sustainability

MDPI

Article

Twitter Bot Detection Using Diverse Content Features and Applying Machine Learning Algorithms

Fawaz Khaled Alarfaj [1,*], Hassaan Ahmad [2], Hikmat Ullah Khan [2,*], Abdullah Mohammaed Alomair [1], Naif Almusallam [1] and Muzamil Ahmed [2]

[1] Management Information Systems, School of Business, King Faisal University, Hofuf 31982, Saudi Arabia
[2] Department of Computer Science, COMSATS University Islamabad, Wah Campus, Wah Cantt 47040, Pakistan
* Correspondence: falarfaj@kfu.edu.sa (F.K.A.); hikmat.ullah@ciitwah.edu.pk (H.U.K.)

Abstract: A social bot is an intelligent computer program that acts like a human and carries out various activities in a social network. A Twitter bot is one of the most common forms of social bots. The detection of Twitter bots has become imperative to draw lines between real and unreal Twitter users. In this research study, the main aim is to detect Twitter bots based on diverse content-specific feature sets and explore the use of state-of-the-art machine learning classifiers. The real-world data from Twitter is scrapped using Twitter API and is pre-processed using standard procedure. To analyze the content of tweets, several feature sets are proposed, such as message-based, part-of-speech, special characters, and sentiment-based feature sets. Min-max normalization is considered for data normalization and then feature selection methods are applied to rank the top features within each feature set. For empirical analysis, robust machine learning algorithms such as deep learning (DL), multilayer perceptron (MLP), random forest (RF), naïve Bayes (NB), and rule-based classification (RBC) are applied. The performance evaluation based on standard metrics of precision, accuracy, recall, and f-measure reveals that the proposed approach outperforms the existing studies in the relevant literature. In addition, we explore the effectiveness of each feature set for the detection of Twitter bots.

Keywords: bot detection; machine learning; deep learning; feature engineering; cyber security

Citation: Alarfaj, F.K.; Ahmad, H.; Khan, H.U.; Alomair, A.M.; Almusallam, N.; Ahmed, M. Twitter Bot Detection Using Diverse Content Features and Applying Machine Learning Algorithms. *Sustainability* **2023**, *15*, 6662. https://doi.org/10.3390/su15086662

Academic Editors: Ming Hour Yang, Vijayalakshmi Murugesan, Mercy Shalinie Selvaraj and Gwanggil Jeon

Received: 13 February 2023
Revised: 29 March 2023
Accepted: 12 April 2023
Published: 14 April 2023

1. Introduction

The social web facilitates its users to generate their own content. The users can share their views, ideas, and opinions regarding diverse topics on the social web channels. The channels create a massive amount of data due to various users' activities [1]. Nowadays, we share and view content on social media regarding politics, news, and other issues. The users follow their favorite celebrities and share their views on various social networks. Microblogging is a mixture of messaging and blogging. It allows the users to create their material through videos, images, and text.

In many cases, the users may not be human but a bot [2] behaving like a human. The bots are the software users of a social web channel known as the social bot. These bots may perform any activity, such as posting new content, liking a post of other users, etc. A bot can also share content through various negative and malicious posts. The classification of posts is essential in identifying users' positive or negative images [3]. Twitter is a widely used microblog launched in 2006. The simplicity of Twitter is the main reason for its popularity among users. It accepts text-based posts commonly known as 'Tweets'. Twitter allows a 280-character limit per tweet. It also allows 'special characters such as @ that help in tagging a person and has tag #, which represents a topic. Twitter is a popular and widely used social network with 330 million monthly active users worldwide. We observe that Twitter has become a successful web application for business promotion [3], political campaigning, and disaster communication. The greatest challenge to understanding the

bots for their detection is to understand the modern bots. Early bots can only perform one activity, and that is the posting of one piece of content automatically. These types of bots were straightforward to detect by focusing on a large volume of generated content. Nowadays, Twitter bot detection is essential for many reasons, including that a bot can share various negative and malicious content. A bot can also post on behalf of some politician or celebrity. In 2011, a honeypot was implemented by James Caverlee's team at Texas A&M University, which had the ability to identify a considerable number of bots at a time [4]. The team created some Twitter bot accounts to generate tweets making no sense. Thus, humans would never be interested in these kinds of tweets. These accounts attracted many followers. After inspection, it was revealed that these were bots that were following random accounts just to enhance their social circles.

Bots can likewise ruin the headway of open strategy by showing the impact of a popular development of opposite views or public popularity of political dialogue seen in social networks [5]. Internet-based life impacts can be changed by them, misleadingly growing the crowd of only a few people [6], and just for business or political purposes. They can also demolish the notoriety of an organization [7]. The informative intensity of example-based grouping may aid in distinguishing support, including bot movement in social administration. At that point, clarification can be utilized to justify the lawful move or formal protest for a record container. Likewise, it can be connected to admonishing client accounts for conceivable abuse. Bots ruin the advancement of public policy by rejecting popular opinions or by taking part in some abstract political discussion [8]. Even bots can change the perception of social media effects, such as fake followers of a few accounts [6], or the company image can be destroyed for marketing or gaining influencers [9]. The strength of pattern-based classification helps report a bot activity and presence on a social platform. Social network users can also be warned about exploiting their accounts. As far as we know, it can never be done without applying pattern-based classification of bot detection [10].

In this research study, we proposed a bot detection model based on using machine learning and deep learning algorithms. The main contributions of this study can be summarized as follows:

1. Preparation of Twitter Bot Data by first scraping data of over 11,000 tweets belonging to Bots as well as humans.
2. Conducting feature engineering by extraction of message-based, part-of-speech-based, special characters-based, word frequency-based, and sentiment-based content-based features from given tweets.
3. Application of feature selection algorithms such as info gain, gain ratio, and relief-F to identify significant feature sets.
4. Employing deep learning and machine learning algorithms including deep neural network (DNN), multilayer perceptron (MLP), random forest (RF), and rule-based classification (RBC) for Twitter bot detection.
5. Evaluation of the proposed model over-extracted tweets dataset to assess the performance of the proposed model and comparison with state-of-the-art Twitter bot detection methods.

The remainder of the paper is organized as follows: Section 2 presents the most related studies on Twitter bot detection, Section 3 presents the background concepts and proposed research methodology, Section 4 presents the dataset and the discussion on experimental results, and finally, Section 5 presents the conclusion based on obtained results.

2. Related Work

In this section, we discuss the most relevant literature to bot detection, which is a generic type of research domain and then review work already done on Twitter bot detection.

2.1. Social Bot Detection

Social bots have become common nowadays. As discussed earlier, social bots are computer programs that continuously post various content over social media, and the approaches usually rely on inspecting the social graph and identification of bots using classification [11]. Social bots are created to achieve some goals and accomplish some motives. Motives could be political, such as election campaigns [12], or stock-exchange manipulations, among other activities. Bots can even carry out complex communication types such as chatting with other people and posting comments on posts. Moreover, social bots are also detected based on the logic when too many other users are added as a friend [13]. The aim of social bots varies; some users might have their most significant objective only to add new friends to enhance their social circle [14].

Like a typical human user, the social bot can play a positive or negative role. Depending upon the usage of social bots, they may be a serious threat to social networks as they can be used to spread false news or information without any credibility and authenticity [15]. Furthermore, bots can be used to spread and generate content that could alter and mimic one's behavior to manipulate ideas [16,17]. These ideas can derive public opinion, which could result in achieving adverse effects on the public [18]. One could witness those social activities over the internet as ubiquitous, and almost everyone with internet access is making a presence there. The increase in users using social platforms increase the risk of getting affected by the opinions of social bots and, consequently, a higher chance of impacting society negatively. One of the major risks of social bots include false information alerts. They have been spreading misinformation (false news) to a greater extent [14].

