Review



# Enabling Public Security Text-Based Analytics: A Survey to Outline Research Directions

Victor Diogho Heuer De Carvalho *, Robério José Rogério Dos Santos, Thyago Celso Cavalcante Nepomuceno *, Thiago Poleto

Posted Date: 1 March 2024

doi: 10.20944/preprints202403.0064.v1

Keywords: Text mining; Public security; Survey; Applications; Opportunities; Future Research Directions

 doi:10.20944/preprints202403.0064.v1



*Review*

# Enabling Public Security Text-Based Analytics: A Survey to Outline Research Directions

**Victor Diogho Heuer de Carvalho [1,*], Robério José Rogério dos Santos [1], Thyago Celso Cavalcante Nepomuceno [2] and Thiago Poleto [3]**

[1] Technologies Axis, Campus do Sertão, Universidade Federal de Alagoas, Delmiro Gouveia 57480-000, Brazil; roberio.santos@delmiro.ufal.br
[2] Department of Statistics, Center for Exact and Natural Sciences, Universidade Federal de Pernambuco, Recife 50740-550, Brazil; thyago.nepomuceno@ufpe.br
[3] Department of Business Administration, Institute of Applied Social Sciences, Universidade Federal do Pará, Balém 66075-110, Brazil; thiagopoleto@ufpa.br
* Correspondence: victor.carvalho@delmiro.ufal.br

**Abstract:** Text mining is a technological trend often highlighted in the continuous exchange of information through interconnected media. Its applicability goes beyond private organizations, as the public sector also requires it to treat textual information regarding services offered. Within this scenario, public security emerges as a prominent user of text mining that seeks to ensure the construction of data and knowledgebases to support decision-making about law enforcement actions to ensure citizen welfare. The primary objectives of this article are: (i) to develop a survey to identify text mining applications, techniques, opportunities, and challenges in public security, and (ii) to outline research directions concerning these topics and provide insights so that interested researchers can develop new studies. The literature was searched within four databases: Scopus, IEEE Xplore, ACM Digital Library, and Web of Science. A filtering process was applied to extract the works most aligned with the target theme, resulting in the selection of 194 of the most relevant works for a literature review. There were identified nineteen key applications of text mining related to public security and the most recurrent techniques and technologies reported between 2014 to 2021, supporting outlining three axes for future directions: one with possible expansion of objectives for new research; another on changes and adaptations in scopes for the methodological context; and the last one on expansions and changes in application scenarios based on the literature.

**Keywords:** text mining; public security; survey; applications; opportunities; future research directions

## 1. Introduction

Text mining is an area of artificial intelligence dedicated to data extraction from unstructured text to find helpful information and develop the knowledge needed to support decision-making [1,2]. Also known as knowledge discovery in textual databases, the process extracts previously unknown, understandable, potential, and practical patterns or knowledge from unstructured textual data [3,4]. Government and public agencies are aware of the relevance of using the social web, mobile services, artificial intelligence, the analytical processes enabled by their use, and the social benefits these tools bring to public administration and citizens [5]. Public health, environmental, and security surveillance, for instance, are areas that greatly benefit from the application of text mining which acts as a digital transformation vector able to assist in obtaining feedback from citizens on services [6–9]. Among other government and public administration areas, Zhang *et al.* [10] highlighted public security as a recipient of artificial intelligence applications, noting that there is interest in studying possible applications to improve management actions.

 doi:10.20944/preprints202403.0064.v1



Specifically, about public security, Han *et al.* [11] commented on the importance of general data mining applications to detect and prevent crimes, such as supporting the detection of fraud, money laundering, and insurance crimes. Hashimi *et al.* [12] stated the advantages of using text mining to provide automatic tasks to deal with incremental amounts of data by retrieving and extracting useful information, making predictions based on the observations and statistics provided, discovering patterns from the data provided for trend detection to support public security agencies in monitoring and analyzing textual data from the social web.

Also, regarding the benefits and advantages of using text mining in matters related to public security, Tseng *et al.* [13] presented a method to mine terms in text collections for crime investigations in their study. Crime investigation is one of the recurrent activities in public security, involving the analysis of various types of material related to the occurrence of crimes, including texts in different formats, whether printed or digital. Criminal databases belonging to police agencies may contain numerous texts that can describe detailed relationships between data such as the type of crime, suspects, victims, and locations so text mining tools aid in relationship discovery. Considering this context, the area of cybersecurity, aimed at combating incidents that could compromise the security of users of cyber systems, has a notable participation, providing strategic plans and tools to guarantee the integrity of systems and the information that flows through them [14].

Another facet of applying text mining in public security is that it enables an understanding of human behavior, as highlighted by the study developed by Tutun *et al.* [15]. They dedicated their study to understanding patterns and relationships in terrorist behavior using text mining techniques. It is noticed that any human activity that can compromise public security in general and that generates textual records, for example, through social networks, chats, emails, or blogs, can be analyzed with text mining techniques [16].

This article presents the results of a survey to extract key information about text mining in the context of public security, including applications, technologies, and directions for future research. A research agenda was outlined based on the directions identified, intending to support researchers in finding research gaps to develop new studies.

The sequence of this article is organized as follows. Section 2 contains a description of the systematic literature review process applied; Section 3 presents the results of the review and the related discussions, and at the end, it outlines future research directions supporting defining a research agenda based on identified opportunities and challenges, according to three axes; Section 4 concludes the work; Section 5 brings updates considering the interval between 2021 and 2024.

## 2. Survey Process

Surveys (or systematic literature reviews) are processed through formally defined bibliographic research protocols [17] that provide researchers with information and insight about a subject of interest, supporting new research [18,19], and reducing the effect of publication redundancy [20]. The literature review presented in this article was performed following the guidelines defined by Kitchenham and Charters [21] in three key steps: planning, conduction, and reporting. While this process appears sequential, it involves iterative steps [21]. The two first steps are presented in the following sections, and the last one represents the results in Section 3.

### *2.1. Methodological summary*

All methodological information and criteria related to the systematic literature review planning and subsequent search execution are presented in Table 1.

 doi:10.20944/preprints202403.0064.v1



**Table 1.** Information and criteria for the survey.

| | |
|---|---|
| Research Questions | RQ1: *What has been researched on application areas for text mining within the context of public security?*<br>RQ2: *What are the most employed text mining techniques and technologies in public security in general and for each application area?*<br>RQ3: *What research opportunities and challenges exist for text mining in public security?* |
| Databases | Scopus, Web of Science, IEEE Xplore, and ACM Digital Library. |
| Search String/Query | The following search string was applied over titles, abstracts, and keywords (or correlated, as each base allows):<br>*("text mining") AND ("public security" OR "crime" OR "terrorism" OR "piracy" OR "drug trafficking" OR "arms trafficking" OR "human trafficking" OR "sexual exploitation" OR "prostitution" OR "pedophilia" OR "rape" OR "homicide" OR "murder" OR "femicide" OR "infanticide" OR "bodily injury" OR "extortion" OR "theft" OR "robbery" OR "assault" OR "burglary" OR "property damage" OR "misappropriation" OR "money laundering" OR "embezzlement" OR "stellionate" OR "receiving" OR "kidnapping" OR "defamation" OR "cybercrime")* |
| Selection Criteria | Period: *2014 to 2021 for broad selection, and 2018 and 2021 to identify works with future research indications.*<br>Type: *Journal Articles, Complete Conference Papers, Book Chapters*<br>Language: *English only* |
| Information Extraction Strategy | Information of interest: *objectives, problems, techniques/methods/technologies, application in public security, keywords, indications about further research.*<br>Bibliographic information: *authors, title, kind of work (journal/conference/book), publication year, journal/conference/book name.* |
| Software | Mendeley software, Python language, and spreadsheets. |

The terms used in the Boolean condition were taken from the current Brazilian penal code (see the official Brazilian Penal Code website: http://www.planalto.gov.br/ccivil_03/decreto-lei/del2848.htm) and compared with those in (i) the American Model Penal Code – see Robinson and Dubber [22] –, and (ii) in the crime information according to the Crown Prosecution Service (see the Crown Prosecution Service website: https://www.cps.gov.uk/cps/crime-info) that covers Wales and England. Finally, Brazilian Jurisprudence (the JusBrasil portal was consulted: https://www.jusbrasil.com.br/) was also consulted, through documents in a public access portal, to further corroborate its validity in the search condition.

### *2.2. Review conduction*

The final search query was applied to the selected databases on June 11th, 2021, and it found 806 articles, including 485 from Scopus, 141 from Web of Science, 76 from IEEE Xplore, and 104 from ACM Digital Library. The subsequent filter eliminated 178 duplicates. The next stage consisted of screening the remaining material by reviewing titles and abstracts, resulting in the exclusion of 376 papers, leaving 252 remaining articles for the information extraction phase. Subsequently, 58 works were excluded because they were considered out of the literature review's scope. With these exclusions, the final number of findings in the literature was 194. Figure 1 represents the literature selection process with the number of works selected and excluded in each step.

 doi:10.20944/preprints202403.0064.v1



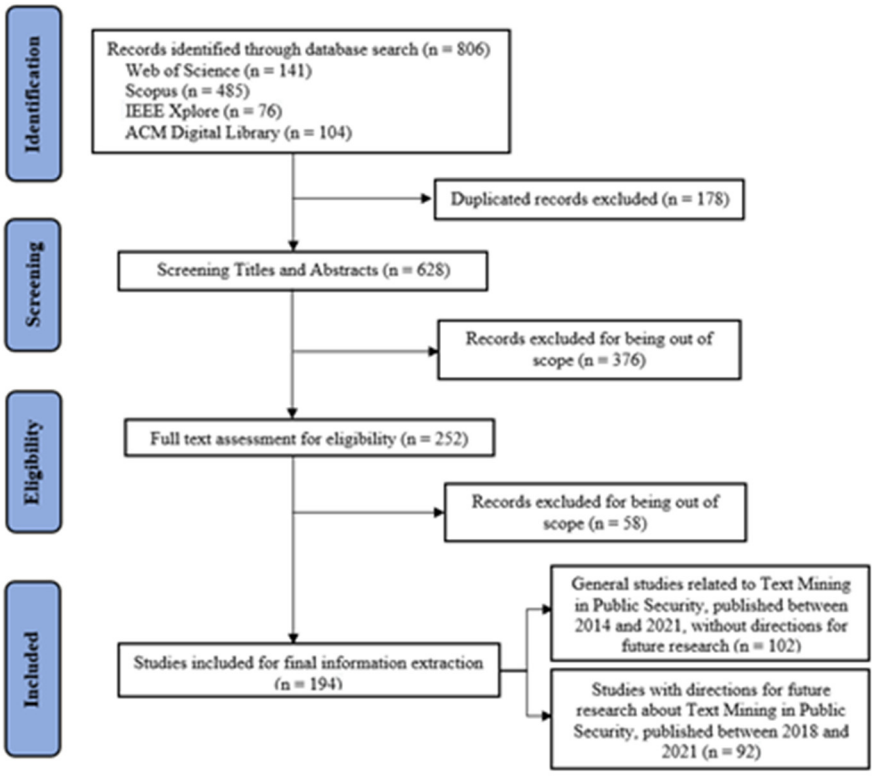


**Figure 1.** Literature selection process.

The third research question (RQ3) required additional filtering for the most recent literature (published between 2018 and the first semester of 2021) to identify works with future research proposals supporting the detection of opportunities and challenges on the research themes. In this final filter, a set of 92 studies was selected, containing the desired information, supporting a research agenda on text mining in public security. The following section will present and discuss the results of the literature review.

## 3. Results and Discussion

This section focuses on (i) identifying the primary applications performed according to the 194 selected articles, (ii) providing a visualization of the methods, techniques, and technologies used in the context of public security, and (iii) discovering the opportunities and challenges for the development of new research. The final part of this section is devoted to comments, based on the literature presented, on the ethical issue related to the use of text mining in public security. A spreadsheet with the information extracted from all selected literature is available in a GitHub public repository (see https://github.com/victorheuer/tm_ps_literature-info). More details about the results presented in this section are presented in de Carvalho and Costa [23].

### *3.1. General Findings*

Following the procedure presented in Section 3, general data were extracted from the 194 selected articles regarding the types of research, publication years, and keywords to enable the identification of interesting information. Figure 2 contains the distribution of these works according to the publication years from 2014 to 2021.

 doi:10.20944/preprints202403.0064.v1



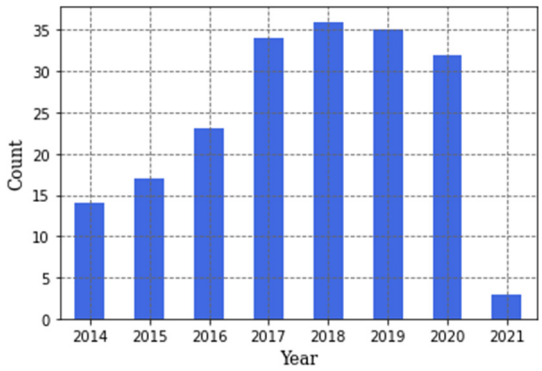


**Figure 2.** The counts of selected works between 2014 and 2021.

The largest number of works occurred in 2018 with 36 works, followed by 2019 with 35, 2017 with 34, and 2020 with 32. These four years added up to 137 works (70.10% of all selected publications). These data suggest that 2017 was a milestone year for the increase in publication numbers on topics of interest, with 47,83% more than 2016. It is important to emphasize that the final number of works for 2021 was not higher because data collection was completed in June 2020. Figure 3 shows the types of work distribution, with most being conference papers, followed by journal publications and book chapters.

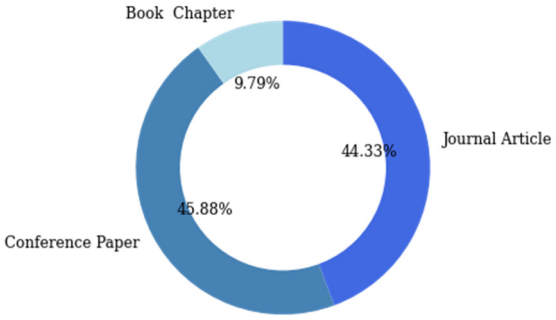


**Figure 3.** Selected literature totals by the type of work.

Most conference papers occurred in 2017, followed by 2016 and 2020, cumulating 50 works. 2018 incurred the largest number of journal articles, followed by 2019 and 2020, cumulating 53 works. Finally, the book chapters occurred at the same number in 2018 and 2019, followed by 2016 and 2020, cumulating 16 works. The distribution of these types by year is listed in Table 2.

**Table 2.** Distribution of the type of work per year.

| Type \ Year | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | Total |
|---|---|---|---|---|---|---|---|---|---|
| Conference Paper | 6 | 10 | 15 | 21 | 11 | 12 | 14 | 0 | 89 |
| Journal Article | 7 | 6 | 5 | 12 | 20 | 18 | 15 | 3 | 86 |
| Book Chapter | 1 | 1 | 3 | 1 | 5 | 5 | 3 | 0 | 19 |
| Total | 14 | 17 | 23 | 34 | 36 | 35 | 32 | 3 | 194 |

From the information extracted about the publication channels, it was possible to list the number of works published on text mining in public security defined according to each journal, conference, and book/series between 2014 and 2021. Table 3 contains the journals with more than one article published on the theme in the defined years.

 doi:10.20944/preprints202403.0064.v1



**Table 3.** List of journals with more than one article about text mining in public security published between 2014 and 2021.

| Journal | Count |
|---|---|
| Procedia Computer Science | 5 |
| Expert Systems with Applications | 4 |
| IEEE Access | 3 |
| Information Sciences | 3 |
| International Journal of Advanced Computer Science and Applications | 3 |
| Journal of Management Information Systems | 3 |
| Journal of Medical Internet Research | 3 |
| Crime Science | 2 |
| Digital Investigation | 2 |
| Information Processing & Management | 2 |
| Knowledge-Based Systems | 2 |
| Telematics and Informatics | 2 |

About these findings, it is important to note that Procedia Computer Science is a journal dedicated to publishing high-quality conference proceedings. The journals in Table 3 were extracted from a list containing 64 journals.

There was an extensive list of 84 different conferences, with only five of them – the 2016 Pacific Asia Conference on Information Systems, the 2017 European Intelligence and Security Informatics Conference, the 2018 IEEE International Conference on Intelligence and Security Informatics, 22nd Americas Conference on Information Systems, 8th International Conference on Computing, Communication, and Networking Technologies – with more than one paper from the selected literature. Finally, among the seventeen books from which the selected chapters came, one of them - Advances in Intelligent Systems and Computing - contained three selected chapters, and all the others contained only one.

### *3.2. Application areas for text mining in public security*

The first research question (RQ1) considers the application areas for text mining within the context of public security. According to what the authors made evident in their texts, the analysis of the selected studies detected nineteen different application areas. These areas and the count of related findings are listed in Table 4.

**Table 4.** The nineteen application areas within public security identified in the selected literature and the count of the related articles.

| Application Area | Count | % |
|---|---|---|
| Cybersecurity | 62 | 31.96 |
| General crime detection/prediction | 29 | 14.95 |
| Fraud detection | 22 | 11.34 |
| Terrorism detection | 16 | 8.25 |
| Cyberbullying detection | 14 | 7.22 |
| Digital / Cyber forensics | 14 | 7.22 |
| Support to the Judiciary power | 6 | 3.09 |
| Support to Law Enforcement agencies' actions | 6 | 3.09 |
| Crimes victims support | 4 | 2.06 |
| Sex-related crimes detection | 4 | 2.06 |
| Drug-related crimes detection | 3 | 1.55 |
| Espionage detection | 3 | 1.55 |
| Information security | 3 | 1.55 |

 doi:10.20944/preprints202403.0064.v1



| | | |
|---|---|---|
| Software piracy detection | 2 | 1.03 |
| Civil unrest detection | 2 | 1.03 |
| Drug-related crime detection and Weapons trafficking detection | 1 | 0.52 |
| Weapons trafficking detection | 1 | 0.52 |
| Armed conflicts solution | 1 | 0.52 |
| Violence against woman analysis | 1 | 0.52 |
| Total | 194 | 100.00 |

The nineteen application areas contained in Table 4 were defined according to the findings in the literature. Each journal article, conference paper, or chapter selected was manually tagged based on the primary area of application of text mining tools in public security, as per the reading of these materials, also seeking this information.

These results suggest that "Cybersecurity" is the area with most studies containing text mining applications. Following this field are "General Crime detection/prediction", "Fraud detection", as well as "Terrorism detection", "Cyberbullying detection", and "Digital/Cyber forensics", collectively representing 80.94% of the selected studies.

While the labels used to designate the corresponding topics seem simple, some areas, such as "General crime detection/prediction", "Support to Law Enforcement agencies' actions", and "Support to the Judiciary power" are more general. Others, such as "Digital/Cyber forensics", "Cyberbullying", and "Information security" are all aligned to "Cybersecurity" but remain separated to ensure more details about these areas.

