On the Characterization of Digital Trolls from Twitter Big Data

Ahmed khudhair Abbas
University of Diyala
Dr.ahmed.k.abbas@uodiyala.edu.iq

Hayder Hassan Safi
Mustansiriyah University
hayder.h.safi@uomustansiriyah.edu.iq

Ali Hasan Taresh
University of Information Technology
and Communications (UoITC)
alihtaresh@uoitc.edu.iq

Abstract

Recently, Twitter becomes one of the most common and effective social media tools in our life. People use Twitter to share their opinions, feeling, and orientations, especially during political unrest or protests. In order to disrupt the protest operations on Twitter or to influence public opinion, electronic flies (Trolls) are widely and effectively used.Accordingly, there is a need to find a method that can automatically and precisely detect these accounts and isolate them from Twitter. Moreover, detecting and characterizing these accounts becomes a significant task to reduce or mitigate its effect on the real general opinion. Thispaper presents an intensive analysis that can be utilized to effectively detect the troll accounts and isolate its bad effect from Twitter.We considered the public trolls accounts datasetspublished by Twitter and we also gathered a new dataset from Twitter that includes tweetsand users' information from different countries to make a fair analysis for the trolls' accounts. The results show thatthe suspicious activities of Twitter troll accounts can be used to detect most of these accounts automatically without using sentiment analysis and opinion mining techniques with accuracy of 95%.To accomplish this task, we propose a set of robust and efficient features that can accurately characterize troll accounts with a relatively small number of features.

***Key Words** : Trolls, Twitter, Big data, social network*

1. Introduction

The Big Data is the raw data and contains trillions of records that include different types of variables such as people opinions from social network, communication devices' data and other organizations data which is not easy to process when a classical database system such as relational is used [1]. Mostly these data are generated from different sources like social networks websites, photos, sensor, smart devices, communication devices, and internet. Big data is processed or analyzed for getting useful information by extracting the raw data to purposeful information which is a difficult undertaking nowadays [2]. Social Networks becomes one of most famous sources of big data where it is the place for communicating with peoples and friends around the world and also expands the business network by implementing connections or following to other users through individual's users, often through social networks sites like Facebook, Twitter, Instagram, LinkedIn, Google Plus and other applications. Social networking connects online users from different locations and cultures to help people make contacts that would be better for social and work life from different perspectives [3].

After the emerging of social networks, SNA which stands for Social network analysis (SNA) becomes one of most important topics in data mining and artificial intelligent. SNA is the method for analyzing social network datasets through the use of networks and graph methods. It represents the network structure in terms of nodes, each sub-node represents different variables with specific values like actors, singer, politician and other people who are on the network that connects them.SNA has many applications in different domains in social networks to study the structures of data generated by people and how we can extract useful information by utilizing different machine leaning and data mining algorithms [4, 5].

Twitter is one of social network websites that gain increasing interest from people and society in last decade [6, 27]. It is used for typing short statements or tweets about daily life activities. People can share their opinion about what they are doing, what's going on in their world. Twitter is basically a network which allows the user to share their opinion, idea, news and any other daily life activities with other users in the network. Each user has its own followers which can get the notification for all the activities by those users[7].

Due to the fast-growing use of the social network like Facebook, Twitter, LinkedIn that allows people to share their posts on the social network, these services generate more than thousands of topics every day which is very useful for predicting future variables in every field of life [28]. It plays a key role in business organizations, politics, crimes, civil organization, Movies, advertisements and marketing decisions which makes the popularity of these topics in a short amount of time. In the same time, many organizations, political parties, and even governments start to exploit social media to affect the general ideas and directions by managing a large number of groups of social media accounts and then use them to post many posts with a pre-planned scenario [8] which called trolling. By looking to the Oxford dictionary, we find that it describes trolling as making "a deliberately offensive or provocative online posting with the aim of upsetting someone or eliciting an angry response from them". Recently, the activities of troll accounts were recognized and noted in much political unrest especially in Twitter [9.

29]. The main problems of existing methods to detect troll accounts are that most of them use a high number of features that needs long time in the extraction process (such as text mining and sentiment analysis methods) and it is developed by analyzing a small set of troll accounts. In this paper, we propose a fast and robust algorithm to detect the troll accounts from twitter website using new derived features and different classification algorithms. The proposed mechanism can work faster than current proposed algorithms since it focuses on highly efficient new proposed features which give the ability to detect troll accounts in real time scenarios.