Automation and ease of access have streamlined the way to carry out malicious activities on social media, therefore, compromising its reliability. Due to this, it has become imperative to identify the social bots since the false information being spread by the social bots can get mixed with true (reliable) information [17]. If not identified, the unprecedented results could further exacerbate the situation if the false news goes viral. Graph-based social bot identification methods are used to identify fake identities and Sybils accounts by examining the structure of the graph. Such methods, such as SybilRank, assume that sybil accounts have limited connections to genuine users, and are mainly associated with other Sybil accounts. These methods are effective in identifying large groups of integrated Sybil accounts [19]. The adequacy of such recognition systems is bound by the conduct of suspicion that genuine users will not connect with obscure accounts. A huge penetration of bots on Facebook demonstrated that over 20% of genuine accounts acknowledge friendship demands indiscriminately, and nearly 60% acknowledge demands from users with a minimum of one mutual friend in common [20]. Globally penetrated bots over a social network are difficult to recognize based on the structure of network information. Manual detection using socially justified users and complementary detection techniques helps in training supervised learning algorithms for bot detection [21].

Some researchers have studied the detection possibility by humans, and they proposed the crowdsourcing of social bot detection with the help of workers [22]. They developed an online platform for social turning tests. Authors assume bot detection is easy for humans, as they can identify conversational nuances and recognize patterns and behavior that a machine cannot understand. Some accounts are shown to crowdsourced workers and their majority vote will be the final decision. However, this solution is not cost-effective due to the massive user-based platform such as Twitter and Facebook. Secondly, an expert will be required to ensure accurate results and this may not suit small social platforms. Moreover, another issue might arise while divulging personal information to other workers, thus, compromising privacy. Features-based social bot detection method involves running machine learning algorithms on different features to identify the bot [23]. Behavioral patterns are identified using features, and through supervised learning algorithms, a bot can be predicted [10]. A study [11] divides bot detection-related work into three parts: crowdsourcing and leveraging human intelligence, social platform data-based and machine

learning-related approaches based on features that distinguish humans and bots. Here, the authors combined these ideas and methods and used them to detect social bots.

2.2. Twitter Bot Detection

Bots on Twitter are utilized with different goals in mind. Spam is dispersed using various bots (spambots) called content polluters or spammers [24]. Some of them generate revenues by attracting clients and selling products. Another category includes harmful bots that spread malicious content over the web. Bot identification on Twitter depends on the fact that a human behaves differently from a machine [25,26]. Twitter bot detection is a more recent area of research [27,28]. The first study detects the user, whether human, a bot, or a cyborg, based on their behavior, tweet content, and account details [2]. Tweet content is the features related to a tweet message. It can be the number of words, capital words and punctuations, order of letters, and resemblance of tweet content with malicious content. Tweet content features may also include the number of URLs and mentions to other users [29].

A research study [30] was carried out to observe if a bot's activity includes the distribution of false content and whether or not these words cross legal boundaries. For example, if the content violates the law or may spread false information, then it is unethical, and they believe that bots should be ethical. They identify specific features [31] that separate humans from bots and D. Zengi [18] suggested a behavior-enhanced deep model (BeDM) for the identification of bots using tweet messages as temporal tweet information or some patterns. In [31], the authors propose using the network, linguistic, and application-oriented variables as possible features and identify specific features that distinguish well between humans and bots. Ratkiewicz, J. and Conover, M. [32] proposed feature-based algorithms to detect bots, relying upon suspicious behaviors which could filter the bots from humans. Moreover, they proposed an extensible framework that enables real-time analysis of meme diffusion in social media through mining, visualization, mapping, classification, and modeling of massive streams of public microblogging events. Hwang, I. Pearce and Nanis, M. [11] suggested that the most predictive feature is the user meta-data, which could separate bot behavior from humans, and classify it as the most interpretable one. Wang, G. Mohanlal and M., Wilson worked on detecting the social bot via a crowdsourcing site, thus creating an online test platform, 'Truing', which was used by humans to identify the bots, assuming that bot detection is more straightforward for humans.

In addition to classification, we found other methods to detect a bot. DeBote [33] found the difference in temporal usage of their accounts using a correlation-based study. The main statement of this research is that, for an extended period of time, a human can never be highly synchronous, while a bot and for the DARPA job for the detection of a bot, the system should be semi-supervised. A research study [34] proposes a model that allows a study member to delete a bot and enhances recall in bot detection. The study shows that the algorithm deletes more bots than previous methods and they applied it to two real-world social media datasets for evaluation. Similarly, another study proposes retweeting social bots detection using a model named Retweet-Buster (RTbust) [32], which needs timestamps of retweets for each analyzed account. Hence, there is no need for complete user timelines or social graphs. The study compared the temporal retweeting pattern of a vast user group. RTbust looks for accounts within the group using distinct and synced patterns [35].

In [36], the authors present a platform, *on-demand* bot detection, that is topic- or geographic location-related bot discovery. One research study also discovered a direct relationship between the Bursty botnet with significant online spamming violence in 2012. Common features are assumed in most techniques of detecting bots that were supposedly shared by all bots [37]. A classifier named Decorate is proposed by the research, which uses Twitter content, user account, and usage features. It claims a performance of 0.88 using the performance evaluation metric F-Measure in its research experiments [38]. Bot detection study [39] used a supervised learning algorithm to conclude that bots have more added

friends than their followers, their tweet posts contain more URL links and mention more users in tweets, irrespective of their relationship.

Another study [40] presents a framework called Bot-AHGCN for bot detection, which addresses the limitations of existing flow-based bot detection approaches. It models network flow objects as a multi-attributed heterogeneous graph and transforms the bot detection problem into a semi-supervised node classification task on the graph. Moreover, a bot detection study [41] uses modern deep learning techniques to automatically learn policies for botnet detection instead of heuristically designed multi-stage detection criteria. They generate training data by synthesizing botnet connections with different communication patterns overlaid on large-scale real networks. Graph neural networks (GNN) are tailored to detect the properties of centralized and decentralized botnets. The experimental results show that GNNs are better at capturing botnet structure than previous non-learning methods, and deeper GNNs are crucial for learning difficult botnet topologies. However, ref. [42] proposed a graph-based machine learning model for botnet detection, which considers the significance of graph features and selects important features for the detection of botnets. They evaluated the proposed model on two botnet datasets including CTU-13 and IoT-23 using several supervised machine learning algorithms, and the results show that the model reduces training time and model complexity while achieving a high bot detection rate and robustness to zero-day attacks.

Twitter bot detection could have been easier if the algorithms could work in multiple directions without focusing on a single detection technique [21]. Alvisi's Renren Sybil detector [43] explored the user activity and its behaviors, further adding another dimension of time spent by the users. Bots do not spend much time seeing others' profiles. They just post the content and make friends or follow others, whereas humans spend much more time viewing other profiles and posts. Renren detector classified the social profiles depending upon the activity and timing information to determine that it is a bot-like or human-like profile. The literature review reveals that there is a need to study the content generated by both human as well as Twitter bot users and then analyze their content using supervised learning models. Moreover, Table 1 presents a comprehensive summary of the literature related to Twitter bot detection. The symbol, tick, shows the type of features explored in the research study. Most authors incorporated message-based features for bot detection with machine learning approaches. However, the novelty of this study lies in the extraction of four types of features including message-based, special character-based, part-of-speech-based and sentiment-based features with feature selection, along with DL methods.

Table 1. Comprehensive summary of existing research studies and comparison with the proposed model.