The label "General crime detection/prediction" concerns an application area not dedicated to a specific type of crime, such as "Fraud detection" or "Drug-related crime detection." The works by Aghababaei and Makrehchi [24], Das and Das [25], Qazi and Wong [26], and Lal *et al.* [27] contain examples of applications in "General crime detection/prediction".

The "Support to Law Enforcement agencies' actions" application area is related to providing complete systems architecture, methodologies, and frameworks for agencies dedicated to ensuring compliance with the laws to maintain public security and social welfare. This set is composed of the works by Badii *et al.* [28] that proposed a system architecture to provide data analytics (including a text mining and analytics module) for supporting decision-making in law enforcement agencies; Bisio *et al.* [29] that proposed an approach to allow law enforcement agencies to detect events, using Twitter traffic monitoring, that compromise public security; Basilio *et al.* [30] that presented a methodology to extract knowledge from police reports for extracting information to support activities related to law enforcement; Behmer *et al.* [31] that proposed a framework to support law enforcement agencies in the investigations and analyzes of organized crime; Basilio *et al.* [32] that developed a method for knowledge discovery in emergency response databases based on police reports; and Hou *et al.* [33] that proposed the Bidirectional Encoder Representation from Transformers based on the Chinese relation extraction algorithm for public security, for security information mining.

The "Support to the Judiciary power" area refers to developing applications that aid judiciary activities since they may also be related to crime judgments and analysis or judicial reports about crimes. This set is composed of the works by Nikolić *et al.* [34] that proposed an e-Government service for extracting information from documents related to laws (Criminal Codes, for instance); Iftikhar *et al.* [35] that proposed a system to support courts' activities with text mining to extract relevant information from legal data; Pina-Sánchez *et al.* [36] that analyzed court sentence databases to detect ethical discrimination; Pina-Sánchez *et al.* [37] that proposed an approach to access data based on mining judiciary sentence records about crime available online; Xia *et al.* [38] that evaluated if judge gender exerted some effect over the sentences concerning rape, and Gomes and Ladeira [39] that applied an empirical evaluation of a framework for jurisprudence retrieval to ease the task of retrieval of other decisions with the same legal opinion.

The "Drug-related crimes detection and Weapons' trafficking detection" combination appears among the application areas, with only one study selected, by Al-Nabki *et al.* [40] that proposed a new feature replacing the use of external sources of knowledge, applying it to recognize named

 doi:10.20944/preprints202403.0064.v1



entities related to suspicious activities related to weapons and drug trafficking through the Tor Darknet. The distribution of the selected works by year is shown in Figure 4, highlighting cybersecurity as the area with the greatest number of selected works over seven years in the defined period.

### *3.3. Text mining techniques and technologies applied in public security*

The selected works' methodological sections were analyzed to answer the second research question (RQ2). In this case, information extraction was performed to identify the terms referring to techniques or technologies, counting their frequencies. Techniques are all the algorithms and methods used to make text mining viable, while technologies can be understood as tools such as programming languages, code libraries, and other software that contain implementations of these techniques.

For each term, the number of occurrences represents the number of works that included a specific technique while recognizing that each work could apply more than one technique or technology. Figure 5 contains the frequency of the 20 more recurrent terms.

The terms "support vector machines", "naïve Bayes", "random forests", "decision trees", "logistic regression", "*k*-nearest neighbors", and "neural networks" represent machine learning techniques applied to classification problems typically related to the detection or prediction of crimes within the context of the types of applications in the security areas presented in the previous section. Of these, "support vector machines" is the most frequent technique, being a discriminative classifier [41] and one of the most effective classification algorithms for general purposes [42].

The term "naïve Bayes" refers to one of the simplest generative machine learning classifiers [43], and its algorithm is based on the Bayes Theorem with independence assumptions between the predictors [44]. It is the second machine learning technique most frequently applied by the literature selected.

The term "random forests" refers is an ensemble technique with excellent predictive performance [42] using unpruned decision trees based on bootstrap samples of the training data [45]. "Decision trees" refer to another popular technique based on a tree data structure that contains a set of nodes and edges to support decision-making [43]. Both "random forests" and "decision trees" occurred the same number of times in the selected literature.

The term "logistic regression" refers to a generalized linear regression model [42] that makes predictions using a binary or multiclass outcoming [46]. The term "*k*-nearest neighbors" refers to a popular technique that assigns elements to a class with their neighbors according to a similarity measure (as in cosine and Jaccard similarities, for instance) [44,47].

The term "neural networks" refers to non-linear machine learning techniques that simulate the human brain to solve problems [43,48]. These networks establish relationships between inputs and outputs, associating input data to their belonging classes through a series of hidden layers and the links between the created nodes [49].

The term "latent Dirichlet allocation" refers to a machine learning technique dedicated to topic modeling. It is a generative probabilistic model used to identify latent topics among the texts in a corpus, modeling each corpus item as a finite mixture over a latent set of topics [50,51]. Topic modeling is the process of discovering hidden topics within semantic structures that contain interrelated concepts [52–54]

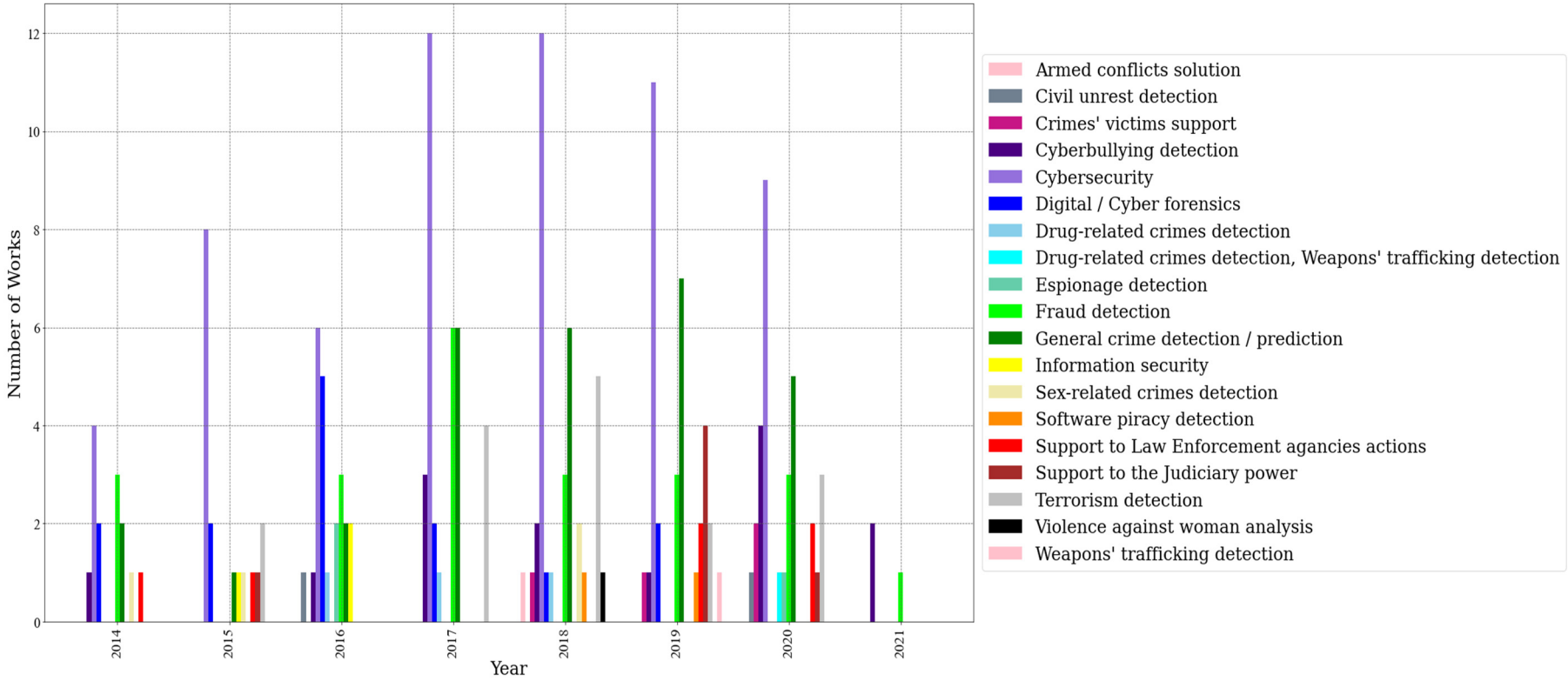


**Figure 4.** The distribution of selected works, based on the type of featured application spanning the period between 2014 and 2021.



doi:10.20944/preprints202403.0064.v1

 doi:10.20944/preprints202403.0064.v1



"Term frequency-inverse document frequency" is a statistical measure applied for feature extraction and selection, which consists of reducing the original set of textual data into a new set, more readable by other techniques, such as machine learning related ones [55–57]. Another related term is "term frequency", simply referring to counting the frequency of words in a text, being a component of "term frequency-inverse document frequency" calculation [54]. In Figure 5, a subset of terms refers to technologies for text mining solutions, such as programming languages, code libraries, and some programs specifically developed to apply data mining.

Among the programming languages, Python is the most recurrent. For instance, Al-Nabki *et al.* [40] applied Python with Keras framework in a neural network architecture to recognize named entities in suspicious Darkweb domains. Birks *et al.* [58] also used this programming language with Gensim wrapper to identify crime clusters. Bozyiğit *et al.* [59] used Python with Scikit-Learn to classify cyberbullying contents using texts extracted from social media.

The "Natural Language Toolkit" and "Scikit-Learn" are libraries developed in Python, the first deals with natural language processing problems, and the second contains pre-built machine learning techniques, such as many of those presented above. The "Scikit-Learn" library includes several machine learning techniques implemented with great flexibility for applications, as demonstrated in the works by Chen *et al.* [60], Dong *et al.* [61], Martín *et al.* [47], and Thao *et al.* [48]. The "Natural Language Toolkit" contains essential functions implemented to perform preprocessing tasks (Dong et al., 2018), "named entity recognition" [25], and to apply "term frequency-inverse document frequency" [55], for example. Preprocessing tasks involve applying natural language processing techniques to treat the texts by eliminating noise that affects the analytical process and formatting the text to perform subsequent processing. Examples of these preprocessing tasks include text cleaning and normalization, removing special characters, numbers, empty or white spaces, stop words, performing case folding, stemming, and lemmatizing, tokenization, and extraction of n-grams as evidenced by the work by Aboluwarin *et al.* [62], Chandra *et al.* [63], Gil *et al.* [64], Martín *et al.* [47], and Savaliya and Philip [65].

The "R language" is the second most recurrent within this set, containing several functions like the Python language and its libraries. The work by Basilio et al. (2019), for instance, applied preprocessing tasks and topic modeling using the R language. In addition to performing preprocessing, Cichosz [42] applied machine learning classification techniques from R language packages. Aboluwarin et al. (2016) applied the R language for preprocessing and several Scikit-Learn functions, using Python, to perform classifications. Other languages were detected but did not appear as regularly as Python and R, including Java [66–69]; Perl [36,37,70]; PHP [66,71]; and C++ [72].

Distinct from these programming languages, "WEKA" and "RapidMiner" are computer programs specifically developed for machine learning and data mining purposes. WEKA is a machine learning platform that contains several implemented techniques (Alothman & Rattadilok, 2017). The research by Das and Das [73,74] used WEKA to compare it with their methodology to process and analyze online newspaper reports covering crime. Almehmadi *et al.* [75] used WEKA to perform preprocessing tasks over a retrieved corpus and to apply a machine learning technique (support vector machines). RapidMiner is a software dedicated to data mining with several functions for data manipulation, statistical analysis, and graphic presentation [76]. Noviantho *et al.* [77] and Samtani *et al.* [78] are examples from the selected literature using RapidMiner and WEKA, the first paper dedicated to cyberbullying classification, the second for identifying and assessing vulnerabilities in Supervisory Control and Data Acquisition (SCADA) systems.

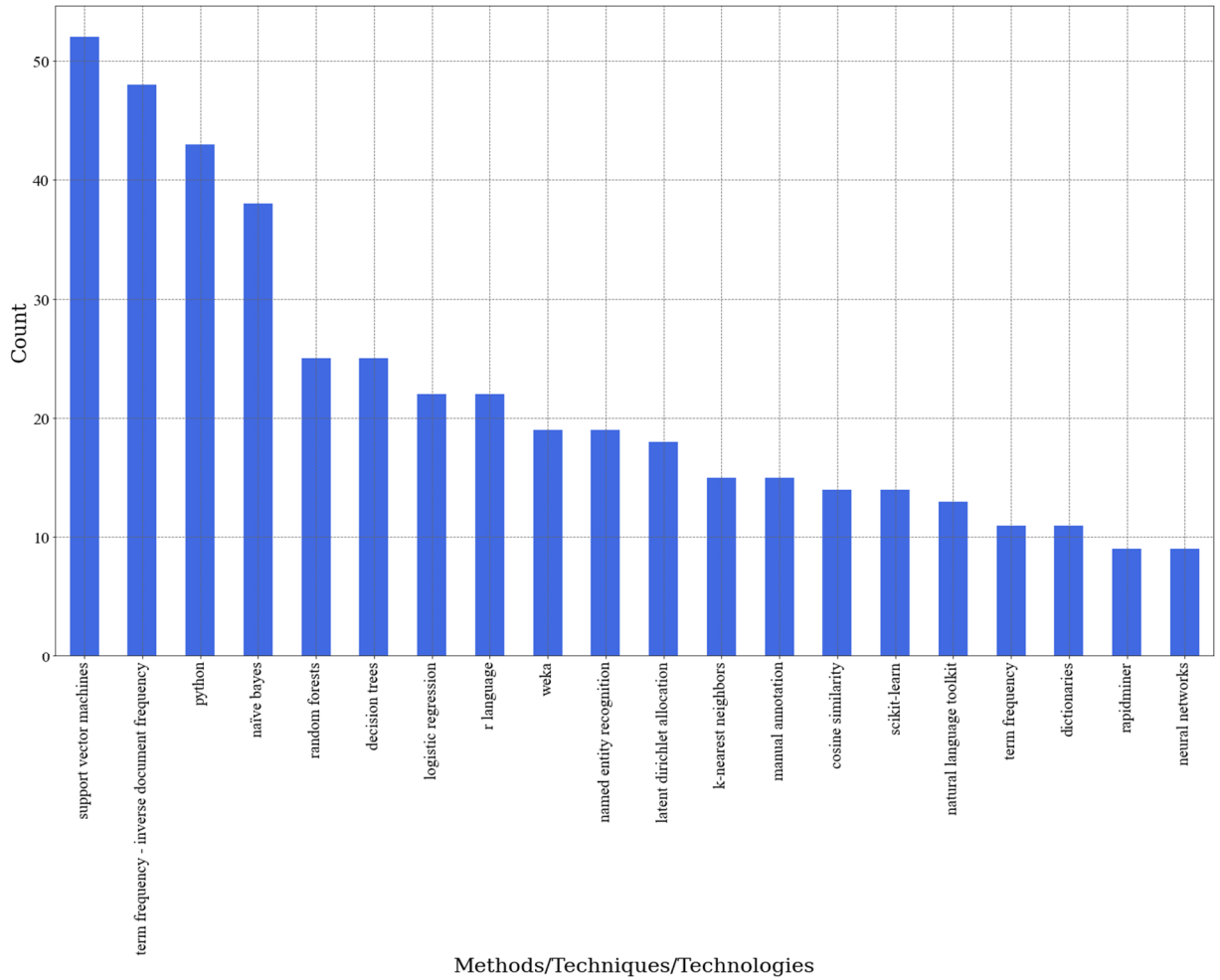


**Figure 5.** The most common terms related to text mining techniques and technologies.



doi:10.20944/preprints202403.0064.v1

 doi:10.20944/preprints202403.0064.v1



Terms like "named entity recognition", "manual annotation", and "dictionaries", refer to natural language processing subjects. The term "named entity recognition" refers to an information extraction task for detecting named entities that are related, such as people, organizations, locations, expressions of time, and money [79,80]. "Dictionaries" are lists composed of keywords extracted from texts with descriptions of the characteristics related to a target term or word [81]. These textual data structures present the sensitivity of a text or document as defined by the experts in the field to which it is related [82]. "Manual annotation" refers to the process of creating a corpus with some labels or tags, such as in sentiments' polarities, using expert people. Petrovskiy and Chikunov [83] and Saini and Bansal [84] applied manual annotation to create corpora, which were later used to train machine learning techniques for performing classifications.

Most frequent techniques and technologies by application area

According to each application area, a separation of the most frequent techniques and technologies was made to provide greater detail in answer to RQ2. Figure 6 contains an assembly with bar plots showing the counts for the most frequent techniques and technologies in the six areas with the most works selected (see Table 4). In this figure, there are cases where there is more than one term associated with a bar in the graph, indicating that each term has precisely the same number of occurrences as represented by the bar.

For “Cybersecurity”, the bar with four occurrences for each term involves: adaboost, named entity recognition, word clouds, and support vector machines. For “General crime detection/prediction”, the bar with two occurrences for each term involves: cluster analysis, georeferencing, logistic regression, natural language toolkit, neural networks, random forests, and rapidminer. For “Fraud detection”, the bar with two occurrences for each term involves: bagging, georeferencing, latent Dirichlet allocation, loss calculation, matlab, neural networks, risk calculation, scikit-learn, cosine similarity, and principal component analysis.

Figure 7 contains a way to visualize the combinations of terms between the six main areas according to the term extraction performed. The dots refer to terms appearing isolated, and dots connected by lines indicate term combinations. The bars on the left side are the number of occurrences among the six main areas, and the bars on the top of the plot are the counts of terms (isolated or in combination with other terms).

The term "naïve bayes" is an interesting case to exemplify the analyses that can be done with Figure 7: it appears five times among the terms listed in all areas, which determines that it is at the intersection of five areas; it also appears twice alone, once in combination with just the term “support vector machines”, once with "term frequency-inverse document frequency", and once with both the terms "r language" and "term frequency-inverse document frequency". The term “python” has similar behavior in this plot: it also appears five times, being in the intersection of five areas; twice it is isolated from other terms; once it is combined with “random forests” and “named entity recognition”; once it is combined with “term frequency-inverse document frequency” and “k-nearest neighbors”; and once it is combined with “support vector machines”, “named entity recognition” and “natural language toolkit”.

