Inteligencia Artificial 25(69), 57-86
doi: 10.4114/intartif.vol25iss69pp57-86

INTELIGENCIA ARTIFICIAL
http://journal.iberamia.org/

Feature extractions and selection of bot detection on Twitter: A systematic literature review

Raad Ghazi Al-Azawi[1,A], Safaa O. AL-mamory[2]

[1] Software Department, College of Information Technology, University of Babylon, Babylon, Iraq.

[A] raad.alazawi@student.uobabylon.edu.iq

[2] College of Business Informatics, University of Information Technology and Communications, Bagdad, Iraq.

Abstract Automated or semiautomated computer programs that imitate humans and/or human behavior in online social networks are known as social bots. Users can be attacked by social bots to achieve several hidden aims, such as spreading information or influencing targets. While researchers develop a variety of methods to detect social media bot accounts, attackers adapt their bots to avoid detection. This field necessitates ongoing growth, particularly in the areas of feature selection and extraction. This research aims at providing an overview of bot attacks on Twitter, shedding light on issues in feature extraction and selection that have a significant impact on the accuracy of bot detection algorithms, and highlighting the weaknesses in training time and dimensionality reduction. To the best of our knowledge, this study is the first systematic literature review based on a preset search strategy that encompasses literature published between 2018 and 2021 which are concerned with Twitter features (attributes). The key findings of this research are threefold. First, the paper provides an improved taxonomy of feature extraction and selection approaches. Second, it includes a comprehensive overview of approaches for detecting bots in the Twitter platform, particularly machine learning techniques. The percentage was calculated using the proposed taxonomy, with metadata, tweet text, and merging (meta and tweet text) accounting for 37%, 31%, and 32%, respectively. Third, some gaps are also highlighted for further research. The first is that public datasets are not precise or suitable in size. Second, the use of integrated systems and real-time detection is uncommon. Third, detecting each bots category identified separately is needed, rather than detecting all categories of bots using one generic model and the same features' values. Finally, extracting influential features that assist machine learning algorithms in detecting Twitter bots with high accuracy is critical, especially if the type of bot is pre-determined.

Keywords: Feature selection, Feature extraction, Social Media, Social Bot, Twitter, Machine Learning.

1 Introduction

In society, there has always been misinformation. Nowadays, technological advancements and the proliferation of social networks, phony newspapers, and blogs have exacerbated the problem by making it easier for malicious news to propagate quickly. This fact makes it easier to use disinformation as a vector of attack against large communities. As a result, procedures for detecting the appearance of this type of news and mitigating its impact

ISSN: 1137-3601 (print), 1988-3064 (on-line)

have been developed. As the Internet has progressed, social media has emerged as one of the primary avenues of dis seminating personal, political, and other information[1].

Social media sites such as Twitter have been gradually evolved to be the most interesting platforms for expressing users' ideas and opinions on a variety of topics. Many businesses are drawn to this data, particularly to study people's thoughts and opinions on a variety of topics such as political events, social events, movies, songs, product reviews, and so on. While legitimate uses for social media exist, many influence seekers and harmful groups utilize it further to their hidden objectives. People and businesses need to establish an impact on societal media to take advantage of its potential, which is why social media bots (SMBs) were created. SMBs are computer algorithms that create content and engage with users on social media platforms[2].

Bots are responsible for a sizable portion of online activity. According to Twitter, bots account for approximately 8.5% of all Twitter users[3]. According to a study on social bots, 9% to 15% of all English-speaking active Twitter users exhibit bot-like behaviors[2]. SMBs have either beneficial, neutral, or harmful intentions[4]. Bots that send out earthquake alerts tweets automatically are instances of benign bots, as are chat bots that interact with users and carry out their needs[5], whereas news bots automatically disseminate articles from news agencies. Bots that post or repost jokes and nonsense are an example of neutral bots[6].

Malicious bots are the most researched category in social media, in which new types are constantly being discovered [7]. Malicious SMBs are typically controlled by a botmaster, who is the human in command of the bots and oversees their assault and actions. A variety of malicious bots are spambots that distribute malicious links and illegal messages [8]. Cashtag piggybacking bots promote low-value which shares by obtaining the benefit of the popularity of elevated items [7], whereas Astroturfing bot creates the appearance of significant assistance for a politician or point of view [9]. The Sybils' pseudonymous are examples of user accounts [10]. Moreover, fake accounts that share posts consist of encrypted commands for a botnet attack [11]. Paybots generate money by stealing content from reputable sources and using it to drive visitors to the site [3]. Social botnets are used in political disagreements [12]. Furthermore, bots of an organization's penetration act apparently like friends [13]. Finally, cyberbullying bots entail the deliberate and aggressive use of information and communication technology by an individual or group with the intent to harm others [14,15].

Therefore, detecting SMBs and secure podium and rightful users against bots is necessary. This initiative is further needed in the Twitter platform since most of its members are celebrities, politicians, and important companies. Besides, the open structure of Twitter has attracted a large number of bots. Detecting bot accounts, whether personally or in groups, from their early phases (account creation) or after they merge into social media, is the most common protection method in the research community.

To eliminate the spread of harmful SMBs, researchers used a variety of detection techniques. Machine learning algorithms are considered one of the most important methods for bot detection because of their, ease, speed of computing time, and ability in handling a large amount of data [16]. The majority of available machine learning techniques use supervised learning algorithms, in which the model is trained with labeled data. However, instead of assessing user social behavior, these approaches focus on statistical attributes (features), the usefulness of the features set, and the training set's efficiency [17]. On the other hand, other algorithms find their way to cluster input data using unsupervised machine learning [18]. This approach does not require labeled data to detect bots and does not rely on the values of specific features to classify each account, so it is based on what is common among groups of accounts because it employs partially labeled data. The other direction is semi-supervised machine learning, which is between supervised and unsupervised techniques. This method uses a large amount of unlabeled data and a small fraction of labeled data to develop classifiers, which can diminish the worth of collecting labeled examples while increasing classification accuracy [19]. Semi-supervised machine learning is an important topic, although the number of publications that use this method is not broad.

This article aims at providing a synopsis of a bot that attacks the Twitter platform, focusing on issues in feature extraction and selection, and highlighting the weaknesses in training time and dimensionality reduction. The study focuses on approaches that extract and select the important features that support machine learning algorithms for bots detection. It is believed that this systematic review can help researchers, particularly those who are unfamiliar with Twitter bots in obtaining enough background on bot detection. It can also help select the important features that have a high impact on bot detection with the ability of SMBs that rapidly evolve, resulting in a modification in the values of the distinguished features. Based on this systematic review, the following contributions are drawn:

1) Proposing a new taxonomy that may lead to the development of novel feature extraction approaches or selection strategies.
2) Including machine learning strategies, where various distribution results, statistics, and the most important classifiers used are provided.

3) Twitter datasets used in previous literature are classified into bots detection, sentiment analysis, and others.
4) Challenges, recommendations, and possible solutions are also provided to ensure that Twitter bots detection techniques are robust.
5) The study presents important features used in previous studies for bot detection on the Twitter platform.

The rest of this work is organized as follows. Section 2 deals with the extraction and selection of Twitter features and provides general background about this area. Section 3 determines the methodology of conducting this systematic review. Section 4 includes an improved taxonomy as well as a discussion of techniques that fall into each category, as well as the key findings of a literature review and evaluation. In section 5, we discuss the common challenges, motivation, and recommendations for future studies. Section 6 shows Twitter datasets labeled by usage that help to understand the demeanor of Twitter bots in comparison to human demeanor. Section 7 discusses the method's effectiveness in bot detection and explained the most used machine learning techniques. Section 8 discusses the Measures of Performance used to evaluate social bot detection. Section 9 highlights common features that are used and provides a description of the vital features' categories that were used in previous literature for bot detection. Section 10 discusses the current challenges and gaps in features selection, extraction, and Twitter bots detection. Section 11 concludes some remarks and suggests future research topics.

2 Features eextraction and selection on Twitter

Twitter is one of the most popular social media sites in the world and this, in turn, leads to recommending a wide variety of efficient feature selection algorithms that can successfully reduce the original data into a low-dimensional space. As such, when attempting to differentiate bot accounts from actual human accounts on Twitter, the questions are: What distinguishes a bot account from a human account, and how are they different? One feature is used by some researchers to recognize bot accounts such as screen names to identify bot accounts [20] and posts' locations [21]. However, because overcoming one vulnerability is not an insurmountable effort, bot-masters can easily design bots that can resist detection by such models. Many researchers relied on a predefined list of attributes along with a labeled set of accounts provided to the computer to tackle this problem. The machine's job is to determine thresholds for feature values that aid in determining whether an account is a bot or not as well as estimating an account's botnets. This is the most common scenario for supervised machine learning approaches. Unsupervised machine learning uses exploited properties as comparison criteria. It computes the degree of similarity between a group of social media users by values of a set of predetermined parameters. The majority of the exploited features may be grouped into four primary categories, namely user profile information, post content, posting behavior, and network structure.

Choosing features to recognize SMBs is important, especially when it comes to machine learning methods. The fast-evolving abilities of SMBs, which cause changes in the values of the distinguished features and failure to identify them, is one of the primary issues in this research area. Thus, identifying robust traits is a hot topic. The question is how many characteristics subsets should be picked to aid machine learning algorithms in detecting hostile bots on Twitter. High-dimensional data has been extended in various domains, including social media. The presence of high dimensional data has several drawbacks such as computational cost, overfitting, and poor overall performance [22]. Filtering away unnecessary and redundant features can help reduce overfitting, save computation time, and improve type correctness.

3 Research methodology

This study was designed based on the guidelines of preferred reporting items for systematic reviews and meta-analyses, as shown in Fig. 1. It is recommended to avoid depending on searching a single database for the literature because no single database contains all relevant references [23]. Previous research [24–26] advises that to cover the bulk of publications, a complete systematic evaluation should be undertaken on many databases. Here, four major digital databases have been chosen and searched to improve the chances of obtaining the best search results, namely (1) Science Direct (SD), which offers access to a variety of journals from various scientific domains; (2) IEEE Xplore digital library, which offers various engineering- and technology-related publications; (3) Taylor & Francis, which offers access to various articles from various domains; and (4) Google Scholar (GS), which offers a simple way to search for scholarly literature.

3.1 Inclusion criteria

Inclusion criteria are generally based on various approaches for identifying relevant material for a review study,

such as searching databases and search engines like Google, study design, and date [27]. The inclusion criteria for the chosen topic are as follows:

1) Survey papers on techniques of feature selection, feature extraction, and bots detection.
2) Papers with experimental consequences.
3) Papers written in the English language and submitted to journals or conferences.
4) Papers that include studies in the area of bots detection on the Twitter platform, and some papers in the public social platforms for its so importance in features selection or extraction methods.
5) The publication date is from 2018 to 2021.

The generally agreed preferred reporting items for systematic review and meta-analysis criteria must be adopted by any SLR. (PRISMA) [27]. The PRISMA flow diagram template is depicted in figure 1.

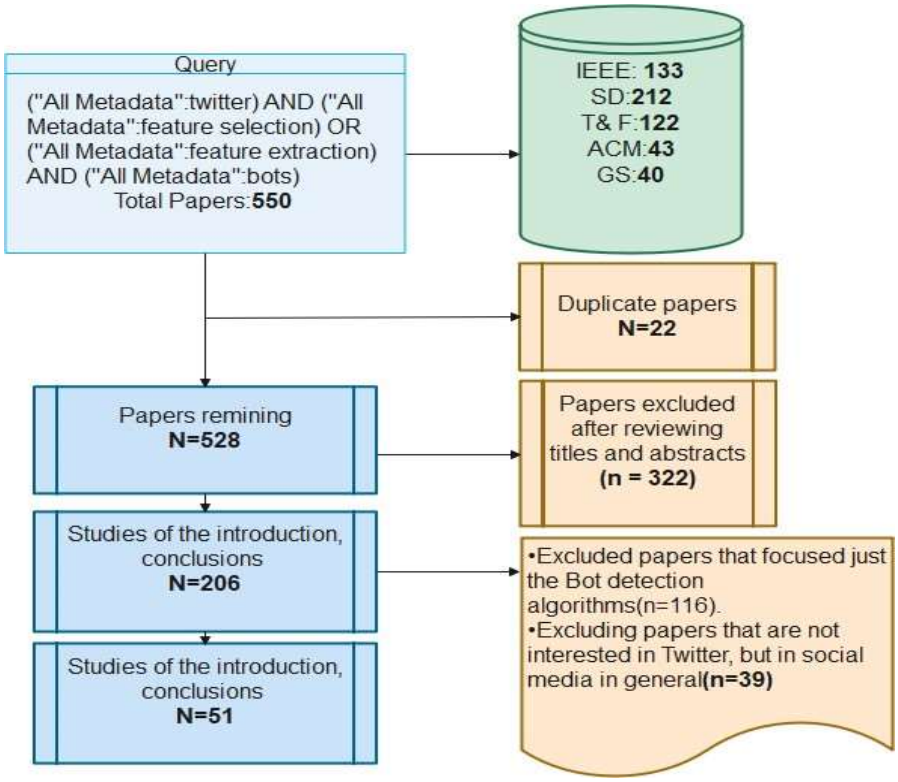

Figure. 1 PRISMA flow diagram of studies' screening and selection.

3.2 Exclusion Criteria

The key criteria this research adopts to exclude irrelevant studies are:

1) Excluding research that offers a detection system or technology, but not for social media bots or any subtype thereof.
2) Excluding articles with no explicit publication information.
3) Excluding studies that focused just on the bot detection algorithms, but not on features selection or extraction.
4) Excluding papers that are not on Twitter.

3.3 Search strategy

Based on the study's aims and research questions, the following search methodologies were used. There are three steps to the search, namely term-based search, crawling-based search, and applying inclusion/exclusion criteria.

Under a term-based search, the search terms are ("All Metadata":twitter), ("All Metadata":feature selection), ("All Metadata":feature extraction), and ("All Metadata":bots). Through a crawling-based search, the literature reviewed was searched based on past research in this area. Finally, inclusion/exclusion criteria are used to ensure that the survey only includes relative works.

Subsequently, all records are accrued into one Endnote library so that duplicates are deleted. All references that have (1) the equal name and writer, and are published within the identical year and (2) the equal name and writer, and published within the identical journal, are deleted. A final set of references was exported to an excel document with vital information for screening. This includes the authors' names, publication year, journal or conference name, DOI, URL link, and summary.

4 Results

About 550 papers from the four databases were retrieved in the initial research selection phase. Following the conclusion of the duplicate screening, a total of twenty-two papers were deleted, leaving 528 publications. The title and abstract scanning were done in the second round of screening, yielding a total of 206 articles. The next step was reading the entire articles. Based on our criteria, 51 articles were reviewed and determined to be relevant to the study, with 37 articles relevant in the Twitter dataset (see Fig. 2). Hence, the goal similarity of selected articles was used to categorize them.