Ref	Message Based	Special Character Based	Part-of-Speech Based	Sentiment Based	Feature Selection	Deep Learning
[1]	✓	-	-	-	-	-
[2]	-	-	✓	✓	✓	-
[14]	✓	✓	-	✓	-	-
[15]	✓	-	-	-	-	-
[17]	-	-	-	-	✓	-
[21]	✓	✓	-	-	-	-
[23]	-	-	-	-	-	✓
[26]	✓	✓	-	-	-	-
[28]	✓	-	-	✓	-	-
Proposed	✓	✓	✓	✓	✓	✓

3. Proposed Research Methodology

We propose a framework for Twitter bot detection that classifies the Twitter users as a bot or a human. The proposed model aims to capture the content-related features such as part-of-speech, special characters, word frequency and sentiments based on four machine learning algorithms. The framework of the proposed model is shown in Figure 1.

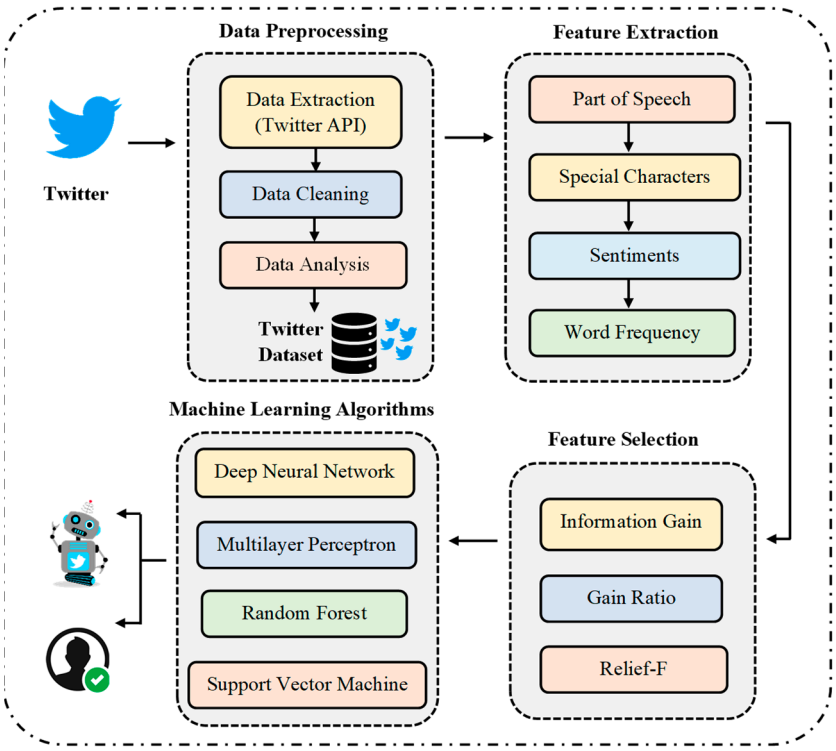

Figure 1. The model of Twitter bot detection using deep learning algorithms.

We selected 50 bots from the Wikipedia Twitter bot list (https://en.wikipedia.org/wiki/Twitter_bot, accessed on 1 December 2022) and 50 real human accounts for data pre-processing. The proposed model extracts tweet data using Python script from Twitter API using the Twitter developer's account credentials. After mining the tweet of both bots and humans, data is then cleaned, and missing data is fixed.

In the features extraction phase, features of the pre-processed Twitter dataset are extracted by features related to content including special characters, sentiments, part-of-speech, and word frequency or message-based features. As many as 18 features are extracted for further processing. Feature engineering techniques for selection such as Gain Ratio (GR), Information Gain (IG) and Relief-F are applied to all extracted features. The top 10 ranked features are then finally selected for running the machine learning algorithm.

To detect Twitter bots, we have chosen a supervised learning approach, and famous deep learning algorithms such as deep learning (DL) and multilevel perceptron (MLP) are applied along with other classification-related algorithms including random forest (RF), naïve Bayes (NB) and rule-based classification (RBC). The MLP is a shallow neural network, whereas DL consists of multiple layers of neurons. MLP has a single input layer, one or more hidden layers, and a single output layer. DL can have multiple hidden layers, and each layer can consist of many neurons. Moreover, 5-fold cross-validation and holdout methods are applied having 70% as training and 30% as test data. The basic purpose of the model is to detect whether a tweet is posted by a bot or a human user. The detail of hyper parameters with their values of all applied models is presented in Table 2.

Table 2. The list of hyper parameters of proposed models for Twitter bot detection.

Model	Hyper Parameter	Value
Deep Learning	Number of Layers	5 layers
	Hidden Layers	512 layers
	Activation Function	ReLU and Sigmoid
	Learning Rate:	0.001
	Dropout:	0.5
	Batch Size:	32
	No. of Epochs:	20
Multilevel Perceptron	Number of Layers	4 layers
	Regularization	L2
	Hidden Layers	512 layers
	Activation Function	ReLU and Sigmoid
	Learning Rate:	0.001
	Dropout:	0.5
	Batch Size:	32
	No. of Epochs:	20
Random Forest	n_estimators	100
	max_depth	10
	min_samples_leaf:	02
Naïve Bayes	Smoothing Parameter	01

3.1. Feature Engineering

The primary purpose of feature engineering is to propose and explore such features sets which can be used for application of machine learning algorithm. Table 3 presents the types of feature set, as well as description and symbolic representation of all the features considered in each feature set. After examining the nature of our data, the feature set obtained from our data is divided into four classes:

1. Message-based feature set;
2. Special character-based feature set;
3. Part-of-speech-based feature set;
4. Sentiment-based feature set.

Table 3. The list of features proposed for Twitter bot detection.

Type	Symbols	Description
Message Based	N_w^t	Number of Words in a Tweet
	N_u^t	Number of URLs
	U_r^t	Number of Retweets
	N_f^t	Number of Favorite Tweets
	U_{sw}^t	Number of Similar Words
Special Character Based	N_{ms}^t	Number of Mention Symbols
	N_{ht}^t	Number of Hashtags
	N_{qm}^t	Number of Question Marks
	N_{ex}^t	Number of Exclamation Marks
	N_{sc}^t	Number of Special Characters
Part-of-Speech Based	N_n^t	Number of Nouns
	N_p^t	Number of Pronouns in a Tweet
	N_v^t	Number of Verbs
	N_a^t	Number of Adverbs
Sentiment Based	S_{Po}^t	Ratio of Positive Words by the Number of Words in a Wweet
	S_{Ne}^t	Ratio of Negative Words by the Number of Words in a Wweet
	S_{Nu}^t	Ratio of Neutral Words by the Number of Words in a Tweet

3.2. Message-Based Features

Features related to tweet messages consist of the No. of URLs, repeated word frequency, total no of words, retweet (RT) frequency and favorite tweet frequency.

- The proposed module has four different attributes;
- Number of words in a post by user;
- Number of mentions used in a post by user u;
- Number of retweets of a post by user u;
- Number of hashtags and mentions in all tweets of a Twitter user.

3.3. Number of Words

The frequency of words gives us information about the total number of words used by a blogger in a tweet. This attribute calculates the rate of production of micro-blogger. Using Oracle SQL query (select length(yourCol) − length (replace (yourcol, ' ', '')) + 1 NumbofWords from yourtable), we calculated the frequency of the word in tweets of each micro-blogger in a dataset.

3.4. Number of Retweets (RT)

The frequency of retweets gives information about the actual number of retweets. Oracle query also has the capability to extract this feature. We search the keyword RT before a tweet to check the number of retweeting tweets to calculate the number of retweets.