The most recurrent term is “support vector machines”, appearing six times, in other words, it is in the intersection of the six main areas. It is followed by “python”, “term frequency-inverse document frequency”, “random forests”, and “naïve bayes”. For more counts, see Figure 7.

 doi:10.20944/preprints202403.0064.v1



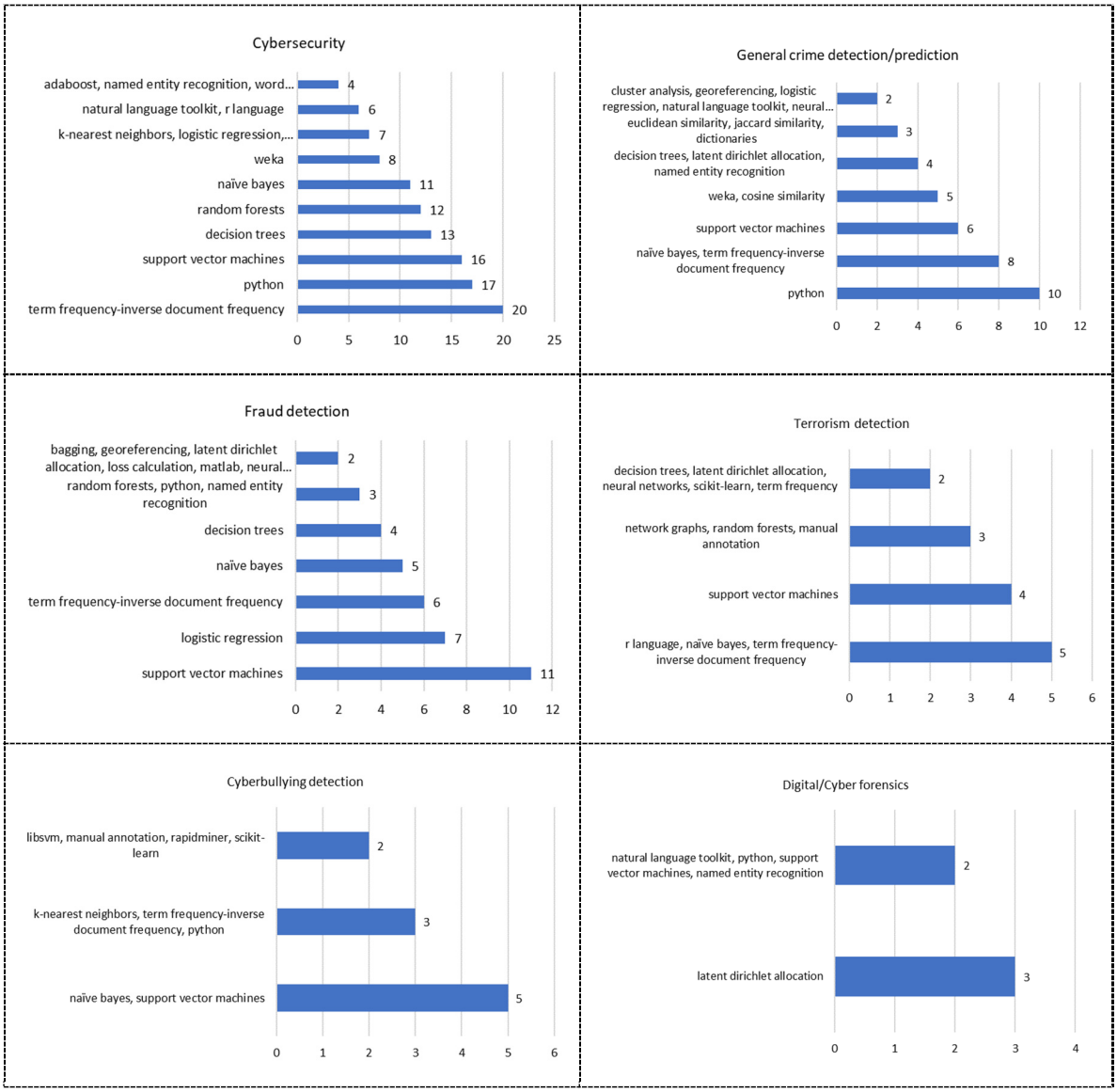


**Figure 6.** Count of the most frequent techniques/technologies in the six areas with the most work in the literature selection.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 March 2024 doi:10.20944/preprints202403.0064.v1



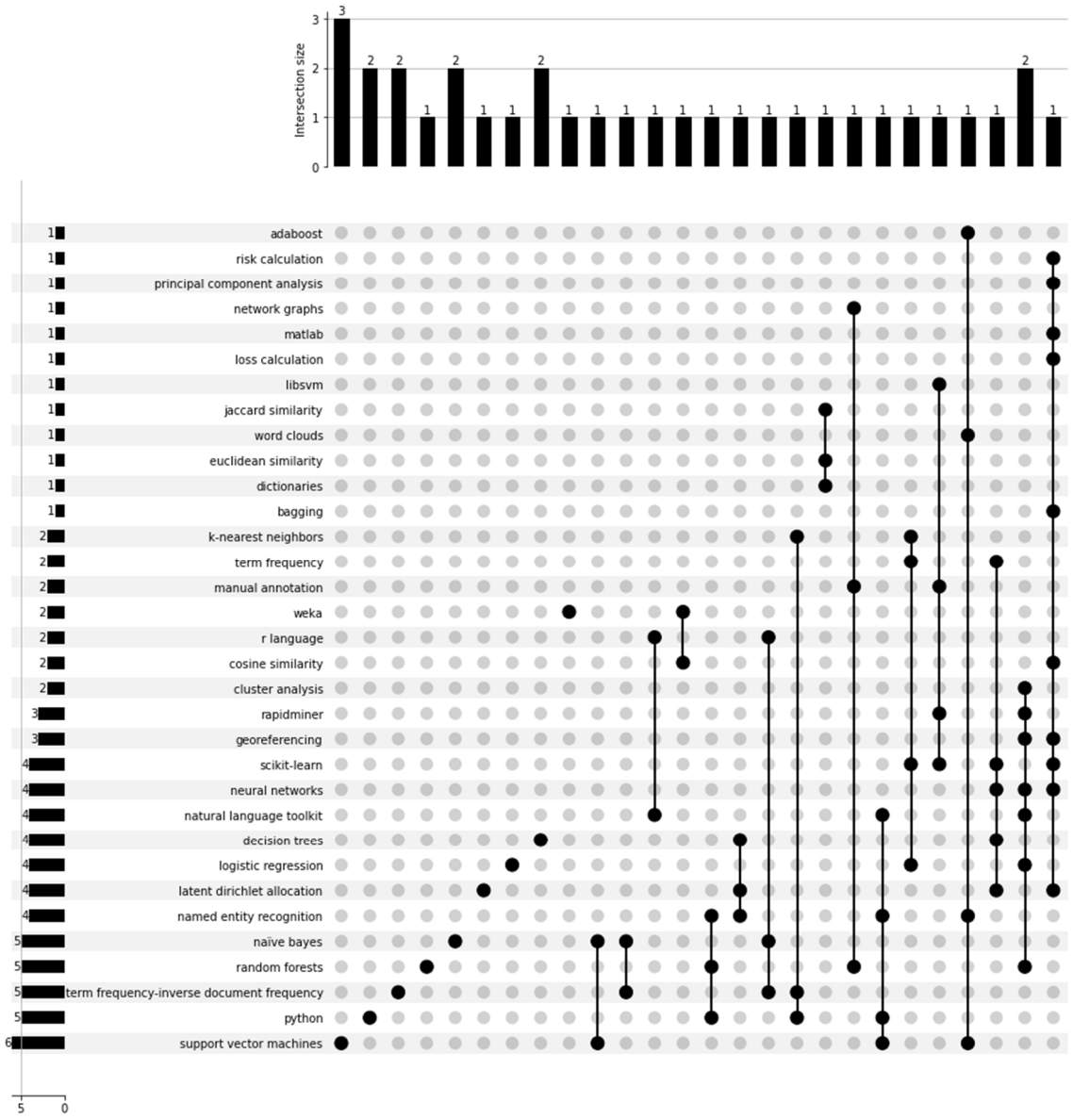


**Figure 7.** Intersections and combinations of terms between the six main areas of text mining applications in public security.

### *3.4. Identifying Opportunities and challenges for text mining in public security-related applications*

To identify opportunities and challenges as directions for text mining in the public security field (RQ3), the conclusion sections or correlates of the selected articles were analyzed during the last filtering process to extract only those that contained proposals or directions for future developments.

This filtering identified 92 works among the 194 selected for determining directions to enhance research already developed. Identifying and extracting these directions are critical to supporting the outline of the research agenda presented in the following subsections. These directions were grouped according to the applications presented in Section 4.2, and Table 5 lists the work counts from 2018 to 2021 that included future research indications by the application area.

 doi:10.20944/preprints202403.0064.v1



**Table 5.** The count of the most recent works, including future directions by application area.

| Application area | Count | | | | |
|---|---|---|---|---|---|
| | 2018 | 2019 | 2020 | 2021 | Total |
| Cybersecurity | 12 | 8 | 8 | 0 | 28 |
| General crime detection/prediction | 6 | 5 | 5 | 0 | 16 |
| Fraud detection | 2 | 3 | 3 | 1 | 9 |
| Terrorism detection | 5 | 1 | 2 | 0 | 8 |
| Digital/Cyber forensics | 1 | 1 | 0 | 0 | 2 |
| Cyberbullying detection | 2 | 1 | 4 | 2 | 9 |
| Support to Law Enforcement agencies actions | 0 | 2 | 2 | 0 | 4 |
| Sex-related crime detection | 1 | 0 | 0 | 0 | 1 |
| Support to the Judiciary power | 0 | 3 | 0 | 0 | 3 |
| Drug-related crime detection | 1 | 0 | 0 | 0 | 1 |
| Information security | 0 | 0 | 0 | 0 | 0 |
| Espionage detection | 0 | 0 | 1 | 0 | 1 |
| Crimes victims support | 1 | 1 | 2 | 0 | 4 |
| Software piracy detection | 1 | 1 | 0 | 0 | 2 |
| Drug-related crime detection and Weapons trafficking detection | 0 | 0 | 1 | 0 | 1 |
| Violence against woman analysis | 0 | 0 | 0 | 0 | 0 |
| Armed conflicts solution | 1 | 0 | 0 | 0 | 1 |
| Weapons trafficking detection | 0 | 1 | 0 | 0 | 1 |
| Civil unrest detection | 0 | 0 | 1 | 0 | 1 |
| Total | 33 | 27 | 29 | 3 | 92 |

The following subsections outline a research agenda referring to some of the 92 selected works seeking to respond to RQ3, thus pointing out the opportunities and challenges within three axes in a more synthetic way.

### 3.4.1 Research Directions: Outlining a research agenda

The opportunities and challenges found in the literature associated with Table 5 demonstrated further research proposals from the most recent related works that outline a research agenda in three interrelated axes. The first axis consists of objectives expansions, the second incorporates methodological extensions, and the third considers scenario extensions and changes.

The interrelationship between these axes involves the influence of the objectives in determining the methodological approach to be used since it is impossible to define a methodology unless the research objectives are identified. Consequently, these two axes' definitions imply the need for changes and extensions in study scenarios, including data collection and the use of techniques and technologies, considering the target subject (public security) and its relations with other subjects of interest. Another important detail is that these relationships between the three axes are related to the need to improve the results of previous studies. Figure 8 represents the three axes in the agenda and the described interrelationship among them.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 March 2024 doi:10.20944/preprints202403.0064.v1



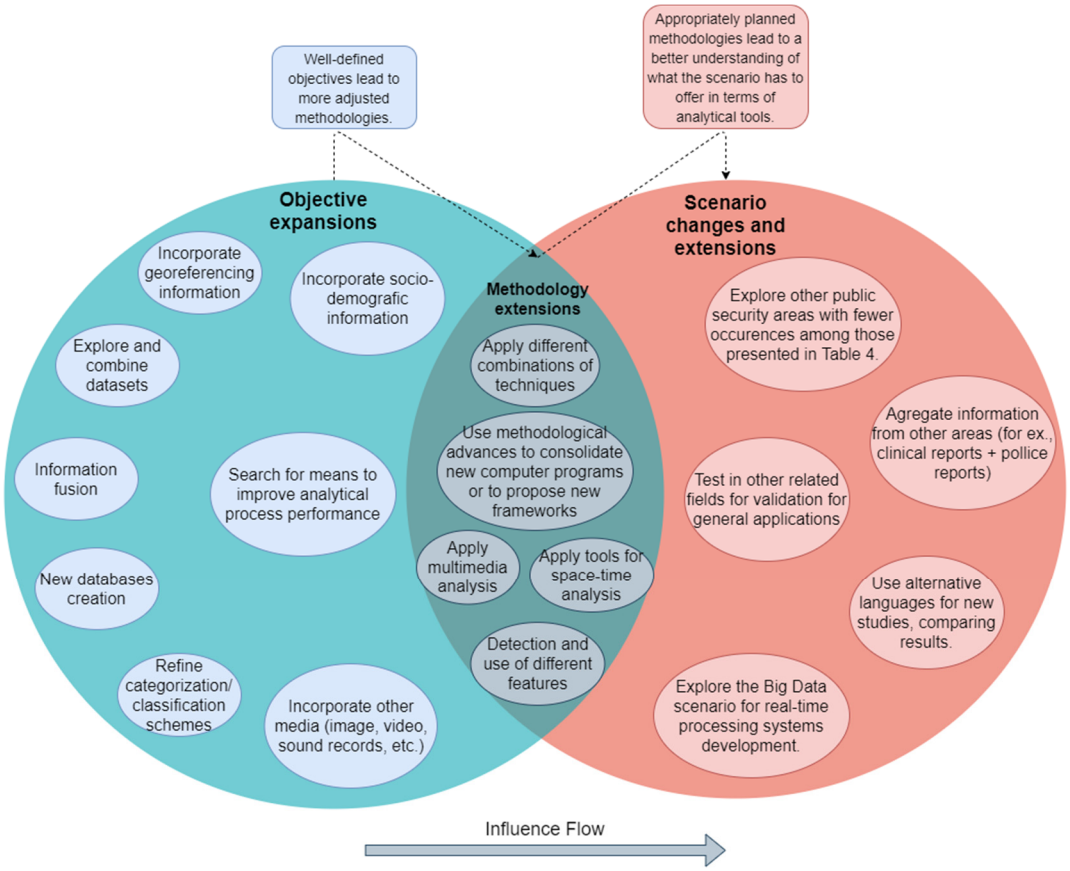


**Figure 8.** Three axes in the research agenda, with related elements.

The diagram in Figure 8 also contains the detected elements presented for each axis, which will be described in the following three sections.

Objectives expansions

The literature explored to support identifying the most recent directions for further research indicated new objectives to enhance the current research. Further analysis and experimental developments using text mining-related technologies in public security indicate the need to extend the current research to incorporate other target elements to be analyzed. The incorporation of socio-demographic and georeferencing information [59,85] and the expansion and combination of datasets [67,68,86,87] are important indications for improving the quality of classifications and predictions. New objectives in this sense may have repercussions on creating new databases, new corpora, performing data fusion, and ensuring more extensive and more diversified datasets for use with mining techniques. The refinement of categorization schemes applied to the objects of study is a natural consequence of the proper use of more extensive databases [78].

The use of visualization tools like maps for spatial analysis of crime occurrences over time, for instance, becomes feasible using georeferenced information (see, for instance, [43,88]). From this perspective, expanding research objectives could incorporate georeferencing, when geographic information is available, providing maps as elements of visualization of some phenomenon or event occurring concerning public security. Virtually every public security application area can employ this information to improve analytics by providing more interesting dashboards and reports for related managers to make their decisions.

Incorporating socio-demographic and georeferencing information is essential for developing new experiments that can enable, for example, the understanding of where the people who express opinions through the analyzed media are located, their age groups, gender, education levels, and positioning according to social classes. Combined with the texts mined and analyzed, this

 doi:10.20944/preprints202403.0064.v1



information can allow managers in public security to filter specific problems and ensure decisions are more focused on the appropriate target audience.

Incorporating, in addition to text analysis, other media such as images, videos, and sound recording, in other words, multimedia analysis [49], can favor areas such as: "Digital / Cyber forensics", for assisting in investigation processes, extracting relevant information from various media related to crimes that occurred; "Support to the Judiciary power" for its ability to facilitate the analysis of evidence in multiple formats to support court decisions; "Crimes' victims support", which can provide police authorities with indications that victims of crime need protection, as they are still under some threat that has been recorded either in text or in other media; and "Terrorism detection", for providing indications of movements of suspicious groups, for example, through video analysis.

Another point that deserves to be commented on is the review of objectives incorporating the improvement of results and performance of the proposed analytical solutions are important considerations for further research indications [55,61] with implications on the methodological approaches the research should adopt, as is explored in the next section.

The performance improvement of the techniques used in text mining, in terms of metrics such as accuracy, precision, recall, and F1-score, is of great relevance for developing or improving frameworks and systems, appearing as a recurring expansion of objectives in proposals for future work. Analyzing techniques' performances allows researchers to incorporate and select different text mining approaches for a range of public security applications, making better decisions about which technique to use [25,86,89–93].

### Methodological extensions

As the objectives lead to the need for adaptations, changes, or extensions in the methodology applied in current research, the corresponding proposals are presented in this part, continuing the research agenda's outline. The proposal, application, and combination of different text mining techniques, algorithms, and technologies, as well as the assessment of their performance, are recurrent methodological recommendations from recent literature, as in AL-Saif and Al-Dossari [43], Alakrot *et al.* [94], Concepción-Sánchez *et al.* [95], Elkhawas and Abdelbaki [41], Husari *et al.* [96], Ruano-Ordás *et al.* [70], Zainal *et al.* [97], Basilio *et al.* [30], Iftikhar *et al.* [35], Mine *et al.* [90], Saini and Bansal [84], Sonowal and Kuppusamy [57], Lal *et al.* [27], and Monish and Pandey [49].

Machine learning techniques are recurrent among the methodological sets found in the selected literature, as observed in the previously listed works. Nonetheless, other groups of methods were recommended for further research developments, such as multicriteria methods for decision support in planning activities for criminal combat [30]; social network analysis to find hidden patterns of illegal activities [84]; traditional statistical methods to support the analysis of opinions about security threats using information collected via questionnaires [98]; bio-inspired algorithms to simulate security threats or risks [97]; the combination of graphs and neural networks to assist in answering key investigation questions [99].

The detection and use of different features [41,45,57], as well as the application of topic modeling ([69,83,84] and named entity recognition [35,100], are recommended to improve the quality of the methods.

The development of computer programs and the proposal of frameworks with advances for methodological enhancements were also suggestions, as these artifacts may perform the complete analytical methodology involved to support investigating security threats and making decisions about related preventive and corrective actions [89,95,96,101].

