



Measuring Cybercrime in Calls for Police Service

Doy Kwon¹ · Hervé Borrión¹ · Richard Wortley¹

Received: 7 March 2024 / Accepted: 24 June 2024
© The Author(s) 2024

Abstract

Conventional police databases contain much information on cybercrime, but extracting it remains a practical challenge. This is because these databases rarely contain labels that could be used to automatically retrieve all cybercrime incidents. In this article, we present a supervised machine learning method for extracting cybercrime incidents in calls for police service datasets. Data from the Korean National Police (2020, 9 months, $N=15$ million call logs) is used for the demonstration. We combined methods of keyword query selection, minority oversampling, and majority voting techniques to develop a classifier. Three classification techniques, including Naïve Bayes, linear SVM, and kernel SVM, were tested, and the kernel model was chosen to build the final model (accuracy, 93.4%; F1-score, 92.4). We estimate that cybercrime only represents 4.6% of the cases in the selected dataset (excluding traffic-related incidents), but that it can be prevalent with some crime types. We found, for example, that about three quarters (76%) of all fraud incidents have a cyber dimension. We conclude that the cybercrime classification method proposed in this study can support further research on cybercrime and that it offers considerable advantages over manual or keyword-based approaches.

Keywords Calls for service · Cyber-enabled crime · Cybercrime · Machine learning · Emergency call · Text classification

Introduction

Digital technology has transformed society. It underpins much of people's routine activities, from accessing and sharing information on social media to using banking services on the move, through ordering goods and having them delivered at home. However, the impact of digital technology is not limited to lawful applications. It has changed the opportunity structure for traditional crimes (Felson & Boba, 2010; Holt & Bossler, 2015).

✉ Doy Kwon
doy.kwon.19@ucl.ac.uk

Hervé Borrión
h.borrión@ucl.ac.uk

Richard Wortley
r.wortley@ucl.ac.uk

¹ University College London Department of Security and Crime Science, 35 Tavistock Square, London WC1H 9EZ, UK

Indeed, stalking, bullying, and fraud can be facilitated by misuse of emails, social media, geo-localisation, and online purchase services. Furthermore, new crimes have emerged in cyberspace, including new crimes that can only occur because of cyber (e.g. hacking, scamming, phishing, smishing) and traditional crimes that are enabled by cyber (e.g. online harassment, online fraud, and online child sexual abuse material).

Extant efforts to measure and understand cybercrime typically rely on three sources of data: traditional police-recorded crime databases (Furnell & Dowling, 2019; Popham et al., 2020; van der Laan & Tollenaar, 2021), databases from recently established cybercrime bureaux (e.g. Action Fraud, the IC3), and cyber-victimisation surveys (DeVoe & Murphy, 2011; Reep-van den Bergh and Junger 2018). However, these sources have significant limitations, as outlined below, that the current study seeks to redress:

- *Police-recorded crime databases* do not always label cybercrime incidents, and those that do rarely contain more than a few cybercrime sub-categories (Furnell & Dowling, 2019). This is because most traditional crime classifications are based on legal definitions of crime that tend to be defined based on a type of harm, not tools (McGuire, 2019). Hence, in police databases that adhere to the legal definition of crime for processing criminal cases within the criminal justice system, most conventional crimes that are enabled by cyber technology are recorded as conventional crimes. As a result, data mining methods (i.e. the ‘incident filtering function’) standardly found on police database management software cannot be used to extract certain types of cybercrime incidents (e.g. cyber sexual harassment or detailed types of cyber frauds). Extraction therefore requires selecting a broader set of crime incidents (e.g. sexual harassment or fraud) and manually identifying cybercrime incidents among them, a process that can be very time consuming.
- *Databases generated by cybercrime bureaux* only contain cybercrimes with pre-defined categories. First, it may not be possible to report all types of cybercrime incidents. Second, non-reporting rate is high in these specialist agencies, mainly because victims are not aware that they can report cybercrime to them. In the UK, Furnell and Dowling (2019) found that the most commonly cited reason for not reporting cybercrime was that people ‘haven’t heard of Action Fraud’—the national self-reporting website for fraud and cybercrime. The same is likely to apply in South Korea.
- *Victimisation surveys* are known to yield better reporting rates, but they do not tap into all facets of cybercrime. Due to time constraints, survey questionnaires usually focus on the most common types of cybercrime, such as online scams (Reep-van den Bergh & Junger, 2018; Whitty & Buchanan, 2012), cyberstalking (Fissel & Reynolds, 2020), and online defamation (Näsi et al., 2015). Notwithstanding the strengths of victimisation surveys, these instruments cannot be used to provide an overview of all cybercrime types currently.

Given these limitations, calls for police service (CPS) datasets may be considered a promising data source to understand the scale and nature of cybercrime. CPS databases include records about not only crimes but also anti-social behaviour and even simple consultations from citizens—all of which may have a cyber component (Boulton et al., 2017). The data also contain more crime events than official crime records datasets which only include crimes classified under legislative categories. Hence, a CPS dataset may encompass many different types of cybercrime incidents. Moreover, a CPS dataset offers practical advantages as analysts can readily access it. They should therefore be able to conduct timely investigations of cybercrime without requesting data from other bodies.

Unfortunately, extracting cybercrime records from CPS datasets is challenging because the records do not have built-in meta-data (e.g. labels, flags) for cybercrime and its sub-categories. Researchers have used a range of methods for extracting information and analysing cybercrime: manual investigation (Montoya et al., 2013), keyword query extraction (Domenie et al., 2009), and supervised machine learning classification (van der Laan & Tollenaar, 2021). Basit (2003) pointed out that the size of the dataset, available resources (time and fund), and the expertise of the researchers are all critical factors in the selection of an appropriate data extraction approach. Manual investigation of large text datasets is a time-consuming process that requires significant resources (van der Laan & Tollenaar, 2021). Keyword query extraction can be useful when the analyst has sufficient knowledge of the topic to suggest appropriate keywords (Sagayam et al., 2012), but it requires manual verification to filter out false positives (van der Laan & Tollenaar, 2021), and this can be time consuming.

In this context, supervised machine learning classification appears a promising approach for efficiently classifying texts from large volumes of documents (Kadhim, 2019). However, there are no publications describing in detail how a cybercrime classifier should be designed and what performance can be expected from it. This study aims to fill this gap and explain how a supervised machine learning method can be developed to identify cybercrime cases in CPS datasets. The classifier was developed and tested using data from the Korean CPS dataset ($N=15$ million, 1 January–30 September 2020).

Background

Cybercrime in Police Crime Database

Crimes involving network technology are generally referred to as ‘cybercrime’. This category is often divided into two groups: cyber-dependent crime and cyber-enabled crime (McGuire & Dowling, 2013). Cyber-dependent crime (CDC) refers to offenses that cannot be committed without computers, networks, or information communication technology (ICT). Examples include hacking, computer viruses, and DDoS attacks. In contrast, cyber-enabled crime (CEC) refers to traditional incidents in which computers, networks, or ICT have an important role, such as romance scams, cyberstalking, and online fraud (McGuire, 2019).