The rest of this paper is organized as follow, in Section 2, we present the recent important research work in analyzing and detecting troll accounts from twitter. Section 3 describes our dataset that will be used in our experiments. In Section 4, we present our feature set to detect troll accounts and give the results of our experiments. Finally, the paper is concluded in Section 5.

2. Related Work

Because of the use of troll accounts started in recent years, this topic isconsidered as a new research and there is no much work in literature regarding it. In this section, we summarize the common research work that dedicated to handle this issue. Note that we focus only on the research work of detecting troll accounts through twitter. Detecting troll accounts in other social media websites can be similar to twitter but it definitely needs different frameworks since there is a different set of features for each social media network or platform. In addition, it is noted that the research work on troll accounts can be found sometimes with different names such as abusive behavior on social media orpolitical propaganda isolation on Twitter [33][34].

The TrollPacifier is considered as one of the most significant and general frameworks that can be used to detect twitter trolls [10]. As the authors called it, it is a holistic algorithm that works by utilizing a large number of features from a specified twitter account to detect if it is a troll account or not. This systemuses features from a variant groupof approachessuch as style of writing, sentiment analysis, behaviors features, user interactions, and timing features. As reported in this paper, this algorithm can achieve high classification accuracy (95.5%). But the main issue of this algorithm is that it needs a long time to analyze and detect troll accounts. This is because the high number of used features and also the use of some features that requires more executing time such as features that require sentiment analysis. Using the same methodology, there is a number of other papers that proposed different methods to detect troll account on twitter as in [11] and [12]

One of most interesting and studied topics in detecting troll accounts is the 2016 US Presidential Election where many groups of Russian trolls tried to affect the results of the election and make propaganda against one of the two sides of the election [13].In [14], the authors proposed a method to detect and hunt the Russian twitter accounts that can be considered as trolls and worked positively toaffect the US election in 2016. In this method, the authors used a label propagation method to identify Twitter trolls ideology through the focus on the news sources that shared by those accounts. According to this work, the proposed algorithmcan classify troll accounts with accuracy of 84%. In [15], another algorithm is developed using an adaptive framework for analyzing Twitter accounts identity using a social sequence analysis algorithmas explained in [16].

To analyze andinvestingthe troll activities in more details during USA Presidential Election and to check if a similar influence on elections can be detected in Europe, Ansgar Kellner et al. analyzed the propaganda on Twitter during the 2017federal election inGermany. It was found that a group of twitter accounts around 79 accounts works as trolls in both German and US elections. These accounts after influenced US election they tried to do a campaign in the German federal election[17]. In addition, in [18]researchers accuse Russian trolls in amplifying the vaccine debate and uses it as weaponized health communication. In another work, Zannettou, Savvas, et al. [19] designed a method to analyze the use of pictures by twitter troll accounts during the US election. More specifically, the authors investigate how trolls utilize imagery to achieve their goals and change or affect public opinions.

Sahmoud et al. 2020 [9], investigated how we candetect suspicious activities of digital Trolls during political crisis. This study considered the recently occurred Iraq unrest and Iraqi people protest as a case study to analyze the activities of Twitter users and detect if there are any external groups try to influence the people’s opinions and orientations during the crisis. Researchers gathered a new related dataset from Twitter (in 2020) that includes tweets and users’ information collected during the crisis. Using this dataset, the authors analyzed the behavior and activities of twitter users by employing different features and tools[9].

3. Datasets Gathering

Data gathering is the first step and one of most significant processes when we aim to analyze and detect the behavior of troll accounts over social media networks. Usually, the popular social media networks such as Facebook and Twitter apply a strict privacy policy to protect the data of their users. These policies make it very difficult to find a public relevant dataset that can be used in training or analyzing processes. As a result, collecting our data directly from social media networks (Twitter in our case) will be first step and the only way to obtaina suitable and relevant dataset.Fortunately, Twitter has designed an API to support and simplify the work of developers and researchers calledTwitter API [20]. To gather our dataset, we used two commonly used tools. The first one called Knime [21] and it works based on Twitter API. The second one called Twint [22] and it works using different web scraping techniques. Knime is a popular data retrieving and machine learning open-source tool that includes a set of easy to use features. Knime is widely used in big data research community when we need to use both machine learning and data mining techniques. It has also a powerful graphical user interface that works using the style of drag and drop [21]. On the other hand, Twint is an advanced easy-to-use and open-source Twitter scraping tool written in Python that simplifies scraping Tweets from Twitter accounts without using Twitter's API [20]. Twint library has many features such as it can fetch almost all Tweets with no limits (Twitter API allow only 3200 Tweets) and the ability to be used anonymously without Twitter sign up.