### Scenario changes and extensions

The axis of scenario changes and extensions is directly linked to the indications in the first two axes. Therefore, it reflects the impact of determining objectives and methodological design for further research. The application in other related fields of the techniques identified or proposed in the literature is in line with the need for testing and validation for possible more generalized applications.

 doi:10.20944/preprints202403.0064.v1



Also, these extensions can offer a concentration of a broader range of issues related to public security and make these methods commercially accessible [102–104].

Information extracted from other areas, such as clinical registers, that may be combined with police reports of sexual or domestic violence [42,67,68,87] or from managerial and financial reports to support the detection of fraud [49,105–107] are necessary for the enrichment of the analytical process. Combining clinical and police reports can help health professionals understand the causes of trauma that generate psychological illness in victims, supporting a more adjusted treatment, and the police to understand sexual or domestic violence patterns to prevent the occurrence or recurrence of related situations. In fraud detection, the development of more in-depth analyses depends on specific documents that make it possible to combine an adequate number of fraudsters' action patterns to ensure better accuracy.

Another direction suggested is concerning the language and location shifts of future work by analyzing alternative languages from the ones used initially in current studies and understanding how local cultures leave their mark on a language that impacts text analysis. The language change effects are interesting for analysis because they can corroborate the power and breadth of the application of the techniques identified or proposed in the reviewed literature, demonstrating the general applicability (or not) of the methods or tools under evaluation [105,108–112].

Giacalone *et al.* [113] suggested that exploring scenarios with more data is an important trend in the context of Big Data that allows the development of systems using real-time processing over the data massively generated on social media. Monish and Pandey [49] commented on using other data formats, such as video and images, and Cichosz [42] mentioned the need to incorporate non-textual attributes to the analysis for extracting from social media streams, reinforcing the importance of Big Data scenarios variety.

The analysis of Table 4 (in Section 4.2) covering text mining applications in public security suggests another insight for scenario expansions. Eleven application areas exist with four or fewer occurrences in the selected works, and these fewer occurrences highlight research opportunities in these areas using new techniques or technologies in addition to those employed in the selected works. Among these eleven, even greater emphasis may be applied to the existing gaps for research related to software piracy detection, violence against woman analysis, armed conflict solution, weapon trafficking detection, and civil unrest detection, as each was represented with only two or one related work.

## 4. Conclusion

The systematic literature review explored literature on text mining in public security to extract information about primary application areas and the most recurrent technologies, opportunities, and challenges. From a total of 194 selected works, 98 most recent contributions from 2018 and 2020 were identified to support outlining a research agenda with general trends to researchers intending to develop new studies about the reported themes.

The detection of nineteen types of applications for text mining in public security enabled the exploration with comments about these works' distributions among the types and the periods of publication, with the recent period of 2017 to 2020 containing a peak in the number of works. The analysis of techniques and technologies supported discovering the 20 most recurrent terms related to techniques of various types, from programming languages and other computer programs dedicated to or containing text mining functions and information extraction approaches. Another interesting discovery is related to the techniques and technologies intersections and combinations among the six areas with the greatest numbers of works in the selections, demonstrating how recurrent some of these elements are for more research and practice.

The research agenda outlined three axes of objective expansions, methodological extensions, and scenario changes and extensions. The first identified general directions about future work recommended by the related literature focused on accomplishing additional details to an existing objective, such as using more data and sources for improving the results of previous analyses and experiments. The second indicated improvements in methodologies, including some combinations

 doi:10.20944/preprints202403.0064.v1



of techniques, to compare them and choose the most suitable method for the occasion. In addition, this extension can guarantee the development of frameworks and software containing the best analytical approaches dedicated to public security. The third axis corroborated the need to expand the applications of the methodologies to other related fields by validating them and reinforcing their effectiveness, as well as indicating an existing trend toward Big Data to enable real-time analysis based on data streaming of the social web and the use of other unstructured data formats in addition to texts.

## 5. Update

The analysis developed and reported in the previous sections focused on a period covering 2014 to the beginning of 2021. Seeking an update on the topics involved, an additional search was carried out to now cover the complete period from January 2021 to February 2024. The additional search returned the following values, in general: in the Scopus database, 161 results; in Web of Science, 88 results; in IEEE Xplore, 57 results; in ACM Digital Library, 300 results. The overall total was therefore 606 results.

These results were joined using Mendeley software, and a database with 577 entries was composed, considering that the software already makes some duplicities joining into one entry. Among this database were detected 156 full conference proceedings that were removed, resulting in 421 entries. However, a total of 51 remaining duplicities were detected and removed, resulting in 370 unique entries.

A screening was applied on titles and abstracts to remove materials out of the research scope. The elimination of out-of-scope articles resulted in a final number of 170 entries, however, three of them were already included in the previous database, with texts from the beginning of 2021 (see [59,107,114]). So, the final number of new works from 2021 to 2024 is 167, distributed as follows: 58 in 2021 (see [115–172]); 61 in 2022 (see [173–233]; 38 in 2023 (see [234–271]); and 10 in 2024 (see [272–281]).

**Author Contributions:** Conceptualization, V.D.H.C.; methodology, V.D.H.C. and R.J.R.S.; software, V.D.H.C.; validation, R.J.R.S., T.C.C.N. and T.P.; formal analysis, V.D.H.C.; investigation, V.D.H.C.; resources, T.P.; data curation, V.D.H.C., R.J.R.S and T.C.C.N.; writing—original draft preparation, V.D.H.C.; writing—review and editing, V.D.H.C., R.J.R.S, T.C.C.N. and T.P.; visualization, V.D.H.C.; supervision, V.D.H.C. and T.P.; project administration, V.D.H.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Supplementary data are available on GitHub: https://github.com/victorheuer/tm_ps_literature-info.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zainol, Z.; Jaymes, M.T.H.; Nohuddin, P.N.E. VisualUrText: A Text Analytics Tool for Unstructured Textual Data. In Proceedings of the Journal of Physics: Conference Series; Computer Science Department, Faculty of Science and Defence Technology, Universiti Pertahanan Nasional Malaysia, Kem Sungai Besi, Kuala Lumpur, Malaysia, 2018; Vol. 1018.
2. de Carvalho, V.D.H.; Costa, A.P.C.S. Towards Corpora Creation from Social Web in Brazilian Portuguese to Support Public Security Analyses and Decisions. *Libr. Hi Tech* **2022**, *in press*. https://doi.org/10.1108/LHT-08-2022-0401.
3. Zhang, Y.; Chen, M.; Liu, L. A Review on Text Mining. In Proceedings of the Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS; School of Computer Science and Engineering, Beihang University, Beijing, China, 2015; Vol. 2015-Novem, pp. 681–685.
4. de Carvalho, V.D.H.; Nepomuceno, T.C.C.; Costa, A.P.C.S. An Automated Corpus Annotation Experiment in Brazilian Portuguese for Sentiment Analysis in Public Security. In *Lecture Notes in Business Information Processing*; 2020; Vol. 384 LNBIP, pp. 99–111 ISBN 9783030462239.

5. Shahzad, F.; Xiu, G.; Shafique Khan, M.A.; Shahbaz, M. Predicting the Adoption of a Mobile Government Security Response System from the User's Perspective: An Application of the Artificial Neural Network Approach. *Technol. Soc.* **2020**, *62*, 101278. https://doi.org/10.1016/j.techsoc.2020.101278.
6. Boulos, M.N.K.; Sanfilippo, A.P.; Corley, C.D.; Wheeler, S. Social Web Mining and Exploitation for Serious Applications : Technosocial Predictive Analytics and Related Technologies for Public Health , Environmental and National Security Surveillance. *Comput. Methods Programs Biomed.* **2010**, *100*, 16–23. https://doi.org/10.1016/j.cmpb.2010.02.007.
7. Kowalski, R.; Esteve, M.; Jankin Mikhaylov, S. Improving Public Services by Mining Citizen Feedback: An Application of Natural Language Processing. *Public Adm.* **2020**, *98*, 1011–1026. https://doi.org/10.1111/padm.12656.
8. de Carvalho, V.D.H.; Nepomuceno, T.C.C.; Poleto, T.; Costa, A.P.C.S. The COVID-19 Infodemic on Twitter: A Space and Time Topic Analysis of the Brazilian Immunization Program and Public Trust. *Trop. Med. Infect. Dis.* **2022**, *7*. https://doi.org/10.3390/tropicalmed7120425.
9. de Carvalho, V.D.H.; Todaro, M.S.F.; dos Santos, R.J.R.; Nepomuceno, T.C.C.; Poleto, T.; Figueiredo, C.J.J.; Turet, J.G.; de Moura, J.A. AI-Driven Decision Support in Public Administration: An Analytical Framework. In *Information Technology and Systems*; Rocha, Á., Ferrás, C., Diez, J.H., Rebolledo, M.D., Eds.; Springer, Cham, 2024; pp. 237–246.
10. Zhang, W.; Zuo, N.; He, W.; Li, S.; Yu, L. Factors Influencing the Use of Artificial Intelligence in Government: Evidence from China. *Technol. Soc.* **2021**, *66*, 101675. https://doi.org/10.1016/j.techsoc.2021.101675.
11. Han, J.; Kamber, M.; Pei, J. Data Mining: Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems); 2011; ISBN 0123814790.
12. Hashimi, H.; Hafez, A.; Mathkour, H. Selection Criteria for Text Mining Approaches. *Comput. Human Behav.* **2015**, *51*, 729–733. https://doi.org/10.1016/j.chb.2014.10.062.
13. Tseng, Y.-H.; Ho, Z.-P.; Yang, K.-S.; Chen, C.-C. Mining Term Networks from Text Collections for Crime Investigation. *Expert Syst. Appl.* **2012**, *39*, 10082–10090. https://doi.org/10.1016/j.eswa.2012.02.052.
14. Poleto, T.; Nepomuceno, T.C.C.; de Carvalho, V.D.H.; Friaes, L.C.B. de O.; de Oliveira, R.C.P.; Figueiredo, C.J.J. Information Security Applications in Smart Cities: A Bibliometric Analysis of Emerging Research. *Futur. Internet* **2023**, *15*, 393. https://doi.org/10.3390/fi15120393.
15. Tutun, S.; Khasawneh, M.T.; Zhuang, J. New Framework That Uses Patterns and Relations to Understand Terrorist Behaviors. *Expert Syst. Appl.* **2017**, *78*, 358–375. https://doi.org/10.1016/j.eswa.2017.02.029.
16. de Carvalho, V.D.H.; Nepomuceno, T.C.C.; Poleto, T.; Turet, J.G.; Costa, A.P.C.S. Mining Public Opinions on COVID-19 Vaccination: A Temporal Analysis to Support Combating Misinformation. *Trop. Med. Infect. Dis.* **2022**, *7*, 256. https://doi.org/10.3390/tropicalmed7100256.
17. Sundermann, C.V.; Domingues, M.A.; Sinoara, R.A.; Marcacini, R.M.; Rezende, S.O. Using Opinion Mining in Context-Aware Recommender Systems: A Systematic Review. *Inf.* **2019**, *10*, 1–45. https://doi.org/10.3390/info10020042.
18. O'Mara-Eves, A.; Thomas, J.; McNaught, J.; Miwa, M.; Ananiadou, S. Using Text Mining for Study Identification in Systematic Reviews: A Systematic Review of Current Approaches. *Syst. Rev.* **2015**, *4*, 5. https://doi.org/10.1186/2046-4053-4-5.
19. Nepomuceno, T.C.C.; Piubello Orsini, L.; de Carvalho, V.D.H.; Poleto, T.; Leardini, C. The Core of Healthcare Efficiency: A Comprehensive Bibliometric Review on Frontier Analysis of Hospitals. *Healthcare* **2022**, *10*, 1316. https://doi.org/10.3390/healthcare10071316.
20. Usai, A.; Pironti, M.; Mital, M.; Aouina Mejri, C. Knowledge Discovery out of Text Data: A Systematic Review via Text Mining. *J. Knowl. Manag.* **2018**, *22*, 1471–1488. https://doi.org/10.1108/JKM-11-2017-0517.
21. Kitchenham, B.; Charters, S. Guidelines for Performing Systematic Literature Reviews in Software Engineering; 2007;
22. Robinson, P.H.; Dubber, M.D. The American Model Penal Code: A Brief Overview. *New Crim. Law Rev.* **2007**, *10*, 319–341. https://doi.org/10.1525/nclr.2007.10.3.319.
23. de Carvalho, V.D.H.; Costa, A.P.C.S. Exploring Text Mining and Analytics for Applications in Public Security: An in-Depth Dive into a Systematic Literature Review. *Socioecon. Anal.* **2023**, *1*, 5–55. https://doi.org/10.51359/2965-4661.2023.259008.
24. Aghababaei, S.; Makrehchi, M. Mining Twitter Data for Crime Trend Prediction. *Intell. Data Anal.* **2018**, *22*, 117–141. https://doi.org/10.3233/IDA-163183.
25. Das, P.; Das, A.K. Graph-Based Clustering of Extracted Paraphrases for Labelling Crime Reports. *Knowledge-Based Syst.* **2019**, *179*, 55–76. https://doi.org/10.1016/j.knosys.2019.05.004.
26. Qazi, N.; Wong, B.L.W. An Interactive Human Centered Data Science Approach towards Crime Pattern Analysis. *Inf. Process. Manag.* **2019**, *56*. https://doi.org/10.1016/j.ipm.2019.102066.
27. Lal, S.; Tiwari, L.; Ranjan, R.; Verma, A.; Sardana, N.; Mourya, R. Analysis and Classification of Crime Tweets. *Procedia Comput. Sci.* **2020**, *167*, 1911–1919. https://doi.org/10.1016/j.procs.2020.03.211.

 doi:10.20944/preprints202403.0064.v1




28. Badii, A.; Tiemann, M.; Adderley, R.; Seidler, P.; Evangelio, R.H.; Senst, T.; Sikora, T.; Panattoni, L.; Raffaelli, M.; Cappel-Porter, M.; et al. MOSAIC: Multimodal Analytics for the Protection of Critical Assets. In Proceedings of the Proceedings of the 2014 International Conference on Signal Processing and Multimedia Applications (SIGMAP); IEEE: Vienna, 2014; pp. 311–320.
29. Bisio, F.; Meda, C.; Zunino, R.; Surlinelli, R.; Scillia, E.; Ottaviano, A. Real-Time Monitoring of Twitter Traffic by Using Semantic Networks. In Proceedings of the Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15; ACM Press: New York, New York, USA, 2015; pp. 966–969.
30. Basilio, M.P.; Pereira, V.; Brum, G. Identification of Operational Demand in Law Enforcement Agencies: An Application Based on a Probabilistic Model of Topics. *Data Technol. Appl.* **2019**, *53*, 333–372. https://doi.org/10.1108/DTA-12-2018-0109.
31. Behmer, E.-J.; Chandramouli, K.; Garrido, V.; Mühlenberg, D.; Müller, D.; Müller, W.; Pallmer, D.; Pérez, F.J.; Piatrik, T.; Vargas, C. Ontology Population Framework of MAGNETO for Instantiating Heterogeneous Forensic Data Modalities. In *IFIP Advances in Information and Communication Technology*; Fraunhofer Institute for Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany, 2019; Vol. 559, pp. 520–531.
32. Basilio, M.P.; Brum, G.S.; Pereira, V. A Model of Policing Strategy Choice. *J. Model. Manag.* **2020**, *15*, 849–891. https://doi.org/10.1108/JM2-10-2018-0166.
33. Hou, J.; Li, X.; Yao, H.; Sun, H.; Mai, T.; Zhu, R. BERT-Based Chinese Relation Extraction for Public Security. *IEEE Access* **2020**, *8*, 132367–132375. https://doi.org/10.1109/ACCESS.2020.3002863.
34. Nikolic, V.; Markoski, B.; Ivkovic, M.; Kuk, K.; Djikanovic, P. Information Retrieval for Unstructured Text Documents in Serbian into the Crime Domain. In Proceedings of the Proceedings of the 16th IEEE International Symposium on Computational Intelligence and Informatics; IEEE: Budapest, 2015; pp. 267–271.
35. Iftikhar, A.; Jaffry, S.W.U.Q.; Malik, M.K. Information Mining from Criminal Judgments of Lahore High Court. *IEEE Access* **2019**, *7*, 59539–59547. https://doi.org/10.1109/ACCESS.2019.2915352.
36. Pina-Sánchez, J.; Roberts, J. V; Sferopoulos, D. Does the Crown Court Discriminate Against Muslim-Named Offenders? A Novel Investigation Based on Text Mining Techniques. *Br. J. Criminol.* **2019**, *59*, 718–736. https://doi.org/10.1093/bjc/azy062.
37. Pina-Sánchez, J.; Grech, D.; Brunton-Smith, I.; Sferopoulos, D. Exploring the Origin of Sentencing Disparities in the Crown Court: Using Text Mining Techniques to Differentiate between Court and Judge Disparities. *Soc. Sci. Res.* **2019**, *84*, 102343. https://doi.org/10.1016/j.ssresearch.2019.102343.
38. Xia, Y.; Cai, T.; Zhong, H. Effect of Judges' Gender on Rape Sentencing: A Data Mining Approach to Analyze Judgment Documents. *China Rev.* **2019**, *19*, 125–149.
39. Gomes, T.; Ladeira, M. A New Conceptual Framework for Enhancing Legal Information Retrieval at the Brazilian Superior Court of Justice. In Proceedings of the Proceedings of the 12th International Conference on Management of Digital EcoSystems; ACM: New York, NY, USA, November 2 2020; pp. 26–29.
40. Al-Nabki, M.W.; Fidalgo, E.; Alegre, E.; Fernández-Robles, L. Improving Named Entity Recognition in Noisy User-Generated Text with Local Distance Neighbor Feature. *Neurocomputing* **2020**, *382*, 1–11. https://doi.org/10.1016/j.neucom.2019.11.072.
41. Elkhawas, A.I.; Abdelbaki, N. Malware Detection Using Opcode Trigram Sequence with SVM. In Proceedings of the Proceedings of the 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM); IEEE, September 2018; pp. 1–6.
42. Cichosz, P. A Case Study in Text Mining of Discussion Forum Posts: Classification with Bag of Words and Global Vectors. *Int. J. Appl. Math. Comput. Sci.* **2018**, *28*, 787–801. https://doi.org/10.2478/amcs-2018-0060.
43. Al-saif, H.; Al-dossari, H. Detecting and Classifying Crimes from Arabic Twitter Posts Using Text Mining Techniques. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 377–387.
44. Alothman, B.; Rattadilok, P. Android Botnet Detection: An Integrated Source Code Mining Approach. In Proceedings of the Proceedings of the 12th International Conference for Internet Technology and Secured Transactions (ICITST); IEEE, December 2017; pp. 111–115.
45. Hadad, T.; Puzis, R.; Sidik, B.; Ofek, N.; Rokach, L. Application Marketplace Malware Detection by User Feedback Analysis. *Commun. Comput. Inf. Sci.* **2018**, *867*, 1–19. https://doi.org/10.1007/978-3-319-93354-2_1.
46. Maktabar, M.; Zainal, A.; Maarof, M.A.; Kassim, M.N. Content Based Fraudulent Website Detection Using Supervised Machine Learning Techniques. In *Hybrid Intelligent Systems*; Abraham, A., Muhuri, P.K., Muda, A.K., Gandhi, N., Eds.; Springer: New Delhi, 2018; Vol. 734, pp. 294–304.
47. Martín, A.; Calleja, A.; Menéndez, H.D.; Tapiador, J.; Camacho, D. ADROIT: Android Malware Detection Using Meta-Information. In Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence (SSCI); 2016; pp. 1–8.
48. Thao, T.P.; Yamada, A.; Murakami, K.; Urakawa, J.; Sawaya, Y.; Kubota, A. Classification of Landing and Distribution Domains Using Whois' Text Mining. In Proceedings of the Proceedings - 16th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 11th IEEE