3.5. Number of Mentions (@)

The number of mentions represents an actual count of mentions or the @ sign in a message tweeted by a user. To calculate the count of mentions of a Twitter user, we search the keyword @ in a tweet to check the number of mentions used in tweets.

3.6. Number of Hashtags (#)

At the start, a hashtag helps the users to follow, find, and contribute to a conversation. The number of hashtags shows the actual total hashtag count used by a Twitter user to participate in different subjects (topics) using the # sign. So, the count of total hashtags used by a Twitter user in all their tweets in the available dataset gives the count of the number of hashtags of a micro-blogger.

3.7. Special Character-Based Features

Special character-based features in the tweets are derived by the number of occurrences of special characters, including mentioning symbols, question marks (?), hashtags (#), exclamation marks (!) and other special characters. We count all of these symbols, which may differentiate the human and bot tweets.

3.8. Part-of-Speech-Based Features

Part-of-speech-based features are also used to differentiate the usage of bots and humans in their tweets by calculating the number of nouns, pronouns, verbs and adverbs used. We have used the Python Natural Language Toolkit (NLTK) for tagging words in a tweet by their category and then counted the total number of nouns, pronouns, verbs and adverbs in tagged categories.

3.9. Sentiment-Based Features

Sentiment-based features rely on the sentiments extracted from the tweet and classify a tweet as positive, negative, or neutral. NLTK is used for computation of the sentiment score for each tweet.

4. Results and Discussion

4.1. Dataset Preparation

In this research study, we selected the top 50 Twitter bot accounts from the Wikipedia Twitter bot list to prepare a Twitter bot detection dataset. In addition, we also selected 50 real human accounts to properly balance the dataset as a negative class. Next, we employed Twitter API to extract tweets data from selected 100 accounts. We have extracted up to 11 k tweets from Twitter bots as well as human accounts. The prepared dataset consists of a huge variety of tweets related to different topics and categories. Table 4 presents statistics of the prepared dataset.

Table 4. Characteristics of the prepared dataset.

Total Twitter Tweets	11,175
Human Tweets	5494
Bot Tweets	5681

4.2. Dataset Visualization

Let us discuss here the data visualization experiments and the results using Twitter tweets dataset. In this research, we have extracted 18 features from the prepared dataset. These 18 features are extracted from every tweet of the given dataset. These feature values are analyzed by a data visualization process. For almost 11,000–11,200 tweets, both human and bots contain a higher number of message-based features than special characters-based, part-of-speech-based and sentiments-based features, whose maximum values for a human are between [3.1–204] and for a bot it is lesser than the human, containing maximum values ranges between [0–155], as in Figure 2.

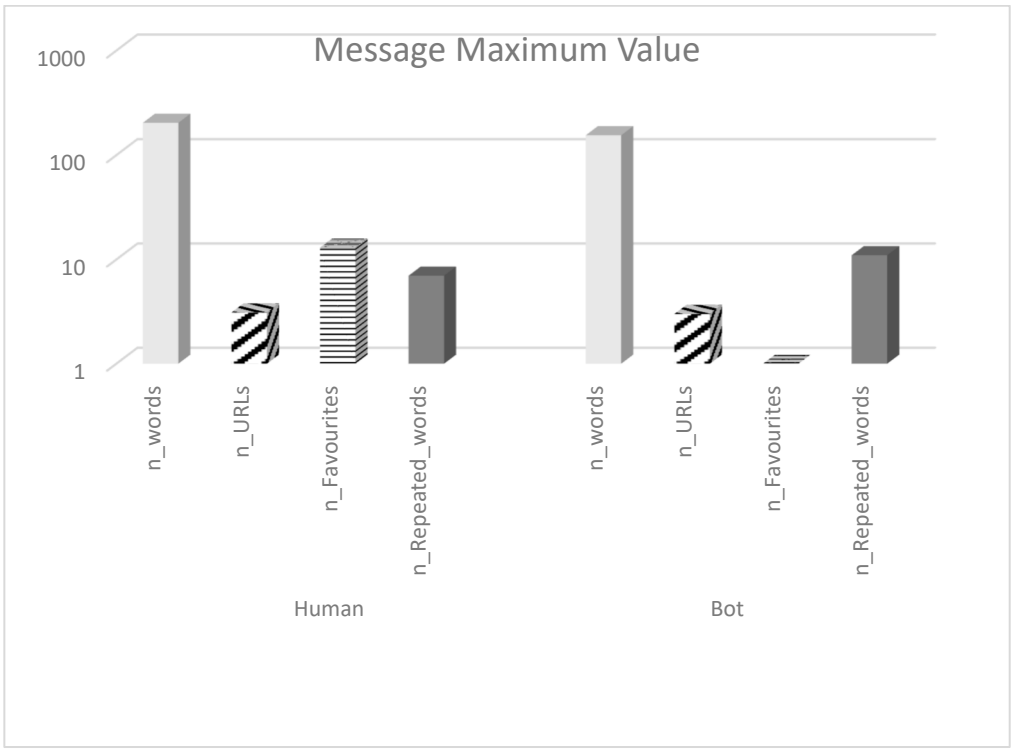

Figure 2. Message-based feature frequency.

For human tweets with special character-based features lie in maximum range between [4–23] and for bot tweets the maximum range is in between [2–17]. While the ratio is not so high for part-of-speech-based features, human and bot tweets, both, almost contain the same number of part-of-speech-based features, which is less than the rate of special

characters-based features, and maximum values for human are in the range of [3, 4] and for bot tweets its maximum value lie in between [2–4], as shown in Figures 3 and 4, respectively. For sentiment features the ratio is same for both human and bot, where human tweets have maximum compound sentiment values in a range of [0.95–1] and bot tweets contain the same maximum range of [0.95–1], as shown in Figure 5.

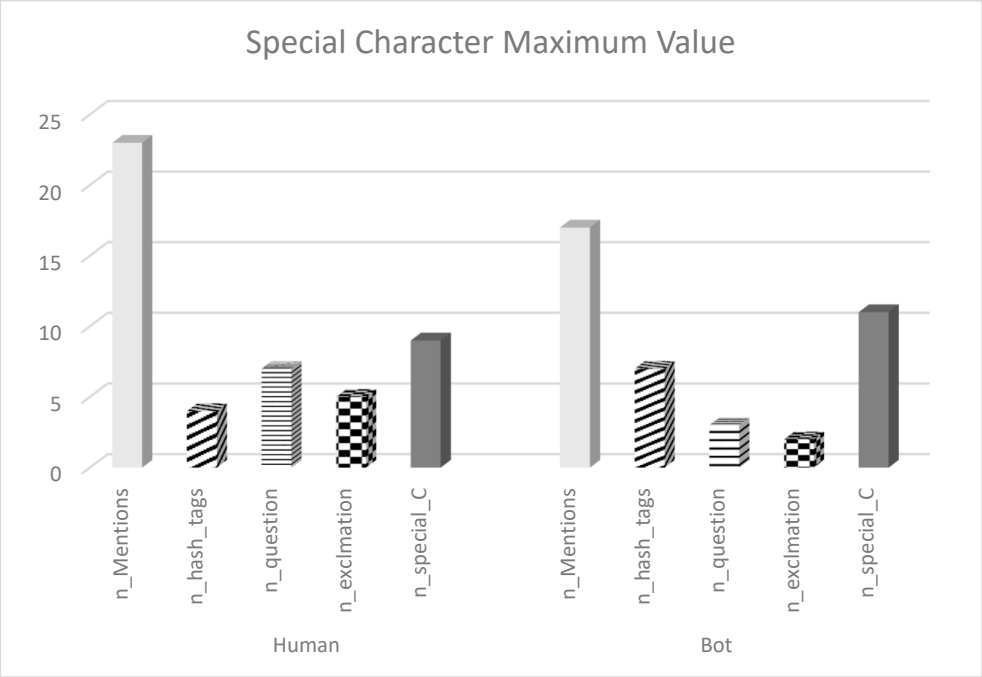

Figure 3. Special characters-based feature frequency.