To analyze the activities of troll accounts, we have to collect the data of two types of twitter accounts which are normal accounts and troll accounts. For normal accounts, we used Knime and Twint library to gather the data of 800 twitter users from 5 countries as described in Table 1. Form those Twitter users; we collected 639,234 tweets as shown in Table 1. To be sure that the gathered accounts in this set are real accounts, we employed a number of collecting strategies such as select accounts with real and known names, real face images, real location information, and a normal number of followers and followings. For troll accounts, we used different sets of twitter archived troll datasets from different countries. These datasets are published publicly to simplify the research work and to support the transparency on public inauthentic campaigns [32]. Table 2 gives the general properties of the different sets of subsets that we considered in this paper including the number of twitter users, the number of tweets, and the release date. Note that we used a variety of datasets from the countries that own the largest number of Twitter troll accounts as classified by Twitter.

Table 1: The dataset collected for normal (non-troll) Twitter user accounts.

Country	Number of twitter users	Number of tweets	Gathering Date
Spain	200	96,561	January 2020
Russia	200	260,000	January 2020
Egypt	200	98,722	January 2020
Ghana	100	95,185	January 2020
Iran	100	88,766	January 2020

Table 2: The dataset collected for troll Twitter user accounts.

Country	Number of twitter users	Number of tweets	Released Date
Spain	259	56,712	August 2019
Russia	416	920,761	January 2019
Egypt and UAE	276	214,898	August 2019
Ghana	71	39,964	March 2020
Iran	104	24,51	September 2020

4. Results and discussion

In this section, the experimental results of comparing different algorithms for the task of identifying Twitter trolls' accounts are presented. First, we explain the features that we utilized to detect troll accounts and its importance. After that, we make a comparison between features. Finally, the results and discussionsfor using different machine learning algorithms to detect troll accounts are given.

4.1 Features Engineering for Twitter Troll Classification

Feature engineering is the process of using domain knowledge to extract features (characteristics, properties, attributes) from raw data [30]. A feature is a property shared by independent units on which analysis or prediction is to be done [31]. In order to efficiently classify troll accounts, extracting and after that selecting a relevant set of features is a veryimportant step which may directly affect the classification algorithms. As shown in related work section, a large number of features have been selected and used in many machine learning algorithms for troll classification and detection. By reviewing the presented work in literature, we found that most of proposed features are irrelevant or it needs very long time to be calculatedwhich is not suitable in online classification scenarios. Therefore, we chose a set of features that avoid the limitations of previous algorithms and are more relevant for the considered task. Some of these features are derived from previousresearch papersand some of them are new derived experimentally. In the following, we explain the used features in detail:-

Retweets of User Tweets: In this feature, we compute the number tweets of the considered user where the other accounts retweeted them. As an example, a value of 2 of this feature means that in average each tweet from this user is retweeted two times. This feature is important since it reflects the people interest of the tweets of this user. Figure 1 shows the scatter plot for the values of feature after evaluating it using our dataset of trolls and non-trolls Twitter accounts.The redpoints represent the non-troll or normalTwitter accounts where bluepoints represent the values of troll accounts.As shown in the figure, the results show that people normally get interested with tweets of the normal accounts more than the tweets of troll accounts.

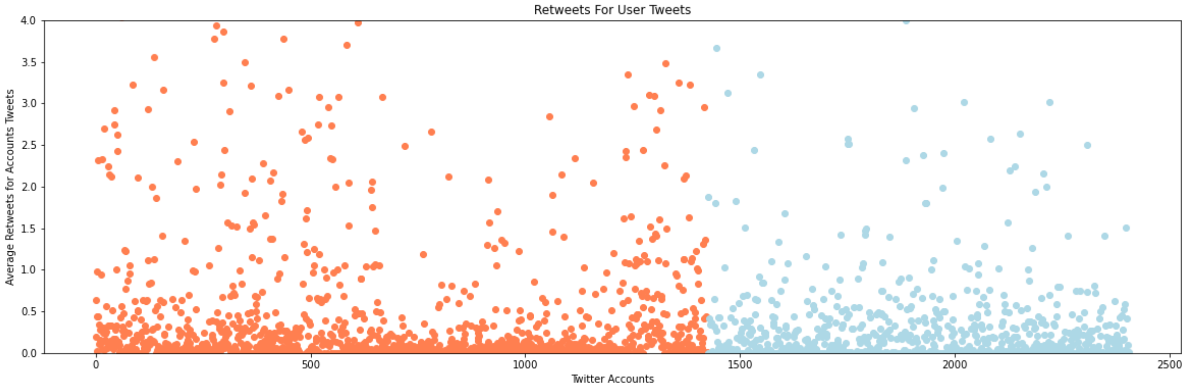

Figure 1: Scatter plot for the “Retweets of User Tweets”feature after applying it on our dataset.