International Conference on Big Data Science and Engineering and 14th IEEE International Conference on Embedded Software and Systems; 2017; pp. 1–8.
49. Monish, H.; Pandey, A.C. A Comparative Assessment of Data Mining Algorithms to Predict Fraudulent Firms. In Proceedings of the Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering; Jaypee Institute of Information Technology, Noida, India, 2020; pp. 117–122.
50. Lee, T.-H.; Sung, W.-K.; Kim, H.-W. A Text Mining Approach to the Analysis of Information Security Awareness: Korea, United States, and China. In Proceedings of the Pacific Asia Conference on Information Systems, PACIS 2016 - Proceedings; 2016.
51. Noel, G.E.; Peterson, G.L. Applicability of Latent Dirichlet Allocation to Multi-Disk Search. *Digit. Investig.* **2014**, *11*, 43–56. https://doi.org/10.1016/j.diin.2014.02.001.
52. Kuang, D.; Brantingham, P.J.; Bertozzi, A.L. Crime Topic Modeling. *Crime Sci.* **2017**, *6*, 12. https://doi.org/10.1186/s40163-017-0074-0.
53. Sundarkumar, G.G.; Ravi, V.; Nwogu, I.; Govindaraju, V. Malware Detection via API Calls, Topic Models and Machine Learning. In Proceedings of the Proceedings of the 2015 IEEE International Conference on Automation Science and Engineering (CASE); IEEE: Gothenburg, 2015; pp. 1212–1217.
54. Yang, B.; Jiang, J.; Li, N. Towards Discovering Covert Communication through Email Spam. In Proceedings of the IFIP Advances in Information and Communication Technology; 2016; Vol. 486, pp. 191–201.
55. Chung, W.; Liu, J.; Tang, X.; Lai, V.S.K. Extracting Textual Features of Financial Social Media to Detect Cognitive Hacking. In Proceedings of the 2018 IEEE International Conference on Intelligence and Security Informatics, ISI 2018; 2018; pp. 244–246.
56. Parapar, J.; Losada, D.E.D.E.; Barreiro, Á. Combining Psycho-Linguistic, Content-Based and Chat-Based Features to Detect Predation in Chatrooms. *J. Univers. Comput. Sci.* **2014**, *20*, 213–239.
57. Sonowal, G.; Kuppusamy, K.S. SmiDCA: An Anti-Smishing Model with Machine Learning Approach. *Comput. J.* **2018**, *61*, 1143–1157. https://doi.org/10.1093/comjnl/bxy039.
58. Birks, D.; Coleman, A.; Jackson, D. Unsupervised Identification of Crime Problems from Police Free-Text Data. *Crime Sci.* **2020**, *9*. https://doi.org/10.1186/s40163-020-00127-4.
59. Bozyiğit, A.; Utku, S.; Nasibov, E. Cyberbullying Detection: Utilizing Social Media Features. *Expert Syst. Appl.* **2021**, *179*, 115001. https://doi.org/10.1016/j.eswa.2021.115001.
60. Chen, W.; Aspinall, D.; Gordon, A.D.; Sutton, C.; Muttik, I. A Text-Mining Approach to Explain Unwanted Behaviours. In Proceedings of the Proceedings of the 9th European Workshop on System Security - EuroSec '16; ACM Press: New York, 2016; pp. 1–6.
61. Dong, F.; Yuan, S.; Ou, H.; Liu, L. New Cyber Threat Discovery from Darknet Marketplaces. In Proceedings of the Proceedings of the 2018 IEEE Conference on Big Data and Analytics (ICBDA); IEEE, November 2018; pp. 62–67.
62. Aboluwarin, O.; Andriotis, P.; Takasu, A.; Tryfonas, T. Optimizing Short Message Text Sentiment Analysis for Mobile Device Forensics. In *Proceedings of the 2th International Conference on Digital Forensics*; New Delhi, 2016; pp. 69–87 ISBN 9783319462783.
63. Chandra, N.; Khatri, S.K.; Som, S. Anti Social Comment Classification Based on KNN Algorithm. In Proceedings of the Proceedings of the 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO); IEEE, September 2017; pp. 348–354.
64. Gil, V.D.; Betancur, J.D.; Puerta, I.C.; Montoya, L.M.; Sepulveda, J.M. The Femicide in Colombia and Mexico: A Text Mining Analysis. *Turkish Online J. Des. Art Commun.* **2018**, *2018*, 170–177. https://doi.org/10.7456/1080MSE/021.
65. Savaliya, B.R.; Philip, C.G. Email Fraud Detection by Identifying Email Sender. In Proceedings of the Proceedings of the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS); IEEE: Chennai, August 2017; pp. 1420–1422.
66. Gowri, S.; Anandha Mala, G.S.; Divya, G. Enhancing the Digital Data Retrieval System Using Novel Techniques. *J. Theor. Appl. Inf. Technol.* **2014**, *66*, 481–489.
67. Karystianis, G.; Adily, A.; Schofield, P.; Knight, L.; Galdon, C.; Greenberg, D.; Jorm, L.; Nenadic, G.; Butler, T. Automatic Extraction of Mental Health Disorders From Domestic Violence Police Narratives: Text Mining Study. *J. Med. Internet Res.* **2018**, *20*, e11548. https://doi.org/10.2196/11548.
68. Karystianis, G.; Adily, A.; Schofield, P.W.; Greenberg, D.; Jorm, L.; Nenadic, G.; Butler, T. Automated Analysis of Domestic Violence Police Reports to Explore Abuse Types and Victim Injuries: Text Mining Study. *J. Med. Internet Res.* **2019**, *21*. https://doi.org/10.2196/13067.
69. Po, L.; Rollo, F. Building an Urban Theft Map by Analyzing Newspaper Crime Reports. *Proc. - 13th Int. Work. Semant. Soc. Media Adapt. Pers. SMAP 2018* **2018**, 13–18. https://doi.org/10.1109/SMAP.2018.8501866.
70. Ruano-Ordás, D.; Fdez-Riverola, F.; Méndez, J.R. Concept Drift in E-Mail Datasets: An Empirical Study with Practical Implications. *Inf. Sci. (Ny).* **2018**, *428*, 120–135. https://doi.org/10.1016/j.ins.2017.10.049.
71. Andriansyah, M.; Purwanto, I.; Subali, M.; Sukowati, A.I.; Samos, M.; Akbar, A. Developing Indonesian Corpus of Pornography Using Simple NLP-Text Mining (NTM) Approach to Support Government Anti-

 doi:10.20944/preprints202403.0064.v1



Pornography Program. In Proceedings of the 2017 Second International Conference on Informatics and Computing (ICIC); IEEE, November 2017; pp. 1–4.
72. Venčkauskas, A.; Karpavičius, A.; Damaševičius, R.; Marcinkevičius, R.; Kapočiūtė-Dzikienė, J.; Napoli, C. Open Class Authorship Attribution of Lithuanian Internet Comments Using One-Class Classifier. In Proceedings of the Proceedings of the 2017 Federated Conference on Computer Science and Information Systems; IEEE: Prague, 2017; Vol. 11, pp. 373–382.
73. Das, P.; Das, A.K. An Application of Strength Pareto Evolutionary Algorithm for Feature Selection from Crime Data. In Proceedings of the Proceedings of the 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT); IEEE: Dept. of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India, July 2017; pp. 1–6.
74. Das, P.; Das, A.K. Crime Analysis against Women from Online Newspaper Reports and an Approach to Apply It in Dynamic Environment. In Proceedings of the Proceedings of the 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC); IEEE: Dept. of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India, March 2017; pp. 312–317.
75. Almehmadi, A.; Joudaki, Z.; Jalali, R. Language Usage on Twitter Predicts Crime Rates. In Proceedings of the Proceedings of the 10th International Conference on Security of Information and Networks; ACM: New York, NY, USA, 2017; pp. 307–310.
76. Margono, H.; Yi, X.; Raikundalia, G.K. Mining Indonesian Cyber Bullying Patterns in Social Networks. In Proceedings of the Proceedings of the 37th Australasian Computer Science Conference; ACS: Auckland, 2014; Vol. 147, pp. 115–124.
77. Noviantho; Isa, S.M.; Ashianti, L. Cyberbullying Classification Using Text Mining. In Proceedings of the Proceedings of the 1st International Conference on Informatics and Computational Sciences (ICICoS); IEEE: Semarang, November 2017; pp. 241–246.
78. Samtani, S.; Yu, S.; Zhu, H.; Patton, M.; Matherly, J.; Chen, H. Identifying SCADA Systems and Their Vulnerabilities on the Internet of Things: A Text-Mining Approach. *IEEE Intell. Syst.* **2018**, *33*, 63–73. https://doi.org/10.1109/MIS.2018.111145022.
79. Das, P.; Das, A.K. A Two-Stage Approach of Named-Entity Recognition for Crime Analysis. In Proceedings of the 8th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2017; 2017.
80. Zaeem, R.N.; Manoharan, M.; Yang, Y.; Barber, K.S. Modeling and Analysis of Identity Threat Behaviors through Text Mining of Identity Theft Stories. *Comput. Secur.* **2017**, *65*, 50–63. https://doi.org/10.1016/j.cose.2016.11.002.
81. Liang, N.; Biros, D. Validating Common Characteristics of Malicious Insiders: Proof of Concept Study. In Proceedings of the Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS); IEEE: Koloa, January 2016; pp. 3716–3726.
82. Nwafor, E.; Chowdhary, P.; Chandra, A. A Policy-Driven Framework for Document Classification and Enterprise Security. Proc. - 13th IEEE Int. Conf. Ubiquitous Intell. Comput. 13th IEEE Int. Conf. Adv. Trust. Comput. 16th IEEE Int. Conf. Scalable Comput. Commun. IEEE Int. **2017**, 949–953. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0149.
83. Petrovskiy, M.; Chikunov, M. Online Extremism Discovering through Social Network Structure Analysis. In Proceedings of the Proceedings of the IEEE 2nd International Conference on Information and Computer Technologies (ICICT); IEEE, 2019; pp. 243–249.
84. Saini, J.K.; Bansal, D. A Comparative Study and Automated Detection of Illegal Weapon Procurement over Dark Web. *Cybern. Syst.* **2019**, *50*, 405–416. https://doi.org/10.1080/01969722.2018.1553591.
85. Zahra, K.; Azam, F.; Butt, W.H.; Ilyas, F. A Framework for User Characterization Based on Tweets Using Machine Learning Algorithms. In Proceedings of the ACM International Conference Proceeding Series; 2018; pp. 11–16.
86. Bhardwaj, A.; Gupta, R. Qualitative Analysis of Financial Statements for Fraud Detection. In Proceedings of the Proceedings of the 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN); IEEE, October 2018; pp. 318–320.
87. Karystianis, G.; Simpson, A.; Adily, A.; Schofield, P.; Greenberg, D.; Wand, H.; Nenadic, G.; Butler, T. Prevalence of Mental Illnesses in Domestic Violence Police Records: Text Mining Study. *J. Med. Internet Res.* **2020**, *22*, e23725. https://doi.org/10.2196/23725.

88. Saldana, M.; Escobar, C.; Galvez, E.; Torres, D.; Toro, N. Mapping of the Perception of Theft Crimes from Analysis of Newspaper Articles Online. In Proceedings of the 2020 15th Iberian Conference on Information Systems and Technologies (CISTI); IEEE: Universidad Arturo Prat, Almirante Juan José Latorre 2901, Faculty of Engineering and Architecture, Antofagasta, 1244260, Chile, June 2020; Vol. 2020-June, pp. 1–7.
89. Lee, P.S.; Owda, M.; Crockett, K. Novel Methods for Resolving False Positives during the Detection of Fraudulent Activities on Stock Market Financial Discussion Boards. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 1–10. https://doi.org/10.14569/IJACSA.2018.090101.
90. Mine, T.; Hirokawa, S.; Suzuki, T. Does Crime Activity Report Reveal Regional Characteristics? In *Advances in Intelligent Systems and Computing*; S., L., R., I., H, C., Eds.; Springer, Cham, 2019; pp. 582–598.
91. Nedeljkovic, S.; Nikolic, V.; Cabarkapa, M.; Misic, J.; Randelovic, D. An Advanced Quick-Answering System Intended for the e-Government Service in the Republic of Serbia. *ACTA Polytech. HUNGARICA* **2019**, *16*, 153–174. https://doi.org/10.12700/APH.16.4.2019.4.8.
92. Roopa, V.; Induja, K. Customized Visualization of Email Using Sentimental and Impact Analysis in R. *Commun. Comput. Inf. Sci.* 2019, *1046*, 144–154.
93. Miranda, E.; Aryuni, M.; Fernando, Y.; Kibtiah, T.M. A Study of Radicalism Contents Detection in Twitter: Insights from Support Vector Machine Technique. In Proceedings of the Proceedings of 2020 International Conference on Information Management and Technology, ICIMTech 2020; School of Information Systems, Bina Nusantara University, Information Systems Department, Jakarta, 11480, Indonesia, 2020; pp. 549–554.
94. Alakrot, A.; Murray, L.; Nikolov, N.S. Dataset Construction for the Detection of Anti-Social Behaviour in Online Communication in Arabic. In Proceedings of the Procedia Computer Science; Elsevier B.V., 2018; Vol. 142, pp. 174–181.
95. Concepcion-Sanchez, J.A.; Molina-Gil, J.; Caballero-Gil, P.; Santos-Gonzalez, I. Fuzzy Logic System for Identity Theft Detection in Social Networks. In Proceedings of the 2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data); IEEE, August 2018; pp. 65–70.
96. Husari, G.; Niu, X.; Chu, B.; Al-Shaer, E. Using Entropy and Mutual Information to Extract Threat Actions from Cyber Threat Intelligence. In Proceedings of the Proceedings of the 2018 IEEE International Conference on Intelligence and Security Informatics (ISI); IEEE, 2018; pp. 1–6.
97. Zainal, K.; Jali, M.Z.; Hasan, A.B. Comparative Analysis of Danger Theory Variants in Measuring Risk Level for Text Spam Messages. In *Advances in Intelligent Systems and Computing*; Springer International Publishing, 2018; Vol. 753, pp. 133–152 ISBN 9783319787527.
98. Silomon, J.A.M.; Roeling, M.P. Assessing Opinions on Software as a Weapon in the Context of (Inter)National Security. In *Transactions on Computational Science XXXII*; Gavrilova, M.L., Tan, C.J.K., Sourin, A., Eds.; Springer Berlin Heidelberg, 2018; Vol. 10830 LNCS, pp. 43–56 ISBN 9783662566718.
99. Henseler, H.; Hyde, J. Technology Assisted Analysis of Timeline and Connections in Digital Forensic Investigations. In Proceedings of the CEUR Workshop Proceedings; Magnet Forensics, Waterloo, Canada, 2019; Vol. 2484, pp. 32–37.
100. Ariffin, N.; Zainal, A.; Maarof, M.A.; Nizam Kassim, M. A Conceptual Scheme for Ransomware Background Knowledge Construction. In Proceedings of the Proceedings of the 2018 Cyber Resilience Conference (CRC); IEEE: School of Computing, Universiti Teknologi Malaysia, Johor, Malaysia, November 2018; pp. 1–4.
101. Pires, M.; Georgieva, P. An Intelligent Tool for Detection of Phishing Messages. In *Advances in Intelligent Systems and Computing*; Department of Electronics Telecommunications and Informatics, University of Aveiro, Aveiro, Portugal, 2020; Vol. 942, pp. 116–125.
102. Al-Khalisy, M.A.E.; Jehlol, H.B. Terrorist Affiliations Identifying through Twitter Social Media Analysis Using Data Mining and Web Mapping Techniques. *J. Eng. Appl. Sci.* **2018**, *13*, 7459–7464. https://doi.org/10.36478/jeasci.2018.7459.7464.
103. Ristea, A.; Kurland, J.; Resch, B.; Leitner, M.; Langford, C. Estimating the Spatial Distribution of Crime Events around a Football Stadium from Georeferenced Tweets. *ISPRS Int. J. Geo-Information* **2018**, *7*, 43. https://doi.org/10.3390/ijgi7020043.
104. Savaş, S.; Topaloğlu, N. Data Analysis through Social Media According to the Classified Crime. *Turkish J. Electr. Eng. Comput. Sci.* **2019**, *27*, 407–420. https://doi.org/10.3906/elk-1712-17.
105. Dastjerdi, A.R.; Foroghi, D.; Kiani, G.H. Detecting Manager's Fraud Risk Using Text Analysis: Evidence from Iran. *J. Appl. Account. Res.* **2019**, *20*, 154–171. https://doi.org/10.1108/JAAR-01-2018-0016.
106. Angenent, M.N.; Barata, A.P.; Takes, F.W. Large-Scale Machine Learning for Business Sector Prediction. In Proceedings of the Proceedings of the 35th Annual ACM Symposium on Applied Computing; ACM: New York, NY, USA, March 30 2020; pp. 1143–1146.
107. Siering, M.; Muntermann, J.; Grčar, M. Design Principles for Robust Fraud Detection: The Case of Stock Market Manipulations. *J. Assoc. Inf. Syst.* **2021**, *22*, 156–178. https://doi.org/10.17705/1jais.00657.
108. Correa, J.C.; García-Chitiva, M.D.P.; García-Vargas, G.R. A Text Mining Approach to the Text Difficulty of Latin American Peace Agreement. *Rev. Latinoam. Psicol.* **2018**, *50*, 61–70. https://doi.org/10.14349/rlp.2018.v50.n1.6.