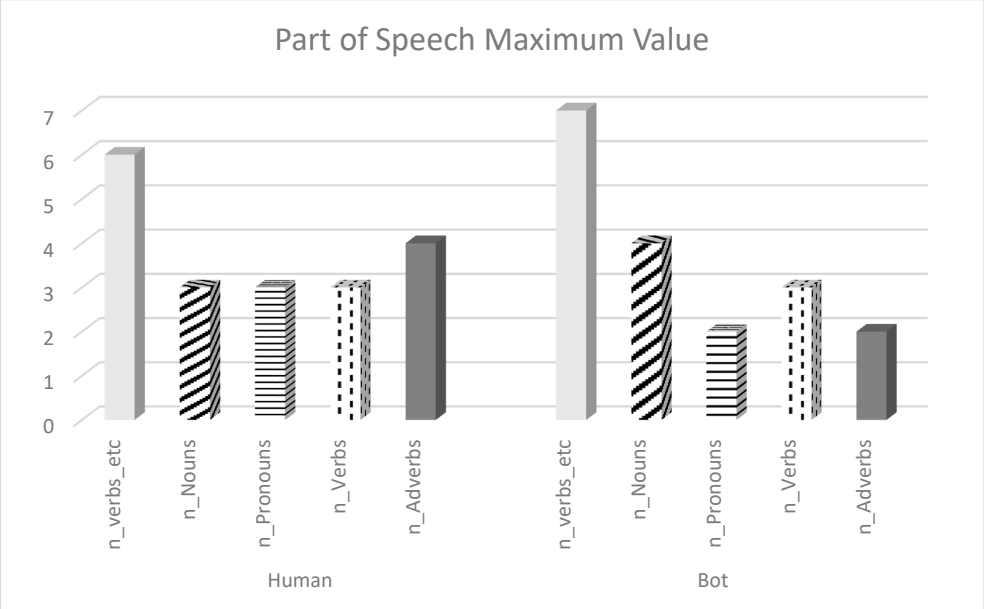

Figure 4. Part-of-speech-based feature frequency.

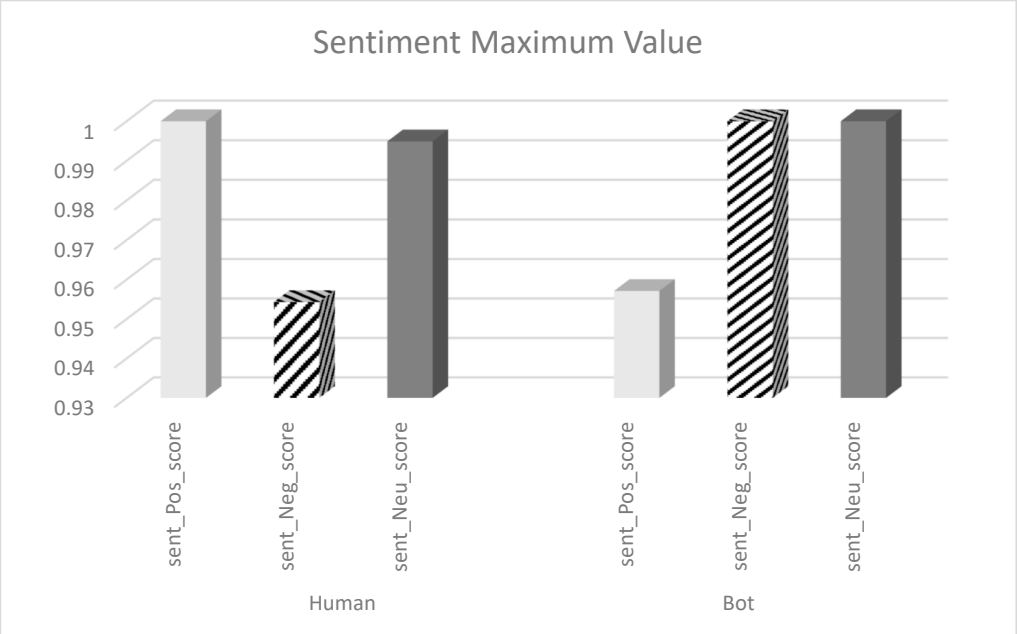

Figure 5. Sentiment-based feature frequency.

4.3. Top 10 Ranked Features

In this research study of identifying bots and humans based on their tweets through content analysis, 17 features are extracted from each tweet in the dataset. Extracted features are labelled in these categories: message, special character, part-of-speech and semantic. To reduce the features sets, three-dimensional reduction strategies are practical for decreasing the feature sets. These famous feature selection techniques include Information Gain (IG), Gain Ratio (GR), and Relief-F. The top ten features are selected among 18 features based on feature ranking values as listed in [31].

4.4. Classifiers Results Based on Hold out and Cross-Validation

On ranked attributes of the dataset mentioned in Table 5, machine learning classifiers are applied using cross-validation, and hold out settings MLP, DL, RF, NB and RBC are applied to the selected feature set. Using a 5-fold cross validation dataset is randomly split up into '5' groups. One of the groups is a test set and the rest are training sets. Data is also then split using a 70%-30% hold-out method. A total of 70% of data is used for training and 30% is for testing. The accuracy and f-measure are used to evaluate the performance of these proposed models.

In 5-fold cross-validation, DL model outperforms all other techniques in terms of F-measure when IG ranked features are applied. RF is slightly above MLP in terms of accuracy. While using both GR and Relief-F ranked feature sets, the MLP shows best results as compared to others shown in Table 6, DL, MLP and RF dominate other machine learning classifiers when applied to the selected feature sets of IG, GR and Relief-F with 5-fold cross-validation. The same algorithms are applied with 70%-30% holdout method, where 70% of data is training set and 30% is testing. The results shown in Table 7 reveal that MLP, RF and DL dominated other classifiers again, when applied with 70%-30% holdout method. Although other classifiers RF and SVM showed good results in the classification process, MLP, RF and DL are the most suitable classifiers as evident from accuracy and f-measure.

Table 5. Feature ranking by IG, GR and Relief-F.

	IG		GR		Relief-F	
Sr. No	**Ranked Features**	**Values**	**Ranked Features**	**Values**	**Ranked Features**	**Values**
1	Favorite Tweet	1.000	Favorite Tweet	1.000	Mentions used in a Tweet	1.000
2	Retweets of a Tweet	0.888	Special Characters in a Tweet	0.967	Special Characters in a Tweet	0.655
3	URLs in a Tweet	0.460	Hashtag in a Tweet	0.888	Words in a Tweet	0.482
4	Mentions used in a Tweet	0.227	Nouns in a Tweet	0.813	Nouns in a Tweet	0.394
5	Nouns in a Tweet	0.130	Negative Sentiment	0.722	Verbs in a Tweet	0.389
6	Words in a Tweet	0.122	Retweets of a Tweet	0.722	Neutral Sentiment	0.283
7	Verbs in a Tweet	0.096	Neutral Sentiment	0.717	Hashtag in a Tweet	0.268
8	Special Characters in a Tweet	0.078	URLs in a Tweet	0.694	Pronoun in a Tweet	0.212
9	Hashtag in Tweet	0.065	Words in a Tweet	0.680	Adverbs in a Tweet	0.156
10	Neutral Sentiment	0.034	Verbs in a Tweet	0.667	Repetitive Words in Tweets	0.138

Table 6. Performance of feature selection on 5-fold cross-validation.