Average TweetsPerDay: In this feature, we compute the average number tweets of the considered user that he posted in a day. As an example, a value of 2.5 of this feature means that in average each day this user posts 2.5 tweets. This feature is expected to give good results in marking troll accounts since it reflects the post frequencyof the twitter user. Figure 2 shows the scatter plot for the values of thisfeature after evaluating it using our dataset of trolls and non-trolls Twitter accounts. As shown in the figure, the results show that normal accounts post more tweets than troll accounts. This result can be misunderstood by some researcher since it is expected larger tweets from troll accounts. But based on our analysis, troll accounts usually concentrate on retweets of other tweets and usually delete the tweets after a specified time to hide their real activities.

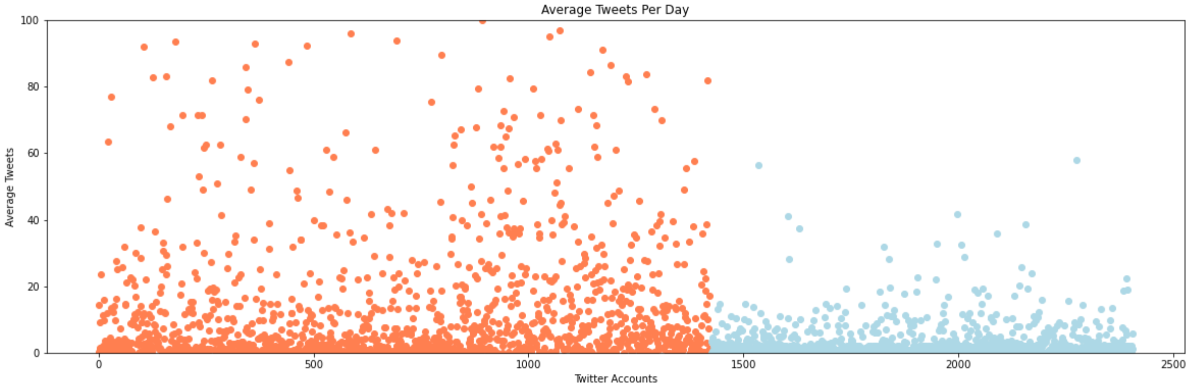

Figure 2: Scatter plot for the “Average Tweets Per Day”feature after applying it on our dataset.

Average Likes PerTweet: In this feature, we compute the average number of likes that every tweet get from other twitter users. As an example, a value of 10 for this feature means that in average each tweet from this twitter user gains 10 likes. This feature reflects the acceptability degree of the posted tweets by other twitter users. It is expected to obtain high values in the case of normal accounts than troll accounts since troll accounts tweets usually come in the inverse of normal social directions and attitudes. Figure 3 shows the scatter plot for the values of thisfeature after evaluating it using our dataset. As shown in the figure, the results show that normal accounts obtain more likes than troll accounts.