Researchers have attempted to measure the scale of cybercrime using police crime databases. Furnell and Dowling (2019) estimated that cybercrime represented about 1% of incidents recorded in UK police databases. van der Laan and Tollenaar (2021) investigated crime records from the Dutch police and reported that CDC and CEC constituted about 0.34% and 5.59% of all crime records in 2016. Interestingly, Mazowita and Vézina (2014) found that 55% of cybercrime incidents (9084) recorded in the Canadian Uniform Crime Report (UCR) dataset of 2012 were fraud cases. The low number of cybercrimes in the UCR implies that traditional crime datasets are unlikely to capture the scale of the problem or its evolving nature. Additionally, cybercrime victims may not see the police as the natural agency to deal with many forms of cybercrime.

Statistics were also generated from cases reported on special cybercrime bureaux’s online platforms. The FBI’s Internet Crime Complaint Centre, for example, reported that an average of 552,000 cybercrime complaints a year were handled in the US between 2017 and 2021 (IC3, 2021: 7). Similarly, the Action Fraud (2021a, 2021b) announced that

31,322 cyber-dependent crimes and 875,622 fraud cases were reported in the UK in 2021. Additionally, the Korean National Police Agency (KNPA, 2020) denoted that an average of 169,802 cybercrime cases were reported in South Korea between 2016 and 2020.

Automatic Extraction of Cases from Police-Recorded Crime Datasets

To date, only a few studies have attempted to automatically classify crime incidents in police datasets. Matos et al. (2022) built a classifier to automate tasks that were manually performed by the Assistant Secretariat for Intelligence and Criminal Analysis (SIAC) in Brazil. Created using a manually labelled dataset, their Convolutional Neural Network achieved an accuracy of 78% in classifying the 20,000 cases of the test dataset into 463 crime categories. This approach, however, cannot be replicated here because cybercrime cases are not labelled in police datasets.

Haleem et al. (2019) took a different approach in their attempt to classify mental illness cases from calls for police service dataset of Greater Manchester Police, UK. To build the training set, two researchers and one police officer independently labelled 391 cases and finalised each label through quorum voting. Their approach, which is based on word-embedding techniques and deep learning including continuous bag of words and distributed bag of words, yielded 89% accuracy. However, the size of the test data set ($N=78$) and lack of details about the method used to build the training set ($N=391$ call logs) raise concerns about the application of the method to large datasets. Additionally, all non-consensus samples were labelled based on the opinions of two out of three coders. However, descriptions in the individual cases of CPS data sometimes lacked information regarding situations, causes, or motives revealed at the scene, which could indicate whether a case is associated with a mental health issue or not. Therefore, some samples may fall into a grey zone and cannot be confidently assigned to either class. This limitation is discussed later in the paper.

Karystianis et al. (2018) used a rule-based approach to identify mental health disorder events in domestic violence (DV) cases in the Australian police database WebCOPS. The 200 cases randomly selected from the DV records ($N=492$ k) were manually labelled by two experts. Labelling was based on lexical patterns identified through manual reading, as well as a list of mental health disorders in the International Classification of Diseases (ICD, 2017). A lexical rule was created to distinguish who, between the offender or the victim, had mental health disorder. The model performance obtained using a separate testing set ($N=100$) showed that the F -scores were about 81% and 87% for victims and offenders, respectively. The final model was applied to the entire DV dataset, and it was estimated that mental disorder-related cases represented 16% ($N=78$ k) of all cases. The researchers provided detailed information about the labelling process, but they did not explain how they resolved disagreements between coders. Furthermore, the testing set was too small to be confident that the performance level would hold with very large datasets.

van der Laan and Tollenaar (2021) developed a supervised classifier and applied it to cybercrime incidents recorded in the Dutch police registration system. Incidents were distributed between three classes of CDC (hacking, ransomware, and DDoS attack) and five classes of CEC (online threat, online stalking, cyber-libel, online identity theft, and online fraud). With 67% accuracy, the classifier exhibited moderate performance, leaving ample room for improvement before it could be reliably used by crime analysts. Furthermore, important details were not provided about the design of the classifier, which makes replication difficult. First, the authors did not explain how they selected the cybercrime-related

keywords in the study. Second, during the process of creating a training set for a supervised classifier, the authors categorised all incidents in the dataset into one of eight cybercrime types based on their criteria. They used cybercrime-related keywords to sample 5000 cases. Presumably, some of them did not correspond to any of the eight classes used in the research. Therefore, the authors should have included a separate class for other types of cybercrime incidents that do not belong to any of the eight categories. For instance, cases related to online child sexual abuse material or various phishing frauds cannot be classified as one of the defined cybercrime categories. It remains unclear how this affected the classification results. Third, without an ‘others’ class, non-cybercrime cases, if any, would have been incorrectly classified. It is not clear how the authors proposed to address this issue.

Data and Proposed Classification Method

This research aims to build a machine learning classifier based on the text information in call logs. The KCPS data were anonymised and employed for this study. The dataset comprises all the calls for police service made in South Korea between the 1 January and 30 September 2020 ($N=15,743$ k). The learning set (which consists of the combinations of training and testing sets) was used to develop the classifier. The latter was then applied to the whole KCPS dataset, and statistics were computed about the identified cybercrime incidents.

The flowchart in Fig. 1 represents the seven-step procedure used in this study. The central methodological part involves building a high-quality learning set. First, all call logs in the KCPS dataset were pre-processed to identify individual event cases by removing redundant information. After that, a learning dataset was built through three procedures, namely, generating a cybercrime-related terms list, performing minority oversampling, and labelling incidents. Next, word features were chosen using the term frequency-inverse document frequency (TF-IDF) scores. The model performance was evaluated using a tenfold cross-validation technique, with the training and testing sets rotated ten times to ensure good validity. The models were trained and tested with the learning set ($N=22,877$ incidents). The model with the greatest combination of accuracy and performance was then selected and applied to the KCPS dataset.

Data Pre-processing

A number of calls were removed from KCPS dataset. These include test calls¹ made by police ($N=18$ k), calls without text ($N=4651$ k), duplicates ($N=1324$ k), and transferred calls² ($N=159$ k). Duplicate cases were identified based on the content of the text column: ‘same as previous call’ (‘동일’ or ‘중첩’). Through this process, the size of the KCPS dataset was reduced by 39% to 9591 k incidents, which were finally used as population data to apply the final model.

¹ Test calls refer to the calls made by police officers to test the call system. These calls do not correspond to real incidents.

² Transferred calls are created when the caller is not located in the same area as the case handler. In this case, the call is transferred to the jurisdictional police of the case location. However, the incident number remains the same, so officers can understand the history of the case.