1) Metadata features are the first significant category with (n = 26) articles in which two subcategories were determined.
2) Under the second major category, tweet text features with a total of (n = 22) articles were provided which includes five sub-categories.
3) The metadata contains two subcategories. These articles were classified and put into a coherent taxonomy based on the observed pattern (see figure 2).
4) Fifteen articles were not included in the analysis of figure 2 because they are not about Twitter datasets.

A total of 37 studies are related to Twitter, but the classification in Fig 2 shows that more than 37 studies exist because some studies are shared by more than one category according to the method used.

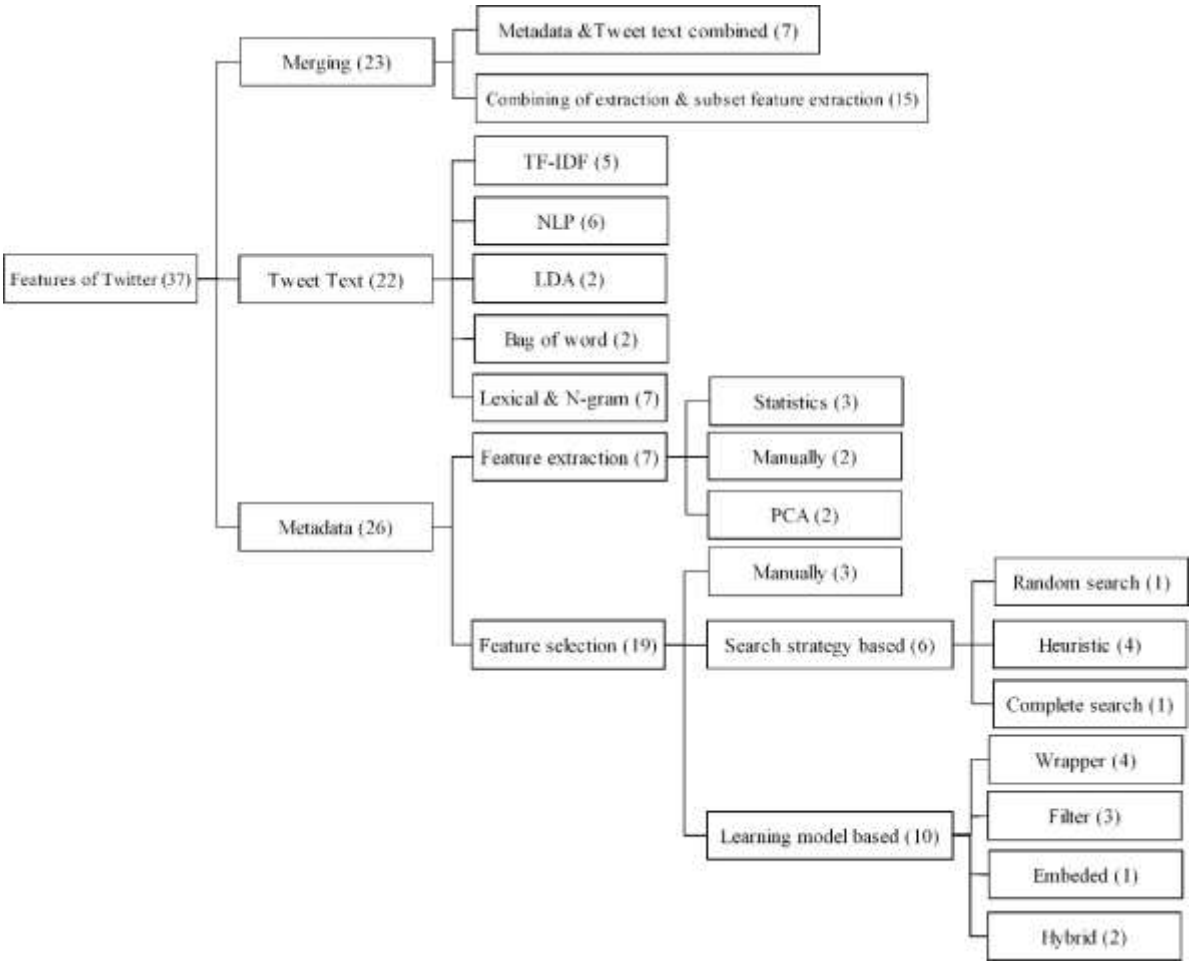

Figure 2. A proposed Taxonomy of feature selection and extraction methods

4.1 Metadata

Twitter metadata explains several events linked to a tweet, such as the time and the location where it was sent. Several metadata-related aspects that are often employed in Twitter-based apps are discussed. The metadata category includes a total of (n = 26/37) articles. Fig 2 shows Twitter's metadata characteristics as well as feature selection and extraction. The following summary divides current research on Twitter's metadata features into three categories.

4.1.1 The Features Extraction (FE) category

Feature extraction (FE) is a technique for reducing dimensionality and improving learning accuracy. It includes two types of algorithms namely, linear and nonlinear techniques. However, the ideal feature extraction-based dimensionality reduction methods are principal component analysis (PCA) [28,29], statistic methods [11,30,31], linear discriminant analysis (LDA) [32], and manually [18,19]. Here, seven out of twenty-six articles that include four parameters are discussed.

Detecting fraudulent accounts in online social networks and creating a model that can precisely describe fake profiles was performed using supervised machine learning techniques and an improved Support Vector Machine (SVM) [28]. The developed method yielded 90% of accuracy in comparison to the Support Vector Machine and Nave Bayes (NB) which achieved 77.4% and 77.3% respectively. A constructed robust features' set was used to detect Twitter spammers in a mix of linear regression and PCA [29]. The performance of the newly constructed features' set revealed an increase in the detection rate and accuracy as well as a low false-positive rate.

A multi-objective hybrid strategy was utilized to discover the most effective feature set for detecting fake accounts on Twitter [11]. Other features were extracted using the typical statistical criteria (entropy and standard deviation). Experiments on two Twitter datasets showed that the proposed approach can reach an accuracy of 98%, 97.6%, and 98% for the random forest, nave Bayes, and SVM techniques respectively. In another research study [30], an ensemble-learning-based approach was conducted to verify the trustworthiness of a large number of tweets by analyzing a huge collection of tweets, particularly for COVID-19-related information. The proposed method divided the data into two categories, namely credible and noncredible. Tweet credibility classifications are based on a variety of factors, including tweet- and user-level features that combine 26 hand-crafted and generic features such as following rate (i.e., followings or account age +1). Several tests were carried out on the labeled and collected dataset. The results showed that the suggested framework was quite good at recognizing credible and noncredible tweets that contain COVID-19 information. In [31], whatever approach was used to propagate false news is ineffective if no one was prepared to believe it. The study analyzed Twitter online accounts with a high number of bots among their friends and called credulous users. Therefore, various characteristics, such as the number of tweets, friends, and followers, which can be easily derived from an account's profile, were statistically significant in distinguishing credulous (C) and non-credulous (NC) individuals. Furthermore, this study proved that using two statistical tests in C users amplifies bot material more than NC users by evaluating the retweets and replies of the accounts.

4.1.2 Features Selection

Feature selection is used to minimize the datasets' dimensionality by identifying a subset of features that efficiently define data [33]. The basic goal of feature selection is to create a limited subset of features that accurately captures the key features of the entire data. Feature selection algorithms are divided into three key categories namely, (i) search strategy-based (n=6/19); (ii) relationship with the learning model-based (n=10/19); and (iii) manual-based (n=3/19) approaches.

4.1.2.1 Manual Feature Selection

Identifying and characterizing the features that are significant for a specific situation, as well as providing a method to pick those characteristics, are all part of manual feature selection. In many cases, having a thorough understanding of the backdrop or domain can assist in making educated decisions about which characteristics are useful.

Three studies used the feature selection manually [18,19], whereas another study [34] proposes that one-class classification can be used to improve Twitter bot detection because it enables the detection of new bot accounts while just requiring samples of legitimate accounts. To define the accounts and distinguish between bots and humans, one-class classifiers have the advantage of not requiring examples of aberrant behavior, which in this case is the behavior of bot accounts. The experiments of this approach revealed that various forms of bots can be reliably detected with a performance above 0.89 as assessed by using Area Under the Curve (AUC) score as a criterion without having any prior knowledge about them. The features include a mix of text, nominal, and numeric data. However, the one-class classifiers were chosen only to deal with numeric data.

4.1.2.2 Search Strategy

The selection of candidate feature subsets in many feature selection algorithms is relying on a search method. Feature selection methods are split into three categories in terms of search strategy namely, complete, randomized, and heuristic search. The complete search involves scouring the whole search space for the best subset of features. Therefore, it is nearly difficult to find the best subset of features in a high-dimensional dataset in a reasonable amount of time. Randomized search methods explore a limited space from the total state space, whereas the size of the subspace depends on the stopping criterion such as the maximum number of iterations. In [35], Genetic Algorithm (GA) and Wrapper Approaches (WA) were integrated to design the proposed method (GAWA) for the best feature selection. It is built on two wrapper techniques for prime feature selection and a changed fitness value for feature minimization in the Genetic algorithm. The wrapping techniques allowed the extraction of 8,243 premier feature sets from Twitter data, which were then pruned by the Genetic algorithm to 3,137 Ideal features.

Each iteration of feature selection algorithms based on heuristic search adds or subtracts one feature from the selected feature set. As a result, their computational cost is substantially lower than that of comprehensive search methods. Many heuristic search algorithms have been developed. In recent years, swarm intelligence-based methods such as the binary grey wolf (BGW), binary moth flame (BMF) [36], particle swarm optimization (PSO),

ant colony optimization (ACO), cuckoo search (CS) [37,38], and genetic algorithm for feature reductions have received increased attention [35]. The global search capability of meta-heuristic search methods is very useful, especially in high-dimensional data [39]. The algorithms can be readily tweaked to fit the task at hand. The major feature of meta-heuristic algorithms is their extraordinary ability to avoid algorithms from converging prematurely. Given the stochastic nature of algorithms, the techniques operate as a black box, avoiding local optima as well as efficiently and effectively exploring the search space. The algorithms make a balance between exploration and exploitation. Where the algorithms completely study the promising search space in the exploration phase, the exploitation phase for the local search of promising area/s is discovered in the exploration phase [40].

4.1.2.3 relationship with the learning model-based

Feature selection is also known as variable selection, attribute selection, or variable subset selection in machine learning and statistics [41]. It refers to the process of selecting a subset of relevant features (variables and predictors) for use in model construction. Feature selection methods are divided into four types namely, wrapper, filter, hybrid, and embedded. Wholly irrelevant and noisy features, weakly relevant and redundant features, weakly relevant and non-redundant features, and strongly relevant features are the four categories of features that can be found in an original set.

Wrapper methods use the performance of a classifier as an evaluation criterion on the selected feature set [35,38,42,43]. Wrapper wraps feature selection around the learning algorithm and uses performance accuracy or classification process error rate as a feature assessment criterion. It also chooses the best discriminative collection of features by lowering the estimation error of a certain classifier. Hence, it can achieve better performance and high accuracy in comparison to the filter algorithm.

Before the learning tasks, the filter method checks the features based on intrinsic qualities and primarily assesses feature properties using four types of measurement criteria namely, information, dependency, consistency, and distance [43–45]. The feature selection procedure is performed independently in the filter method. Furthermore, this technique outperforms the wrapper technique in terms of performance and efficiency as it is scalable in high-dimensional datasets. The main disadvantage of this strategy is ignoring the relationship between the selected subset and the induction algorithm's performance.

Hybrid and ensemble methods [11,45] can be developed either by integrating two various methods (e.g., wrapper and filter), two methods with the same criteria or two feature selection approaches. The advantages of both strategies can be inherited in the hybrid method by combining their complementary capabilities [46]. The most popular hybrid method is a combination of filter and wrapper methods [43].

The embedded method is a built-in feature selection mechanism that embeds feature selection in the learning process and leverages its properties to guide feature evaluation. In [29], a combination of binomial linear regression and PCA was achieved. In terms of computation, the embedded technique outperforms the wrapper method. This is because the embedded technique eliminates the need to run the classifier many times and examine each feature subset.

4.2 Tweet text

Recently, a lot of attention has been paid to research on tweet text. Dealing with unstructured data makes extracting features from text a challenging process. However, Natural Language Processing (NLP), Latent Dirichlet Allocation (LDA), Term Frequency-Inverse Document Frequency (TF-IDF), and information extraction (IE) techniques are used to evaluate a large number of texts to gather the features of comments posted by various Twitter users.

4.2.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) refers to the ability of a computer program to understand human language as spoken and written, which is referred to as natural language. In our proposed taxonomy, (n=6/22) methods depend on NLP [35,47–51]. In [47], chi-square was used in conjunction with NLP to obtain high efficiency and accuracy. In [50], the unigram, bigram, and n-gram were combined with POS tags, such as adjectives, adverbs, verbs, and nouns with NLP. The accuracy of NB was 86, SVM was 74.6, and the maximum entropy was 82.6. NB had the highest accuracy and could be considered the baseline learning approach, while the maximum entropy methods can be quite useful in specific instances. For feature extraction, NB and NLP were used in [48]. The accuracy of NB was 63.50% which is lower than that of NLP (72.28%). The processing performance of NB was approximate-

ly 5.4 times higher than that of the NLP technique. Word embedding was used to encode tweets in [51]. On Twitter, a pre-trained GloVe word vectors dataset was used and built on two billion tweets, 27 billion tokens, and a 1.2 million-word lexicon, which includes the Stanford CoreNLP toolkit for tokenization and sentence splitting. Experiments on the Cresci-2017 dataset demonstrate that the approach can compete with state-of-the-art bot detection systems such as test-1 (Mathews correlation coefficient (MCC) = 0.920, F-measure=0.963, accuracy=0.961, recall=0.976, and precision=0.940) using NLP approaches and the extraction of opinion terms in [49]. If this is compared with a strategy that only uses semantic similarity without fuzzy logic, it can be concluded that this approach enhances the percentage of classification rate (from 74% to 86%) and decreases the mistake rate (from 26% to 14%). In [35], feature reductions are achieved by combining text preprocessing and relationship between the Genetic Algorithm(GA) and Wrapper Approaches (WA) techniques to design the proposed method (GAWA). The best accuracy was observed with GAWA and the classifier NB with the genetic method (92 %). In [52], two methods for detecting bots are presented, both rely on Natural Language Processing (NLP) to distinguish regular users from bots. A feature extraction technique is proposed in the first method for detecting accounts that send automated messages. A deep learning architecture is proposed in the second method to determine if tweets were posted by real users or generated by bots. The accuracy with ANOVA F-Value and SVM classifier is 0.9898.