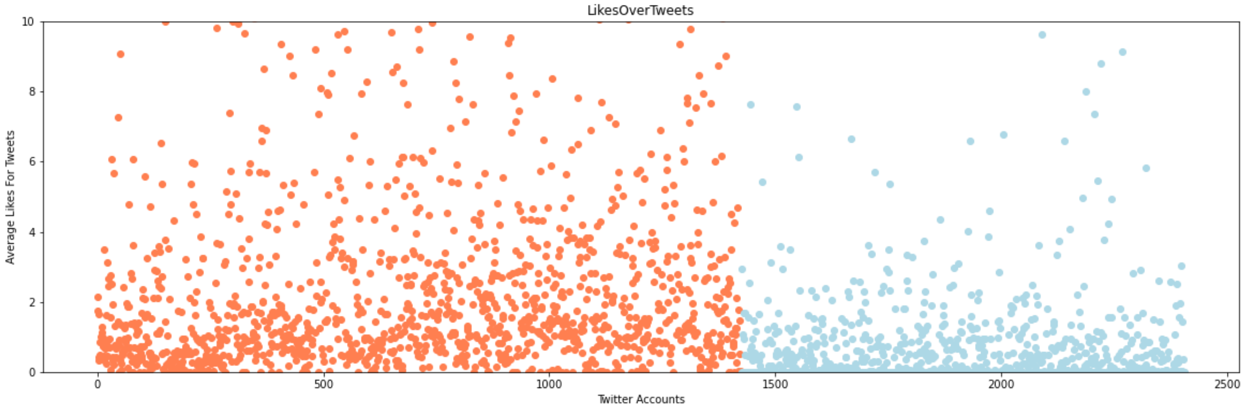

Figure 3: Scatter plot for the "Average Likes per Tweet"feature after applying it on our dataset.

Tweets with Zero Likes: In this new feature, we compute the average number of user tweets that do not get any likes or we can mention it as zero-like tweets. To normalize the values of this feature, we divided the zero-like tweets number by the total number of original tweets. As a result, we get values normalized between zero and one. As an example, a value of 0.6 for this feature means that in average 60% of the tweets of this user do not receive any like from other twitter users. This feature strongly reflects the degree of interaction between this user and other users. To the best of our knowledge, this feature is not used before and this is the first use of it in detecting troll accounts. Figure 4 shows as expected the effectiveness of this feature where it can significantly characterize between troll and normal accounts.

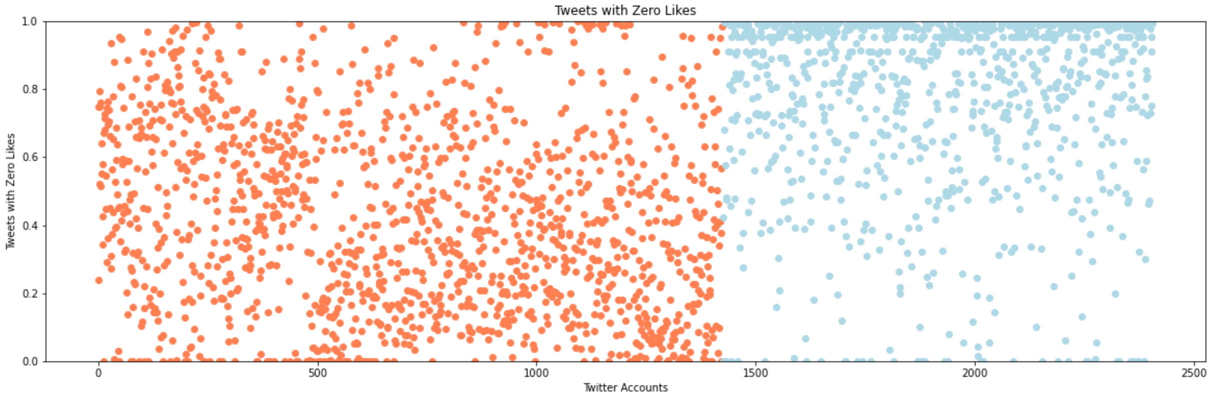

Figure 4: Scatter plot for the “Tweets with Zero Likes”feature after applying it on our dataset.

Tweets with Hashtags: In this feature, we compute the number of user tweets that has a hashtag inside. To normalize the values of this feature, we divided number of tweets with hashtags over the total number of original tweets. As a result, we get values normalized between zero and one. As an example, a value of 0.9 for this feature means that in average 90% of the tweets of this user has a hashtag inside its text. This feature has the ability to measure the degree of interaction between this user and the general topics and hashtags in the society. Figure 5shows the scatter plot for the values of thisfeature after evaluating it using our dataset. As shown in the figure, the results confirm the stability and robustness of this feature since there is a big and easy to note difference between the normal and troll accounts. The results of this feature approve that troll accounts use hashtags heavily much more than normal accounts.