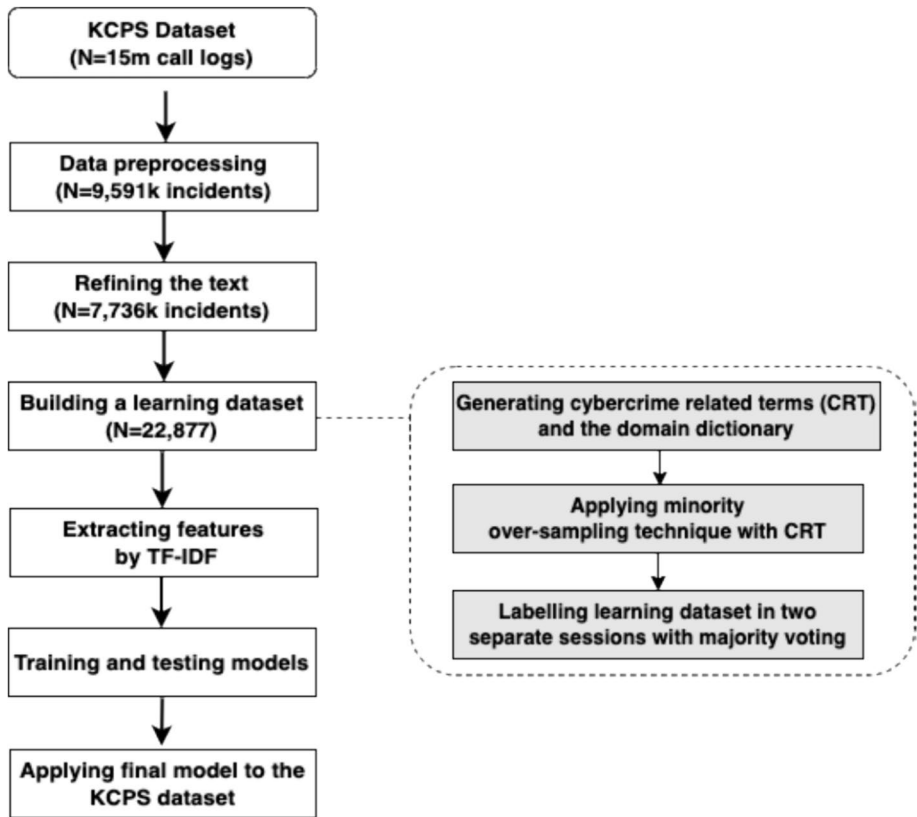


Fig. 1 Flowchart representing the development process of the cybercrime classifier

As Korean sentences become shorter during morphological analysis (Han & Palmer, 2004), very short case descriptions may not possess enough features after lemmatisation for accurate classification. The primary researcher conducted a manual review of the text in the dataset and arbitrarily determined that a minimum of nine characters were necessary to effectively convey meaningful information for describing a crime case. Records containing less than nine characters ($N=1855$ k) were therefore removed. The resulting dataset ($N=7736$ k) was used to build a learning dataset.

Refining the Text

The dataset was refined by merging columns, removing stop words,³ and improving spelling. First, each incident was assigned a unique document number for the study. Then, the three columns that contain information about the cases (i.e. report, end comment, and reference comment) were merged into a single column. Populated by the KCPS call centre

³ Stop words refer to commonly used words that are considered insignificant or lacking substantial meaning (e.g. the, a, an, of).

operators, the ‘report’ column summarises the issues raised by callers (including citizens or other agencies). The text in the ‘end comment’ column is recorded by the officer dispatched at the scene once they finished dealing with the case. It contains a chronological record of events and actions taken during the incident. ‘Reference comments’ were recorded by the operator when they wanted to log further information deemed relevant to the case.

As there is no standard list of ‘stop words’ in Korean, we drew a list of elements without specific lexical meaning and removed them from the dataset. The elements, including numbers (10), letters (26), and punctuation marks (29), can be found in the Appendix. English letters in the dataset (e.g. CCTV) were used for the morphological analysis. As operators are expected to quickly elicit, understand, and synthesise information from a large number of calls, spelling accuracy is often secondary in their job. After manual reading, misspelt terms deemed to be related to cybercrime ($N=30$) were then edited (Appendix Table 7).

Building the Learning Dataset

Cybercrime-Related Terms

As previously noted, Furnell and Dowling (2019) estimated that cybercrime accounts for about 1% of all recorded crimes in the UK police data. If the proportion of cybercrime is similarly low in the KCPS dataset, we could expect random sampling to yield a highly imbalanced learning dataset. To address this challenge, we carried out keyword search of incidents involving digital technology. In total, 88 cybercrime-related terms (CRT) were selected that relate to cyber, online, computer, Internet, Internet of Things (IoT), ICT, and networks (Appendix Table 8). The CRT was searched using three different approaches and finally selected when a term was found in the KCPS dataset:

- A random number generator in the R software was used to select 20,000 incidents. The primary researcher identified 50 terms that meet the above inclusion criterion.
- TF-IDF keywords selection method was applied to the dataset (Aizawa, 2003), and 4 additional CRT were selected among the top 200 keywords.
- Another set of 29 terms were identified in searching from two Korean online portal sites (i.e. Naver and Daum).

Although the R package NLP4kec (Namyoun, 2020) is a useful tool for performing morphological analysis on Korean text data, it has limitations in identifying specific terms that are not included in its built-in dictionary. To address this, we iteratively created a ‘domain dictionary’ consisting of 712 relevant acronyms and bigrams frequently found in the learning set. Throughout the labelling process, we continuously updated this dictionary. For instance, the phrase ‘컴퓨터 (computer) + 조회 (search)’ is recognised as two separate words by a machine learning model because they are treated as a noun and a verb in the built-in dictionary. However, within the samples in the learning set, the bigram ‘컴퓨터조회 (computer search)’ was consistently observed when officers used it to refer to searching relevant information for a case in the intranet system. This bigram conveys a different meaning (and context) within the Korean police domain compared to when the words are separated. To illustrate, consider the sentences ‘a man uses a computer to search a victim online’ and ‘officers use a computer search for the case in their intranet’. Both sentences contain the words ‘computer’ and ‘search’. By treating ‘computer search’ as a distinct term in the latter sentence, the features in the learning set can be improved. Hence, a model can

recognise this bigram as a noun by including it in the domain dictionary. By incorporating the domain dictionary into morphological analysis, we retained non-standard terminologies in the dataset instead of discarding them during lemmatisation. This was possible because the computer can automatically recognise the terms listed in the domain dictionary as nouns.