109. Mansour, S. Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining. In Proceedings of the Procedia Computer Science; 2018; Vol. 140, pp. 95–103.
110. Öztürk, N.; Ayvaz, S. Sentiment Analysis on Twitter: A Text Mining Approach to the Syrian Refugee Crisis. *Telemat. Informatics* **2018**, *35*, 136–147. https://doi.org/10.1016/j.tele.2017.10.006.
111. Balim, C.; Gunal, E.S. Automatic Detection of Smishing Attacks by Machine Learning Methods. In Proceedings of the 2019 1st International Informatics and Software Engineering Conference (UBMYK); IEEE: Afyon Kocatepe University, Dept. of Computer Programming, Sandikli Vocational School, Afyonkarahisar, Turkey, November 2019; pp. 1–3.
112. Alatrista-Salas, H.; Morzán-Samamé, J.; Nunez-del-Prado, M. Crime Alert! Crime Typification in News Based on Text Mining. In *Lecture Notes in Networks and Systems*; Universidad del Pacífico, Av. Salaverry, Lima, 2020, Peru, 2020; Vol. 69, pp. 725–741.
113. Giacalone, M.; Cusatelli, C.; Romano, A.; Buondonno, A.; Santarcangelo, V. Big Data and Forensics: An Innovative Approach for a Predictable Jurisprudence. *Inf. Sci. (Ny).* **2018**, *426*, 160–170. https://doi.org/10.1016/j.ins.2017.10.036.
114. Choi, Y.-J.; Jeon, B.-J.; Kim, H.-W. Identification of Key Cyberbullies: A Text Mining and Social Network Analysis Approach. *Telemat. Informatics* **2021**, *56*, 101504. https://doi.org/10.1016/j.tele.2020.101504.
115. Adily, A.; Karystianis, G.; Butler, T. Text Mining Police Narratives to Identify Types of Abuse and Victim Injuries in Family and Domestic Violence Events. *TRENDS ISSUES CRIME Crim. JUSTICE* **2021**, 1–12. https://doi.org/10.52922/ti04923 WE - Emerging Sources Citation Index (ESCI).
116. Aldera, S.; Emam, A.; Al-Qurishi, M.; Alrubaian, M.; Alothaim, A. Online Extremism Detection in Textual Content: A Systematic Literature Review. *IEEE Access* **2021**, *9*, 42384–42396. https://doi.org/10.1109/ACCESS.2021.3064178.
117. Alharbi, A.; Dong, H.; Yi, X.; Tari, Z.; Khalil, I. Social Media Identity Deception Detection: A Survey. *ACM Comput. Surv.* **2021**, *54*. https://doi.org/10.1145/3446372.
118. Al-Wesabi, F.N. An Optimized English Textwatermarking Method Based on Natural Language Processing Techniques. *Comput. Mater. Contin.* **2021**, *69*, 1519–1536. https://doi.org/10.32604/cmc.2021.018202.
119. Asante, A.; Feng, X. Content-Based Technical Solution for Cyberstalking Detection. In Proceedings of the 2021 3rd International Conference on Computer Communication and the Internet, ICCCI 2021; Institute of Electrical and Electronics Engineers Inc.: Catholic University College of Ghana, Department of Computing and Information Sciences, Fiapre-Sunyani, Ghana, 2021; pp. 89–95.
120. Avgerinos Loutsaris, M.; Lachana, Z.; Alexopoulos, C.; Charalabidis, Y. Legal Text Processing: Combing Two Legal Ontological Approaches through Text Mining. In Proceedings of the DG.O2021: The 22nd Annual International Conference on Digital Government Research; Association for Computing Machinery: New York, NY, USA, 2021; pp. 522–532.
121. Canossa, A.; Salimov, D.; Azadvar, A.; Harteveld, C.; Yannakakis, G. For Honor, for Toxicity: Detecting Toxic Behavior through Gameplay. *Proc. ACM Hum.-Comput. Interact.* **2021**, *5*. https://doi.org/10.1145/3474680.
122. Chen, H.; Hossain, M. Developing a Google Chrome Extension for Detecting Phishing Emails. In Proceedings of the EPiC Series in Computing; R., W., F., H., A., R., Eds.; EasyChair: University of Minnesota Crookston, MN, United States, 2021; Vol. 77, pp. 13–22.
123. Corrêa, I.T.; Faria, E.R. An Analysis of Violence against Women Based on Victims' Reports. In Proceedings of the Proceedings of the XVII Brazilian Symposium on Information Systems; Association for Computing Machinery: New York, NY, USA, 2021.
124. de Oliveira, G.A.; de Sousa, R.T.; Albuquerque, R.D.; Villalba, L.J.G. Adversarial Attacks on a Lexical Sentiment Analysis Classifier. *Comput. Commun.* **2021**, *174*, 154–171. https://doi.org/10.1016/j.comcom.2021.04.026.
125. Degadwala, S.; Vyas, D.; Hossain, M.R.; DIder, A.R.; Ali, M.N.; Kuri, P. Location-Based Modelling and Analysis of Threats by Using Text Mining. In Proceedings of the Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESC 2021; Institute of Electrical and Electronics Engineers Inc.: Sigma Institute of Engineering, Gujarat, Vadodara, India, 2021; pp. 1940–1944.
126. Desmet, C.; Cook, D.J. Recent Developments in Privacy-Preserving Mining of Clinical Data. *ACM/IMS Trans. Data Sci.* **2021**, *2*. https://doi.org/10.1145/3447774.
127. Febriany, A.; Utama, D.N. Analysis Model for Identifying Negative Posts Based on Social Media. *Int. J. Emerg. Technol. Adv. Eng.* **2021**, *11*, 96–103. https://doi.org/10.46338/IJETAE1021_12.
128. Franchina, L.; Ferracci, S.; Palmaro, F. Detecting Phishing E-Mails Using Text Mining and Features Analysis. In Proceedings of the CEUR Workshop Proceedings; A., A., M., C., Eds.; CEUR-WS: HERMES Bay S.R.L, 2021; Vol. 2940, pp. 106–119.
129. Gradoń, K.T.; Hołyst, J.A.; Moy, W.R.; Sienkiewicz, J.; Suchecki, K. Countering Misinformation: A Multidisciplinary Approach. *Big Data Soc.* **2021**, *8*. https://doi.org/10.1177/20539517211013848.

 doi:10.20944/preprints202403.0064.v1




130. Gryaznova, E.; Kirina, M. Defining Kinds of Violence in Russian Short Stories of 1900-1930: A Case of Topic Modelling With LDA and PCA. In Proceedings of the CEUR Workshop Proceedings; R.V., B., N.V., B., A.V., C., D.E., P., A.E., V., V.P., Z., Eds.; CEUR-WS: National Research University, Higher School of Economics, 123 Griboyedova emb., St. Petersburg, 190068, Russian Federation, 2021; Vol. 3090, pp. 281–290.
131. Guo, W.Y.; Zeng, Q.T.; Duan, H.; Ni, W.J.; Liu, C. Process-Extraction-Based Text Similarity Measure for Emergency Response Plans. *Expert Syst. Appl.* **2021**, *183*. https://doi.org/10.1016/j.eswa.2021.115301.
132. Hajarian, M.; Khanbabaloo, Z. Toward Stopping Incel Rebellion: Detecting Incels in Social Media Using Sentiment Analysis. In Proceedings of the 2021 7th International Conference on Web Research (ICWR); IEEE, May 19 2021; pp. 169–174.
133. Hamza, M.; Jamila, M.; Lunn, J.; Aljumaili, W. Crime Geo Analytics Tool. In Proceedings of the 2021 14th International Conference on Developments in eSystems Engineering (DeSE); 2021; pp. 577–581.
134. Herbert, F.; Schmidbauer-Wolf, G.M.; Reuter, C. Who Should Get My Private Data in Which Case? Evidence in the Wild. In Proceedings of the Proceedings of Mensch Und Computer 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 281–293.
135. Husain, F.; Uzuner, O. A Survey of Offensive Language Detection for the Arabic Language. *ACM Trans. Asian Low-Resource Lang. Inf. Process.* **2021**, *20*, 1–44. https://doi.org/10.1145/3421504.
136. Ignaczak, L.; Goldschmidt, G.; Costa, C.A. Da; Righi, R.D.R. Text Mining in Cybersecurity: A Systematic Literature Review. *ACM Comput. Surv.* **2021**, *54*. https://doi.org/10.1145/3462477.
137. Jiang, Y.; Atif, Y. An Approach to Discover and Assess Vulnerability Severity Automatically in Cyber-Physical Systems. In Proceedings of the 13th International Conference on Security of Information and Networks; Association for Computing Machinery: New York, NY, USA, 2021.
138. Kusuma, N.; Suhardi Detection of Online Prostitution in Twitter Platform Using Machine Learning Approach. In Proceedings of the 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT); 2021; pp. 55–60.
139. Langton, S.; Bannister, J.; Ellison, M.; Haleem, M.S.; Krzemieniewska-Nandwani, K. Policing and Mental Ill-Health: Using Big Data to Assess the Scale and Severity of, and the Frontline Resources Committed to, Mental Ill-Health-Related Calls-for-Service. *POLICING-A J. POLICY Pract.* **2021**, *15*, 1963–1976. https://doi.org/10.1093/police/paab035.
140. Lee, M.-C.; Vajiac, C.; Kulshrestha, A.; Levy, S.; Park, N.; Jones, C.; Rabbany, R.; Faloutsos, C. INFOSHIELD: Generalizable Information-Theoretic Human-Trafficking Detection. In Proceedings of the Proceedings - International Conference on Data Engineering; IEEE Computer Society: National Chiao Tung University, 2021; Vol. 2021-April, pp. 1116–1127.
141. Liu, T.; Wang, S.; Fu, J.; Chen, L.; Wei, Z.; Liu, Y.; Ye, H.; Xu, L.; Wang, W.; Huang, X. Fine-Grained Element Identification in Complaint Text of Internet Fraud. In Proceedings of the International Conference on Information and Knowledge Management, Proceedings; Association for Computing Machinery: Fudan University, Shanghai, China, 2021; pp. 3268–3272.
142. Meilani, Z.D.; Nizar, I.M.; Sunandar, M.F.; Hidayati, S.C.; IEEE Lawyering Social: Navigating Legal Issues on Social Media Posts with a Low-Cost Data Algorithm. *2021 5TH Int. Conf. INFORMATICS Comput. Sci. (ICICOS 2021)* 2021.
143. Meraliyev, B.; Kongratbayev, K.; Sultanova, N. Content Analysis of Extracted Suicide Texts From Social Media Networks by Using Natural Language Processing and Machine Learning Techniques. In Proceedings of the 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST); 2021; pp. 1–6.
144. Min, M.; Lee, J.J.; Park, H.; Lee, K. Honeypot System for Automatic Reporting of Illegal Online Gambling Sites Utilizing SMS Spam. In Proceedings of the World Automation Congress Proceedings; IEEE Computer Society: Korea University, Department of Information Security School of Cybersecurity, Seoul, South Korea, 2021; Vol. 2021-Augus, pp. 180–185.
145. Mitra, S.; Tasnim, T.; Islam, M.A.R.; Khan, N.I.; Majib, M.S. A Framework to Detect and Prevent Cyberbullying from Social Media by Exploring Machine Learning Algorithms. In Proceedings of the 6th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, IC4ME2 2021; Institute of Electrical and Electronics Engineers Inc.: Military Institute of Science and Technology, Department of Computer Science and Engineering, Dhaka, 1216, Bangladesh, 2021.
146. Mladenović, M.; Ošmjanski, V.; Stanković, S.V. Cyber-Aggression, Cyberbullying, and Cyber-Grooming: A Survey and Research Challenges. *ACM Comput. Surv.* **2021**, *54*. https://doi.org/10.1145/3424246.
147. Peng, H.; Li, J.; Song, Y.; Yang, R.; Ranjan, R.; Yu, P.S.; He, L. Streaming Social Event Detection and Evolution Discovery in Heterogeneous Information Networks. *ACM Trans. Knowl. Discov. Data* **2021**, *15*. https://doi.org/10.1145/3447585.
148. Permana, M.A.; Thohir, M.I.; Mantoro, T.; Ayu, M.A. Crime Rate Detection Based on Text Mining on Social Media Using Logistic Regression Algorithm. In Proceedings of the 7th International Conference on Computing, Engineering and Design, ICCED 2021; Institute of Electrical and Electronics Engineers Inc.: NusaPutra University, School of Computer Science, Sukabumi, Indonesia, 2021.

149. Qureshi, K.A.; Sabih, M. Un-Compromised Credibility: Social Media Based Multi-Class Hate Speech Classification for Text. *IEEE Access* **2021**, *9*, 109465–109477. https://doi.org/10.1109/ACCESS.2021.3101977.
150. Rahman, T.; Rohan, R.; Pal, D.; Kanthamanon, P. Human Factors in Cybersecurity: A Scoping Review. In Proceedings of the Proceedings of the 12th International Conference on Advances in Information Technology; Association for Computing Machinery: New York, NY, USA, 2021.
151. Rizwan, K.; Babar, S.; Nayab, S.; Hanif, M.K. HarX: Real-Time Harassment Detection Tool Using Machine Learning. In Proceedings of the International Conference of Modern Trends in ICT Industry: Towards the Excellence in the ICT Industries, MTICTI 2021; Institute of Electrical and Electronics Engineers Inc.: University of Faisalabad, Department of Computer Science, Faisalabad, Pakistan, 2021.
152. Rovera, M.; Nanni, F.; Ponzetto, S.P. Event-Based Access to Historical Italian War Memoirs. *J. Comput. Cult. Herit.* **2021**, *14*. https://doi.org/10.1145/3406210.
153. Sajid, M.S.I.; Wei, J.; Abdeen, B.; Al-Shaer, E.; Islam, M.M.; Diong, W.; Khan, L. SODA: A System for Cyber Deception Orchestration and Automation. In Proceedings of the Proceedings of the 37th Annual Computer Security Applications Conference; Association for Computing Machinery: New York, NY, USA, 2021; pp. 675–689.
154. Samtani, S.; Li, W.; Benjamin, V.; Chen, H. Informing Cyber Threat Intelligence through Dark Web Situational Awareness: The AZSecure Hacker Assets Portal. *Digit. Threat.* **2021**, *2*. https://doi.org/10.1145/3450972.
155. Sasaki, S.; Miyamoto, Y. Source-Oriented POV Visualization for Multidimensional Analysis of International Conflicts and Terrorist Incidents with 5D World Map System. In Proceedings of the International Electronics Symposium 2021: Wireless Technologies and Intelligent Systems for Better Human Lives, IES 2021 - Proceedings; A.A., Y., A., K.N., H., H., P.A.M., P., F., G., M., R., Y.R., P., M., R., Eds.; Institute of Electrical and Electronics Engineers Inc.: Musashino University, Department of Data Science, 3-3-3 Ariake, Koto-ku, 135-8181, Japan, 2021; pp. 323–330.
156. Seyler, D.; Li, L.; Zhai, C. Semantic Text Analysis for Detection of Compromised Accounts on Social Networks. In Proceedings of the Proceedings of the 12th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; IEEE Press, 2021; pp. 417–424.
157. Sharma, C.; Ramakrishnan, R.; Pendse, A.; Chimurkar, P.; Talele, K.T. CYBER-BULLYING DETECTION VIA TEXT MINING AND MACHINE LEARNING. In Proceedings of the 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021; Institute of Electrical and Electronics Engineers Inc.: Electronics Engineering Department, Sardar Patel Institute of Technology, Mumbai, India, 2021.
158. Sheng, Q.; Zhang, X.; Cao, J.; Zhong, L. Integrating Pattern- and Fact-Based Fake News Detection via Model Preference Learning. In Proceedings of the Proceedings of the 30th ACM International Conference on Information & Knowledge Management; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1640–1650.
159. Shi, T.; Wang, N.; Zhang, L. LDA-CBOW-Based Mining Model for Risky Driving Behavior in Traffic Accidents. In Proceedings of the Journal of Physics: Conference Series; IOP Publishing Ltd: Beijing Police College, Beijing, 102202, China, 2021; Vol. 2138.
160. Singh Yadav, A.K.; Sora, M. Fraud Detection in Financial Statements Using Text Mining Methods: A Review. In Proceedings of the IOP Conference Series: Materials Science and Engineering; Dept. of CSE, NERIST, Itanagar, Arunachal Pradesh, India, 2021; Vol. 1020.
161. Solanke, A.A.; Chen, X.; Ramirez-Cruz, Y. Pattern Recognition and Reconstruction: Detecting Malicious Deletions in Textual Communications. In Proceedings of the Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021; Y., C., H., L., Y., T., U., F., X., Z., X.T., H., S., B., X., L., J., Z., S., P., V., P., J., W., A., C., C., O., Eds.; Institute of Electrical and Electronics Engineers Inc.: University of Bologna, Cirsfid-AlmaAI, Bologna, Italy, 2021; pp. 2574–2582.
162. Thaipisutikul, T.; Tuarob, S.; Pongpaichet, S.; Amornvatcharapong, A.; Shih, T.K. Automated Classification of Criminal and Violent Activities in Thailand from Online News Articles. In Proceedings of the KST 2021 - 2021 13th International Conference Knowledge and Smart Technology; Faculty of Information and Communication Technology, Mahidol University, Thailand, 2021; pp. 170–175.
163. Toubes, D.R.; Araújo-Vila, N. The Treatment of Language in Travel Advisories as a Covert Tool of Political Sanction. *Tour. Manag. Perspect.* **2021**, *40*. https://doi.org/10.1016/j.tmp.2021.100866.
164. Tundis, A.; Melnik, M.; Naveed, H.; Mühlhäuser, M. A Social Media-Based over Layer on the Edge for Handling Emergency-Related Events. *Comput. Electr. Eng.* **2021**, *96*. https://doi.org/10.1016/j.compeleceng.2021.107570.
165. Ul Rehman, Z.; Abbas, S.; Khan, M.A.; Mustafa, G.; Fayyaz, H.; Hanif, M.; Saeed, M.A. Understanding the Language of ISIS: An Empirical Approach to Detect Radical Content on Twitter Using Machine Learning. *C. Mater. \& Contin.* **2021**, *66*, 1075–1090. https://doi.org/10.32604/cmc.2020.012770.

