Webology (ISSN: 1735-188X)
Volume 19, Number 3, 2022

Social Engagement Analysis For Detection Of Fake News On Twitter Using Machine Learning

Ali Imran[1] , Moiz Uddin Ahmed (Ph.D)[2] , Shahid Hussain (Ph.D)[3] , Javed Iqbal[4]

[1]Research Scholar - MS (CS), Department of Computer Science, Allama Iqbal Open University, Islamabad, Pakistan.

[2]Assistant Professor, Department of Computer Science, Allama Iqbal Open University, Islamabad, Pakistan.

[3]Assistant Professor, Department of Mass Communication, Allama Iqbal Open University, Islamabad, Pakistan.

[4]Ph.D. Scholar, Department of Mass Communication, University of Karachi, Pakistan.

Abstract

Fake news alludes to the news containing deluding or manufactured information that is groundless, which is purposefully misrepresented and spread among a community. It can mutilate reality and become a cause of many social problems and misunderstandings. Many studies have been conducted to investigate the engagement features. However, the choice of parameters is diverse, secondly, the spreading of news is a continuous process that needs to be evaluated. This study proposes a model for analyzing social engagement features on a news dataset of Twitter which is one of the most important platforms for spreading news. The news data was collected from an open-source data store of Kaggle and LIAR. The data was preprocessed and transformed into proportional values with six engagement features i.e. Retweets, Likes, Comments, Quoted Retweets, Multimedia and Images. These features were divided into 05 classification models with a combination of two engagement features. Based on the results four models were selected for further analysis. Models 1 (Retweet and QRT) and Model 5 (Likes and QRT) showed an imbalance of accuracy metrics on fake and real news data. Model 3 (Comments and QRT) and Model 4 (Retweets and Likes) obtained a good balance on accuracy metrics. Results showed that social engagement features can be used to estimate the credibility of news on Twitter.

Keywords: Fake News, Social Media, Social Engagement, Machine Learning, Data Analysis

Introduction

The social networking sites (SNS)are specialized web applications that provide a platform to people for virtual social relationships with other people having similar careers, backgrounds, interests and activities (Kumar & Revathy, 2020). These sites provide people the opportunity to share news, political affairs, current affairs, or their opinions in almost real-time (Roumani & Nwankpa, 2017). People from all generations use the social media such as Facebook and Twitter (Sayed & Dafoulas, 2019), where they publish, share and tweet about diverse affairs such as economy, entertainment, showbiz, politics, events, sports, educational and personal stories (Kim & Lee, 2019). Twitter is one of the most popular social media of the contemporary era (Nyow & Chua, 2019; Maryam & Ali, 2019). People mostly pay attention to shared content and make opinions through social engagement. Social engagement is an individual or collective involvement on a particular post/topic on social media. It is also defined as an individual reaction on social media post through comments, like, dislike and retweeting etc. (Yang & Shu, 2019). Social engagement provides very helpful information regarding user opinion on specific information, however, the judgment of authenticity of any post needs careful analysis to determine whether the shared information is accurate or fake (Shu & Wang, 2018).

Fake News refers to the news which contains groundless, fabricated or misleading information. It is usually defined as false material intentionally generated for malicious intentions or unfair political polarization (Shu & Sliva, 2017). Undoubtedly, the fake piece of information spreads openly across the networks or social sites and most people share this piece of information with others without verification. The phenomenon of fake news is on the rise with rise in the social media like Twitter and Facebook (Vogel & Meghana, 2020). Fake news sometimes can be generated and propagated by the hostile actors for getting unfair means in the country. The content related to social crises easily lead to believe and people blindly consider it true and ultimately spread fake news to others by showing social engagement (Shu & Zhou et al., 2019). The severity of such action varies among people, but in some conditions, these can bring very severe consequences. The recent acceleration in spreading of fake news can be a peril to democracy and can push countries into wars. The motives of such kind of fake news are usually to hurt or disrepute the opponents, to manufacture the public opinion, and even to propagate the ideological and national biases (Sharma, Qian, Jiang, Ruchansky, Zhang, & Liu, 2019). The Machine Learning (ML) is the best approach for the detection of fake news on social media with the analysis of social engagement.