4.2.2 Term Frequency Inverse Document Frequency (TF-IDF)

Term Frequency Inverse Document Frequency (TF-IDF) is a technique to quantify a word in documents, where the weight of each word is computed, which signifies the importance of the word in the document and corpus. This method is a widely used technique in information retrieval and text mining. The (n=5/22) methods relied on TF-IDF [36,53–56]. In [36], TF-IDF is used in feature extraction. BGW and BMF are implemented for feature selection. On the SemEval 2016 benchmark dataset, SVM with binary grey wolf optimizer achieves the maximum accuracy of 76.5 %. In [56], a formula was proposed to analyze harmoniously two heterogeneous data sources together by (1) fusing a decimal TF-IDF value with an integer value of the number of related articles and (2) considering the impact of related articles in the time domain. The F-measure, precision, and recall of the experimental results were on average at 0.711, 0.711, and 0.883, respectively. In [54], topic models (TM) use latent semantic indexing (LSI) to create a term-document matrix (TDM) and using TF- IDF for weighting schema to assign weights and LDA to identify topics based on the text of each tweet and take advantage of neighborhood overlap NOV. Feature augmentation was achieved using clusters with strongly connected nodes. This approach yields a 0.92 F-measure in comparison to 0.80 and 0.84 when using the word embedding (WEM) approach [57]. Even before the data reduction step, this approach yields better results with 0.87 for F-measure. In [55], features were extracted using bigram, unigram, and trigram and were weighted by their TF-IDF. The accuracy was 92% and 95% of recall to detect offensive language with NB and 90% of accuracy and 92% of recall with linear SVM. In [53], for feature extraction, the TF-IDF word level was used alongside with N-gram on the SS-Tweet dataset of sentiment analysis. The experiment shows that the logistic regression was the best algorithm for sentiment analysis and both feature extraction techniques are good enough.

4.2.3 Lexical and N-gram

In probability and statistical natural language processing, N-gram models are commonly utilized. Simpleness and scalability are two advantages of n-gram models and with a bigger n, a model can contain more context with a well-understood space-time tradeoff, allowing modest experiments to scale up efficiently. The two types of words are lexical and non-lexical. Lexical words are those that have independent meaning such as a noun (N), verb (V), adjective (A), adverb (Adv), or preposition (P). The (n=7/22) methods depended on lexical and N-gram [17,50,53,58–61]. In [59], for feature extraction, the unigram technique was used, whereas, for features selection, information gain (IG) and Pearson's correlation (PC) were adopted. Performance of the classifiers in terms of AUC (F-measure) using tenfold cross-validation is a class association and attribute relevancy based imputation algorithm (CAARIA) with NB (0.72=0.7), and CAARIA with SVM (0.7=0.7) average for three datasets. In [58], the N-Grams technique was used for features extraction. On the other side, wrapper methods, Top-k, and chi-square test as the parameter for scoring function, forward selection, and backward elimination techniques were used for features selection. The best classification accuracy of the top-k feature selection method was obtained for all bigram features which was 0.77 when classified with logistic regression classifier. In [60], unigrams, bigrams, and parts of speech (POS) were used. The study used tweets with emoticons for distant supervised learning. The maximum entropy (MaxEnt) with both unigrams and bigrams achieved an accuracy of 83% compared with the

NB with an accuracy of 82.7%. In [17], an algorithm called LA-based malicious social bot detection (LA-MSBD) was proposed that integrates a trust computational model with a set of URL-based features for the detection of malicious social bots. The proposed algorithm achieved precisions of 95.37% and 91.77%. In [61], extract linguistic features, POS, and n-grams were integrated with stylometric features and features from pre-trained lexica were used. Researchers in sociolinguistics derived lexicons of words and phrases that correlate with different age groups. The result of the convolutional neural networks (CNN)-based classifier, when compared with baseline models, yields an improvement of up to 12.3% for the Dutch dataset, 9.8% for the English1 dataset, and 6.6% for the English2 dataset in the micro-averaged F1 score. This study examined the effect of adding features incrementally and concluded that the proposed model outperforms the baseline by 12.3%, 9.8%, and 6.6% for Dutch (27), English1 (46), and English2 (30) datasets, respectively.

4.2.4 Latent Dirichlet Allocation (LDA)

LDA is a tool for topic modeling that classifies or categorizes the text in a document and the words per topic using Dirichlet distributions and processes. The (n=2/22) methods depended on LDA, [32,54]. This study analyzed people's conversations on Twitter when they mentioned AI in advertising for two years (2018 and 2019). The results of the LDA-based topic modeling indicated that Twitter users discussed AI in advertising from eight primary aspects, including advertising targeting, social media campaigns, human-AI interaction, trends, marketing tools, content creation, business applications, and related techniques. Unsupervised LDA-based and LDA Mallet are used because they can provide a better quality of topics than Gensim's. Gensim (Generate Similar) is a popular open-source natural language processing (NLP) library used for unsupervised topic modeling. This study built multiple LDA Mallet models with different values of several topics (k). The k value was set from 2 to 20. The results indicated that k = 8 generated the highest coherence score (0.5293).

4.2.5 Bag of Word (BOW)

The BOW model is a representation of NLP and information retrieval that simplifies things. A text is represented as a bag of its words in this approach, which ignores syntax and even word order while maintaining multiplicity. BOW is a text representation that describes the frequency with which words appear in a document. The (n=2/22) methods of feature extraction for tweet text methods were based on BOW [14,62]. In [14], the study explores machine learning approaches using word embeddings such as a distributed bag of words (DBOW), distributed memory means (DMM), and the performance of Word2vec convolutional neural networks (CNNs) to classify online hate, Word2Vec is defined as a distributed representation of words in a vector space that is used to aid learning algorithms in NLP tasks by grouping similar phrases. The Continuous Bag-of-Words (CBOW) and Skip-gram architectures are used by the Word2Vec model to learn word representations. The neural network achieved an accuracy of 95.33% for Dataset 1 and an accuracy of 96.38% for Dataset 2. In [62], hyper-partisan news was shared from two angles: (1) the features that make hyper-partisan content shareable and (2) the user motivations that drive the process. The study looks at one week's worth of Infowars.com content that was shared on Twitter and it was discovered that human interest and conflict in news stories drive the sharing process from a content standpoint, using both manual coding news material and semi-automated clustering of Twitter account descriptions. The results show in terms of accuracy (>0.8), precision (>0.7), recall (0.8), and F1-score (>0.8).

4.3 Merging

Two ways of merging were described here which were grouped based on the analysis of various pieces of literature. The category includes a total of (n = 23/37) articles out of all those screened. The first is based on merging metadata and tweet text (n=7), while the second is based on extracting tweet text features and selecting subset features (n=15), as explained in the next sections. Despite the growth of fusion approaches, the literature continues to rely on a single criterion. The existing techniques have a common theme of evaluating features based on certain criteria and selecting the highest performing feature subset.

4.3.1 A combination of Metadata and tweet text

In the hybrid methods, features from tweet text extracted by NLP methods were merged with Twitter metadata to help machine learning techniques for predicting malicious bots with high accuracy. From the reviewed papers, seven out of twenty three (n=7/24) relied on such method [11,18,30,43,61,63,64]. In [63], a review was provided on various tweet-based bot detection methods that use shallow and deep learning techniques to distinguish human

and bot accounts. According to the study, there is no standard set of features that can guarantee good performance. However, each study introduced some set of features that were thought to be ideal for the chosen classifier. It should be clear that the selection of features is critical, as combining metadata accounts with tweets features may lead to high accuracy and performance [65]. The accuracy and computational cost of bot detection are highly dependent on feature selection because poor feature selection can cause high computational cost, high dimensionality in data, over-fitting, and decay in predictor performance. As a result, for better bot classification and feature selection, 59 features were summarized in [18] for building a feature model, including 36 features related to tweet text and an additional 8 features related to tweet date and time and Twitter account metadata. In [66], experiments were conducted on three different types of new social bot datasets from the real world, using a deep learning model that consists of three stages: social bot detection based on tweet combined features, social bot detection based on tweet user information temporal features, and features fusing. The proposed model achieved nearly perfect detection accuracy (more than 99%). Because social bot identification based on deep learning achieves nearly flawless accuracy on diverse datasets, it necessitates a vast number of tweets and integrating more than one dataset.

Another research study [43] characterizes and classifies the features into four categories according to their attributes: user characteristics, microblog characteristics, network structure characteristics, and user interaction characteristics. Then, such features were formally expressed and quantified to obtain numerical features. To solve the differences between the types and size of the eigenvalues, the maximum and minimum normalization method was used to map the values of all features to [0,1] interval, and this, in turn, led to obtaining the complete set of features. Experimental results demonstrated that the model had the highest precision and F1 score than NByes, logistic regression, random forest, and SVM, while the F1 score reached 0.885. In [64], a multilingual strategy was proposed for addressing the bot identification task in Twitter using deep learning (DL) approaches to assist end-users in determining the legitimacy of a particular Twitter account. Therefore, a series of experiments were carried out using state-of-the-art multilingual language models to generate an encoding of the user account's text-based features, which were then concatenated with the rest of the metadata to create a potential input vector on top of a Dense Network called Bot-DenseNet. The result of the bot-dense model in terms of F1-Score=0.77 is determined.

4.3.2 A Combination of features' extraction and selection

Such methods extract features from tweet text using various natural language processing methods and feature selection methods to select subset features. The main advantages of applying these techniques are reducing the high dimensionality and selecting the important features, which can help ML to obtain high accuracy. Fifteen out of twenty three of the reviewed papers (n=15/23) depended on the extraction of tweet text features and the selection of subset features [8,11,35–38,43,44,47,50,58–61,64,67]. The study of [36] used population-based meta-heuristic algorithms, feature extraction using term frequency-inverse document frequency (TFIDF), feature selection using binary grey wolf (BGW), and BMF for feature optimization, the highest accuracy of 76.5% is observed for SVM with binary grey wolf optimizer on SemEval 2016 benchmark dataset. According to [67], NB has faster training data on the airline dataset. In this airline dataset, the SVM linear classifier has the highest classification accuracy. The features with higher mutual information (MI) calculation value than other features contain more essential information. The results revealed that the training data with features selection using complementary information was better. The best classifier for both datasets was the linear SVM, where the accuracy was 72.66. In [8], most of the theory-based feature selection methods focus on retaining the features that contain more information and removing features that contain less information. However, this may lead to information loss. Moreover, features that are individually less significant may be useful when they are combined with other features. Fuzzy cross-entropy was used in [8] with two datasets namely, T1 and T2 to preserve information that is contained in the selected feature subset to be equal to the information contained in the full feature set. The accuracy obtained with Random Forest was 95.3 for dataset T1 and 90.88 for dataset T2. In [68], deep learning is defined by the automatic feature selection process in models that implemented a deep architecture. To determine opinion polarity, the impact of earlier data refinement in the pre-processing step before using deep learning was examined. This enhancement incorporated a traditional textual content process as well as a popular feature selection technique. This study showed that combining feature selection with a basic preprocessing step to improve data quality can yield good results when using Deep Belief Networks. The experiments surpassed the results of the earlier literature with the Deep Belief Network application in opinion classifications. The obtained accuracy was 81.7 for movies and 76.0 for books [68].

5 Discussion

This paper examined some of the approaches that have been employed to detect the activities of social bots on Twitter. A social bot is more deceptive than ever before, and it is difficult to develop systems that can detect bots. To make progress in this area, there is a need to consider the main challenges and factors that impact social bot detection activities. Understanding such challenges can help address many issues. Therefore, we discuss different techniques of features selection and extraction that are used in bot detection, with an explanation of those techniques within each category (see table 1).

It can be concluded that when solely PCA is used to extract features [29], essential features may be lost, and it is preferred to assign each feature a function score and then select the set of features with the highest score. Furthermore, the calculation of the detection rate can be increased by building better machine learning algorithms that perform correlation among fresh feature sets and specify more successful variables in the future.

The study [34] proposes that one-class classification can be used to improve Twitter bot detection depending on just requiring samples of legitimate accounts. The problem with this technique is that if the cyborgs (accounts that combine bot and human behavior) exhibit similar tendencies to valid accounts, it is a loophole that cyborgs can be used to create a new group that is similar to valid accounts that are difficult to detect.

In feature selection algorithms based on a search method, the best accuracy achieved from GAWA [35] with Naive Bayes (NB) and genetic algorithm is 92%. Although a compromise between speed of convergence and optimality of the outcome was made via parameter setting, we believe the algorithm still tends to become locked in a local optimum. Randomized search-based algorithms have lower computational complexity than complete search-based algorithms.

The literature [36] [37,38] [39] focuses mostly on two goals namely, maximizing accuracy and decreasing the number of selected features. In addition, multi-objective feature selection should also consider computing time, complexity, stability, and scalability.

Relationship with the learning model-based such as wrapper methods [35,38,42,43], when compared with the filter strategy, has the disadvantages of computational complexity and increased sensitivity to over-fitting. Because most wrapper approaches are multidimensional, they require long computation periods to reach convergence and can be intractable for large datasets. The embedded technique combines the benefits of both the filter and wrapper methods in one package and picks features during the mining algorithm construction. This leads to reducing the computational expenses.

Based on research conducted on Natural Language Processing (NLP) [35,47–51], to extract feelings from social networks, we must perform large-scale opinion mining on the data. This, however, could be a difficult operation because social network texts are typically short, full of idioms, with peculiar grammatical structures, and a variety of other issues. For this problem, it is recommended to combine metadata features with tweet text features.

Empirical research shows that the logistic regression algorithm using TF-IDF [36,53–56] to extract features without removing stop words was the best algorithm for sentiment analysis where compared with Lexical and N-gram, and Bag of Word (BOW).

In the hybrid methods [11,18,30,43,61,63,64], features from tweet text extracted by NLP methods were merged with Twitter metadata to help machine learning techniques predict malicious bots in high accuracy, but the main challenge in such methods is the prediction time.

Table 1. Challenges, motivation, and study tips in a nutshell

Ref	*Brief description*	*Motivation*	*Challenges*	*Recommendation*
[69]	For bot detection, this study uses clustering algorithms. The selection of characteristics in clustering is difficult since some features are critical for clustering while others may obstruct the clustering process. This research focuses on the characteristics that distinguish bot users.	The motivation is to identify bots using feature extraction and clustering for an unlabeled dataset.	The major challenge with the non-labeled dataset is figuring out how to extract features that aid in bot detection.	More features must be used, as well as the utilization of the tweet's sentence, because Feature selection is a vital step in unsupervised learning, and it is critical to select characteristics that aid in the detection process.