Minority Oversampling

As explained above, the dataset contains a relatively small number of cybercrime incidents. Our approach therefore involved ‘oversampling the minority class’ and ‘undersampling the majority class’ in the learning dataset (see Chawla et al., 2002). The CRT was used to partition the dataset into two groups: Group A ($N=423$ k), whose elements contained the CRT, and Group B ($N=7313$ k), whose elements did not. Oversampling from Group A enabled us to obtain a learning set with similar proportions of cybercrime and non-cybercrime cases. Datasets A’ ($N=13,000$) and B’ ($N=10,000$) were generated through random sampling from Datasets A and B, respectively. Note that the sample size of dataset A’ was purposely greater than that of dataset B’ because some of the incidents in dataset A’ may not actually be cybercrime. The dataset ($N=23,000$) obtained by combining the two samples (A’ and B’) constitutes the learning dataset.

Labelling the Learning Dataset

To label the learning dataset, we recruited 16 Korean police officers (13 males and three females) with direct field experience. This ensured that the coders had sufficient understanding of the Korean language, KNP call recording system, and crime in Korea. The duty stations of officers were spread across the country; the officers’ average length of service was 12.7 years.

The labelling procedure consisted of two sessions. In the first session, the learning dataset ($N=23,000$) was divided into 16 subsets. Each subset was independently labelled by two people: the primary researcher and an officer (note: a different one for each subset). This reduced the burden on individual officers and the risk of systematic bias. The officers received guidelines (see Appendix Table 9) that explained the concept of cyber-enabled and cyber-dependent crime. After the first session, the officers and the researcher disagreed on 820 incidents (3.56%).

In the second labelling session, these 820 incidents were split into eight subsets, and each subset was independently re-labelled by two more officers. Majority voting (Sheng et al., 2017) between the researcher and three officers was adopted to decide on the final label. After the second review, only 123 unresolved incidents (0.53%) remained for which there was no majority vote. These were discarded. The final learning dataset comprised 10,261 cybercrime incidents (44.85%) and 12,616 non-cybercrime incidents (55.15%).

Extracting Features (TF-IDF)

The morphological analysis that lemmatises and extracts nouns, verbs, and adjectives from the dataset was performed using the R package NLP4kec (Namyoun, 2020). The resulting corpus was transformed to a document term matrix (DTM) with TF-IDF weighting. Infrequent terms, occurring less than once in 1000 cases, were eliminated from the DTM. As a result, the number of terms in the DTM decreased from 30,436 to 2135.

Training and Testing the Models

The learning dataset ($N=22,877$) was randomly divided into training ($N=18,000$) and testing ($N=4877$) sets. Finally, tenfold cross-validation was used to validate and optimise the hyper-parameters of the final model. The linear SVM, kernel SVM, and Naïve Bayes algorithms were employed to build the model. The final model that showed the best performance among three algorithms was applied to classify cybercrime from the KCPS dataset ($N=9591$ k). The results are described in the next section.

Results

Distribution of Crime Categories in the Datasets

Table 1 illustrates how the cases in the KCPS, training, and testing sets are distributed across the seven crime categories. The sex crime category consists of calls regarding sexual violence, sexual harassment, dating violence, and stalking. Violent crime refers to violence, domestic violence, child abuse, extortion, blackmail, property damage, robbery, homicide, dispute, and aggressive behaviour in public. The fraud category consists of traditional fraud and phishing fraud. Juvenile delinquency comprises school violence and other misbehaviours typically carried out by young people. An example would be a group of students calling out one student on Facebook messenger and harassing him while filming a video. Theft relates to larceny, burglary, and snatch theft. Traffic consists of traffic accidents, traffic violations, hit-and-runs, intoxicated driving, and traffic inconveniences. The 'others' field covers a wide range of situations, including criminal offenses such as illegal drugs, illegal gambling, miscellaneous criminal offences, other minor offenses, and wanted persons, as well as non-criminal situations that require police intervention for security or safety reasons, such as consultation, disaster, emergency bell alarms, fire, intoxicated person, missing persons, noise complaints, preventing a dangerous situation, requests from private security company, rescue operations, police service request, suicide, and unconfirmed cases.

As expected, the distribution of incidents varies between the three datasets due to minority oversampling. In particular, the proportion of incidents classified as 'fraud' is seven to eight times higher in the training (13.7%) and testing sets (14.9%), respectively, than in the KCPS dataset (1.9%). This implies that fraud incidents are much more likely to

Table 1 Distribution of incidents by crime types in the KCPS, training, and testing datasets

Crime types	KCPS ($N=9591$ k)	Training set ($N=18$ k)	Testing set ($N=4.8$ k)
Sex crime	0.7%	0.6%	0.7%
Violent crime	9.9%	8.0%	8.1%
Fraud	1.9%	13.6%	14.9%
Juvenile delinquency	0.6%	0.5%	0.5%
Theft	4.6%	3.4%	3.8%
Others	71.7%	67.5%	65.2%
Traffic	10.7%	6.4%	6.7%

Table 2 Proportions of cybercrime in the training and testing sets by crime types

Crime types	Training set (<i>N</i> = 18,000)		Testing set (<i>N</i> = 4877)	
	Cybercrime	Non-cybercrime	Cybercrime	Non-cybercrime
Total	44.7%	55.3%	44.4%	55.6%
Sex crime	29.5%	70.5%	19.4%	80.6%
Violent crime	6.5%	93.5%	5.4%	94.6%
Fraud	95.0%	5.0%	95.2%	4.8%
Juvenile delinquency	5.4%	94.6%	6.4%	93.7%
Theft	18.7%	81.8%	20.3%	79.7%
Others	45.9%	54.3%	47.6%	52.4%
Traffic	0.5%	99.5%	0%	100%

have a cyber component than other types of incidents. In contrast, the proportion of traffic incidents is 0.6 times lower in the learning datasets (6.4% and 6.7%) than in the KCPS dataset (10.7%). This reflects the fact that traffic incidents are comparatively less likely to be related to cybercrime. Sex crime, violent crime, juvenile delinquency, and ‘others’ have relatively similar proportions among datasets, suggesting that cybercrime prevalence is not significantly higher or lower in these categories compared to frauds and traffics in the dataset. Notably, incidents classified under ‘others’ account for a substantial proportion of the datasets (65–72%).

The Proportion of Cybercrime in the Datasets

Table 2 shows the proportion of cybercrime incidents in the learning set (45%) and across different crime categories. Two categories, fraud (~95%) and ‘others’ (46 to 47%), have higher proportions of cybercrime than the other five crime types (0.5–29.5%). Sex crime contains about 30% and 20% of cybercrime in the training and testing sets, respectively. Given that sex crime represents only a small fraction of recorded incidents (~0.7%), a 10-percentage point difference can be considered a coincidental consequence of random sampling. There are no statistically significant differences among the ratios of sex cybercrime in learning, training, and testing sets, as determined by the proportion test (p -value=0.58). Cybercrime represents about 5–6% of violent crime and juvenile delinquency incidents, which is notably lower than its proportion in the theft cases (20%). The ratios of cybercrime in both datasets fell within the 95% confidence interval (43.57 to 46.74%), as determined by the proportion test.

As explained above, traffic incidents do not typically include cyber elements.⁴ This is confirmed by the results in Table 2: only 0.5% of the traffic incidents in the training set were labelled as cybercrime. For this reason, the classification accuracy cannot be determined in this category.

