection using H2O framework. *Procedia Computer Science* 113 (2017), 154–161.
[162] Nan Sun, Jun Zhang, Paul Rimba, Shang Gao, Leo Yu Zhang, and Yang Xiang. 2019. Data-driven cybersecurity incident prediction: A survey. *IEEE Communications Surveys Tutorials* 21, 2 (Second quarter 2019), 1744–1772.
[163] Ahmet Ali Süzen. 2020. A risk-assessment of cyber attacks and defense strategies in industry 4.0 ecosystem. *International Journal of Computer Network & Information Security* 12, 1 (2020).
[164] Romilla Syed. 2018. Enterprise reputation threats on social media: A case of data breach framing. *The Journal of Strategic Information Systems* 28, 3 (2018), 257–274.
[165] Romilla Syed, Maryam Rahafrooz, and Jeffrey M. Keisler. 2018. What it takes to get retweeted: An analysis of software vulnerability messages. *Computers in Human Behavior* 80 (2018), 207–215.
[166] Symantec. 2019. Internet Security Threat Report. Retrieved from https://www.symantec.com/content/dam/symantec/docs/reports/istr-24-2019-en.pdf.
[167] Chee-Wooi Ten, Govindarasu Manimaran, and Chen-Ching Liu. 2010. Cybersecurity for critical infrastructures: Attack and defense modeling. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 40, 4 (2010), 853–865.
[168] Kutub Thakur, Juan Shan, and Al-Sakib Khan Pathan. 2018. Innovations of phishing defense: The mechanism, measurement and defense strategies. *International Journal of Communication Networks and Information Security* 10, 1 (2018), 19–27.
[169] Dirk Thorleuchter and Dirk Van den Poel. 2012. Improved multilevel security with latent semantic indexing. *Expert Systems with Applications* 39, 18 (Dec. 2012), 13462–13471.
[170] Monica Todros. 2018. Artificial Intelligence in Black and White. Retrieved from https://www.recordedfuture.com/artificial-intelligence-information-security/.
[171] Andeep S. Toor, Harry Wechsler, Michele Nappi, and Kim-Kwang Raymond Choo. 2018. Visual question authentication protocol (VQAP). *Computers & Security* 76 (2018), 285–294.
[172] Wiem Tounsi and Helmi Rais. 2018. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security* 72 (2018), 212–233.
[173] Alper Kursat Uysal and Serkan Gunal. 2014. The impact of preprocessing on text classification. *Information Processing & Management* 50, 1 (2014), 104–112.
[174] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*. 5998–6008.
[175] Ike Vayansky and Sathish A. P. Kumar. 2020. A review of topic modeling methods. *Information Systems* 94 (2020), 101582.
[176] Maria Vergelis, Tatyana Shcherbakova, and Tatyana Sidorina. 2019. Spam and phishing in Q2 2019. Retrieved November 30, 2019 from https://securelist.com/spam-and-phishing-in-q2-2019/92379/.

[177] Sokratis Vidros, Constantinos Kolias, Georgios Kambourakis, and Leman Akoglu. 2017. Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset. *Future Internet* 9, 1 (March 2017), 6.
[178] G. Vinodhini and R. M. Chandrasekaran. 2016. A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. *Journal of King Saud University-Computer and Information Sciences* 28, 1 (2016), 2–12.
[179] Rossouw Von Solms and Johan Van Niekerk. 2013. From information security to cyber security. *Computers & Security* 38 (2013), 97–102.
[180] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.
[181] Tara Walker. 2017. Hello Amazon Macie: Automatically Discover, Classify, and Secure Content at Scale. Retrieved from https://aws.amazon.com/blogs/aws/launch-amazon-macie-securing-your-s3-buckets/.
[182] Peng Wang, Bo Xu, Jiaming Xu, Guanhua Tian, Cheng-Lin Liu, and Hongwei Hao. 2016. Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification. *Neurocomputing* 174 (2016), 806–814.
[183] Qihua Wang and Hongxia Jin. 2011. Data leakage mitigation for discretionary access control in collaboration clouds. In *Proceedings of the 16th ACM Symposium on Access Control Models and Technologies*. Association for Computing Machinery, 103–112. https://doi.org/10.1145/1998441.1998457
[184] Tawei Wang, Karthik N. Kannan, and Jackie Rees Ulmer. 2013. The association between the disclosure and the realization of information security risk factors. *Information Systems Research* 24, 2 (June 2013), 201–218.