[38]	The study compared different feature subset evaluators for Twitter sentiment categorization. Filter feature selection based on Information Gain was computed before the application of EC-based feature subset selection to lower the size of unigram feature space and reduce the computing time required for the wrapper evaluators.	Selecting strategies that have proven successful in handling the feature selection problem is motivation.	Several challenges arise while performing sentiment analysis on Twitter data. High-dimensional space is one of the most difficult problems to solve.	Although EC feature selection methods achieve higher performance, they face several difficulties, the most significant of which is the computational cost. There is a need to speed up the search technique and the evaluation measure.
[8]	The Community Inspired Firefly Algorithm for Spam Detection (CIFAS) is proposed in this paper to handle the combination search for features with good performance utilizing fuzzy cross-entropy as the fitness function.	Designing a Spam detection system that can handle a combined search for attributes that work well.	The challenge is to reduce the number of features that help Spam detection but at the same time preserve the information without any loss.	Hardcore feature extraction is used in the suggested method, which necessitates domain knowledge. As a result, in this Big Data era, the suggested algorithm is constrained by its low adaptiveness and high cost. Automated feature extraction methods based on deep learning may be able to tackle the problem.
[70]	A model for the classification of suicidal tweets is constructed with the goal of suicide prevention by detection, motivated by the increasing association between the expression of suicidal ideation on social media and suicide rates.	The motivation is to create a comprehensive classification system that can accurately identify suicidal intent, separate it from non-suicidal suicide-related communication, and prevent suicide.	The fundamental challenge is the lack of specific feature selection-based approaches to train robust suicidal ideation classification models.	Because there is no explicit suicide ideation in tweets, as opposed to the suicidal rhetoric utilized in training, they represent a practical challenge.
[55]	The purpose of this study is to provide a method for detecting inflammatory language in Twitter data. For this challenge, two strategies were chosen: Linear SVM and Naive Bayes, both of which are ML algorithms.	The motivation is to build an algorithm that can recognize inappropriate language in tweets better.	The challenge in automated detection of offensive language	The Linear SVM requires a well-balanced input to produce effective results. As a result, the parameter for this approach is a bit tricky, and it is preferable to use another technique.
[59]	This research created and tested a new algorithm called CAARIA (class association and attribute relevancy based imputation algorithm) to improve the quality of classification for Twitter sentiment analysis.	The goal is to handle a variety of sparse matrices problems that arise while converting input text into some feature representation. Dimensionality reduction can be done feature-wise (i.e., feature selection) or sample-wise.	Twitter sentiment analysis is a difficult undertaking. As a result, one of the most difficult issues is to employ intelligent approaches for automatic Twitter sentiment analysis.	Must use other classification techniques including deep recurrent neural networks to be more investigated and further evaluated, with large datasets of Twitter.
[67]	This study uses a dataset derived from a collection of tweets regarding US Airlines that already contains numerous metadata, allowing for a simple feature selection experiment.	Sentiment analysis is required to collect sentiment classification for the company via feature extraction and feature selection from the body of tweets.	The most difficult aspect of sentiment analysis is transforming unstructured and organized data before applying classification methods.	From start to finish, this research examines the incremental Mutual Information value. As a result, more effective strategies for obtaining features with high Mutual Information values are required.
[17]	To distinguish between genuine and malicious tweets, features derived from the posted URLs (in the tweets) are used to examine the malicious behavior of participants.	The goal is to create a model that can detect dangerous social bots with greater accuracy and recall.	Extracting social relationship-based information takes a long time. As a result, distinguishing malevolent social bots from real users on the Twitter network is a difficult challenge.	Must investigate the among the features and their impact on bots detection for another dataset.

[31]	The research provided in this study was done specifically to see how human-operated accounts reacted to bot actions.	The primary objective is automatically to identify legitimate online users and limit deceptive actions carried out by malevolent entities such as social bots.	The challenge is to limit deceptive actions carried out by malevolent entities such as social bots.	To better detection, analyze the nature (real vs. bots) of individuals who have begun to be followed, stopped being followed, and who stay on the followees lists for extended periods.
[47]	The purpose of this study is to use sentiment analysis methods on natural data to investigate the effects of social media themes in digital money markets.	The motivation is to determine whether or not the use of semi-supervised feature selection methods in sentiment analysis helped the classification results for the problem found in digital money markets.	Given the globalization of languages, adding new terms to the languages we use every day can have a negative impact on the success of machine learning algorithms.	The semi-structured approaches are likely to be more supervised, based on the findings. For a better learning rate, we propose at least 5,000 tagged comments for the first system training.
[61]	This research offers a unique method for age prediction of Twitter users that incorporates characteristics extracted from hashtags and URLs from tweets.	The motivation is to leverage language-related traits and Twitter metadata to categorize individuals into age groups.	The challenge is some users modulate their communication strategies to protect their privacy.	The issue is that some users adjust their communication tactics to protect their privacy, we advocate using graph theory to identify persons who participate in the one-on-one chat on a public forum.
[34]	This study proposes that one-class classification be used to improve Twitter bot detection since it enables the discovery of unique bot accounts while just requiring samples of genuine accounts.	Given that bot types will continue to evolve in the future, and that bot creators will modify behavior to evade detection, a new technique for automatic bot detection is required.	The challenge is that supervised classifiers may struggle to detect new forms of bots if the behaviors observed in the training examples are too dissimilar.	The issue with this strategy is that if the cyborgs all have similar tendencies, it could result in a new group of semi-automated accounts or a simpler real account. The proposed method could be used to categorize the different types of Twitter bots that have been found.
[29]	A newly built robust feature set is used in this study to detect Twitter spammers using a linear statistic technique. To detect evasion strategies used by Twitter spammers, an in-depth analysis of features are performed.	The goal of this study is to extract features from our data using a hybrid approach that combines logistic regression with a dimensional reduction technique called principal component analysis.	Twitter is facing significant challenges as a result of spammers who have tarnished the website's reputation, causing many users to abandon it.	A more thorough examination of the various types of spammers, as well as their evasion strategies, is required by building better machine learning algorithms.
[51]	This research offers a recurrent neural network (RNN) model, namely BiLSTM, with word embeddings to identify Twitter bots from human accounts, and it focuses on the categorization of human and spambot accounts on Twitter.	The goal is to create a model that uses word embeddings to detect bots that only use tweets and does not require extensive feature engineering.	The challenge in the automated programs used in Twitter is that automation is a double-edged sword between Twitter legitimate bots and malicious bots that have been widely exploited to spread spam or malicious content.	In this study RNN model uses only the contextual content of tweets as the input to the mode, It is better for more accuracy with a multi-feature approach, including features on the profile, user behavior, friendship networks, and the timeline of an account.
[32]	This study analyzed people's conversations on Twitter when they mentioned AI in advertising over two years (2018 and 2019), including advertising targeting, social media campaigns, human-AI interaction, trends, marketing tools, and business applications.	The motivation is to create a study that will be useful for academic research on AI advertising as well as the practical application of AI in advertising.	Understanding people's opinions of AI advertising is a challenge because it still has several constraints, such as artificial study settings and a small number of respondents	To contextualize the research findings, future research should seek to determine the existence of distinct Twitter user types (e.g., business vs. individual tweets).
[62]	This study looks at hyper-partisan news sharing from two angles: (1) the features that make hyper-partisan content shareable, and (2) the user motives that drive the process.	The motivation is to uncover key aspects of news that drive news sharing, as well as to comprehend key user motivations that drive news sharing.	The major challenge is to figure out what makes hyper-partisan news more or less social media shareable.	This strategy has to be tested on a larger number of hyper-partisan news outlets over a longer period, to generate a more detailed range of account clusters.

[49]	This work provides a hybrid technique based on fuzzy logic and information retrieval system (IRS) concepts with the usage of semantic similarity to classify tweets into three categories (positive, negative, and neutral).	To get better results than a typical technique to improve the quality of the categorization of the tweets, must find new approaches and methods to enhance the quality of the classification of the tweets.	The challenge is extracting opinions, emotions, and attitudes from social networks' data such as Facebook comments or tweets.	Because it is critical for the system to retain its relevance and value over time, the model must be trained on datasets that include newer terms.
[63]	To combat tweet-based botnets and reliably discriminate between human and tweet-based bot accounts, this study focuses on large data analytics, particularly shallow and deep learning.	The motivation is at providing an overview of different tweet-based bot detection methods.	One of the challenges faced in evaluating bot detection approaches is that the ground-truth data is insufficient. some challenges still need further investigation.	To aid in the universal evaluation of detection approaches, datasets with various sets of social bots must be built.
[71]	The goal of this article is to look into the consequences of spam, particularly in terms of excessive participation inequality. In contrast to such complex and resource-intensive machine learning detection methods, this study advocate for a method that focuses on educational researchers' practicality.	The motivation is to investigate the consequences of spam, particularly in terms of excessive participation inequality.	The subjective factor in classifying spam is a basic challenge that hampers decisions about whether to exclude specific tweets or people from a collection.	Yet, little attention has been made to spam's prevalence and effects on online educational communities. Most examples of spam removal deal with user-level identification, however for some studies, identifying spam at the tweet level may be more relevant or beneficial.
[44]	The usage of a contrast pattern-based classifier for bot detection in Twitter is proposed in this paper.	On the social network Twitter, contrast pattern-based classifiers are being used to detect bot behavior.	Bot detection may be used to perform more complex activities, such as sending brand new messages or faking human engagement.	Improving the filtering approach to get a smaller number of high-quality patterns for bot identification.
[19]	This study presents a preliminary result based on a sample of Twitter accounts that were later analyzed using machine learning models to determine whether or not a Twitter account is a bot using SVM and Random Forest.	Finding the best machine learning algorithm for determining a social bot, as well as which features benefit the algorithm the most, is the motivation.	One of the issues on social media is the usage of social bots, which are used to persuade a human user to believe a bot's opinion.	To enhance bot detection in larger datasets, we used a tweet similarity between each user over time and a description similarity between each user.
[42]	This study shows how to use two separate ways to minimize feature subset size and enhance classification accuracy by combining the filter method with wrapper-based feature selection methods.	Creating a feature selection strategy to reduce feature set size and improve classifier accuracy.	In the field of sentiment categorization, the challenge is how to select a suitable feature set.	Experiments using additional evolutionary methodologies, such as differential evolution and genetic algorithms, should be included, and they should be applied to diverse sorts of datasets, such as Twitter.
[15]	A supervised machine learning strategy for recognizing and combating cyberbullying is proposed in this research. To train and recognize bullying behaviors, a variety of classifiers are utilized.	The goal of this research is to present a supervised machine learning strategy for recognizing and combating cyberbullying.	Given the negative effects of cyberbullying on victims, it's critical to identify effective ways to detect and prevent it.	To increase the performance, more cyberbullying data is required. As a result, deep learning techniques will be appropriate for larger data because they have been shown to outperform machine learning algorithms on larger datasets.
[72]	This research introduces a new rapid hybrid dimension reduction approach that combines multi-strategy feature selection and grouped feature extraction.	The motivation is to combine multi-strategy feature selection and grouped feature extraction	Extrapolate important information from large amounts of data presents us with enormous challenges. The curse of dimensionality, in comparison to the challenge of data reduction, may be more difficult to overcome.	We only keep the first principal component and discard all other components in the grouped PCA method, which may inevitably result in additional information loss, such as adding priority weights to different feature groups, which may help us preserve more effective information.

[14]	The study investigates the performance of Word2vec Convolutional Neural Networks (CNNs) and machine learning algorithms that use word embeddings such as DBOW (Distributed Bag of Words) and DMM (Distributed Memory Mean) to classify online hatred.	Designing a supervised machine learning strategy for automatically detecting and preventing new cyberbullying incidents.	Bullies can use social media to attack victims because they create a rich environment for them to do so. Given the negative effects of cyberbullying on victims, it's critical to identify effective ways to detect and prevent it.	To increase the performance, more cyberbullying data is required. As a result, deep learning techniques will be appropriate for larger data because they have been shown to outperform machine learning algorithms on larger datasets.
[58]	Using a new dataset of 6903 tweets taken from Twitter, the methodology focuses on word-level language detection. Various n-gram profiles are investigated using a variety of feature selection strategies across a large number of classifiers.	The motivation is to create feature selection algorithms for a variety of learning algorithms to investigate the impact of the method as well as the number of features on language identification performance.	Language identification has become a difficult problem as a result of the dramatic increase in data generated by social media platforms such as Twitter, where a large number of socially connected people communicate in an informal language.	For language classification, the evaluation of a mix of filter and wrapper approaches might be explored. The filter method can be used as a preprocessing step to remove features that don't have anything to do with the language classification model, and then the wrapper method can collect the best set of features for the learning model.
[36]	The application of two swarm intelligence algorithms, binary grey wolf and binary moth flame, for feature optimization to improve sentiment classification performance accuracy, is demonstrated in this study. ((The basic goal is to identify distinctive features that can be used to classify data into positive, negative, or neutral categories, resulting in enhanced	The basic goal is to identify distinctive features that may be used to classify data into positive, negative, or neutral categories, resulting in increased sentiment classification accuracy.	A lot of uncertainty is generally associated with the micro-blog content, primarily due to the presence of noisy, heterogeneous, structured, or unstructured data which may be high-dimensional, ambiguous, vague, or imprecise. This makes feature engineering for predicting the sentiment arduous and challenging	The study can further be extended to analyze the use of other bio-inspired and swarm-inspired algorithms for improving the sentiment classification accuracy. The use of different filter methods that are other than TFIDF can be explored to give interesting insights into this filter-wrapper arrangement for sentiment classification.
[28]	The goal of this research is to address the challenge of detecting fake profiles with the suggested model (ISVM) by extracting appropriate features that can accurately distinguish fake and real profiles.	The motivation is to address the issue of detecting fraudulent profiles in online social networks and to design a model that can properly identify phony profiles.	The challenge is to address the issue of detecting fraudulent profiles in online social networks	Using PCA to extract features, you'll miss out on vital details. It is preferable to assign each feature a function score and then select the set of features with the highest score.
[35]	Sentiment analysis or opinion mining is the key to natural language processing for the extraction of useful information from the text documents of numerous sources. A novel method (named GAWA) is proposed for the optimal feature selection.	Designing a method for feature selection to select the premier features and reduce the size of the premier features.	One of the biggest challenges in Sentiment analysis is accuracy regarding the massive volume of features.	It is better to examine the proposed algorithm with multiple datasets from various sources to select the best features with various categories of syntactic and stylistic features.
[48]	Using the Nave Bayes algorithm and natural language processing, the study offers a system that can extract human sentiment information from large amounts of unstructured big data from social media sites (NLP).	To extract relevant information from large data, a rapid processing technique is required as the volume of processing rises.	Obtaining varied information from unstructured data makes data processing more difficult.	Because the dataset used in the study is unclear, it's best to compare it to other classifiers. The author is not only reliant on speed but also on the accuracy of the outcome.
[56]	This work suggested a method for evaluating the credibility of Twitter-based event detection that can examine both tweets and an external trustworthy data source harmonically.	The motivation is a design method that evaluates the credibility of Twitter-based event detection by analyzing both tweets and an external reliable data source.	The severe issue is that if there is too much bogus information in the system, it will fail to recognize events appropriately.	This study evaluates the credibility of the event detection result just with the proposed formula whereas is better if compared with another ML classifier with TF-IDF values.
[37]	New clustering techniques based on K-means and DENCLUE have been developed in this work to assess the sentiments of tweets.	The aim is to create a suitable clustering method that produces an appropriate number of clusters in a reasonable amount of time.	One of the important challenges is reducing exploited information on Twitter by using sentiment analysis tools.	It will be also interesting to cluster the sentiments of the tweets based on emoticons.