⁴ An example of cybercrime found in the traffic field was a dispute during a face-to-face meeting between the seller and customer of a secondhand motorcycle. The cyber dimension of the crime lies in that the description of the item and the sale took place on an online platform.

Classifier Performance

Three classification methods (Naïve Bayes, linear SVM, and kernel SVM) were applied using the aforementioned learning datasets. The training sample size was selected by differentiating sample sizes using learning curves (see details in Appendix Table 6). As shown in Table 3, the kernel SVM classifier was identified as the best classifier, considering accuracy (93.45%), recall (88.35%), precision (96.94%), and F1-score (92.44%). The optimal parameters for the kernel SVM classifier were obtained using the `e1071` package in R (Meyer et al., 2020). The average accuracy of tenfold cross-validation is over 90%, and the best optimised hyper-parameters for gamma and cost are 0.0005 and 1, respectively.

Table 4 presents an overview of the kernel SVM classifier's performance on various sub-crime categories in the testing set. The total accuracy and F1-score of the model are 93.5% and 92.44%, respectively. The recall rate for all crime categories shows that 88.4% of cybercrime cases in the testing set were correctly identified as such by the model. The precision rate of the model indicates that 96.9% of the predicted cybercrime cases were real cybercrime cases in the testing dataset.

While the average performance of the classifier (93.5%) is promising, it is important to examine its performance for individual crime sub-types. With an accuracy over 90% for all crime types except for theft (82%), the model should be suitable for many applications. We can see that it performs best with fraud incidents, with an accuracy, recall, precision, and F1-score of 98.7%, 98.8%, 99.8%, and 99.3%, respectively. The classifier also performs well in identifying cybercrime in the 'others' category, with

Table 3 Confusion matrix for the three classifiers

	Naïve Bayes	Linear SVM	Kernel SVM
Accuracy	88.44%	90.96%	93.45%
95% CI*	87.53–89.32%	90.13–91.74%	92.74–94.12%
<i>p</i> -value	$< 2.2e - 16$	$< 2.2e - 16$	$< 2.2e - 16$
Recall	90.62%	90.77%	88.35%
Precision	84.95%	89.47%	96.94%
F1-score	87.69%	90.12%	92.44%

*Confidence interval of accuracy

Table 4 Performance of the kernel SVM classifier on the crime categories of the testing dataset

Crime categories	Proportions (%)	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
Total	100.0	93.5	88.4	96.9	92.4
Sex crime	0.7	90.3	83.3	71.4	76.9
Violent crime	8.1	96.7	54.6	92.3	68.6
Fraud	14.9	98.7	98.8	99.8	99.3
Juvenile delinquency	0.5	94.4	50.0	100.0	66.7
Theft	3.8	82.2	21.9	70.0	33.3
Others	65.2	93.5	90.7	96.2	93.4
Traffic	6.7	100.0	-	-	-

similarly high scores: 93.5%, 90.7%, 96.2%, and 93.4%. As the ‘others’ field constitutes over two-thirds of the KCPS dataset, the overall performance of the classifier is strongly linked to the performance for this crime type. The excellent results are therefore encouraging, but they should be interpreted with care as we cannot confirm that the ‘others’ cases in the testing set are perfectly representative of the ‘others’ cases in the original dataset.

We found that the performance of the model is good, but not as high for sex crime, violent crime, juvenile delinquency, and theft incidents. For these, the F1-scores are about 77%, 69%, 67%, and 33%, respectively. As shown in Table 4, the proportion of cybercrime is smaller than 20% for these four crime categories in the testing set. Given that it takes into account the trade-offs between false positives and false negatives, the F1-score is a useful benchmark for evaluating the classifier’s performance on imbalanced datasets (Huilgol, 2019).

Proportion of Cybercrime in the KCPS Dataset

This section concludes our demonstration with statistics about the classification results. Table 5 details the proportion of cybercrime incidents in the KCPS dataset. After removing traffic-related incidents⁵ ($N=1022$ k), the classifier labelled about 394 k (4.6%) of the remaining 8.5 million incidents as cybercrime. The greatest proportion of cybercrime in the dataset was found for fraud incidents (76.3%). The lowest proportion of cybercrime concerns juvenile delinquency and violent crime, with less than 1%. Although juveniles are usually considered more tech-savvy, the low percentage of cybercrime in juvenile delinquency is presumed to be due to their having other options to report cyber incidents (e.g. school teachers). Cybercrime cases were slightly more prevalent in the following three categories: theft (5.7%), sex crime (4%), and ‘others’ (3.2%).

Discussion

The purpose of this study was to evaluate the potential value of using supervised classification techniques to classify cybercrime from calls for police service data. In this section, we go back to some of the issues encountered during the development of the dataset and discuss future research directions.

Building a Quality Training Set

When using a supervised model with natural language processing (NLP), feature engineering, in which researchers use their domain knowledge to manually define and extract noteworthy features from the texts, is an essential technique to improve a training dataset (Liu et al., 2023). In other words, text features in the training set should be representative of the population it aims to classify, and mislabelled samples in the dataset must be handled appropriately (Sánchez et al., 2003) as well as maintaining balances between classes (Chawla et al., 2002).

⁵ The traffic field has no cybercrime in the testing set, so recall and precision values for the field are unavailable. Hence, the traffic field was omitted when applying the model to the KCPS dataset.

Table 5 The proportion of cybercrime and non-cybercrime in the KCPS dataset (1 January–30 September 2020)

Crime types	Total ($N=8569$ k, 100%)	Cybercrime ($N=394$ k, 4.6%)	Non-cybercrime ($N=8175$ k, 95.4%)
Sex crime	62,342 (0.7%)	2494 (4.0%)	59,848 (96.0%)
Violent crime	946,632 (11.0%)	7573 (0.8%)	939,059 (99.2%)
Fraud	182,229 (2.1%)	139,041 (76.3%)	43,188 (23.7%)
Juvenile delinquency	60,424 (0.7%)	363 (0.6%)	60,061 (99.4%)
Theft	443,105 (5.2%)	25,257 (5.7%)	417,848 (94.3%)
Others	6,874,830 (80.2%)	219,995 (3.2%)	6,654,835 (96.8%)

In this context, the current study demonstrated how minority oversampling can be applied to police-recorded textual big data. From the perspective of classifying cybercrime, the KCPS dataset is highly imbalanced. Hence, this study sought to balance classes by using the CRT and the minority oversampling technique. This phase increased the likelihood of selecting cybercrime incidents when sampling the learning dataset from the two partitioned datasets. Although the dataset with the CRT ($N=13,000$) is randomly sampled more than the other dataset ($N=10,000$), the proportion of cybercrime and non-cybercrime in the learning dataset is 45:55. This fact implies that relying solely on keyword-query request classification is not suitable for document classification because many incidents with keywords are not classified as cybercrime cases in the learning dataset. Moreover, as shown in Table 1, the proportions of the crime categories can be varied after applying the minority oversampling technique to the dataset. This result indicates that researchers should consider the potential changes in crime categories' proportions when sampling the dataset to build a learning set.