[185] Zuoguang Wang, Hongsong Zhu, and Limin Sun. 2021. Social engineering in cybersecurity: Effect mechanisms, human vulnerabilities and attack methods. *IEEE Access* 9 (2021), 11895–11910.
[186] Tao Wen, Yuqing Zhang, Ying Dong, and Gang Yang. 2015. A novel automatic severity vulnerability assessment framework. *Journal of Communications* 10, 5 (2015), 320–329. https://doi.org/10.12720/jcm.10.5.320-329
[187] Tao Wen, Yuqing Zhang, Qianru Wu, and Gang Yang. 2015. ASVC: An automatic security vulnerability categorization framework based on novel features of vulnerability data. *Journal of Communications* 10, 2 (2015), 107–116.
[188] Mark A. Williams, Sumi Dey, Roberto Camacho Barranco, Sheikh Motahar Naim, M. Shahriar Hossain, and Monika Akbar. 2019. Analyzing evolving trends of vulnerabilities in national vulnerability database. In *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data'18)*. Institute of Electrical and Electronics Engineers Inc., 3011–3020.
[189] Alex S. Wilner. 2018. Cybersecurity and its discontents: Artificial intelligence, the Internet of Things, and digital misinformation. *International Journal* 73, 2 (2018), 308–316.
[190] Lingyun Xiang, Xingming Sun, Gang Luo, and Bin Xia. 2012. Linguistic steganalysis using the features derived from synonym frequency. *Multimedia Tools and Applications* 71, 3 (Dec. 2012), 1893–1911.
[191] Jiaming Xu, Bo Xu, Peng Wang, Suncong Zheng, Guanhua Tian, and Jun Zhao. 2017. Self-taught convolutional neural networks for short text clustering. *Neural Networks* 88 (2017), 22–31.
[192] Yueshen Xu, Lei Li, Honghao Gao, Lei Hei, Rui Li, and Yihao Wang. 2020. Sentiment classification with adversarial learning and attention mechanism. *Computational Intelligence* (2020).
[193] Jianwu Yang and Xiaoou Chen. 2002. A semi-structured document model for text mining. *Journal of Computer Science and Technology* 17, 5 (2002), 603–610.
[194] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2019. Attention-based convolutional approach for misinformation identification from massive and noisy microblog posts. *Computers and Security* 83 (2019), 106–121.
[195] Munwar Ali Zardari and Low Tang Jung. 2016. Data security rules/regulations based classification of file data using TsF-kNN algorithm. *Cluster Computing* 19, 1 (Feb. 2016), 349–368.
[196] ChengXiang Zhai and Sean Massung. 2016. *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*. Morgan & Claypool.
[197] Rui Zhao and Kezhi Mao. 2017. Fuzzy bag-of-words model for document representation. *IEEE Transactions on Fuzzy Systems* 26, 2 (2017), 794–804.
[198] Zhixuan Zhou, Huankang Guan, Meghana Moorthy Bhat, and Justin Hsu. 2019. Fake news detection via NLP is vulnerable to adversarial attacks. Retreived on April 2021 from https://arxiv.org/abs/1901.09657.
[199] Tiantian Zhu, Hongyu Gao, Yi Yang, Kai Bu, Yan Chen, Doug Downey, Kathy Lee, and Alok N. Choudhary. 2016. Beating the artificial chaos: Fighting OSN spam using its own templates. *IEEE/ACM Transactions on Networking* 24, 6 (Dec. 2016), 3856–3869.
[200] Elżbieta Zielińska, Wojciech Mazurczyk, and Krzysztof Szczypiorski. 2014. Trends in steganography. *Communications of the ACM* 57, 3 (2014), 86–95.
[201] Raed Abu Zitar and Adel Hamdan. 2011. Genetic optimized artificial immune system in spam detection: A review and a model. *Artificial Intelligence Review* 40, 3 (Nov. 2011), 305–377.

Received March 2020; revised August 2020; accepted April2021