[50]	Different machine learning techniques for sentiment analysis on Twitter, such as Nave Bayes, SVM, and the Maximum Entropy Method, are reviewed.	The motivation is to find the most accurate machine learning method for sentiment analysis on Twitter.	The challenge is The sentiment analysis of Twitter from unstructured text to help for detection of malicious bots.	To compare performance measurements, more machine learning methodologies are needed in this study. It's also a good idea to explain the dataset's size, such as the number of Tweets, users, and so on.
[11]	A multi-objective hybrid strategy is utilized in this study to discover the most effective feature set for detecting bogus accounts on the Twitter social network.	Using a multi-objective hybrid feature selection strategy that aids feature set selection while providing the best classification performance.	The frequency of fake accounts or social bots is considered one of the serious challenges of online social networks.	Applying the features used in this study to check performance for detection of fake accounts for another dataset
[64]	This paper presents an approach for addressing the bot using state-of-the-art Multilingual Models to generate an encoding of the text-based features of the user account. The models are then concatenated with the rest of the metadata to build a potential input vector on top of a Dense Network.	Building robust automatic systems to improve the quality of experience of consumers by reducing their privacy risks as well as increasing trustworthiness.	Generating an encoding of the text-based features of the user account and concatenating with the rest of the metadata.	Comparing the performance for work described in this paper with the latest Transformers such as the GPT-3 and T5.
[66]	Experiments were conducted on three different types of new social bot data sets from the real world, using a deep learning model that consists of three stages: social bot detection based on tweet combined features, social bot detection based on tweet user information temporal features, and features fusing.	Better detection of the malicious behavior of increasingly complex social bots.	Use the user tweets and information to detect social bots in as little time as possible while ensuring a high detection rate.	Because social bot identification based on deep learning achieves nearly flawless accuracy on diverse data sets, it necessitates a vast number of tweet information from the user, it must employ more than one dataset, at least three large datasets.
[68]	This research proves that combining feature selection with a basic preprocessing step, aiming to increase data quality, might achieve promising results with Deep Belief Network implementation.	Use Deep Belief Network to demonstrate the benefits of pre-processing methodologies by assessing the impact of data refinement, the use of a classic text pre-processing, and a feature selection methodology on polarity classification (DBN).	Providing an analysis of the impact of data refinement, the use of a classical text pre-processing, and a feature selection technique, exerts on polarity classification with the Deep Belief Network (DBN), this study demonstrates the bene ts of pre-processing techniques.	It is preferable to apply this methodology to additional datasets relating to opinion classification, as well as the prospect of investigating a specific Deep Learning extension that has recently been recommended for Sentiment Analysis.
[52]	The first method proposes a feature extraction methodology for detecting accounts that send automated messages. In a second way, a deep learning architecture is proposed to assess whether tweets were submitted by actual users or generated by bots.	The early detection of bots in social media is quite essential.	More sophisticated techniques need to be exploited, to mitigate bot activity in social media. In this framework.	To improve the performance of the deep neural networks used, it is preferable to use neural language models based on transformers.

6 Twitter datasets labeled by usage

To investigate and comprehend the behavior of bots in social networks it is necessary to use datasets that include both human and bot accounts. Researchers face three challenges when working with datasets. First, obtaining recent public datasets on which to run experiments takes time. This is especially when gathering data with sufficient size and containing enriched content. The second issue is creating a trained dataset that is diverse in terms of the bot accounts' content. The need for such a dataset is especially important in studies that use machine learning algorithms in detecting bots. Thus, labeling a sample that is well defined in terms of size and content can be challenging. Many researchers used human annotation of a reasonable training sample to perform this task [11,37,53], but this is a slow procedure that does not produce a large number of labeled data. Moreover, experts may incorrectly classify bot accounts due to human error or because they can be deceptive. Third, because bots are con-

stantly evolving, and up-to-date sufficient dataset is even more critical. Thus, researchers were unable to devise a method for achieving a stable labeling strategy. Undoubtedly, the lack of an absolute ground truth dataset is a significant disadvantage in this field of study.

The reviewed studies focused on machine learning methods and the datasets that consisted of a combination of private and public accounts. Some researchers used publicly available datasets as a ground truth baseline for testing their techniques [11,28,34,51,54,62]. In general, most studies used the Twitter API to collect datasets, except for [49], which used their API called Twitter4j to collect data. Apache Flume is used to extract and store tweets directly in the Hadoop framework. The key information about Twitter datasets used in previous literature was summarized in table 2. This can help researchers find a suitable dataset for their future research.

Table 2. Datasets of social media bots detection, sentiment analysis, and others.

Ref.	Dataset	Description	Labeled
[43]	FU, 2020	Extract data from the Sina Microblog platform (the largest Microblog platform in China)	Factors that affect the forwarding of microblogs
[58]	Ansari, 2020	Tweets Hindi Count (3854) English code mixed (3049) Total (6903)	Identification of languages
[28]	Collected from Kaggle.	Datasets got from Twitter about 37 countries with over one hundred and twenty thousand instances dataset with 34 attributes.	Fake profiles
[35]	ABDUR RASOOL,2020	Twitter Stream API is deployed with a python-based crawler. As a result, received 66,177 tweets in one week.	Sentiment analysis
[60]	Alec Go, 2009	Tweets were collected using Twitter API, from the period between April 6, 2009, to June 25, 2009.	Sentiment analysis
[11]	Cresci et al., 2019	Test Set-1: Genuine accounts + Social Spam Bot, Accounts=1982, Tweets=4061598. Test Set 2: Genuine accounts + Social Spam Bot, Accounts=928, Tweets=2628181	Fake accounts
[50]	Mandloi, 2020	Using API, key consumer, key access token, and their secret key.	Sentiments analysis
[37]	Hajar Rehioui, 2019	Twitter-airline sentiment: collected in February 2015 and classified into PO, NE, neutral, and Twitterdataset: November 17, 2014, to Dec 10, 2014, and Twitter sentiment Corpus-3.	Sentiment analysis
[56]	SATO, 2018	Twitter streaming API(Olympic games)	Credibility of events
[30]	Mabrook S. Al-Rakhami,2020	Collected data from January 15 to April 15, 2020: Keyword (Covid-19, Covid19, Covid_19, Coronavirus, Covid, and Corona, No. of Tweets=1,145,802)	Twitter misinformation
[53]	Dataset Tweet SS	The dataset is annotated manually and contains a total of 4242 tweets, 1037 are negative tweets and +G23+H23	Sentiment analysis
[38]	Stanford sentiment	A total of 140 datasets 1.6 m tweets with two labels, namely, positive and negative.	Sentiment analysis
[8]	Elakkiya E,2020	T1(Twitter's Streaming API t with over 600 million tweets, including more than 6.5 million spam tweets)	Spam detection
[55]	Gabriel Araujo De Souza, 2019	Contains 24783 tweets. From these, 1,430 are classified as hate speech, 19,190 as offensive language, and 4,163 as normal language	Offensive language detection
[67]	Public from Kaggle	Tweets.csv(airline-sentiment, 2015, 1.13MB,14641 x 15)	Sentiment analysis
[17]	Rashmi Ranjan Rout, 2020	Two Twitter datasets, The fake Project data set (legitimate users =3474, malicious social bots= 1000, legitimate tweets= 8377522, malicious tweets=145094), and Social Honeypot dataset (legitimate users =19276, malicious social bots= 22223, legitimate tweets= 3259693, malicious tweets=2353473).	Bots
[47]	Firat AKBA, 2020	Collected data between 2010 and 2019, for comments, were: "Bitcoin", "Bitcoin Price", "Bitcoin Forecast", "BTC", "BTC/USD", and "BTCUSD".	Classify the semantics of tweets
[61]	Pandya, 2020	Three existing datasets Dutch (age 0-40, 2150 users), English1 (age 13-40, 1074 users), English2 (13-25, 1794 users)	Age prediction

[34]	Cresci-2017	Consisting of 3,474 human accounts 8.4 million and 1,455 bots 3 million tweets.	Bots
[29]	Murugan, 2018	The dataset contains 17 million users' tweets with 159 features included	Twitter spammers
[51]	Cresci-2017	Consisting of 3,474 human accounts 8.4 million and 1,455 bots 3 million tweets.	Bots
[32]	Collected Twitter data using Brandwatch.	A total number of 43,908 tweets were collected on 8/8/2020 and combined all the tweets were as one document. This resulted in 16,520 documents.	AI Twitter conversations
[62]	Magdalena Wischnewski, 2021	Collected tweets using the Global Database of Events, Language, and Tone, and identified 169 Infowars articles during the period of 23 to 29 September 2019.	Hyperpartisan news sharing behavior
[54]	Publicly datasets on the Kaggle Data Science community	The first dataset contains about 17k tweets from 112 unique pro-ISIS Twitter accounts. The second dataset consists of 122 K tweets representing 95,725 accounts..tweets divided into 17,000 Pro-ISIS and 77,813 An-ti-ISIS.	Identifying accounts of terrorists
[49]	Youness Madani, 2019	Use a Twitter API called Twitter4j (between June 2015 and June 2017), also use the Apache Flume, to extract and store the tweets directly in the Hadoop framework with its distributed file system (HDFS).	Classify the semantics of tweets
[71]	#Edchat dataset	The dataset contains 482,251 public tweets and retweets for educators who discuss current trends in teaching with technology, collected between Feb 1, 2018, and Apr 4, 2018	Spam and educators
[44]	Octavio Loyo-la-Gonzalez, 2019	51,457 tweets from which 31,654 belong to humans and the remaining (19,804) belong to bots.	Bots
[19]	Pratama, 2019	Tweets are gathered from the presidential candidates, from February 2019.	Candidate's Supporters
[14]	Dataset of Uni-versity of Mary-land	The first dataset consists of over 30,000 tweets, whereas the second dataset by Davidson et al., contains roughly 25,000 tweets.	Automated hate speech
[64]	Twitter API	Dataset is composed of 37438 Twitter accounts, where 25013 were an-notated as human accounts and the remaining 12425 are bots.	Bots
[73]	Chen2018	Twitter 185.922 (bots)	Bots
[74]	DeBot	9,134-bots	Bots
[75]	Beskow2019	Twitter 235k-bots , Bots Their own method	Bots
[76]	Campos2018	Campos2018 Twitter 267-accounts Human, legitimate & malicious bot	Manual annotation
[77]	Chew2018	Twitter 57,888 accounts	Bots
[66]	Cresci-2017	Consisting of 3,474 human accounts 8.4 million and 1,455 bots 3 million tweets.	Malicious social bots
[68]	Ingo Jost 2018	Pang and Lee [78] and on another four datasets of different types of products from Amazon: books (BOO), DVDs (DVD), electronics (ELE), and kitchen appliances (KIT)	Opinion polarity
[52]	Cresci-2017	Consisting of 3,474 human accounts 8.4 million and 1,455 bots 3 million tweets.	Malicious social bots

7 Methods effectiveness

The effectiveness of bot detection depends on the method used with features selection to produce good accuracy. table 3 explains the classification methods used in the literature.

Based on our review, researchers are more likely to use machine learning methods. The majority of the reviewed papers used tree-based approaches and the Bayes theorem. The SVM classifier was the most frequently used classifier among the methods examined, figure 3 illustrates a summary of classifiers used in the reviewed literature. SVM is based on the kernel and parameters that are chosen. Furthermore, a significant disadvantage of this classifier is that it highly relies on a large training set to improve performance. Similar to SVM, the effectiveness of a neural network depends on the sample size; thus when the sample size is large, the vector performs well. The Random Forest (RF) classifier has been used by many researchers. The benefits of this classifier are: (1) less complexity of tuning and (2) achieving more accurate results. However, as with most decision tree algorithms, the complexity of the tree could lead to the issue of over-fitting. In earlier literature, the Bayes-theorem and RF were widely used. As a statistical theorem, the Bayes theorem is quick in terms of training and prediction time. However, when datasets contain a small number of features, this classifier performs better.

Even though all of the literature reviewed had tested and measured their methods' performance, it is inequitable to use these measures to judge method performance because some factors that contributed to unstable performance are labeling accuracy and the types of bots included in the datasets, and discriminate features. For ex-

ample, in supervised learning, performance is highly dependent on two aspects namely, training and testing datasets and the selection of important subset features. When the same method is applied to different datasets or features, the performance of the models varies dramatically.