	IG		GR		Relief-F	
	Accuracy (%)	**F-Measure (%)**	**Accuracy (%)**	**F-Measure**	**Accuracy (%)**	**F-Measure**
DL	71.44	65.78	61.68	60.05	59.24	59.28
MLP	76.62	64.97	76.07	65.47	76.89	64.74
RF	76.73	64.53	65.98	52.82	65.75	49.65
NB	49.47	59.54	47.75	58.94	65.72	37.09
RBC	56.39	47.56	55.94	41.89	56.40	43.04

Table 7. Performance of feature selection on 70%-30% holdout.

	IG		GR		Relief-F	
	Accuracy (%)	**F-Measure (%)**	**Accuracy (%)**	**F-Measure**	**Accuracy (%)**	**F-Measure**
DL	76.23	70.07	51.97	56.24	63.54	61.16
MLP	74.29	58.91	75.19	64.08	77.09	66.27
RF	76.31	69.30	67.14	55.63	61.54	54.76
NB	52.10	62.03	44.34	55.51	63.84	43.68
RBC	59.22	12.18	59.29	49.75	54.71	42.68

4.5. Classifiers Results Using Proposed Feature Set

After cross-validation and hold out settings, the results are separated based on proposed features apart from ranking standards. DL, MLP and NB have beaten up all other techniques when applied to all combined feature sets without any feature selection techniques, evident from precision, accuracy, recall, and f-measure. Moreover, categorized features are separated for more experiments to check the capability of selected machine learning techniques as given in Table 8.

Table 8. Results of machine learning algorithms on feature set.

Feature Set	**Machine Learning Algorithms**	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F-Measure (%)**
All Features	DL	77.15	68.34	74.07	71.09
	MLP	75.23	70.34	59.98	64.75
	RF	58.03	47.26	91.81	62.40
	NB	51.08	43.43	95.81	59.77
	RBC	60.53	48.09	51.34	49.66
Message	DL	51.32	43.65	97.45	60.29
	MLP	59.88	48.06	71.81	57.58
	RF	61.18	49.38	94.59	64.89
	NB	47.54	41.89	98.95	58.86
	RBC	41.48	39.30	99.73	56.38

Table 8. *Cont.*

Feature Set	Machine Learning Algorithms	Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
Special Character	DL	57.54	46.57	81.27	59.21
	MLP	58.90	47.06	67.08	55.32
	RF	52.12	43.13	82.42	56.63
	NB	62.34	51.20	15.04	23.24
	RBC	37.93	37.93	98.00	54.99
Part-of-Speech	DL	42.28	39.40	97.01	56.04
	MLP	61.33	41.35	4.69	8.43
	RF	58.91	45.49	42.05	43.70
	NB	61.89	49.31	17.11	25.41
	RBC	37.93	37.93	98.00	54.99
Sentiment	DL	62.09	52.23	0.32	54.99
	MLP	62.09	52.23	0.32	0.64
	RF	38.37	38.05	99.56	55.06
	NB	61.18	42.08	6.26	10.90
	RBC	62.71	52.16	20.10	29.02

The idea of separating all the categories-based features and computing the results on them explicitly are to compare the performance of classifiers on each set of proposed features. The results show that DL's accuracy and f-measure are higher when all feature sets are combined. While RF is the best choice for a message-based feature in the absence of feature selection techniques. For other features, NB and RBC work best for the separated category. Moreover, when features are separated RF, NB and RBC outperform all other algorithms and works best for the accuracy of each separated category shown in Figure 6.

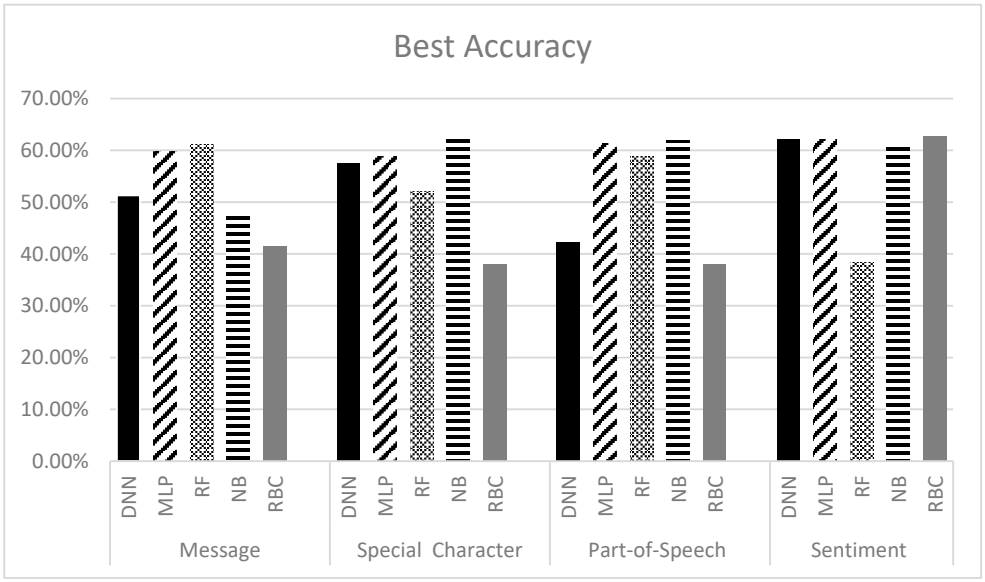

Figure 6. Best average classification accuracy of each feature set.

4.6. Set Rule-Based Classifiers Results Using Feature Set

Based on if else conditions, rules are identified by applying rule-based classifier. Results depict that message and sentiments-based feature are used in the first rule as shown in Table 9. If the condition matches a tweet, then it is a bot. The second rule includes sentiment, message, and special character-based conditions. If the second condition is satisfied for a tweet, then it is written as a bot. The third and fourth rule includes part-of-speech related features, sentiment, and message-based features to satisfy a bot. The fifth condition contains the message and sentiment-related rule. If it is true for a tweet,

then it is a bot tweet. All tweets that do not match these conditions are considered human tweets. The four rules are learned using rule induction (rule-based classification) for the identification of bots are given in [30].

Table 9. Rule induction for the identification of Twitter bots.

Rules	Antecedent	Consequent
R1	(n_Retweets $\leq$ 0.006 and sent_Neu_score $\leq$ 0.782 and n_Favourites $\leq$ 0.001 and sent_Neg_score > 0.007 and sent_Neg_score $\leq$ 0.160 and n_words $\leq$ 0.171)	Bot
R2	(sent_Neu_score $\leq$ 0.795 and n_Favourites $\leq$ 0.001 and n_Retweets $\leq$ 0.002 and sent_Neg_score > 0.084 and sent_Neg_score $\leq$ 0.161 and n_exclmation > 0.071)	Bot
R3	(sent_Neu_score $\leq$ 0.782 and n_Favourites $\leq$ 0.001 and sent_Neg_score > 0.274 and sent_Neg_score $\leq$ 0.352 and n_verbs_etc > 0.188 and sent_Neg_score > 0.303)	Bot
R4	n_Retweets $\leq$ 0.010 and sent_Neu_score $\leq$ 0.782 and n_Favourites $\leq$ 0.001 and sent_Neu_score > 0.600 and n_Retweets $\leq$ 0.001 and sent_Neg_score $\leq$ 0.025 and sent_Neu_score > 0.736 and n_Retweets $\leq$ 0.000 and n_words > 0.288	Bot

While considering the performance of classifiers on each separate set of features shown in Figure 2, RF, NB and RBC outperformed other techniques in terms of accuracy. The overall accuracy score of the five models (DL, MLP, RF, NB and RBC) applied on combined features sets is shown in Figure 7. DL gives better accuracy (77.15%) as compared to other classification techniques when all features are combined.