In the second labelling phase for majority voting, 123 cases were discarded from the learning set. This happened when two coders labelled the case as cybercrime, while the other two identified it as non-cybercrime. Lack of consensus can be due to the coders' experience, level of practice, risk-taking tendencies, and the clarity of guidelines (Demirhan & Yilmaz, 2023). Additionally, some cases are presumed to be inherently ambiguous. Furthermore, individuals have different levels of comprehension when it comes to interpreting information, inferring omitted contents (e.g. motives), and envisioning the actions that would have taken place based on the narratives (Borrior et al., 2017). Given the amount of text information in the KCPS dataset, not every crime event can be described perfectly. Although some cases contain sufficient information, others do not. Each text record in the KCPS dataset represents a crime event with narratives according to the time flow. Consequently, crime-type information may be occasionally omitted while recording the text. Therefore, despite efforts to prevent mislabelling of samples in the training set, it is necessary to exclude samples on which coders do not agree regarding the labels to create a high-quality training dataset.

Machine Learning Model's Performance

The models demonstrated very good overall performance (Table 3). This is partly due to the size and quality of the learning set. First, the labelling procedure involved 16 police officers (with considerable professional experience) independently labelling each case in

the learning set. Cases without consensus were labelled using the principle of majority voting. This labelling process contributes to reducing biases in the learning set. Second, the classes used in this study are mutually exclusive and form a *complete set*: cybercrime and non-cybercrime. Most models in the previous studies did not contain the ‘non-cybercrime’ or ‘others’ class in their learning dataset, which may have resulted in lower performance.

We found that the kernel SVM model has the highest performance among the three classifiers (F1-score, 92.45%). However, the model performs differently in the seven crime categories. This empirical finding demonstrates that although a classifier performs very well overall, it may perform poorly for specific sub-categories. In turn, when we try to improve the quality of the learning dataset, we should consider sampling more incidents from those sub-categories with a low F1-score. Here, the final model may not be useful to identify cyber incidents in the theft and traffic crimes categories (F1-scores, 33.3 and N/A). However, we noted that the classifier performs well for the ‘others’ category, which contains around 80% of incidents in the KCPS dataset.

Previous studies that examined machine learning classifiers for crime datasets (Hal-eem et al., 2019; Karystianis et al., 2018; Matos et al., 2022; van der Laan & Tollenaar, 2021) were also primarily focused on methodological issues. While these studies were useful in explaining how to classify incidents from textual data, the performance of the classifiers was limited to around 60 to 80%.

Assessing Accuracy Acceptability of the Model

In addition to evaluating the accuracy of the machine learning model, we need to consider the ‘acceptability of accuracy’ based on the intended application of the model (Germini et al. 2022). Kay et al. (2015) proposed that we can assess accuracy acceptability by examining the perceived usefulness and sensitivity for the intended application.

The function of the model is to classify cybercrime from the KCPS dataset, for the purpose of analysing cybercrime. The data extracted by the model can be used to measure the scope, trends, police workloads, and patterns of cybercrime. Errors in classifying individual incidents are not as problematic as in other sectors, such as cancer detection. However, the model’s error rate (~6%) needs to be assessed in light of the specific applications for which it is being used.

Limitations

This study has several limitations. The first limitation is that it only analyses a single dataset (the KCPS dataset) which is limited to police-recorded cybercrime incidents in South Korea. Therefore, we cannot be sure the extent to which the findings can be generalised to other regions and other datasets. Future studies can expand the analysis to other datasets to confirm the effectiveness of the proposed classification method. Second, since the CPS data presents the results of calls and first responses by the police, without involving a criminal investigation, we cannot ascertain whether the cases are related to cross-border cybercrime with the CPS data alone. Consequently, the results cannot be used to draw conclusions about the global characteristics of cybercrime. For this, the study should be replicated using databases from other countries. Third, in the text data pre-processing phase, misspelt terms (Appendix Table 7) related to cybercrime were corrected. Since the bag-of-words method was applied in this study, identifying more commonly misspelt

The

set

ana

Table 7 The representative list of misspelled words with correct words

No	Misspelled Korean words	Correct words	Meaning in English
1	몰래 카메라, 몰래 카메라	몰카	Voyeur camera
2	결제	결제	Payment
3	인스타그램, 인스타그램, instagram, insta	인스타	Instagram
4	카톡, 카톡, 카카오톡, kakaotalk	카카오톡	KakaoTalk (WhatsApp)
5	메세지, 메시지, 매세지, 매세지, 메세지, 멧세지	메시지	Message
6	매신저, 매신저, 매신저, 매신저, 매신저	메신저	Messenger
7	에스엔에스, 에스엔에스	sns	SNS
8	왓츠앱, 왓츠앱, 왓츠앱, 왓츠앱, 왓츠앱	왓츠앱	WhatsApp
9	앱, 어플리케이션, app, application	어플	Application
10	휴대전화	휴대폰	Mobile phone
11	스마트폰, 스마트폰, 스마트폰	스마트폰	Smartphone
12	데이트폭력, 데이팅폭력, 데이팅폭력	데이트폭력	Date violence
13	네이버, naver	네이버	NAVER (Korean Google)
14	해킹, 해킹, 해킹, 해킹, 해킹, hacking	해킹	Hacking
15	계좌	계좌	Bank account
16	몸캠, 몸캠, 몸캠, 몸캠	몸캠	Body camera
17	마사지, 마사지, 마사지	마사지	Massage
18	관제센터	관제센터	Monitoring centre
19	씨씨티비, 씨시티비, 시시티비	cctv	CCTV
20	트위터, 트위터, 트윗터, twitter	트위터	Twitter
21	텔레그램, 텔레그램, 텔레그램	텔레그램	Telegram
22	유튜브, 유튜브, 유튜브, 유튜브	youtube	YouTube

Table 7 (continued)

No	Misspelled Korean words	Correct words	Meaning in English
23	악성프로그래, 멀웨어, 악성코드	멀웨어	Malicious ware
24	태블릿, 태브릿, 타블릿, 타블렛	테블릿	Tablet
25	랜섬웨어, 랜섬웨어, 랜섬웨어	랜섬웨어	Ransomware
26	스팸	스팸	Spam
27	인터넷	인터넷	Internet
28	채팅, 채팅, 채팅	채팅	Chatting
29	사이버	사이버	Cyber
30	보이스	보이스	Voice