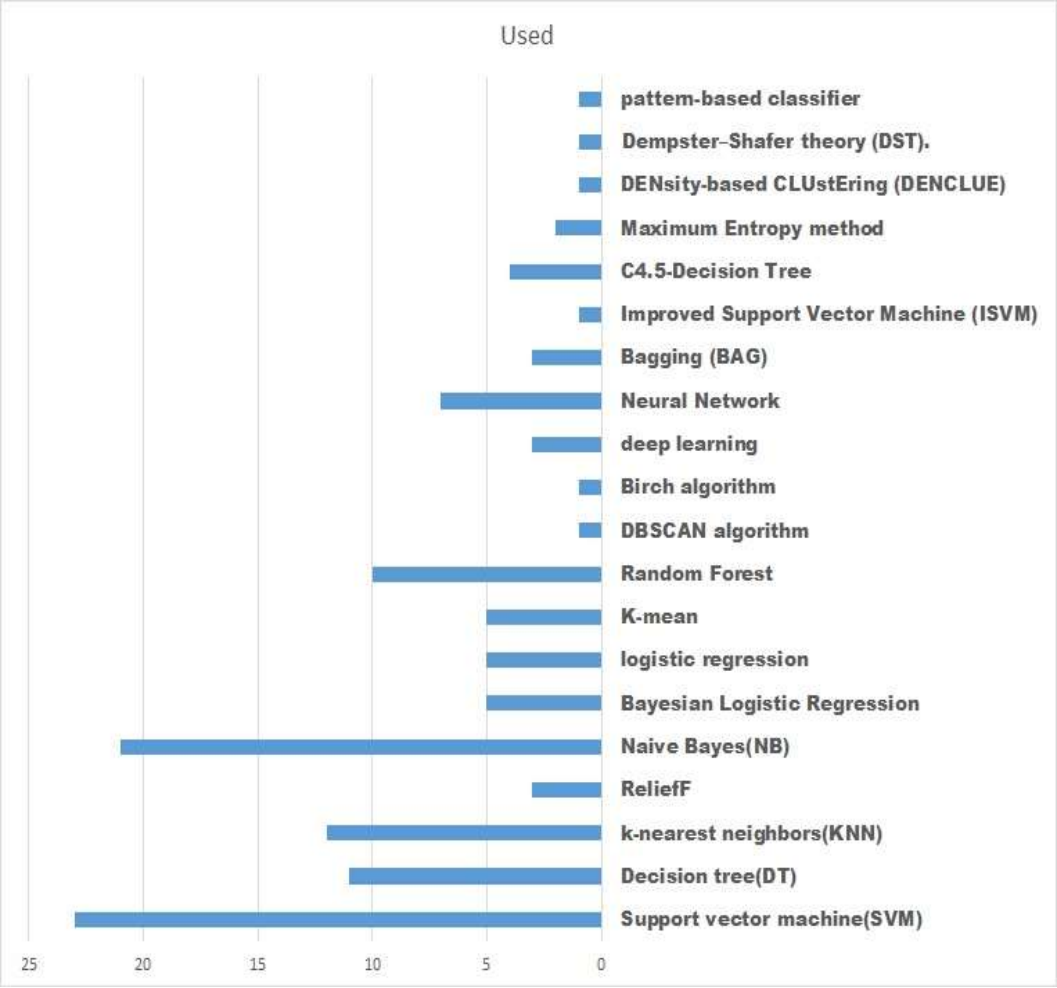

Figure. 3 Summary of classifiers used in the reviewed literature

Table 3. Classification methods used in the reviewed literature

Ref	Bots type	Algorithm	Type	Result
[69]	Bots user	dbscan and k-mean	Unsupervised	97.7% accuracy, 91% precision, 98% recall.
[8]	Spam	SVM, KNN, and RF (RF)	Supervised	T1 RF accuracy= 95.3, T2 RF accuracy =90.88
[55]	Offensive language	SVM and NB	Supervised	NB accuracy=92%, recall =95%, SVM accuracy=90%,recall =92%
[17]	Fake accounts	Bayesian learning and Dempster–Shafer theory (DST).	Supervised	The proposed LA-MSBD algorithm has achieved precisions of 95.37% and 91.77% for MSBD, respectively.
[31]	Fake news	HMM, IBk, BN, NB, VP, SMO, MLP, JRip, 1R, 0R, J48, HT, RT, J48C, J48g, LAD, LMT, REP, and RF	Supervised	ClassA's features with 1R classifier alg. 93.27 accuracy, Classification results for bot detection task with ClassA's features with RF classifier alg. 95.84 accuracy
[34]	Bots	One-class classifiers, Bagging-TPMiner and Bag-	Supervised	Detect different types of bots with a performance above 0.89 measured using AUC, without re-

		ging-RandomMiner, OCKRA.		quiring previous information about them.
[29]	Spam	RF, DT C4.5, Bayes Network, K-NN, and SVM.	Supervised	The best result in RF ,Detection Rate=0.913, F-1 Measure=0.92.
[51]	Spam	RNN model, BiLSTM	Supervised	Test-1:Accuracy=0.961 test2 : Accuracy=0.929
[79]	Malicious social bots	Constrained seed K-means algorithm	Semi super-vised	The experimental result shows, the recall rate is 97.5%, and the F1 Score is 95.2%.
[63]	Botnets	shallow and deep learning methods	Unsupervised	Combined with tweets to produce a high accuracy of 96%.
[71]	Spam	A practical and holistic Ap-proach		
[44]	Malicious activity	pattern-based classifier	Supervised	Results over 0.90 of AUC and 0.91 of MCC for all tested combinations
[19]	Fake news	SVM and RF with cross-validation as a training	Supervised	RF: Accuracy = 0.74 AUC= 0.75, SVM: Accu-racy=0.7
[15]	Cyberbullying	SVM and NN classifiers	Supervised	NN accuracy=92.8% and SVM= 90.3
[14]	Cyberbullying	Convolutional neural net-works (CNNs)	Supervised	95.33% for Dataset 1 and an accuracy of 96.38% for Dataset 2.
[28]	Fake profiles	SVM, NB, and improved Support Vector Machine (ISVM)	Supervised	SVM, ISVM and NB: Accuracy SVM =0.774, Naïve Bayes = 0.773, ISVM= 0.900
[30]	Credibility of tweets	NB, and BN, models, kNN, DT that mainly consisted of C4.5 and RF models and SVM.	Supervised	The accuracy of C4.5 achieved the highest level of accuracy. Average Accuracy C4.5 classifi-er=95.11%
[56]	Fake Infor-mation	TF-IDF 3), with the proposed formula	Unsupervised	Experimental results, F-measure=0.711, preci-sion=0.711, recall =0.88
[11]	Fake accounts	RF, Naïve Bayes and SVM	Supervised	Accuracy (RF= 98%, Naïve Bayes=97.6%, SVM=98%).
[64]	The credibility of the Twitter account	Deep Neural Network (DNN) to as Bot-DenseNet	Hybrid	Bot-Dense model in terms of F1-Score=0.77
[66]	Malicious activity	Deep Neural Network (DNN), LSTM neural network	Hybrid	Achieved nearly perfect detecting accuracy (more than 99%).
[68]	opinion polari-ty	Deep Neural Network (DNN)	Hybrid	Accuracy (%) for movies dataset (81.7), for books dataset (76.0).
[52]	Malicious activity	Deep Neural Network (DNN), NLP	Hybrid	Accuracy ANOVA F-Value and SVM is 0.9898

8 Measures of Performance

Various performance measurements were used to evaluate social bot detection. An accuracy rate is a common approach used to measure performance. It refers to the percentage of accounts that are correctly classified in comparison to the entire sample. However, relying solely on the accuracy rate to evaluate the chosen classifier is insufficient because it does not distinguish between the numbers of correctly classified examples of different classes and this may lead to erroneous conclusions. To validate the results, the majority of the previous literature used classifiers with tenfold and/or fivefold cross-validations. F-measure, precision, and recall were also used in some studies as other measurements of performance assessment. Because the bot detection problem is ultimately a binary classification, such performance measurements are appropriate.

9 Features

Social bot detection relies on a group of selected features to categorize accounts as legitimate or bot accounts. This paper highlights how common features are used to detect social bot accounts in previous work. This includes characteristics such as timing, text usage, and sentiment. It should be clear that a social bot cannot be assumed reliant on a single feature without considering the others [80]. Table 4 summarizes the common features extracted from a full set of features in the reviewed papers to determine whether an account is a human or bot. However, bot-masters can easily develop bots that elude detection by the prediction models based on the use of a few fea-tures. Depending on the bot's objectives, each sort of bot should have distinct qualities.

In general, the extracted features can be used to identify community features by addressing network features. Users' social connections can also be determined and ranked according to the performing content and behavioral analysis. For example, if an account is verified or protected, it is a logical indicator that it is not a bot account. Profile features that can be extracted from metadata such as profile image, screen name, and description may also

indicate the accounts' nature. For example, a default profile image indicates a new user or a bot account. If the temporal pattern such as the average of tweeting and re-tweeting ratios occurs at small intervals, this can be a sign of bot activity. Therefore, using an entropy component as part of the classification system to detect behavior is essential. Furthermore, the frequency with which similar content with URL is posted can be an indicator of a spammer. In other words, the URL feature can be used to detect link farming activity, which is commonly used by spammers and bot accounts. Such features can be used in conjunction with other attributes such as URL and number of links.

The entropy of tweets can also indicate a bot account with malicious intent as shown in [63]. Furthermore, if the number of followers is high in a new account, this may indicate that a such number of followers is fake and the account is a bot. Instead of finding thresholds [81], to find patterns in features, some researchers used a pattern recognition component. This is because algorithms are used to keep track of bot accounts that will eventually leave a pattern that can be utilized to spot bots [44]. Moreover, in unsupervised machine learning, by using the values of a preset set of features, the degree of similarity between a group of social media users is estimated. This survey also reveals that the majority of the exploited features can be classified into four major categories namely, user profile information, post content, posting behavior, and network structure.

Table 4. The important features used in the previous studies for bot detection.

Ref.	Feature	Description	Taxonomy	Type
[10,17,76,27–30,33,43,59,61]	FolloweeFollower	Mean of the no. of followers of a user's followees	Metadata	Account information
	FolloweeFollowerMedian	Median of the no. of followers of a user's followees	Metadata	Account information
	FolloweeFollowerStdDev	Deviation of the no. of followers of a user's followees	Metadata	Account information
	FolloweeFollowerEntropy	The entropy of the no. of followers of a user's followees	Metadata	Account information
	MentionCountRatio	No. of mentions or total no. of tweets	Metadata	Account information
	MentionUniqueRatio	No. of unique mentions/total no. of mentions	Metadata	Account information
	ffratio	Friends-to-followers ratio	Metadata	Account information
	listed	Number of listed tweets in the account	Metadata	Account information
	Volume of tweeting	One spam indicator is unusually high-volume tweeting, which is often bot-generated. This could be measured by a raw count of tweets, or the percentage of tweets posted to a hashtag by a single user.	Metadata	Account information
	Friends_Count	The number of users this account is following	Metadata	Account information
	AccountBackground	Whether the user profile has a background image	Metadata	Account usage
	AccountSourceTweets	Whether the user is the source tweet's author	body of tweet	Account usage
	URL in profile	True if a URL is specified in the account's profile	Metadata	Account usage
	has biography	True if the biography is specified in the account's profile	Metadata	Account usage
	Retweets	The ratio between retweet count and tweet count	Metadata	Account usage
	Replies	The ratio between reply count	Metadata	Account usage
	Favorite	The ratio between the favorite tweet and tweet count	Metadata	Account usage
	Hashtag	The ratio between hashtag count and tweet count	Metadata	Account usage
	url	The ratio between URL count and tweet count	Metadata	Account usage
	Favorites	Number of tweets favorited in this account	Metadata	Account usage
	Language_Code	The BCP 47 code for the user's self-declared user interface language.Justification: Fake accounts tend to have different language_codes on their interface.	Metadata	Account usage
	Char_account	No. of characters in the user's name including white space	Metadata	Account usage
	Coordinates	It represents the geographic location of the tweet.	Metadata	Location
	AccountAge	Total duration from since account created till now	Metadata	Temporal
	AverageTweetCount	No. of tweets/account age	Metadata	Temporal
	ProbWeekend	Probability of a user tweeting on the weekend	Metadata	Temporal
	ProbMorning	Probability of a user tweeting in the morning	Metadata	Temporal
	ProbAfternoon	Probability of a user tweeting in the afternoon	Metadata	Temporal
	ProbEvening	Probability of a user tweeting in the evening	Metadata	Temporal
	ProbNight	Probability of a user tweeting at night	Metadata	Temporal
	Hour-x	Probability of a user tweeting at hour x	Metadata	Temporal
	Weekday-x	Probability of a user tweeting on day x	Metadata	Temporal
	intertime	Average seconds between postings	Metadata	Temporal
	id_created days	No of days id created	Metadata	Temporal
	tweet_year	The year when the tweet was created	Metadata	Temporal
	tweet_month	The month when the tweet was created	Metadata	Temporal
	tweet_day	The day when the tweet was created	Metadata	Temporal
	tweet_hour	The hour when the tweet was created	Metadata	Temporal

	user_created_year	The year when the Twitter account was created	Metadata	Temporal
	user_created_month	The month when the Twitter account was created	Metadata	Temporal
	user_created_day	The day when the Twitter account was created	Metadata	Temporal
	user_created_hour	The hour when the Twitter account was created	Metadata	Temporal
	The count of total words in a tweet;		Bodyof tweet	Tweet-level features
	The count of exclamations;		Bodyof tweet	Tweet-level features
	Entity extraction	URLs or hashtags, particularly photos, etc	Bodyof tweet	Tweet-level features
	Twitter hashtag features	No. of tweets that have at least one hashtag, Avg no. of hashtags per tweet	Bodyof tweet	Tweet-level features
	Agreement	Whether the text has an agreement text.	Bodyof tweet	Tweet-level features
	HashtagAve	Hashtag count / Tweet count	Bodyof tweet	Tweet-level features
	LinkAve	Link count / Tweet count	Bodyof tweet	Tweet-level features

table 2 provides a brief description of the main features' categories that were used in previous literature for bot detection. Some of the researchers used only metadata of Twitter's accounts to detect bots [29,31,34,63], and others used metadata with the content of tweet features to detect bots [29,43,59,10].

As previously stated, one of the major challenges in this research area is the fast-evolving abilities of bots, which causes changes in the values of the discriminating features and failure to detect them. Therefore, discovering robust features is of great interest to detect bots early. Finding vulnerable victims is an example of robust features because their features are relatively stable. Victims are directly connected to fake accounts; they represent a natural "borderline" that separates real accounts from fakes [83]. Figure 4 shows the percentage of methods types for feature extraction, feature selection, and a combination of features extraction and selection used in the reviewed literature.

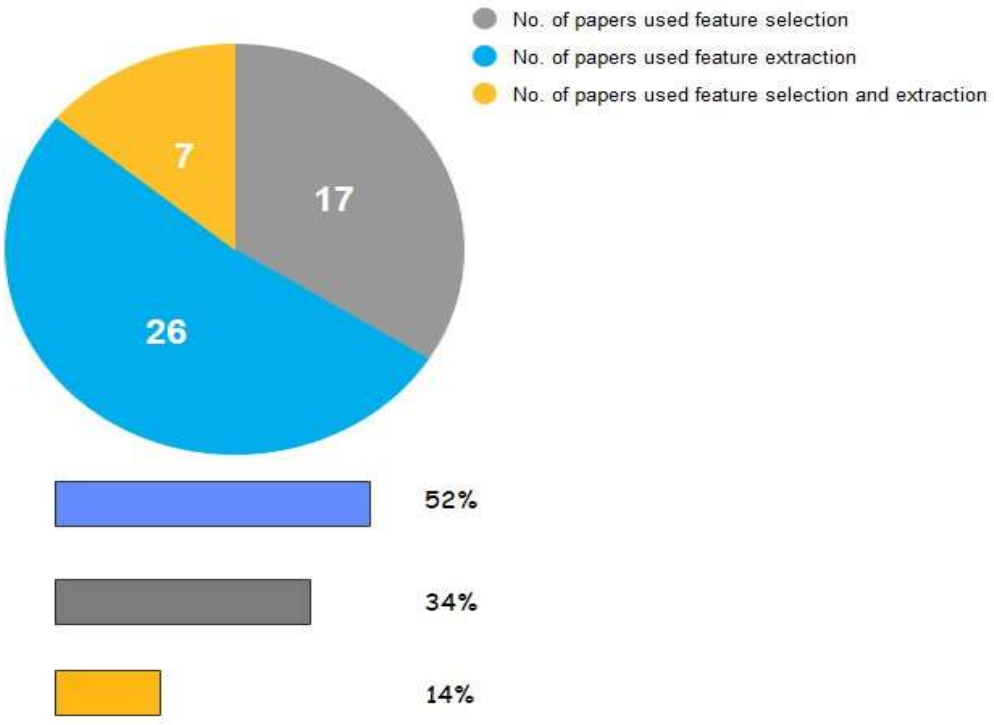

Figure 4. Summary of feature selection, extraction, and combined extraction and selection methods used in the reviewed literature

10 Research gaps

Table 5 shows that the supervised approach is the most commonly used method in which 26 out of 37 studies have adopted it. Unsupervised learning was used in only seven out of 37 studies. Moreover, six out of 27 studies used only a combination of features metadata and text tweet. **"**table 5**"** also presents that the mean value of the eight concepts is 35.81% which can be considered low for such a hot topic. The mean value is calculated using "Eq. (1)" [84]:

$$mean=(total\ frequence)/(8\ key\ concepts\times numer\ of\ ref)\times 100 \quad (1)$$

$$mean=107/(8\times 37)\times 100=36.14\%$$

The impact of eight principles on bot detection is investigated as a wider framework for Twitter bots to gain deeper insights into the researched area. This method promotes SMB-detection development and significantly improves feature selection and extraction. The average result is 35.81%, which has a slight significant impact on SMB- detection. Thus, all eight concepts must be applied completely.