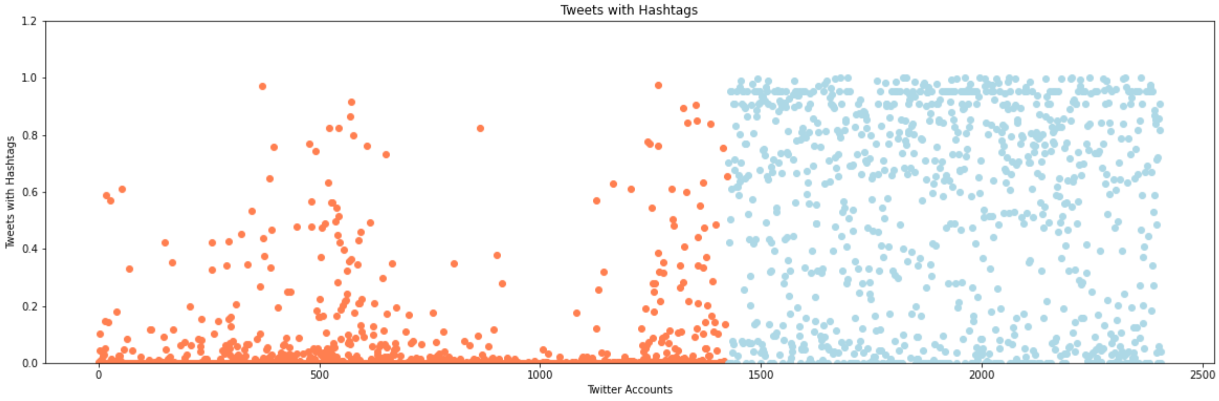

Figure 5: Scatter plot for the “Tweets with Hashtags”feature after applying it on our dataset.

Tweets with Mentions: In this feature, we compute the number of user tweets that has a user mention inside. To normalize the values of this feature, we divided number of tweets with user mentions over the total number of original tweets. As a result, we get values normalized between zero and one. As an example, a value of 0.1 for this feature means that in average 10% of the tweets of this user has a mention to other users inside its text. As the previous feature, this feature has the ability to measure the degree of interaction between this user and the other users in the society. Figure 6 shows the scatter plot for the values of thisfeature after evaluating it using our dataset. As shown in the figure, the results confirm the efficiency of this feature where there is again a large enough difference between the normal and troll accounts.

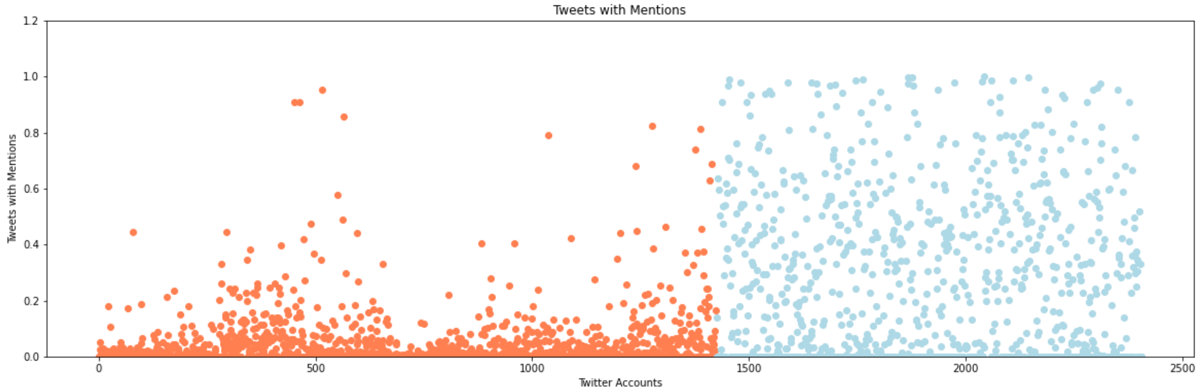

Figure 6: Scatter plot for the “Tweets with Mentions”feature after applying it on our dataset.

4.2 Selected Classification Algorithms for Twitter Troll Classification

In our experimental study, four machine learning algorithms for the troll classification are utilized and compared: K-Nearest Neighbors Algorithm (KNN) [23], Logistic Regression (LR) [24], Support Vector Machine (SVM) [25], and Classification and Regression Algorithm (CART) [26]. These algorithms are implemented in Python using sklearn library. KNN is a popular algorithm in machine learning that works by finding the distances between a point and all other examples/points in the data, a specified (K) closest examples are selected, then votes for the most frequent label (in the case of classification tasks).It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. The second algorithm called logistic regression is one of the most popular supervised learning and machine learning algorithms.LR is used for predicting the categorical dependent variable using a given set of independent variables. The outcome of LR algorithm must be a categorical or discrete value. The SVM is another supervised machine learning algorithm that is capable of performing classification, regression and outlier detection. For example, the linear SVM classifier works by drawing a straight line between two classes to classify the considered examples. The SVM algorithm finds the closest points to the line from both the classes andthese points are called support vectors.The last algorithm CART is a predictive model, which predicts an outcome variable's values based on other values. The output of CART is a decision tree where each fork is a split in a predictor variable and each end node contains a prediction for the outcome variable. In the feature selection step, a wrapper method called Backward Elimination is utilized which starts with all the features and the least significant feature is removed.