166. Wang, A.; Potika, K. Cyberbullying Classification Based on Social Network Analysis. In Proceedings of the Proceedings - IEEE 7th International Conference on Big Data Computing Service and Applications, BigDataService 2021; Institute of Electrical and Electronics Engineers Inc.: Department of Computer Science, San Jose State University, San Jose, United States, 2021; pp. 87–95.
167. Wu, D.; Shi, W.; Ma, X. A Novel Real-Time Anti-Spam Framework. *ACM Trans. Internet Technol.* **2021**, *21*. https://doi.org/10.1145/3423153.
168. Wu, T.; Ma, W.; Wen, S.; Xia, X.; Paris, C.; Nepal, S.; Xiang, Y. Analysis of Trending Topics and Text-Based Channels of Information Delivery in Cybersecurity. *ACM Trans. Internet Technol.* **2021**, *22*. https://doi.org/10.1145/3483332.
169. Yao, S. Application of Data Mining Technology in Financial Fraud Identification. In Proceedings of the ACM International Conference Proceeding Series; Association for Computing Machinery: Business School, Nanjing University, Nanjing, 210023, China, 2021; pp. 2919–2922.
170. Yin, X.; Zhu, Y.; Hu, J. A Comprehensive Survey of Privacy-Preserving Federated Learning: A Taxonomy, Review, and Future Directions. *ACM Comput. Surv.* **2021**, *54*. https://doi.org/10.1145/3460427.
171. Zaimi, R.; Hafidi, M.; Lamia, M. A Literature Survey on Anti-Phishing in Websites. In Proceedings of the Proceedings of the 4th International Conference on Networking, Information Systems & Security; Association for Computing Machinery: New York, NY, USA, 2021.
172. Shrestha, A.; Akrami, N.; Kaati, L. Introducing Digital-7: Threat Assessment of Individuals in Digital Environments. In Proceedings of the Proceedings of the 12th IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; IEEE Press, 2021; pp. 720–726.
173. Pandya, D.D.; Amarawat, G.; Jadeja, A.; Degadwala, S.; Vyas, D. Analysis and Prediction of Location Based Criminal Behaviors Through Machine Learning. In Proceedings of the 2022 International Conference on Edge Computing and Applications (ICECAA); 2022; pp. 1324–1332.
174. Chi, X. Research On Crime Feature Mining Based On Extraction of Spatio-Temporal Elements of Cases. In Proceedings of the 2022 15th International Symposium on Computational Intelligence and Design (ISCID); 2022; pp. 282–285.
175. Labanan, R.M.; Muñoz, N.D.S. A Study on the Usability of Text Analysis on Web Artifacts for Digital Forensic Investigation. In Proceedings of the 2022 2nd International Conference in Information and Computing Research (iCORE); 2022; pp. 54–59.
176. Tsur, A.M.; Nadler, R.; Sorkin, A.; Lipkin, I.; Gelikas, S.; Chen, J.; Benov, A. Patterns in Vehicle-Ramming Attacks. *Isr. Med. Assoc. J.* **2022**, *24*, 579–583.
177. Raja, K.; Bakaraniya, P. A Review On Social Media Crime Related Users Prediction Methodology. In Proceedings of the 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS); 2022; pp. 698–703.
178. Torres-Berru, Y.; López Batista, V.F. Data and Text Mining for the Detection of Fraud in Public Contracts: A Case Study of Ecuador's Official Public Procurement System. In Proceedings of the Lecture Notes in Electrical Engineering; S., B., K., A., Eds.; Springer Science and Business Media Deutschland GmbH: Universidad de Salamanca Plaza de la Merced, Salamanca, Spain, 2022; Vol. 846 LNEE, pp. 116–127.
179. Fridlund, M.; Brodén, D.; Jauhiainen, T.; Malkki, L.; Olsson, L.-J.; Borin, L. Trawling and Trolling for Terrorists in the Digital Gulf of Bothnia: Cross-Lingual Text Mining for the Emergence of Terrorism in Swedish and Finnish Newspapers, 1780-1926. In *CLARIN: The Infrastructure for Language Resources*; De Gruyter: Centre for Digital Humanities, Department of Literature, History of Ideas and Religion, University of Gothenburg, Gothenburg, Sweden, 2022; pp. 781–801 ISBN 978-311076737-7 (ISBN); 978-311076734-6 (ISBN).
180. Liu, Y.; Xu, Z. An Empirical Analysis of Rape Sentence Based on SPSS. In Proceedings of the Proceedings of the 2022 5th International Conference on Software Engineering and Information Management; Association for Computing Machinery: New York, NY, USA, 2022; pp. 182–187.
181. Anastasiadis, M.; Aivatoglou, G.; Spanos, G.; Voulgaridis, A.; Votis, K. Combining Text Analysis Techniques with Unsupervised Machine Learning Methodologies for Improved Software Vulnerability Management. In Proceedings of the 2022 IEEE International Conference on Cyber Security and Resilience (CSR); 2022; pp. 273–278.
182. Balasaraswathi, V.R.; Mary Shamala, L.; Hamid, Y.; Pachhaiammal Alias Priya, M.; Shobana, M.; Sugumaran, M. An Efficient Feature Selection for Intrusion Detection System Using B-HKNN and C2 Search Based Learning Model. *Neural Process. Lett.* **2022**, *54*, 5143–5167. https://doi.org/10.1007/s11063-022-10854-1.
183. Müller, W.; Mühlenberg, D.; Pallmer, D.; Zeltmann, U.; Ellmauer, C.; Carrasco, F.J.P.; Garcia, A.G.; Demestichas, K.; Peppes, N.; Touska, D.; et al. Knowledge Engineering for Crime Investigation. In Proceedings of the Proceedings of World Multi-Conference on Systemics, Cybernetics and Informatics, WMSCI; N.C., C., J., H., B., S., M., S., Eds.; International Institute of Informatics and Cybernetics: Fraunhofer IOSB, Karlsruhe, 76131, Germany, 2022; Vol. 3, pp. 64–69.

184. Nisha, M.; Jebathangam, J. Detection and Classification of Cyberbullying in Social Media Using Text Mining. In Proceedings of the 6th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2022 - Proceedings; Institute of Electrical and Electronics Engineers Inc.: Vistas, Department of Computer Science, Chennai, India, 2022; pp. 856–861.
185. Michell, C.; Winarto, C.N.; Bestari, L.; Ramdhan, D.; Chowanda, A. Systematic Literature Review of E-Wallet: The Technology and Its Regulations in Indonesia. In Proceedings of the 2022 International Conference on Information Technology Systems and Innovation, ICITSI 2022 - Proceedings; Institute of Electrical and Electronics Engineers Inc.: School of Computer Science, Bina Nusantara University, Computer Science Department, Jakarta, 11480, Indonesia, 2022; pp. 64–69.
186. Edlund, J.; Brodén, D.; Fridlund, M.; Lindhé, C.; Olsson, L.-J.; Ängsal, M.P.; Öhberg, P. A Multimodal Digital Humanities Study of Terrorism in Swedish Politics: An Interdisciplinary Mixed Methods Project on the Configuration of Terrorism in Parliamentary Debates, Legislation, and Policy Networks 1968–2018. In Proceedings of the Lecture Notes in Networks and Systems; K., A., Ed.; Springer Science and Business Media Deutschland GmbH: KTH Speech, Music & Hearing, Department of Intelligent Systems, KTH Royal Institute of Technology, Stockholm, 100 44, Sweden, 2022; Vol. 295, pp. 435–449.
187. Zhou, T.; Zhao, H.; Zhang, X. Keyword Extraction Based on Random Forest and XGBoost - An Example of Fraud Judgment Document. In Proceedings of the 2022 European Conference on Natural Language Processing and Information Retrieval (ECNLPIR); 2022; pp. 17–22.
188. Bahaweres, R.B.; Nugrahanti, D.A. Implementation of Text Association Rules about Terrorism on Twitter in Indonesia. In Proceedings of the 2022 10th International Conference on Cyber and IT Service Management, CITSM 2022; Institute of Electrical and Electronics Engineers Inc.: Informatics Uin Syarif Hidayatullah Jakarta South, Tangerang, Indonesia, 2022.
189. Husák, M.; Čermák, M. SoK: Applications and Challenges of Using Recommender Systems in Cybersecurity Incident Handling and Response. In Proceedings of the Proceedings of the 17th International Conference on Availability, Reliability and Security; Association for Computing Machinery: New York, NY, USA, 2022.
190. Li, J. Analyse of Influence of Adversarial Samples on Neural Network Attacks with Different Complexities. In Proceedings of the Proceedings - 2022 2nd International Signal Processing, Communications and Engineering Management Conference, ISPCEM 2022; Institute of Electrical and Electronics Engineers Inc.: Chongqing University of Posts and Telecommunications, Chongqing, China, 2022; pp. 329–333.
191. Kovalchuk, O.; Banakh, S.; Masonkova, M.; Berezka, K.; Mokhun, S.; Fedchyshyn, O. Text Mining for the Analysis of Legal Texts. In Proceedings of the Proceedings - International Conference on Advanced Computer Information Technologies, ACIT; West Ukrainian National University, Department of Applied Mathematics, Ternopil, Ukraine, 2022; pp. 502–505.
192. Wieck, F.; Stein, N. V; Lower, M. Improving Safety in Europe-Detecting Counterfeit Certificates from FFP2 Masks: A Text-Mining Approach. In Proceedings of the ISPCE 2022 - IEEE International Symposium on Product Compliance Engineering; Institute of Electrical and Electronics Engineers Inc.: University of Wuppertal Wuppertal, Dept. of Product Safety and Quality, Germany, 2022.
193. Febro-Naga, J.-D.; Tinam-isan, M.-A.-C. Exploring Cyber Violence against Women and Girls in the Philippines through Mining Online News. *Comunicar* **2022**, *30*, 121–133. https://doi.org/10.3916/C70-2022-10.
194. Min, M.; Lee, J.J.; Lee, K. Detecting Illegal Online Gambling (IOG) Services in the Mobile Environment. *Secur. Commun. Networks* **2022**, *2022*. https://doi.org/10.1155/2022/3286623.
195. de Morais, J.P.M.; Merschmann, L.H. de C. A Cascade Approach for Gender Prediction from Texts in Portuguese Language. In Proceedings of the Proceedings of the Brazilian Symposium on Multimedia and the Web; Association for Computing Machinery: New York, NY, USA, 2022; pp. 142–149.
196. Panggabean, S.; Gata, W.; Setiawan, T.A. Analysis of Twitter Sentiment Towards Madrasahs Using Classification Methods. *J. Appl. Eng. Technol. Sci.* **2022**, *4*, 375–389. https://doi.org/10.37385/jaets.v4i1.1088.
197. Khandelwal, S.; Chaudhary, A. COVID-19 Pandemic & Cyber Security Issues: Sentiment Analysis and Topic Modeling Approach. *J. Discret. Math. Sci. Cryptogr.* **2022**, *25*, 987–997. https://doi.org/10.1080/09720529.2022.2072421.
198. Aguerri, J.C.; Miró-Llinares, F.; Vila-Viñas, D. When Social Media Feeds Classic Punitivism on Media: The Coverage of the Glorification of Terrorism on XXI. *Criminol. Crim. JUSTICE* **2022**. https://doi.org/10.1177/17488958221133467.
199. Wilson, M.; Spike, E.; Karystianis, G.; Butler, T. Nonfatal Strangulation During Domestic Violence Events in New South Wales: Prevalence and Characteristics Using Text Mining Study of Police Narratives. *Violence Against Women* **2022**, *28*, 2259–2285. https://doi.org/10.1177/10778012211025993.
200. Withall, A.; Karystianis, G.; Duncan, D.; Hwang, Y.I.; Kidane, A.H.; Butler, T. Domestic Violence in Residential Care Facilities in New South Wales, Australia: A Text Mining Study. *Gerontologist* **2022**, *62*, 223–231. https://doi.org/10.1093/geront/gnab068.

201. Algefes, A.; Aldossari, N.; Masmoudi, F.; Kariri, E.; IEEE A Text-Mining Approach for Crime Tweets in Saudi Arabia: From Analysis to Prediction. *2022 7TH Int. Conf. DATA Sci. Mach. Learn. Appl. (CDMA 2022)* 2022, 109–114.
202. Boukabous, M.; Azizi, M. Multimodal Sentiment Analysis Using Audio and Text for Crime Detection. *2022 2ND Int. Conf. Innov. Res. Appl. Sci. Eng. Technol.* 2022, 803-807 WE-Conference Proceedings Citation Inde.
203. Gunarathne, P.; Rui, H.X.; Seidmann, A. Racial Bias in Customer Service: Evidence from Twitter. *Inf. Syst. Res.* **2022**, *33*, 43–54. https://doi.org/10.1287/isre.2021.1058.
204. Feng, Y.Y.; Li, G.W.; Sun, X.L.; Li, J.P. Identification of Tourists' Dynamic Risk Perception-the Situation in Tibet. *Humanit. Soc. Sci. Commun.* **2022**, *9*. https://doi.org/10.1057/s41599-022-01335-w WE - Social Science Citation Index (SSCI) WE - Arts & Humanities Citation Index (A&HCI).
205. Unlu, A.; Yilmaz, K. Online Terrorism Studies: Analysis of the Literature. *Stud. Confl. Terror.* **2022**. https://doi.org/10.1080/1057610X.2022.2122108.
206. Bridgelall, R. An Application of Natural Language Processing to Classify What Terrorists Say They Want. *Soc. Sci.* **2022**, *11*. https://doi.org/10.3390/socsci11010023 WE - Emerging Sources Citation Index (ESCI).
207. Ryu, H.; Kim, C.; Kim, J.; Lee, S.J.; Lee, J.Y. Understanding the Road Rage Behavior and Implications: A Textual Approach Using Legal Cases in Korea. *Transp. Res. Interdiscip. Perspect.* **2022**, *16*. https://doi.org/10.1016/j.trip.2022.100712 WE - Emerging Sources Citation Index (ESCI).
208. Mothe, J.; Ullah, M.Z.; Okon, G.; Schweer, T.; Jursenas, A.; Mandravickaite, J. Instruments and Tools to Identify Radical Textual Content. *INFORMATION* **2022**, *13*. https://doi.org/10.3390/info13040193 WE - Emerging Sources Citation Index (ESCI).
209. Charmanas, K.; Mittas, N.; Angelis, L. Predicting the Existence of Exploitation Concepts Linked to Software Vulnerabilities Using Text Mining. In Proceedings of the Proceedings of the 25th Pan-Hellenic Conference on Informatics; Association for Computing Machinery: New York, NY, USA, 2022; pp. 352–356.
210. Li, Y. Towards Forecasting Internet Financial Frauds Based on Advertising. In Proceedings of the Proceedings - 2022 8th International Conference on Big Data and Information Analytics, BigDIA 2022; Institute of Electrical and Electronics Engineers Inc.: Peking University, China, 2022; pp. 5–11.
211. Tampus-Siena, M. Analyzing the Discussion of Gregorio Murder on Twitter Using Text Mining Approach. *Comput. Hum. Behav. Reports* **2022**, *8*. https://doi.org/10.1016/j.chbr.2022.100248.
212. Borah, A.R.; Krishna, B.V.S.; Shrinidhi, M.H.; Gouda, D.N.; Poojary, P. The Soul Safety. In Proceedings of the Proceedings of the 2022 3rd International Conference on Communication, Computing and Industry 4.0, C2I4 2022; Institute of Electrical and Electronics Engineers Inc.: New Horizon College of Engineering, Department of Computer Science and Engineering, Bangalore, India, 2022.
213. Chaudhary, M.; Bansal, D. Open Source Intelligence Extraction for Terrorism-Related Information: A Review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2022**, *12*. https://doi.org/10.1002/widm.1473.
214. Lindstadt, C.; Boyer, B.P.; Ciszek, E.; Chung, A.; Wilcox, G. Drunk Girl: A Brief Thematic Analysis of Twitter Posts about Alcohol Use and #MeToo. *Qual. Res. Reports Commun.* **2022**, *23*, 90–104. https://doi.org/10.1080/17459435.2021.2016919.
215. Alnazzawi, N. Using Twitter to Detect Hate Crimes and Their Motivations: The HateMotiv Corpus. *DATA* **2022**, *7*. https://doi.org/10.3390/data7060069 WE - Emerging Sources Citation Index (ESCI).
216. Zhao, F.; Skums, P.; Zelikovsky, A.; Sevigny, E.L.; Swahn, M.H.; Strasser, S.M.; Huang, Y.; Wu, Y. Computational Approaches to Detect Illicit Drug Ads and Find Vendor Communities Within Social Media Platforms. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **2022**, *19*, 180–191. https://doi.org/10.1109/TCBB.2020.2978476.
217. Hsieh, H.-P.; Jiang, J.; Yang, T.-H.; Hu, R.; Wu, C.-L. Predicting the Success of Mediation Requests Using Case Properties and Textual Information for Reducing the Burden on the Court. *Digit. Gov. Res. Pr.* **2022**, *2*. https://doi.org/10.1145/3469233.
218. Sadlek, L.; Čeleda, P.; Tovar\vnák, D. Current Challenges of Cyber Threat and Vulnerability Identification Using Public Enumerations. In Proceedings of the Proceedings of the 17th International Conference on Availability, Reliability and Security; Association for Computing Machinery: New York, NY, USA, 2022.
219. Ferreira, F.; Duarte, J.; Ugulino, W. Automated Statistics Extraction of Public Security Events Reported Through Microtexts on Social Networks. In Proceedings of the ACM International Conference Proceeding Series; W., S., V.V., G.N., A., de L.F., R.C.G., B., A.R., G., Eds.; Association for Computing Machinery: Instituto Militar de Engenharia Rio de Janeiro, Rio de Janeiro, Brazil, 2022; Vol. Par F18047.
220. Xu, Y.; Chen, G.; Liu, Q.; Xu, W.; Zhang, L.; Wu, J.; Fan, X. A Phishing Website Detection and Recognition Method Based on Naive Bayes. In Proceedings of the IEEE 6th Information Technology and Mechatronics Engineering Conference, ITOEC 2022; B., X., K., M., Eds.; Institute of Electrical and Electronics Engineers Inc.: Institute for Cyberspace Intelligence and Crime Governance, Zhejiang Police College, Hangzhou, China, 2022; pp. 1557–1562.
221. de Carvalho, V.D.H.; Costa, A.P.C.S. Towards Corpora Creation from Social Web in Brazilian Portuguese to Support Public Security Analyses and Decisions. *Libr. Hi Tech* **2022**. https://doi.org/10.1108/LHT-08-2022-0401.