Many challenges obstruct the advancement of research into detecting SMBs. Collecting genuine accurate information, extracting strong characteristics, discovering an effective way to distinguish bots from real users, and evaluating the efficacy of current methods are just some of these obstacles. Some gaps in the earlier literature on bot detection are discussed here.

First, the majority of research focuses on detecting bots in supervised learning approaches, as shown in table 5. However, this can share a common disadvantage, relating to the difficulty of preparing a reliable training dataset. In particular, this issue is clear with assigning accurate class labels to sessions of camouflaged robots. On the contrary, this limitation does not affect unsupervised learning techniques because they attend to learn intrinsic data properties from unlabeled training samples. The adaptive adversary is perhaps another problem for supervised learning techniques as for rules-based implementations of bot detection techniques. This problem arises because new bots may be evolved unlike the available bots [77], whereas unsupervised learning can adapt along with any adversary. Unsupervised learning solves the problem by learning the data and classifying it without any labels. The labels can be added after the data has been classified which is much easier. It is very helpful in finding patterns in data, which are not possible to find using normal methods [85]. Based on these research outcomes, a few studies classified sessions based on unsupervised learning in which the majority of the used approaches are focused on investigating the ability to partition bots and humans into separate clusters and explore session properties depending on this particular cluster [85].

Second, previous literature needs to release the datasets used to be reachable by the research community. This can aid in the training, testing, and evaluation of new models. In addition, fresh public datasets are required, as well as new research that implements existing detection models and tests them on the same real-world dataset.

Third, vague areas that require more investigation are noticed. A new direction that starts attracting the attention of researchers is detecting each bots category independently. However, detecting all bots types using one general model and the same or similar features may produce better and more accurate findings. Furthermore, it is unclear if detecting each type of bot separately is adequate. Some techniques, such as supervised learning techniques, have been intensively investigated in this field. Many methods, on the other hand, require ongoing investigation to better comprehend, develop, or discover new findings. The scientific community is urged to develop new methodologies and/or improve on existing techniques.

Fourth, some of the offered-mentioned approaches have only been trained and tested on a limited dataset. This could affect its overall scalability. Aside from classification methods, clustering approaches are used in several studies in the literature reviewed to identify spam in bulk. Although clustering approaches do not require pre-labeled data, they are limited by the growing scale of social networks, which is a major open challenge in detecting spam campaigns.

The fifth issue is spam drift. This phenomenon means that features of spam tweets are changing over time, as evidenced by the retrieved Twitter data in earlier research. Spam drift can happen due to spammers' frequent modification of attack tactics to avoid being detected by spam detectors. Unfortunately, machine learning algorithms are not updated with various spam tweets, resulting in a significant drop in classifier performance. Some papers attempted to provide solutions to this problem, but it still needs further effort.

Finally, extracting the influential features that may help machine learning algorithms detect Twitter bots with high accuracy is important, especially if a predetermining of the type of bot exists. Thus, feature extraction is focused on one direction and not on all types of bots. This situation can help identify and extract important features. Based on the reviewed papers, the attention was focused on the metadata of Twitter's accounts as well as the application of hybrid methods in the process of selecting features, which in turn, consumes a long time. Using complex methods in the process of determining features is still a question. A balance between accuracy and speed in bot detection must be considered. Furthermore, to avoid falling into the curse of dimensions when selecting features from metadata and the tweets' text, dimensionality reduction is used to reduce the feature space with consideration of a set of principal features.

Table 5. A Summary of bot detection approaches using features extraction and selection methods

Ref	FS: Feature Selection	FE: Feature Extraction	FS&FE	Text tweet	Metadata	Text and metadata	Supervised	Unsupervised
[36]			✓	✓			✓	
[59]			✓	✓			✓	
[67]			✓	✓				✓
[47]			✓	✓				
[58]	✓			✓			✓	
[38]	✓			✓			✓	
[8]	✓			✓			✓	
[60]		✓		✓			✓	
[50]		✓		✓			✓	
[37]		✓		✓				✓
[56]		✓		✓				✓
[48]		✓		✓			✓	
[14]		✓		✓			✓	
[17]		✓		✓			✓	
[51]		✓		✓			✓	
[32]		✓		✓				✓
[62]		✓		✓				
[54]		✓		✓				✓
[49]		✓		✓				✓
[44]		✓		✓			✓	
[55]		✓		✓			✓	
[71]		✓		✓			✓	
[63]		✓		✓				
[35]			✓	✓			✓	
[53]		✓		✓				✓
[18]			✓			✓	✓	
[43]			✓			✓	✓	
[11]		✓				✓	✓	
[30]		✓				✓	✓	
[61]		✓				✓	✓	
[64]		✓				✓		
[34]	✓				✓		✓	
[45]	✓				✓		✓	

[28]		✓			✓		✓	
[31]		✓			✓		✓	
[29]		✓			✓		✓	
[19]		✓			✓		✓	
Frequency	**5**	**25**	**7**	**25**	**6**	**6**	**26**	**7**
Percentage	**13.5%**	**67.5%**	**18.9%**	**67.5%**	**16.2%**	**16.2%**	**70.2%**	**18.9%**

11 Conclusions

Twitter Bots are abused to manipulate public opinion and gain social media power due to the massive increase in social media influence on people's opinions. Even though researchers developed powerful models to detect social media bot accounts, bot-masters are rapidly developing new bots to avoid detection. This study discussed various aspects of bot detection methods used in recent studies on Twitter and highlighted the key Twitter features that can be selected or extracted to detect bots accurately. Moreover, this research aimed to investigate deeply the main features of the previous literature based on four criteria namely, datasets used, features, classifiers, and performance measures. This objective was achieved with a focus on reshaping a new architecture in the techniques of features extrapolation because significant relationships between malicious bots and features architecture on the improvement of existing machine learning applications exist and the development of high standards of Twitter bot detection. This study also provided detailed descriptive information about Twitter features that interact with the malicious bots classification. Based on the inclusion criteria of this research, the included studies were thoroughly and systematically examined to highlight the benefits, challenges, gaps, and recommendations related to bot detection. Thus, solutions to the challenges and issues raised in this study were presented. Therefore, this research provided a wider grounding to the Twitter malicious bot by exploring the impact of features, annotated datasets, methods, and various machine learning techniques to encourage the application of artificial intelligence on Twitter bot detection.

Acknowledgments

My great thanks to some of my colleagues at the University of Babylon for their kind help that leads to addressing some issues in this work.

References

[1] M.M. Ibañez, R.R. Rosa, L.N.F. Guimarães, Sentiment analysis applied to analyze society's emotion in two different context of social media data, Intel. Artif. 23 (2020) 66–84. https://doi.org/10.4114/INTARTIF.VOL23ISS66PP66-84.

[2] B.Y.E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, P96-Ferrara, (2016).

[3] Subrahmanian et al. (2016), The DARPA TWITTER BOT CHALLENGE, 49 (2016) 38–46. https://doi.org/10.1109/MC.2016.183.

[4] S. Stieglitz, F. Brachten, B. Ross, A.K. Jung, Do social bots dream of electric sheep? a categorisation of social media bot accounts, Proc. 28th Australas. Conf. Inf. Syst. ACIS 2017. (2017) 1–11.

[5] A. Xu, Z. Liu, Y. Guo, V. Sinha, R. Akkiraju, A new chatbot for customer service on social media, Conf. Hum. Factors Comput. Syst. - Proc. 2017-May (2017) 3506–3510. https://doi.org/10.1145/3025453.3025496.

[6] A. Wilkie, M. Michael, M. Plummer-Fernandez, Speculative method and Twitter: Bots, energy and three conceptual characters, Sociol. Rev. 63 (2015) 79–101. https://doi.org/10.1111/1467-954X.12168.

[7] S. Cresci, F. Lillo, D. Regoli, S. Tardelli, M. Tesconi, Cashtag piggybacking: Uncovering spam and bot activity in stock microblogs on twitter, ACM Trans. Web. 13 (2019) 1–26. https://doi.org/10.1145/3313184.

[8] E. Elakkiya, S. Selvakumar, R.L. Velusamy, CIFAS: Community Inspired Firefly Algorithm with fuzzy

cross-entropy for feature selection in Twitter Spam detection, 2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020. (2020). https://doi.org/10.1109/ICCCNT49239.2020.9225321.

[9] J. Ratkiewicz, M. Meiss, M. Conover, B. Gonçalves, A. Flammini, F. Menczer, Detecting and Tracking Political Abuse in Social Media, Proc. Fifth Int. AAAI Conf. Weblogs Soc. Media. (2011) 297.

[10] M. Alrubaian, M. Al-Qurishi, S.M.M. Rahman, A. Alamri, A novel prevention mechanism for Sybil attack in online social network, 2015 2nd World Symp. Web Appl. Networking, WSWAN 2015. (2015). https://doi.org/10.1109/WSWAN.2015.7210347.

[11] R.R. Rostami, S. Karbasi, Detecting fake accounts on twitter social network using multi-objective hybrid feature selection approach, Webology. 17 (2020). https://doi.org/10.14704/WEB/V17I1/A204.

[12] N. Abokhodair, D. Yoo, D.W. McDonald, Dissecting a social Botnet: Growth, content and influence in twitter, CSCW 2015 - Proc. 2015 ACM Int. Conf. Comput. Coop. Work Soc. Comput. (2015) 839–851. https://doi.org/10.1145/2675133.2675208.

[13] A. Elyashar, M. Fire, D. Kagan, Y. Elovici, Guided socialbots: Infiltrating the social networks of specific organizations' employees, AI Commun. 29 (2016) 87–106. https://doi.org/10.3233/AIC-140650.

[14] L. Ketsbaia, B. Issac, X. Chen, Detection of hate tweets using machine learning and deep learning, Proc. - 2020 IEEE 19th Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. 2020. (2020) 751–758. https://doi.org/10.1109/TrustCom50675.2020.00103.

[15] J. Hani, M. Nashaat, M. Ahmed, Z. Emad, E. Amer, A. Mohammed, Social media cyberbullying detection using machine learning, Int. J. Adv. Comput. Sci. Appl. 10 (2019) 703–707. https://doi.org/10.14569/ijacsa.2019.0100587.

[16] E. Alothali, N. Zaki, E.A. Mohamed, H. Alashwal, Detecting Social Bots on Twitter: A Literature Review, Proc. 2018 13th Int. Conf. Innov. Inf. Technol. IIT 2018. (2019) 175–180. https://doi.org/10.1109/INNOVATIONS.2018.8605995.

[17] R.R. Rout, G. Lingam, D.V.L.N. Somayajulu, Detection of Malicious Social Bots Using Learning Automata with URL Features in Twitter Network, IEEE Trans. Comput. Soc. Syst. 7 (2020) 1004–1018. https://doi.org/10.1109/TCSS.2020.2992223.

[18] N. Patel, M. Panchal, Survey of Feature-based Bot Detection Methodologies, 2020 (2020) 1–4.

[19] P.G. Pratama, N.A. Rakhmawati, Social bot detection on 2019 Indonesia president candidate's supporter's tweets, Procedia Comput. Sci. 161 (2019) 813–820. https://doi.org/10.1016/j.procs.2019.11.187.

[20] D.M. Beskow, K.M. Carley, Its all in a name: detecting and labeling bots by their name, Comput. Math. Organ. Theory. 25 (2019) 24–35. https://doi.org/10.1007/s10588-018-09290-1.

[21] J. Echeverria, S. Zhou, Discovery, retrieval, and analysis of the'Star wars' botnet in twitter, Proc. 2017 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2017. (2017) 1–8. https://doi.org/10.1145/3110025.3110074.

[22] Z. Manbari, F. AkhlaghianTab, C. Salavati, Hybrid fast unsupervised feature selection for high-dimensional data, Expert Syst. Appl. 124 (2019) 97–118. https://doi.org/10.1016/j.eswa.2019.01.016.

[23] C. Cooper, A. Booth, J. Varley-Campbell, N. Britten, R. Garside, Defining the process to literature searching in systematic reviews: A literature review of guidance and supporting studies, BMC Med. Res. Methodol. 18 (2018) 1–14. https://doi.org/10.1186/s12874-018-0545-3.

[24] W.M. Bramer, M.L. Rethlefsen, J. Kleijnen, O.H. Franco, Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study, Syst. Rev. 6 (2017) 1–12. https://doi.org/10.1186/s13643-017-0644-y.

[25] M. Gusenbauer, N.R. Haddaway, Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources, Res. Synth. Methods. 11 (2020) 181–217. https://doi.org/10.1002/jrsm.1378.

[26] G.M. Tawfik, K.A.S. Dila, M.Y.F. Mohamed, D.N.H. Tam, N.D. Kien, A.M. Ahmed, N.T. Huy, A step by step guide for conducting a systematic review and meta-analysis with simulation data, Trop. Med. Health. 47 (2019) 1–9. https://doi.org/10.1186/s41182-019-0165-6.

[27] H. Kamioka, Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement, Japanese Pharmacol. Ther. 47 (2019) 1177–1185.

[28] A. K. Ojo, Improved Model for Detecting Fake Profiles in Online Social Network: A Case Study of Twitter, J. Adv. Math. Comput. Sci. 33 (2019) 1–17. https://doi.org/10.9734/jamcs/2019/v33i430187.