4.3 Results and Discussion

To measure the classification accuracy of algorithms we used the proposed features as described in Section 4.1 and our gathered dataset as described in Section 3. The performance of algorithms are compared using the accuracy measure which it is the most common performance metric used to evaluate classification algorithms. It is computed simply by computing the average classification accuracy as a ratio of correctly predicted observation to the total observations using the following Equation.

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN)$$

where: TP = True positive; FP = False positive; TN = True negative; FN = False negative

Table 3 shows the results of comparing the four considered classification algorithms by using a different number of features in every experiment. Four different feature sets are tested 4, 6, 8,10 and all features. The results of this table demonstrate the effectiveness of derived and used features to detect and characterize twitter troll accounts even when a small number of features are used.

Table 3: The performance of classification algorithms using different set of features

Classifiers	4 best features (%)	6 best features (%)	8 best features (%)	10 best features (%)	All features (%)
KNN	**92.39**	90.73	89.48	88.99	88.10
LR	91.19	**94.19**	**94.05**	92.94	90.23
SVM	78.01	70.12	67.08	58.92	59.01
CART	92.12	93.49	92.80	**94.19**	**95.57**

From Table 3 we note that in the case of small number of features KNN and LR algorithms gets the best classification results where LR is a little better than KNN algorithm. Unfortunately, the performance of the KNN algorithm gradually decreases as the features dimensionality increase which reflects the nature of this algorithm. In other side, LR obtained the best results in two cases and gets the best results when the 6 best features are used. When we have a higher number of features, the CART algorithm obtained the best results comparing to other algorithms and also achieve the best classification accuracy overall tested cases (95.57%). These results show that CART algorithm is more robust in dealing with troll detection problem since it does not affected by scaling, normalizing, or missing data. Another reason for this is that even in small number of features, the CART algorithm achieved also acceptable results. Moreover, it is noted that the performance of the CART and the LR algorithms are greater than 90% in all test cases. The SVM obtained the worst results comparing with other classifiers. In general, the results demonstrate the power of the new proposed features to detect troll twitter accounts even in the case of small number of features. As an example, using only the best 4 features (Tweets with Mentions, Tweets with Hashtags, Tweets with Zero Likes, and Average Tweets Per Day) can achieve a 92% classification accuracy using KNN and CART algorithms. As a result, the classification algorithm may be able to work very fast and with acceptable accuracy in the online classification scenarios during media campaigns and unrest political situations.

5. Conclusion

Aftertwitter becomes one of the most popular and widely used social media websites, people start to use it frequently to share their opinions, feeling, and orientations, especially during political unrest or protests. As a results, many governments, political parties, and organizations in order to disrupt the protest operations on Twitter or to influence public opinion, electronic flies (Trolls) are widely and effectively used. As a result, detecting and characterizing these bot accounts becomes a significant task to reduce or mitigate its effect on the real general opinion. This paper presents an intensive analysis that can be utilized to effectively detect the troll accounts and isolate its bad effect from Twitter. We considered the public trolls accounts datasets published by Twitter and we also gathered a new dataset from Twitter that includes tweets and users' information from different countries to make a fair analysis for the trolls' accounts. The results show that the suspicious activities of Twitter troll accounts can be used to detect most of these accounts automatically even without using sentiment analysis and opinion mining techniques with accuracy of 95%. To accomplish this task, we propose a set of robust and efficient features that can accurately characterize troll accounts even when a small number of features are considered.

References

[1] Buhl, Hans Ulrich, et al. "Big data." (2013): 63-68.

[2] Gil Press, May 9, 2013, 09:45am, A Very Short History Of Big Data, available: https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/#8a6dc8665a18

[3] Banks, David L., and Nicholas Hengartner. "Social networks." Encyclopedia of Quantitative Risk Analysis and Assessment 4 (2008).

[4] Kolleck, Nina. "Social network analysis in innovation research: using a mixed methods approach to analyze social innovations." European Journal of Futures Research 1, no. 1 (2013): 25., https://doi.org/10.1007/s40309-013-0025-2

[5] Knoke, David, and Song Yang. Social network analysis. Sage Publications, 2019.

[6] Twitter, Inc. "Twitter." URL: https://twitter. com/SenBlumenthal/status/1175102777351122945 (Data obrashcheniya: 20.10. 2019) (2010).