222. Bridgelall, R. Applying Unsupervised Machine Learning to Counterterrorism. *J. Comput. Soc. Sci.* **2022**, *5*, 1099–1128. https://doi.org/10.1007/s42001-022-00164-w.
223. Mantoro, T.; Permana, M.A.; Ayu, M.A. Crime Index Based on Text Mining on Social Media Using Multi Classifier Neural-Net Algorithm. *Telkomnika (Telecommunication Comput. Electron. Control.* **2022**, *20*, 570–579. https://doi.org/10.12928/TELKOMNIKA.v20i3.23321.
224. Casimiro, G.R.; Digiampietri, L.A. Authorship Attribution with Temporal Data in Reddit. In Proceedings of the Proceedings of the XVIII Brazilian Symposium on Information Systems; Association for Computing Machinery: New York, NY, USA, 2022.
225. Mansour Khoudja, A.; Loukam, M.; Belkredim, F.Z. Towards Author Profiling from Modern Standard Arabic Texts: A Review. In Proceedings of the Lecture Notes in Networks and Systems; X., Y., S., S., N., D., A., J., Eds.; Springer Science and Business Media Deutschland GmbH: Hassiba Benbouali University of Chlef, Chlef, Algeria, 2022; Vol. 235, pp. 745–753.
226. Cascavilla, G.; Catolino, G.; Ebert, F.; Tamburri, D.A.; Heuvel, W.J. van den "When the Code Becomes a Crime Scene" Towards Dark Web Threat Intelligence with Software Quality Metrics. In Proceedings of the 2022 IEEE International Conference on Software Maintenance and Evolution (ICSME); 2022; pp. 439–443.
227. Gupta, A.; Matta, P.; Pant, B. Identification of Cybercriminals in Social Media Using Machine Learning. In Proceedings of the 2022 International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON 2022; Institute of Electrical and Electronics Engineers Inc.: Graphic Era Deemed to Be University, Graphic Era Hill University, Dehradun, India, 2022.
228. Deguara, N.; Arshad, J.; Paracha, A.; Azad, M.A. Threat Miner - A Text Analysis Engine for Threat Identification Using Dark Web Data. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data); 2022; pp. 3043–3052.
229. Bifari, E.; Alhalabi, W. Exploring Narrative Court Documents for Use in Police Academic Education. In Proceedings of the Proceedings - 2022 14th IEEE International Conference on Computational Intelligence and Communication Networks, CICN 2022; Institute of Electrical and Electronics Engineers Inc.: King Abdulaziz University, Computer Science Department, Jeddah, Saudi Arabia, 2022; pp. 41–45.
230. Aldossari, N.; Algefes, A.; Masmoudi, F.; Kariri, E. Data Science Approach for Crime Analysis and Prediction: Saudi Arabia Use-Case. In Proceedings of the 2022 Fifth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU); 2022; pp. 20–25.
231. Biswas, B.; Mukhopadhyay, A.; Bhattacharjee, S.; Kumar, A.; Delen, D. A Text-Mining Based Cyber-Risk Assessment and Mitigation Framework for Critical Analysis of Online Hacker Forums. *Decis. Support Syst.* **2022**, *152*. https://doi.org/10.1016/j.dss.2021.113651.
232. Lyu, Y.; Wang, Z.; Ren, Z.; Ren, P.; Chen, Z.; Liu, X.; Li, Y.; Li, H.; Song, H. Improving Legal Judgment Prediction through Reinforced Criminal Element Extraction. *Inf. Process. Manag.* **2022**, *59*. https://doi.org/10.1016/j.ipm.2021.102780.
233. dos Santos, A.R.S.; Rodrigues, C.M. de O.; de Melo, H.B.S. Identifying Xenophobia in Twitter Posts Using Support Vector Machine with TF/IDF Strategy. In Proceedings of the Proceedings of the XVIII Brazilian Symposium on Information Systems; Association for Computing Machinery: New York, NY, USA, 2022.
234. Büsgen, A.; Klöser, L.; Kohl, P.; Schmidts, O.; Kraft, B.; Zündorf, A. From Cracked Accounts to Fake IDs: User Profiling on German Telegram Black Market Channels. In Proceedings of the Communications in Computer and Information Science; A., C., O., G., S., H., C., Q., Eds.; Springer Science and Business Media Deutschland GmbH: Aachen University of Applied Sciences, Aachen, 52066, Germany, 2023; Vol. 1860 CCIS, pp. 176–202.
235. Al-Alawi, A.I.; A-Lmansouri, A.M. Artificial Intelligence in the Judiciary System of Saudi Arabia: A Literature Review. In Proceedings of the 2023 International Conference On Cyber Management And Engineering (CyMaEn); 2023; pp. 83–87.
236. Karteris, A.; Tzanos, G.; Papadopoulos, L.; Soudris, D. Detection of Cyber Security Threats through Social Media Platforms. In Proceedings of the 2023 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW); 2023; pp. 820–823.
237. Chauhan, R.; Upadhyay, S.; Vaidya, H. Fake News Detection Based on Machine Learning Algorithm. In Proceedings of the 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT); 2023; pp. 1–5.
238. Aaditya Anil, K.; Sasikumar, K.; Nambiar, R.K.; Rohith, K.P.; Viji Rajendran, V. Unmasking Cyberbullies on Social Media Platforms Using Machine Learning. In Proceedings of the 2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023; Institute of Electrical and Electronics Engineers Inc.: Nss College of Engineering, Dept. of Computer Science, Kerala, Palakkad, India, 2023.
239. Saravanan, S.; Menon, A.; Saravanan, K.; Hariharan, S.; Nelson, L.; Gopalakrishnan, J. Cybersecurity Audits for Emerging and Existing Cutting Edge Technologies. In Proceedings of the 2023 11th International Conference on Intelligent Systems and Embedded Design (ISED); 2023; pp. 1–7.

240. Sun, P.; Zuo, Y.; Wang, Y. Classification Model for NAVTEX Navigational Warning Messages Based on Adaptive Weighted TF-IDF. In Proceedings of the Proceedings of the 10th Multidisciplinary International Social Networks Conference; Association for Computing Machinery: New York, NY, USA, 2023; pp. 133–142.
241. Vajiac, C.; Lee, M.-C.; Kulshrestha, A.; Levy, S.; Park, N.; Olligschlaeger, A.; Jones, C.; Rabbany, R.; Faloutsos, C. DeltaShield: Information Theory for Human- Trafficking Detection. *ACM Trans. Knowl. Discov. Data* **2023**, *17*. https://doi.org/10.1145/3563040.
242. Saeed, A.; Khan, H.U.; Shankar, A.; Imran, T.; Khan, D.; Kamran, M.; Khan, M.A. Topic Modeling Based Text Classification Regarding Islamophobia Using Word Embedding and Transformers Techniques. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2023**. https://doi.org/10.1145/3626318.
243. Santiago, N.; Mendez, J. Analysis of Common Vulnerabilities and Exposures to Produce Security Trends. In Proceedings of the Proceedings of the 2022 International Conference on Cyber Security; Association for Computing Machinery: New York, NY, USA, 2023; pp. 16–19.
244. Rahmat, R.F.; Aziira, A.H.; Purnamawati, S.; Pane, Y.M.; Faza, S.; Nadi, F. Classifying Indonesian Cyber Crime Cases under ITE Law Using a Hybrid of Mutual Information and Support Vector Machine. *Int. J. Saf. Secur. Eng.* **2023**, *13*, 835–844. https://doi.org/10.18280/ijsse.130507.
245. Correia, F.A.; Nunes, J.L.; Alves, P.H.; Lopes, H. Dynamic Topic Modeling with Tensor Decomposition as a Tool to Explore the Legal Precedent Relevance Over Time. In Proceedings of the Proceedings of the ACM Symposium on Document Engineering 2023; Association for Computing Machinery: New York, NY, USA, 2023.
246. Pejic-Bach, M.; Jajic, I.; Kamenjarska, T. A Bibliometric Analysis of Phishing in the Big Data Era: High Focus on Algorithms and Low Focus on People. In Proceedings of the Procedia Computer Science; R., M., R., R., M.M., C.-C., D., D., E., P., Eds.; Elsevier B.V.: Faculty of Economics and Business, University of Zagreb, Square of John F. Kennedy 6, Zagreb, 10 000, Croatia, 2023; Vol. 219, pp. 91–98.
247. Berhoum, A.; Meftah, M.C.E.; Laouid, A.; Hammoudeh, M. An Intelligent Approach Based on Cleaning up of Inutile Contents for Extremism Detection and Classification in Social Networks. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2023**, *22*. https://doi.org/10.1145/3575802.
248. Sivanantham, K.; Blessington Praveen, P.; Deepa, V.; Mohan Kumar, R. Cybercrime Sentimental Analysis for Child Youtube Video Dataset Using Hybrid Support Vector Machine with Ant Colony Optimization Algorithm. In *Studies in Computational Intelligence*; Springer Science and Business Media Deutschland GmbH: HCL Technologies, Tamilnadu, Coimbatore, India, 2023; Vol. 1080, pp. 175–193 ISBN 1860949X (ISSN).
249. Gómez-Camacho, A.; Hunt-Gómez, C.I.; Núñez-Roman, F.; Esteban, A.N. "Not All Motherfuckers Are MENA, but Most MENA Are Motherfuckers": Hate Speech on Twitter against Unaccompanied Foreign Minors. *J. Lang. Aggress. Confl.* **2023**, *11*, 256–278. https://doi.org/10.1075/jlac.00083.gom.
250. Zhen, Z.; Gao, J. Chinese Cyber Threat Intelligence Named Entity Recognition via RoBERTa-Wwm-RDCNN-CRF. *Comput. Mater. Contin.* **2023**, *77*, 299–321. https://doi.org/10.32604/cmc.2023.042090.
251. Parker, M.A.; Valdez, D.; Rao, V.K.; Eddens, K.S.; Agley, J. Results and Methodological Implications of the Digital Epidemiology of Prescription Drug References Among Twitter Users: Latent Dirichlet Allocation (LDA) Analyses. *J. Med. Internet Res.* **2023**, *25*. https://doi.org/10.2196/48405.
252. Aljohani, E.J.; Yafooz, W.M.S.; Alsaeedi, A. Cyberbullying Detection Approaches: A Review. In Proceedings of the Proceedings of the 5th International Conference on Inventive Research in Computing Applications, ICIRCA 2023; Institute of Electrical and Electronics Engineers Inc.: College of Computer Science and Engineering, Taibah University, Department of Computer Science, Madinah, Saudi Arabia, 2023; pp. 1310–1316.
253. Goldstein, E. V; Mooney, S.J.; Takagi-Stewart, J.; Agnew, B.F.; Morgan, E.R.; Haviland, M.J.; Zhou, W.; Prater, L.C. Characterizing Female Firearm Suicide Circumstances: A Natural Language Processing and Machine Learning Approach. *Am. J. Prev. Med.* **2023**, *65*, 278–285. https://doi.org/10.1016/j.amepre.2023.01.030.
254. Ptaszek, G.; Yuskiv, B.; Khomych, S. War on Frames: Text Mining of Conflict in Russian and Ukrainian News Agency Coverage on Telegram during the Russian Invasion of Ukraine in 2022. *Media, War Confl.* **2023**. https://doi.org/10.1177/17506352231166327.
255. Kovalchuk, O.; Banakh, S.; Kasianchuk, M.; Moskaliuk, N.; Kaniuka, V. Associative Rule Mining for the Assessment of the Risk of Recidivism. In Proceedings of the CEUR Workshop Proceedings; T., H., O., S., P.T., P., S., L., Eds.; CEUR-WS: West Ukrainian National University, 11 Lvivska str., Ternopil, 46009, Ukraine, 2023; Vol. 3373, pp. 376–387.
256. Sood, P.; Sharma, C.; Nijjer, S.; Sakhuja, S. Review the Role of Artificial Intelligence in Detecting and Preventing Financial Fraud Using Natural Language Processing. *Int. J. Syst. Assur. Eng. Manag.* **2023**, *14*, 2120–2135. https://doi.org/10.1007/s13198-023-02043-7.

257. Hui, V.; Eby, M.; Constantino, R.E.; Lee, H.; Zelazny, J.; Chang, J.C.; He, D.; Lee, Y.J. Examining the Supports and Advice That Women With Intimate Partner Violence Experience Received in Online Health Communities: Text Mining Approach. *J. Med. Internet Res.* **2023**, *25*. https://doi.org/10.2196/48607.
258. Qiu, M.; Zhang, X.; Wang, X. An Ex-Convict Recognition Method Based on Text Mining. *Int. J. Secur. Networks* **2023**, *18*, 10–18. https://doi.org/10.1504/IJSN.2023.129905.
259. Bashar, M.A.; Nayak, R.; Knapman, G.; Turnbull, P.; Fforde, C. An Informed Neural Network for Discovering Historical Documentation Assisting the Repatriation of Indigenous Ancestral Human Remains. *Soc. Sci. Comput. Rev.* **2023**, *41*, 2293–2317. https://doi.org/10.1177/08944393231158788.
260. Salama, R.; Al-Turjman, F.; Altrjman, C.; Kumar, S.; Chaudhary, P. A Comprehensive Survey of Blockchain-Powered Cybersecurity- A Survey. In Proceedings of the 2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN); 2023; pp. 774–777.
261. Rahman, M.R.; Hezaveh, R.M.; Williams, L. What Are the Attackers Doing Now? Automating Cyberthreat Intelligence Extraction from Text on Pace with the Changing Threat Landscape: A Survey. *ACM Comput. Surv.* **2023**, *55*. https://doi.org/10.1145/3571726.
262. D., Y.; Panduro-Ramirez, J.; Buddhi, D.; Vekariya, V.; Pillai, B.G.; Tida, N. The Effective Role of Cyber Security in Supply Chain to Enhance Supply Chain Performance and Collaboration. In Proceedings of the 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE); 2023; pp. 838–842.
263. Rathor, K.; Vidya, S.; Jeeva, M.; Karthivel, M.; Ghate, S.N.; Malathy, V. Intelligent System for ATM Fraud Detection System Using C-LSTM Approach. In Proceedings of the 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC); 2023; pp. 1439–1444.
264. Karthika, I.; Boomika, G.; Nisha, R.; Shalini, M.; Srivarshini, S.P. A Survey on Detecting and Preventing Hateful Comments on Social Media Using Deep Learning. In Proceedings of the Smart Innovation, Systems and Technologies; J., C., P., M., T., P., A., J., Eds.; Springer Science and Business Media Deutschland GmbH: Department of Computer Science and Engineering, M.Kumarasamy College of Engineering, Tamil Nadu, Karur, 639113, India, 2023; Vol. 312, pp. 285–298.
265. Stalidi, C.; Popovici, E.-C.; Suciu, G. Preliminary Architecture and a Pilot Implementation for a Malicious Emails Detection Solution. In Proceedings of the 15th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2023 - Proceedings; Institute of Electrical and Electronics Engineers Inc.: Politehnica University of Bucharest, Beia Consult International, Faculty of Electronics, Telecommunications, and Information Technology, Telecommunications Dept., R&d Department, Bucharest, Romania, 2023.
266. Chatzimarkaki, G.; Karagiorgou, S.; Konidi, M.; Alexandrou, D.; Bouras, T.; Evangelatos, S. Harvesting Large Textual and Multimedia Data to Detect Illegal Activities on Dark Web Marketplaces. In Proceedings of the 2023 IEEE International Conference on Big Data (BigData); 2023; pp. 4046–4055.
267. S., G.; Chandrasekaran, D.; Sre, M.D.; Sathiyanarayanan, M. Predicting Abnormal User Behaviour Patterns in Social Media Platforms Based on Process Mining. In Proceedings of the 2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE); 2023; pp. 204–209.
268. Guo, Z.; Wang, P.; Cho, J.H.; Huang, L.F.; ACM Text Mining-Based Social-Psychological Vulnerability Analysis of Potential Victims To Cybergrooming: Insights and Lessons Learned. *COMPANION WORLD WIDE WEB Conf. WWW 2023* 2023, 1381–1388.
269. Lee, C.S.; Jang, A. Questing for Justice on Twitter: Topic Modeling of #StopAsianHate Discourses in the Wake of Atlanta Shooting. *CRIME Delinq.* **2023**, *69*, 2874–2900. https://doi.org/10.1177/00111287211057855.
270. Bera, D.; Ogbanufe, O.; Kim, D.J. Towards a Thematic Dimensional Framework of Online Fraud: An Exploration of Fraudulent Email Attack Tactics and Intentions. *Decis. Support Syst.* **2023**, *171*. https://doi.org/10.1016/j.dss.2023.113977.
271. Zhuchkova, S.; Kazun, A. Exploring Gender Bias in Homicide Sentencing: An Empirical Study of Russian Court Decisions Using Text Mining. *HOMICIDE Stud.* **2023**. https://doi.org/10.1177/10887679231217159.
272. Kumar, S. Negative Stances Detection from Multilingual Data Streams in Low-Resource Languages on Social Media Using BERT and CNN-Based Transfer Learning Model. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2024**, *23*. https://doi.org/10.1145/3625821.
273. Jones, G.M.; Santhiya, P.; Winster, S.G.; Sundar, R. An Intelligent Analysis of Mobile Evidence Using Sentimental Analysis. In Proceedings of the Lecture Notes in Electrical Engineering; S.J., P., N.K., C., B.N., G., S.S., I., Eds.; Springer Science and Business Media Deutschland GmbH: Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India, 2024; Vol. 1075 LNEE, pp. 317–330.
274. Fiesler, C.; Zimmer, M.; Proferes, N.; Gilbert, S.; Jones, N. Remember the Human: A Systematic Review of Ethical Considerations in Reddit Research. *Proc. ACM Hum.-Comput. Interact.* **2024**, *8*. https://doi.org/10.1145/3633070.

275. Shen, B. Analysis and Comparison of Limitation Interests in Civil Procedure Law in the Context of Information Technology. *Appl. Math. Nonlinear Sci.* **2024**, *9*. https://doi.org/10.2478/amns.2023.2.00946.
276. Karami, A.; Swan, S.C.; White, C.N.; Ford, K. Hidden in Plain Sight for Too Long: Using Text Mining Techniques to Shine a Light on Workplace Sexism and Sexual Harassment. *Psychol. Violence* **2024**, *14*, 1–13. https://doi.org/10.1037/vio0000239 WE - Social Science Citation Index (SSCI).
277. Karystianis, G.; Chowdhury, N.; Sheridan, L.; Reutens, S.; Wade, S.; Allnutt, S.; Kim, M.T.; Poynton, S.; Butler, T. Text Mining Domestic Violence Police Narratives to Identify Behaviours Linked to Coercive Control. *CRIME Sci.* **2024**, *13*. https://doi.org/10.1186/s40163-024-00200-2 WE - Emerging Sources Citation Index (ESCI).
278. Albayari, R.; Abdallah, S.; Shaalan, K. Cyberbullying Detection Model for Arabic Text Using Deep Learning. *J. Inf. Knowl. Manag.* **2024**. https://doi.org/10.1142/S0219649224500163.
279. Jiao, J.; He, P.; Zha, J. Factors Influencing Illegal Dumping of Hazardous Waste in China. *J. Environ. Manage.* **2024**, *354*. https://doi.org/10.1016/j.jenvman.2024.120366.
280. Zhang, Y.; Zhai, Y.; Fu, S.; Shi, M.; Jiang, X. Quantitative Analysis of Maritime Piracy at Global and Regional Scales to Improve Maritime Security. *Ocean Coast. Manag.* **2024**, *248*. https://doi.org/10.1016/j.ocecoaman.2023.106968.
281. Salley, C.J.; Mohammadi, N.; Taylor, J.E. Safeguarding Infrastructure from Cyber Threats with NLP-Based Information Retrieval. In Proceedings of the Proceedings of the Winter Simulation Conference; IEEE Press, 2024; pp. 853–862.