Table 8 Cybercrime-related terms list in Korean calls for police service dataset

No	Terms	Meaning	No	Terms	Meaning	No	Terms	Meaning
1	계좌이체	Bank transfer	31	상품권	Gift card	61	악성코드	Malicious code
2	금융	Finance	32	스미싱	Smishing	62	이메일	Email
3	녹화	Recording	33	세이클럽	Sayclub	63	이베이	eBay shopping
4	동영상	Video	34	인스타그램	Instagram	64	통신판매	Mail order
5	메시지	Message	35	유투브	Youtube	65	저작권	Copyright
6	모바일	Mobile	36	아이디	ID	66	톡톡	Tok-tok
7	사이버	Cyber	37	음란영상	Obscene video	67	파밍	Pharming
8	사이트	Site	38	종고나라	Used goods market	68	코인	Coin
9	사칭	Impersonating	39	종고장터	Used goods market	69	가상화폐	Virtual assets
10	스마트	Smart	40	조건만남	Compensated dating	70	사기문자	Fraud text
11	신용카드	Credit card	41	티머니	T-money	71	디씨인이드	DC Inside
12	어플	Application	42	페이스북	Facebook	72	다음	Daum
13	영상	Video	43	피싱	Phishing	73	커뮤니티	Community
14	인터넷	Internet	44	스토킹	Stalking	74	말웨어	Malware
15	체크카드	Check card	45	트위터	Twitter	75	랜섬웨어	Ransomware
16	카드결제	Card payment	46	프로그램	Program	76	메일	Mail
17	카메라	Camera	47	번개장터	Used goods market	77	협박전화	Threatening call
18	카카오톡	KakaoTalk	48	도둑	(Identity) theft	78	협박	Threatening
19	컴퓨터	Computer	49	알몸	Naked	79	바이러스	Virus
20	텔레비전	TV	50	개인정보	Personal information	80	유출	Leak
21	통장	Bank book	51	드론	Drone	81	불법복제	Illegal copy
22	핫라인	Hot line	52	폰팅	Phone date	82	대포폰	Burner phone

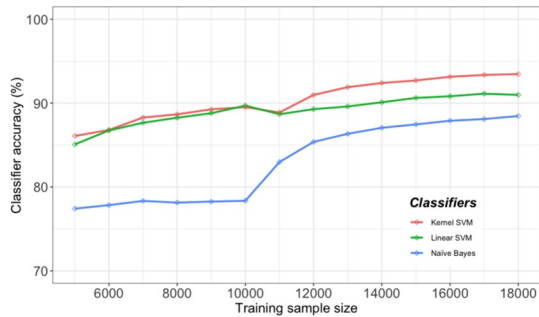
Table 8 (continued)

No	Terms	Meaning	No	Terms	Meaning	No	Terms	Meaning
23	해킹	Hacking	53	유포	Circulated	83	아이템	Item
24	구글	Google	54	도장	Wiretapping	84	소액결제	Micropayment
25	네이버	Naver	55	포르노	Porn	85	메신저	Messenger
26	네트워크	Network	56	야동	Porn	86	네이트	Nate
27	몸캠	Body camera	57	채팅	Chatting	87	네이트 스팸	Spam
28	몰카	Voyeur camera	58	댓글	Reply comment	88	카페	café
29	밴드	Band	59	지격	Sniping			
30	블로그	Blog	60	악플	Bad comments			

Table 9 The main guidelines provided to coders for the labelling of crime cases

No	Cybercrime class	Examples
1	The call should be related to cybercrime, but guilty or innocent is irrelevant, as a call reporting log is not a criminal court sentence	<p>X Asking websites address</p> <p>✓ Reporting voice phishing number, but no damage</p>
2	When the target of crime is an IT/digital device, then label the case as cybercrime	<p>✓ Stealing a smartphone, laptop</p> <p>✓ Deleting computer files</p>
3	When the crime commission process has a cyber aspect (IT, IoT, ICT, online, digital, etc.), then label the case as cybercrime	<p>✓ A man called a woman with Facebook messenger and attacked her</p> <p>✓ Reporting online used goods market fraud</p>
4	When cyber-related activities are the primary cause of a crime (e.g. provocation), then label it as cybercrime	<p>✓ Domestic violence: parent hit a child due to excessive computer games usage</p> <p>✓ A husband assaulted his wife after finding evidence of her cheating on her smartphone</p>
5	When the only role of cyber technology is to protect the victim (i.e. guardianship), then label the case as non-cybercrime. When criminals use it for their protection, label it as cybercrime (e.g. utilising IT tech to commit or facilitate illegal activities)	<p>X Reporting evidence of a crime scene with CCTV recording or digital pictures</p> <p>X A smartwatch, which was provided to a victim of a past crime, is now sending a signal to the police (unless the result of the call is about cybercrime)</p> <p>✓ An illegal brothel installed CCTVs around their property and monitored them online to evade law enforcement, gain advance knowledge of police patrols, or identify potential illegal clients</p>

Fig. 2 Classifier accuracy as a function of training sample size



Learning curves were used to compare the performance of three models, including Naïve Bayes, linear SVM, and kernel SVM, with different training sample sizes. The learning curve is an efficient method to monitor a subject's performance in repetitive tasks (Anzanello & Fogliatto, 2011). The learning curves of the classifiers' accuracy by different training sample sizes are illustrated in Fig. 2. The Y-axis (%) indicates the accuracy of the classifiers, and the X-axis refers to the training sample size.

To sum up, the accuracy of classifiers increases as the training sample size increases. The kernel SVM classifier shows a more accurate performance than other models. When the sample size is 18,000, the kernel SVM model performs best (94%). The accuracy levels for linear SVM and Naïve Bayes models are about 91% and 88%, respectively. In the end, the performance of each classifier was estimated for a training sample size of 18,000 and a testing sample size of 4877.

Acknowledgements We are very grateful to Dr. Enrico Mariconti for his valuable comments.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Action Fraud. (2021a). Cyber Crime Trends 2020–2021. National Fraud Intelligence Bureau. Available online at <https://data.actionfraud.police.uk/cms/wp-content/uploads/2021/07/CYBER-Dashboard-Assessment-20-21.pdf>
- Action Fraud. (2021b). Fraud Crime Trends 2020–2021. National Fraud Intelligence Bureau. Available online at <https://data.actionfraud.police.uk/cms/wp-content/uploads/2021/07/2020-21-Annual-Assessment-Fraud-Crime-Trends.pdf>
- Aizawa, A. (2003). An information-theoretic perspective of TF–IDF measures. *Information Processing & Management*, 39(1), 45–65.
- Anzanello, M. J., & Fogliatto, F. S. (2011). Learning curve models and applications: Literature review and research directions. *International Journal of Industrial Ergonomics*, 41(5), 573–583.