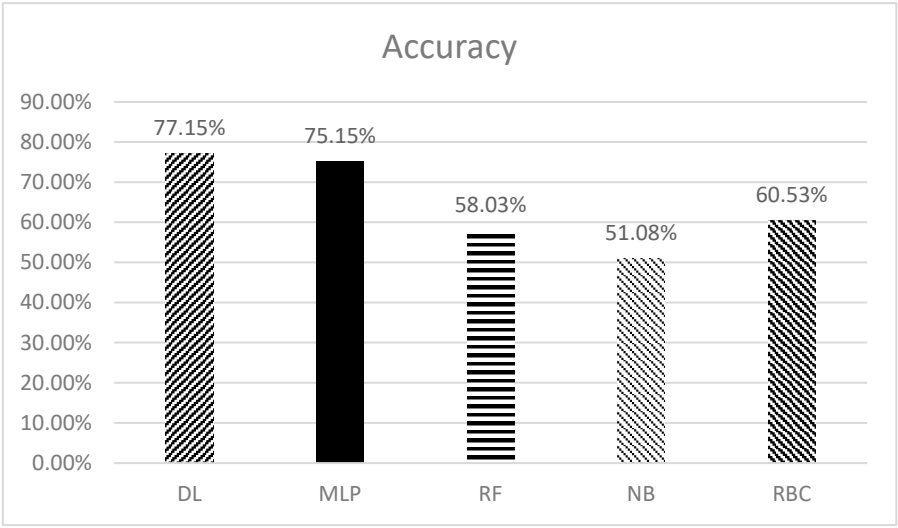

Figure 7. Accuracy score achieved by using five classification techniques.

5. Conclusions

In this research study, a Twitter dataset is studied to evaluate the classification potential of different classifiers. The proposed features are based on the message, special characters part-of-speech, and sentiment to classify and detect Twitter users, whether bots or humans. These features are intended to efficiently increase the classification precision of classifiers for human and bot detection. The proposed features are extracted from the tweet dataset mined using Twitter API. The bot tweets ratio in the dataset is 51% approximately.

The proposed approach uses content analysis and computed content-related features, such as special characters, word frequency, part-of-speech, and sentiments. Previous

studies are focused on behavioral features only and use sentiment analysis, while this study introduced content-based analysis using deep learning algorithm. Among several machine learning techniques, deep learning (DL), multilayer perceptron (MLP), random forest (RF), naïve Bayes (NB) and rule-based classification (RBC) are used. The content-related features are used to identify Twitter users as human or bot.

To estimate the performance of machine learning techniques (DL, MLP, RF, NB and RBC) on the proposed feature sets, selection methods are applied to pick the most capable features of a tweet. Eighteen proposed features are shortlisted by ranking them using three feature selection techniques (IG, GR, Relief-F) and ten features are selected by each technique for further processing. On the top 10 ranked features, cross-validation settings are applied to divide the data into test and training sets. A 5-fold and 70%-30% holdout settings are used. The classifiers' performance is assessed by the performance evaluation measures accuracy, F-measure, recall, and precision. Considering all features, DL outperformed all classifiers in terms of accuracy and F-measure, and MLP performs best in terms of precision. These machine learning algorithms are applied for Twitter message analysis. It is observed that the technique is not broadly used for human and bot detection using Twitter tweets.

This research study proposed a different way to assess the performance of all applied techniques by several feature sets selected by feature selection techniques. Message-based, sentiment-based, special character-based and part-of-speech-based features are also analyzed separately. RF, NB, and RBC performed better than other techniques in measuring the accuracy of separated features. When all features are combined, DL is the best choice for accuracy and f-measure. Hence, it is observed that combined content-related feature sets for the detection of human and bot tweets are computed with DL with higher accuracy than others. Deep learning (DL) is more promising than all state-of-the-art methods when applied with all the proposed feature sets for accuracy and f-measure. Hence, this research study shows that the implementation of DL performed better by including all features set in identifying Twitter bots and humans using content analysis. The future work includes the consideration of images in detecting Twitter bots because images can contain valuable information and analyzing them could potentially improve the accuracy of bot detection models.

Author Contributions: Methodology, H.A.; Software, H.A.; Validation, H.U.K.; Formal analysis, H.U.K. and N.A.; Investigation, F.K.A. and M.A.; Resources, F.K.A. and M.A.; Writing – original draft, H.A.; Writing – review & editing, H.U.K.; Visualization, M.A.; Supervision, F.K.A., H.U.K., A.M.A. and N.A.; Project administration, A.M.A. and N.A.; Funding acquisition, F.K.A., A.M.A. and N.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to the Deanship of Scientific Research at King Faisal University for funding this work through Research Grant no.1858.

Institutional Review Board Statement: This study was conducted on Twitter which was properly prepared followed Data Anonymization rules and regulations, and no humans or animals were involved in this study.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors have prepared the datasets and they are freely available for research purposes based on the condition that this research work will be added as a reference, as well as added in the Acknowledgements section of their manuscripts or any other studies. The data is available at the following link: https://github.com/HikmatNiazi/Twitter-Bot-Data, accessed on 10 February 2023.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jiang, W. Graph-based deep learning for communication networks: A survey. *Comput. Commun.* **2022**, *185*, 40–54. [CrossRef]
2. Chu, Z.; Gianvecchio, S.; Wang, H.; Jajodia, S. Who is tweeting on Twitter: Human, bot, or cyborg? In Proceedings of the 26th Annual Computer Security Applications Conference, Austin, TX, USA, 6–10 December 2010; pp. 21–30.