[7] Lewis et al., 2011 P. Lewis, R. Rezaie, R. Brown, N. Roberts, R.I.M. Dunbar Ventromedial prefrontal volume predicts understanding of others and social network size Neuroimaging, 57 (4) (2011), pp. 1624-1629

[8] Alsmadi, Izzat, and Michael J. O'Brien. "How Many Bots in Russian Troll Tweets?." Information Processing & Management 57.6 (2020): 102303.

[9] Sahmoud, Shaaban, and Hayder Safi. "Detecting Suspicious Activities of Digital Trolls During the Political Crisis." 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). IEEE, 2020.

[10] Fornacciari, Paolo, et al. "A holistic system for troll detection on Twitter." Computers in Human Behavior 89 (2018): 258-268.

[11] Alhazbi, Saleh. "Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter." IEEE Access 8 (2020): 195132-195141.

[12] Engelin, Martin, and Felix De Silva. "Troll detection: A comparative study in detecting troll farms on Twitter using cluster analysis." (2016).

[13] Badawy, Adam, et al. "Characterizing the 2016 Russian IRA influence campaign." Social Network Analysis and Mining 9.1 (2019): 1-11.

[14] Ghanem, Bilal, Davide Buscaldi, and Paolo Rosso. "TexTrolls: Identifying Russian trolls on Twitter from a textual perspective." arXiv preprint arXiv:1910.01340 (2019).

[15] Kim, Dongwoo, et al. "Analysing user identity via time-sensitive semantic edit distance (t-SED): a case study of Russian trolls on Twitter." Journal of Computational Social Science 2.2 (2019): 331-351.

[16] Abbott, Andrew, and Angela Tsay. "Sequence analysis and optimal matching methods in sociology: Review and prospect." Sociological methods & research 29.1 (2000): 3-33.

[17] Kellner, Ansgar, Christian Wressnegger, and Konrad Rieck. "What's all that noise: analysis and detection of propaganda on Twitter." Proceedings of the 13th European workshop on Systems Security. 2020.

[18] Broniatowski, David A., et al. "Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate." American journal of public health 108.10 (2018): 1378-1384.

[19] Zannettou, Savvas, et al. "Characterizing the use of images by state-sponsored troll accounts on Twitter." arXiv preprint arXiv:1901.05997 (2019).

[20] Makice, Kevin. Twitter API: Up and running: Learn how to build applications with the Twitter API. " O'Reilly Media, Inc.", 2009.

[21] Berthold, Michael R., et al. "KNIME-the Konstanz information miner: version 2.0 and beyond." AcM SIGKDD explorations Newsletter 11.1 (2009): 26-31.

[22] Website: https://github.com/twintproject/twint, Date: 10/12/2019

[23]Kozma, Laszlo. "k Nearest Neighbors algorithm (kNN)." Helsinki University of Technology (2008).

[24]Kleinbaum, David G., et al. Logistic regression. New York: Springer-Verlag, 2002.

[25]Noble, William S. "What is a support vector machine?." Nature biotechnology 24.12 (2006): 1565-1567.

[26]Steinberg, Dan. "CART: classification and regression trees." The top ten algorithms in data mining. Chapman and Hall/CRC, 2009. 193-216.

[27] Kumar, Akshi, and Teeja Mary Sebastian. "Sentiment analysis on twitter." International Journal of Computer Science Issues (IJCSI) 9.4 (2012): 372.

[28] Jamali, Mohsen, and Hassan Abolhassani. "Different aspects of social network analysis." 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06). IEEE, 2006.

[29] Paavola, Jarkko, et al. "Understanding the trolling phenomenon: The automated detection of bots and cyborgs in the social media." Journal of Information Warfare 15.4 (2016): 100-111.

[30] Zheng, Alice, and Amanda Casari. Feature engineering for machine learning: principles and techniques for data scientists. " O'Reilly Media, Inc.", 2018.

[31] Dong, Guozhu, and Huan Liu, eds. Feature engineering for machine learning and data analytics. CRC Press, 2018.

[32] Twitter Transparency Project, https://transparency.twitter.com/ retrieved in 20-12-2019

[33] Albahli, Approach Saleh, et al. "COVID-19 public sentiment insights: A text mining approach to the gulf countries." Cmc-Computers Materials & Continua (2020): 1613-1627.

[34] Albahli, Saleh, et al. "Predicting the type of crime: Intelligence gathering and crime analysis." Computers, Materials & Continua 66.3 (2020): 2317-2341.