- Basit, T. (2003). Manual or electronic? The role of coding in qualitative data analysis. *Educational Research*, 45(2), 143–154.
- Borrion, H., Dehghanniri, H., & Li, Y. (2017). Comparative analysis of crime scripts: One CCTV footage—twenty-one scripts. *2017 European Intelligence and Security Informatics Conference (EISIC)* (pp. 115–122). IEEE.
- Boulton, L., McManus, M., Metcalfe, L., Brian, D., & Dawson, I. (2017). Calls for police service: Understanding the demand profile and the UK police response. *The Police Journal*, 90(1), 70–85.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Demirhan, H., & Yilmaz, A. E. (2023). Detection of grey zones in inter-rater agreement studies. *BMC Medical Research Methodology*, 23(1), 1–15.
- DeVoe, J., & Murphy, C. (2011). *Student reports of bullying and cyber-bullying: Results from the 2009 School Crime Supplement to the National Crime Victimization Survey*. Web Tables (NCES 2011–336). National Center for Education Statistics. U.S. Department of Education. Retrieved from <https://nces.ed.gov/pubs2011/2011336.pdf>
- Domenie, M. M. L., Leukfeldt, E. R., Toutenhoofd-Visser en, M. H., Stol, W. Ph. (2009). *Werkaanbod cyber crime bij de politie. Een verkennend onderzoek naar de omvang van het geregistreerde werkaanbod cyber crime*. [Registered cyber crimes by the police in the Netherlands] Leeuwarden: NHL.
- Felson, M., & Boba, R. (2010). *Everyday technology and everyday crime* (pp. 203–206). SAGE Publications, Inc. <https://doi.org/10.4135/9781483349299.n11>
- Fissel, E. R., & Reyns, B. W. (2020). The aftermath of cyberstalking: School, work, social, and health costs of victimization. *American Journal of Criminal Justice*, 45(1), 70–87.
- Furnell, S., & Dowling, S. (2019). Cyber crime: a portrait of the landscape. *Journal of Criminological Research, Policy and Practice*, 5(1), 13–26.
- Germini, F., Noronha, N., Debono, V. B., Philip, B. A., Pete, D., Navarro, T., ... & Iorio, A. (2022). Accuracy and acceptability of wrist-wearable activity-tracking devices: Systematic review of the literature. *Journal of Medical Internet Research*, 24(1), e30791
- Haleem, M. S., Han, L., Harding, P. J., & Ellison, M. (2019). An automated text mining approach for classifying mental-ill health incidents from police incident logs for data-driven intelligence. In 2019 IEEE international conference on systems, man and cybernetics (SMC) (pp. 2279–2284). IEEE.
- Han, C. H., & Palmer, M. (2004). A morphological tagger for Korean: Statistical tagging combined with corpus-based morphological rule application. *Machine Translation*, 18(4), 275–297
- Holt, T. J., & Bossler, A. M. (2015). *Cybercrime in progress*. Routledge.
- Huigol, P. (2019). Accuracy vs. F1-score. Available online at <https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2> Accessed on 4th May 2022.
- IC3. (2021). Federal Bureau of Investigation Internet Crime Report 2021. Internet Crime Complaint Center. Available online at https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3Report.pdf
- ICD. (2017). Mental, Behavioural and neurodevelopmental disorders F01-F99, International Classification of Diseases. Available online at <https://www.icd10data.com/ICD10CM/Codes/F01-F99>
- Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, 52(1), 273–292.
- Karystianis, G., Adily, A., Schofield, P., Knight, L., Galdon, C., Greenberg, D., ... & Butler, T. (2018). Automatic extraction of mental health disorders from domestic violence police narratives: Text mining study. *Journal of medical internet research*, 20(9), e11548.
- Kay, M., Patel, S. N., & Kientz, J. A. (2015). How good is 85%? A survey tool to connect classifier evaluation to acceptability of accuracy. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 347–356).
- KNPA. (2020). Total cybercrime cases reported to KNPA. Korean National Police Agency. Available online at <https://www.police.go.kr/www/open/public/public0204.jsp>
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35.
- Matos, H., Souza, S., Santos, R., Costa, J. W., & Costa, C. (2022, November). A supervised classifier for police reports at the state of Pará, Brazil. In *Anais da II Escola Regional de Alto Desempenho Norte 2 e II Escola Regional de Aprendizado de Máquina e Inteligência Artificial Norte 2* (pp. 21–24). SBC.
- Mazowita, B., & Vézina, M. (2014). Police-reported cybercrime in Canada, 2012. *Juristat*, 34(1), 1–24. Available online at <http://www.statcan.gc.ca/pub/85-002-x/2014001/article/14093-eng.pdf>
- McGuire, M., & Dowling, S. (2013). *Cybercrime: A review of the evidence*. Research report 75. Home Office, London.
- McGuire, M. (2019). It ain't what it is, it's the way that they do it? Why we still don't understand cybercrime. *The human factor of cybercrime* (pp. 3–28). Routledge.

- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., & Lin, C. C. (2020). *e1071: misc functions of the department of statistics, probability theory group*. TU Wien. R package version 1.7-4.
- Montoya, L., Junger, M., & Hartel, P. (2013). How “digital” is traditional crime? *2013 European Intelligence and Security Informatics Conference* (pp. 31–37). IEEE.
- Namyoun. (2020). NLP4kec: NLP for Korean, English, Chinese. Namyoun. R package version 1.4.0. Available online at <https://github.com/NamyounKim/NLP4kec>
- Näsi, M., Oksanen, A., Keipi, T., & Räsänen, P. (2015). Cybercrime victimization among young people: A multi-nation study. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 16(2), 203–210.
- Popham, J., McCluskey, M., Ouellet, M., & Gallupe, O. (2020). Exploring police-reported cybercrime in Canada: Variation and correlates. *Policing: An International Journal*, 43(1), 35–48.
- Reep-van den Bergh, C. M., & Junger, M. (2018). Victims of cybercrime in Europe: A review of victim surveys. *Crime Science*, 7(1), 1–15.
- Sagayam, R., Srinivasan, S., & Roshni, S. (2012). A survey of text mining: Retrieval, extraction and indexing techniques. *International Journal of Computational Engineering Research*, 2(5), 1443–1446.
- Sánchez, J. S., Barandela, R., Marqués, A. I., Alejo, R., & Badenas, J. (2003). Analysis of new techniques to obtain quality training sets. *Pattern Recognition Letters*, 24(7), 1015–1022.
- Scott, S., & Matwin, S. (1999). Feature engineering for text classification. In 16th International Conference on Machine Learning, 99, 379–388. Available online at <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=781a61f26f806208bcbefa7bd733d4b3d9804bc5>
- Sheng, V. S., Zhang, J., Gu, B., & Wu, X. (2017). Majority voting and pairing with multiple noisy labelling. *IEEE Transactions on Knowledge and Data Engineering*, 31(7), 1355–1368.
- Van der Laan, A. M., & Tollenaar, N. (2021). Text mining for cybercrime in registrations of the Dutch police. *Cybercrime in context: The human factor in victimization, offending, and policing* (pp. 327–350). Springer International Publishing.
- Whitty, M. T., & Buchanan, T. (2012). The online romance scam: A serious cybercrime. *Cyber Psychology, Behavior, and Social Networking*, 15(3), 181–183.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.