[29] N.S. Murugan, G.U. Devi, Feature extraction using LR-PCA hybridization on twitter data and classification accuracy using machine learning algorithms, Cluster Comput. 22 (2019) 13965–13974. https://doi.org/10.1007/s10586-018-2158-3.

[30] M.S. Al-Rakhami, A.M. Al-Amri, Lies Kill, Facts Save: Detecting COVID-19 Misinformation in Twitter,

IEEE Access. 8 (2020) 155961–155970. https://doi.org/10.1109/ACCESS.2020.3019600.
[31] A. Balestrucci, R. De Nicola, M. Petrocchi, C. Trubiani, A behavioural analysis of credulous Twitter users, Online Soc. Networks Media. 23 (2021). https://doi.org/10.1016/j.osnem.2021.100133.
[32] L. Wu, N.A. Dodoo, T.J. Wen, L. Ke, Understanding Twitter conversations about artificial intelligence in advertising based on natural language processing, Int. J. Advert. (2021). https://doi.org/10.1080/02650487.2021.1920218.
[33] M.B. Abdulrazzaq, J.N. Saeed, A Comparison of Three Classification Algorithms for Handwritten Digit Recognition, 2019 Int. Conf. Adv. Sci. Eng. ICOASE 2019. (2019) 58–63. https://doi.org/10.1109/ICOASE.2019.8723702.
[34] J. Rodríguez-Ruiz, J.I. Mata-Sánchez, R. Monroy, O. Loyola-González, A. López-Cuevas, A one-class classification approach for bot detection on Twitter, Comput. Secur. 91 (2020). https://doi.org/10.1016/j.cose.2020.101715.
[35] A. Rasool, R. Tao, M. Kamyab, S. Hayat, GAWA-A feature selection method for hybrid sentiment classification, IEEE Access. 8 (2020) 191850–191861. https://doi.org/10.1109/ACCESS.2020.3030642.
[36] A. Kumar, A. Jaiswal, Swarm intelligence based optimal feature selection for enhanced predictive sentiment accuracy on twitter, Multimed. Tools Appl. 78 (2019) 29529–29553. https://doi.org/10.1007/s11042-019-7278-0.
[37] H. Rehioui, A. Idrissi, New clustering algorithms for twitter sentiment analysis, IEEE Syst. J. 14 (2020) 530–537. https://doi.org/10.1109/JSYST.2019.2912759.
[38] N.K. Suchetha, A. Nikhil, P. Hrudya, Comparing the wrapper feature selection evaluators on twitter sentiment classification, ICCIDS 2019 - 2nd Int. Conf. Comput. Intell. Data Sci. Proc. (2019). https://doi.org/10.1109/ICCIDS.2019.8862033.
[39] M. Amoozegar, B. Minaei-Bidgoli, Optimizing multi-objective PSO based feature selection method using a feature elitism mechanism, Expert Syst. Appl. 113 (2018) 499–514. https://doi.org/10.1016/j.eswa.2018.07.013.
[40] P. Agrawal, H.F. Abutarboush, T. Ganesh, A.W. Mohamed, Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019), IEEE Access. 9 (2021) 26766–26791. https://doi.org/10.1109/ACCESS.2021.3056407.
[41] M. Iqbal, M.M. Abid, M.N. Khalid, A. Manzoor, Review of feature selection methods for text classification, Int. J. Adv. Comput. Res. 10 (2020) 138–152. https://doi.org/10.19101/ijacr.2020.1048037.
[42] G. Ansari, T. Ahmad, M.N. Doja, Hybrid Filter–Wrapper Feature Selection Method for Sentiment Classification, Arab. J. Sci. Eng. 44 (2019) 9191–9208. https://doi.org/10.1007/s13369-019-04064-6.
[43] C. Fu, Y. Du, B. Lyu, Q. Zhou, R. Hu, P. Jia, Y. Zhou, Forwarding behavior prediction based on microblog user features, IEEE Access. 8 (2020) 95170–95187. https://doi.org/10.1109/ACCESS.2020.2995411.
[44] O. Loyola-Gonzalez, R. Monroy, J. Rodriguez, A. Lopez-Cuevas, J.I. Mata-Sanchez, Contrast Pattern-Based Classification for Bot Detection on Twitter, IEEE Access. 7 (2019) 45800–45817. https://doi.org/10.1109/ACCESS.2019.2904220.
[45] F. Thabtah, F. Kamalov, S. Hammoud, S.R. Shahamiri, Least Loss: A simplified filter method for feature selection, Inf. Sci. (Ny). 534 (2020) 1–15. https://doi.org/10.1016/j.ins.2020.05.017.
[46] M. Monirul Kabir, M. Monirul Islam, K. Murase, A new wrapper feature selection approach using neural network, Neurocomputing. 73 (2010) 3273–3283. https://doi.org/10.1016/j.neucom.2010.04.003.
[47] F. Akba, I.T. Medeni, M.S. Guzel, I. Askerzade, Assessment of iterative semi-supervised feature selection learning for sentiment analyses: Digital currency markets, Proc. - 14th IEEE Int. Conf. Semant. Comput. ICSC 2020. (2020) 459–463. https://doi.org/10.1109/ICSC.2020.00088.
[48] B.H. Back, I.K. Ha, Comparison of sentiment analysis from large twitter datasets by naive bayes and natural language processing methods, J. Inf. Commun. Converg. Eng. 17 (2019) 239–245. https://doi.org/10.6109/jicce.2019.17.4.239.
[49] Y. Madani, M. Erritali, J. Bengourram, F. Sailhan, A multilingual fuzzy approach for classifying Twitter data using fuzzy logic and semantic similarity, Neural Comput. Appl. 32 (2020) 8655–8673. https://doi.org/10.1007/s00521-019-04357-9.
[50] L. Mandloi, R. Patel, Twitter sentiments analysis using machine learninig methods, 2020 Int. Conf. Emerg. Technol. INCET 2020. (2020) 1–5. https://doi.org/10.1109/INCET49848.2020.9154183.
[51] F. Wei, U.T. Nguyen, Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings, Proc. - 1st IEEE Int. Conf. Trust. Priv. Secur. Intell. Syst. Appl. TPS-ISA 2019. (2019) 101–109. https://doi.org/10.1109/TPS-ISA48467.2019.00021.

[52] L. Ilias, I. Roussaki, Detecting malicious activity in Twitter using deep learning techniques, Appl. Soft Comput. 107 (2021) 107360. https://doi.org/10.1016/j.asoc.2021.107360.

[53] R. Ahuja, A. Chug, S. Kohli, S. Gupta, P. Ahuja, The impact of features extraction on the sentiment analysis, Procedia Comput. Sci. 152 (2019) 341–348. https://doi.org/10.1016/j.procs.2019.05.008.

[54] A. Aleroud, N. Abu-Alsheeh, E. Al-Shawakfa, A graph proximity feature augmentation approach for identifying accounts of terrorists on twitter, Comput. Secur. 99 (2020) 102056. https://doi.org/10.1016/j.cose.2020.102056.

[55] G.A. De Souza, M. Da Costa-Abreu, Automatic offensive language detection from Twitter data using machine learning and feature selection of metadata, Proc. Int. Jt. Conf. Neural Networks. (2020). https://doi.org/10.1109/IJCNN48605.2020.9207652.

[56] K. Sato, J. Wang, Z. Cheng, Credibility Evaluation of Twitter-Based Event Detection by a Mixing Analysis of Heterogeneous Data, IEEE Access. 7 (2019) 1095–1106. https://doi.org/10.1109/ACCESS.2018.2886312.

[57] M. Fernandez, H. Alani, Online Misinformation: Challenges and Future Directions, Web Conf. 2018 - Companion World Wide Web Conf. WWW 2018. (2018) 595–602. https://doi.org/10.1145/3184558.3188730.

[58] M.Z. Ansari, T. Ahmad, A. Fatima, Feature Selection on Noisy Twitter Short Text Messages for Language Identification, Int. J. Recent Technol. Eng. 8 (2019) 10505–10510. https://doi.org/10.35940/ijrte.d4360.118419.

[59] M. Bibi, M.S.A. Nadeem, I.H. Khan, S.O. Shim, I.R. Khan, U. Naqvi, W. Aziz, Class association and attribute relevancy based imputation algorithm to reduce twitter data for optimal sentiment analysis, IEEE Access. 7 (2019) 136535–136544. https://doi.org/10.1109/ACCESS.2019.2942112.

[60] A. Go, R. Bhayani, L. Huang, Twitter Sentiment Classification using Distant Supervision, Processing. (2009) 1–6.

[61] A. Pandya, M. Oussalah, P. Monachesi, P. Kostakos, On the use of distributed semantics of tweet metadata for user age prediction, Futur. Gener. Comput. Syst. 102 (2020) 437–452. https://doi.org/10.1016/j.future.2019.08.018.

[62] M. Wischnewski, A. Bruns, T. Keller, Shareworthiness and Motivated Reasoning in Hyper-Partisan News Sharing Behavior on Twitter, Digit. Journal. 9 (2021) 549–570. https://doi.org/10.1080/21670811.2021.1903960.

[63] A. Derhab, R. Alawwad, K. Dehwah, N. Tariq, F.A. Khan, J. Al-Muhtadi, Tweet-Based Bot Detection Using Big Data Analytics, IEEE Access. 9 (2021) 65988–66005. https://doi.org/10.1109/ACCESS.2021.3074953.

[64] D. Martin-Gutierrez, G. Hernandez-Penaloza, A.B. Hernandez, A. Lozano-Diez, F. Alvarez, A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers, IEEE Access. 9 (2021) 54591–54601. https://doi.org/10.1109/ACCESS.2021.3068659.

[65] S. Kudugunta, E. Ferrara, Deep neural networks for bot detection, Inf. Sci. (Ny). 467 (2018) 312–322. https://doi.org/10.1016/j.ins.2018.08.019.

[66] H. Ping, S. Qin, A social bots detection model based on deep learning algorithm, Int. Conf. Commun. Technol. Proceedings, ICCT. 2019-Octob (2019) 1435–1439. https://doi.org/10.1109/ICCT.2018.8600029.

[67] H. Utama, Sentiment analysis in airline tweets using mutual information for feature selection, 2019 4th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2019. (2019) 295–300. https://doi.org/10.1109/ICITISEE48480.2019.9003903.

[68] I. Jost, J.F. Valiati, Deep learning applied on refined opinion review datasets, Intel. Artif. 21 (2018) 91–102. https://doi.org/10.4114/INTARTIF.VOL21ISS62PP91-102.

[69] H. Khalil, M.U.S. Khan, M. Ali, Feature Selection for Unsupervised Bot Detection, 2020 3rd Int. Conf. Comput. Math. Eng. Technol. Idea to Innov. Build. Knowl. Econ. ICoMET 2020. (2020) 1–7. https://doi.org/10.1109/iCoMET48670.2020.9074131.

[70] R. Sawhney, R.R. Shah, V. Bhatia, C.T. Lin, S. Aggarwal, M. Prasad, Exploring the Impact of Evolutionary Computing based Feature Selection in Suicidal Ideation Detection, IEEE Int. Conf. Fuzzy Syst. 2019-June (2019) 1–6. https://doi.org/10.1109/FUZZ-IEEE.2019.8858989.

[71] J.P. Carpenter, K.B. Staudt Willet, M.J. Koehler, S.P. Greenhalgh, Spam and Educators' Twitter Use: Methodological Challenges and Considerations, TechTrends. 64 (2020) 460–469. https://doi.org/10.1007/s11528-019-00466-3.

[72] M. Li, H. Wang, L. Yang, Y. Liang, Z. Shang, H. Wan, Fast hybrid dimensionality reduction method for

classification based on feature selection and grouped feature extraction, Expert Syst. Appl. 150 (2020) 113277. https://doi.org/10.1016/j.eswa.2020.113277.

[73] Z. Chen, D. Subramanian, An Unsupervised Approach to Detect Spam Campaigns that Use Botnets on Twitter, (2018) 1–7. http://arxiv.org/abs/1804.05232.

[74] N. Chavoshi, H. Hamooni, A. Mueen, DeBot: Twitter Bot Detection via Warped Correlation, (2017) 817–822. https://doi.org/10.1109/icdm.2016.0096.

[75] D.M. Beskow, K.M. Carley, Its all in a name: detecting and labeling bots by their name, Comput. Math. Organ. Theory. 25 (2019) 24–35. https://doi.org/10.1007/s10588-018-09290-1.

[76] S. Barbon, G.F.C. Campos, G.M. Tavares, R.A. Igawa, M.L. Proença, R.C. Guido, Detection of human, legitimate bot, and malicious bot in online social networks based on wavelets, ACM Trans. Multimed. Comput. Commun. Appl. 14 (2018). https://doi.org/10.1145/3183506.

[77] P.A. Chew, Searching for unknown unknowns: Unsupervised bot detection to defeat an adaptive adversary, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 10899 LNCS (2018) 357–366. https://doi.org/10.1007/978-3-319-93372-6_39.

[78] B. Pang, L. Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, (2004). http://arxiv.org/abs/cs/0409058.

[79] P. Shi, Z. Zhang, K.K.R. Choo, Detecting Malicious Social Bots Based on Clickstream Sequences, IEEE Access. 7 (2019) 28855–28862. https://doi.org/10.1109/ACCESS.2019.2901864.

[80] Z. Chu, S. Gianvecchio, H. Wang, S. Jajodia, Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?, IEEE Trans. Dependable Secur. Comput. 9 (2012) 811–824. https://doi.org/10.1109/TDSC.2012.75.

[81] A. Rauchfleisch, J. Kaiser, The False positive problem of automatic bot detection in social science research, PLoS One. 15 (2020). https://doi.org/10.1371/journal.pone.0241045.

[82] R.J. Oentaryo, J.W. Low, E.P. Lim, Chalk and cheese in Twitter: Discriminating personal and organization accounts, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 9022 (2015) 465–476. https://doi.org/10.1007/978-3-319-16354-3_51.

[83] Yazan Boshmaf, Íntegro-Leveraging Victim Prediction for Robust Fake Account Detection in Large Scale OSNs.pdf, (2016) 142–168. https://doi.org/https://doi.org/10.1016/j.cose.2016.05.005.

[84] Z.T. Al-Qaysi, M.A. Ahmed, N.M. Hammash, A.F. Hussein, A.S. Albahri, M.S. Suzani, B. Al-Bander, M.L. Shuwandy, M.M. Salih, Systematic review of training environments with motor imagery brain–computer interface: Coherent taxonomy, open issues and recommendation pathway solution, Health Technol. (Berl). 11 (2021) 783–801. https://doi.org/10.1007/s12553-021-00560-8.

[85] S. Rovetta, G. Suchacka, F. Masulli, Bot recognition in a Web store: An approach based on unsupervised learning, J. Netw. Comput. Appl. 157 (2020). https://doi.org/10.1016/j.jnca.2020.102577.

.