3. Ain, Q.T.; Ali, M.; Riaz, A.; Noureen, A.; Kamran, M.; Hayat, B.; Rehman, A. Sentiment analysis using deep learning techniques: A review. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*. [CrossRef]
4. Lee, K.; Eoff, B.; Caverlee, J. Seven months with the devils: A long-term study of content polluters on twitter. In Proceedings of the International AAAI Conference on Web and Social Media, Barcelona, Spain, 17–21 July 2011; Volume 5, pp. 185–192.
5. Conover, M.; Ratkiewicz, J.; Francisco, M.; Gonçalves, B.; Menczer, F.; Flammini, A. Political polarization on twitter. In Proceedings of the International aaai Conference on Web and Social Media, Barcelona, Spain, 17–21 July 2011; Volume 5, pp. 89–96.
6. Edwards, C.; Edwards, A.; Spence, P.R.; Shelton, A.K. Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Comput. Hum. Behav.* **2014**, *33*, 372–376. [CrossRef]
7. Messias, J.; Schmidt, L.; Oliveira, R.; Benevenuto, F. You followed my bot! Transforming robots into influential users in Twitter. *First Monday* **2013**, *18*. [CrossRef]
8. Khan, H.U.; Nasir, S.; Nasim, K.; Shabbir, D.; Mahmood, A. Twitter trends: A ranking algorithm analysis on real time data. *Expert Syst. Appl.* **2021**, *164*, 113990. [CrossRef]
9. Iqbal, S.; Khan, R.; Khan, H.U.; Alarfaj, F.K.; Alomair, A.M.; Ahmed, M. Association Rule Analysis-Based Identification of Influential Users in the Social Media. *Comput. Mater. Contin.* **2022**, *73*, 6479–6493. [CrossRef]
10. Zeng, Z.; Li, T.; Sun, J.; Sun, S.; Zhang, Y. Research on the generalization of social bot detection from two dimensions: Feature extraction and detection approaches. *Data Technol. Appl.* 2022; *ahead-of-print*. [CrossRef]
11. Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; Flammini, A. The rise of social bots. *Commun. ACM* **2016**, *59*, 96–104. [CrossRef]
12. Kantepe, M.; Ganiz, M.C. Preprocessing framework for Twitter bot detection. In Proceedings of the 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 5–8 October 2017; pp. 630–634. [CrossRef]
13. Ratkiewicz, J.; Conover, M.; Meiss, M.; Gonçalves, B.; Flammini, A.; Menczer, F. Detecting and tracking political abuse in social media. In Proceedings of the International AAAI Conference on Web and Social Media, Barcelona, Spain, 17–21 July 2011; Volume 5, pp. 297–304.
14. Hwang, T.; Pearce, I.; Nanis, M. Socialbots: Voices from the Fronts. *Interactions* **2012**, *19*, 38–45. [CrossRef]
15. Aiello, L.M.; Deplano, M.; Schifanella, R.; Ruffo, G. People are strange when you're a stranger: Impact and influence of bots on social networks. In Proceedings of the International AAAI Conference on Web and Social Media, Dublin, Ireland, 4–6 June 2012; Volume 6, pp. 10–17.
16. Gupta, A.; Lamba, H.; Kumaraguru, P. $1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on Twitter. In Proceedings of the 2013 APWG eCrime Researchers Summit, San Francisco, CA, USA, 17–18 September 2013; pp. 1–12. [CrossRef]
17. Subrahmanian, V.; Azaria, A.; Durst, S.; Kagan, V.; Galstyan, A.; Lerman, K.; Zhu, L.; Ferrara, E.; Flammini, A.; Menczer, F. The DARPA Twitter Bot Challenge. *Computer* **2016**, *49*, 38–46. [CrossRef]
18. Cai, C.; Li, L.; Zengi, D. Behavior enhanced deep bot detection in social media. In Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, China, 22–24 July 2017; pp. 128–130.
19. Cao, Q.; Sirivianos, M.; Yang, X.; Pregueiro, T. Aiding the detection of fake accounts in large scale social online services. In Proceedings of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12), San Jose, CA, USA, 25–27 April 2012; pp. 197–210.
20. Boshmaf, Y.; Muslukhov, I.; Beznosov, K.; Ripeanu, M. Design and analysis of a social botnet. *Comput. Netw.* **2012**, *57*, 556–578. [CrossRef]
21. Alvisi, L.; Clement, A.; Epasto, A.; Lattanzi, S.; Panconesi, A. SoK: The Evolution of Sybil Defense via Social Networks. In Proceedings of the 2013 IEEE Symposium on Security and Privacy, San Francisco, CA, USA, 19–22 May 2013; pp. 382–396. [CrossRef]
22. Wang, G.; Mohanlal, M.; Wilson, C.; Wang, X.; Metzger, M.; Zheng, H.; Zhao, B.Y. Social turing tests: Crowdsourcing sybil detection. *arXiv* **2012**, arXiv:1205.3856.
23. Fields, J. Botnet campaign detection on Twitter. *arXiv* **2018**, arXiv:1808.09839.
24. Dorri, A.; Abadi, M.; Dadfarnia, M. SocialBotHunter: Botnet Detection in Twitter-Like Social Networking Services Using Semi-Supervised Collective Classification. In Proceedings of the 2018 IEEE 16th International Conference on Dependable, Autonomic and Secure Computing, 16th International Conference on Pervasive Intelligence and Computing, 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Athens, Greece, 12–15 August 2018; pp. 496–503. [CrossRef]
25. Chu, Z.; Gianvecchio, S.; Wang, H.; Jajodia, S. Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg? *IEEE Trans. Dependable Secur. Comput.* **2012**, *9*, 811–824. [CrossRef]
26. Beskow, D.M.; Carley, K.M. Bot Conversations are Different: Leveraging Network Metrics for Bot Detection in Twitter. In Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain, 28–31 August 2018; pp. 825–832. [CrossRef]
27. Yip, M.; Shadbolt, N.; Webber, C. Structural analysis of online criminal social networks. In Proceedings of the 2012 IEEE International Conference on Intelligence and Security Informatics, Washington, DC, USA, 11–14 June 2012; pp. 60–65. [CrossRef]
28. Yardi, S.; Romero, D.; Schoenebeck, G.; Boyd, D. Detecting spam in a twitter network. *First Monday* **2010**, *15*. [CrossRef]
29. Wang, B.; Zubiaga, A.; Liakata, M.; Procter, R. Making the most of tweet-inherent features for social spam detection on Twitter. *arXiv* **2015**, arXiv:1503.07405.

30. Salge, C.A.D.L.; Berente, N. Is that social bot behaving unethically? *Commun. ACM* **2017**, *60*, 29–31. [CrossRef]
31. Dickerson, J.P.; Kagan, V.; Subrahmanian, V.S. Using sentiment to detect bots on Twitter: Are humans more opinionated than bots? In Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Beijing, China, 17–20 August 2014; pp. 620–627. [CrossRef]
32. Ratkiewicz, J.; Conover, M.; Meiss, M.; Gonçalves, B.; Patil, S.; Flammini, A.; Menczer, F. Truthy: Mapping the spread of astroturf in microblog streams. In Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 249–252.
33. Chavoshi, N.; Hamooni, H.; Mueen, A. Debot: Twitter bot detection via warped correlation. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016; pp. 817–822.
34. Morstatter, F.; Wu, L.; Nazer, T.H.; Carley, K.M.; Liu, H. A new approach to bot detection: Striking the balance between precision and recall. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 18–21 August 2016; pp. 533–540.
35. Mazza, M.; Cresci, S.; Avvenuti, M.; Quattrociocchi, W.; Tesconi, M. Rtbust: Exploiting temporal patterns for botnet detection on twitter. In Proceedings of the 10th ACM Conference on Web Science, Boston, MA, USA, 30 June–3 July 2019; pp. 183–192.
36. Chavoshi, N.; Hamooni, H.; Mueen, A. On-Demand Bot Detection and Archival System. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; pp. 183–187. [CrossRef]
37. Echeverria, J.; Besel, C.; Zhou, S. Discovery of the twitter bursty botnet. In *Data Science for Cyber-Security*; World Scientific: Singapore, 2019; pp. 145–159.
38. Lee, K.; Caverlee, J.; Webb, S. Uncovering social spammers: Social honeypots+ machine learning. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 19–23 July 2010; pp. 435–442.
39. Cresci, S.; Di Pietro, R.; Petrocchi, M.; Spognardi, A.; Tesconi, M. The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 April 2017; pp. 963–972. [CrossRef]
40. Zhao, J.; Liu, X.; Yan, Q.; Li, B.; Shao, M.; Peng, H. Multi-attributed heterogeneous graph convolutional network for bot detection. *Inf. Sci. (N. Y.)* **2020**, *537*, 380–393. [CrossRef]
41. Zhou, J.; Xu, Z.; Rush, A.M.; Yu, M. Automating Botnet Detection with Graph Neural Networks. *arXiv* **2020**, arXiv:2003.06344.
42. Alharbi, A.; Alsubhi, K. Botnet Detection Approach Using Graph-Based Machine Learning. *IEEE Access* **2021**, *9*, 99166–99180. [CrossRef]
43. Wang, G.; Konolige, T.; Wilson, C.; Wang, X.; Zheng, H.; Zhao, B.Y. You are how you click: Clickstream analysis for sybil detection. In Proceedings of the 22nd USENIX Security Symposium (USENIX Security 13), Washington, DC, USA, 14–16 August 2013; pp. 241–256.