Machine learning is an important area of artificial intelligence which helps in accurate prediction through the use of software applications and historical data. The machine learning algorithms can be used to automatically detect problems without performing human tasks manually (Ray, 2019). With the massive increase in the use of mobile phone devices, now the population gets news very fast and on real time. So, as a result, today there is a big chance of being deceived by the fake or wrong material on social media (Pomerleau & Rao, 2019). Mostly the people are unaware and don’t understand that the news they read is fake or fabricated, which has become a big challenge in today’s media and information industry. There are regular posters, who are most culpable for the proliferation of the false news or material online, which makes the situation more and more frightening and alarming (Vosoughi & Aral, 2018).

This research is designed to investigate the user’s response and study its impact on the inaccuracy of the news. It proposes an analysis model by using social engagement parameters quote, tweet, retweet. Social engagement features and attributes of the fake and real posts have been collected through an open dataset of Kaggle and LIAR.

Twitter

The term Twitter has been derived from the chirping sound of bird called a ‘tweet’. First launched in the year 2006, Twitter has now revolutionized as one of the most popular social media across the globe. The figures of Twitter usage in the year 2021 show that about 200 million users globally are using the application with uncountable data sharing. This application gives the users with the liberty to openly communicate with the celebrities along with many convenient features for data sharing on equal terms. Famous people from almost all the professions like sports, athletes, movie stars, politicians and bureaucrats are exchanging the views in a real-time environment. People can follow each other and important official twitter accounts for updates and surveillance. A careful scrutiny of the twitter account of a person can elaborate important information about the social recognition, prestige, public relations, and influence in a particular field. Number of followers of a twitter user tells about his/her recognition in a particular field. Marketing companies and new media are also using twitter for information dissemination in real time and at a fast speed. So, in this way the possibilities of using the twitter for almost innumerable areas are vast and it can be used for leisure, media, health, medical treatment etc.

Social Engagement

Social Engagement is also being defined as an individual reaction on social media post through comments, like, dislike and retweeting etc (Yang & Shu, 2019). Social engagement is an individual or collective involvement on a particular post/topic on social media community. Social engagement provides very helpful information regarding user opinion on specific information (Shu & Mahudeswaran, 2018). A user implicitly expresses his/her point of view through social engagement in the shape of like, share, comments etc these are very helpful to find the authenticity of the content on social media. The users implicitly show their interest in particular information on social media by using social context features (Alhayan & Pennington, 2018). Figure 1 shows the structure of social engagement features for Twitter.

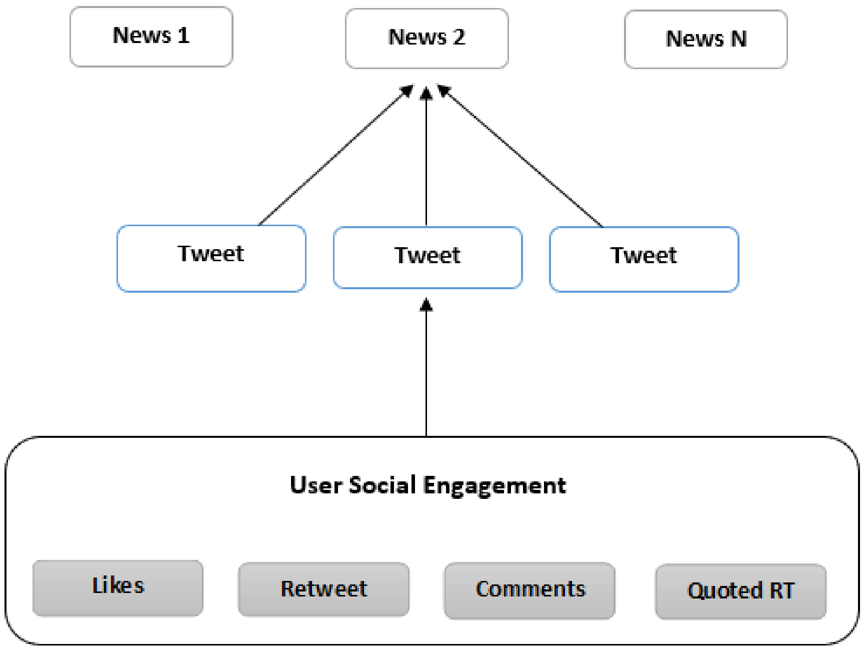

Figure 1: Structure of Social Engagement for Twitter

It is more reliable feature to detect accurately the opinion of the user against specific piece of news. Social engagement included Retweets, Likes, Comments, and Quoted Retweets. It represents clearly how users interact with real and fake news on social media and depicts the viewpoint of the users against a particular piece of information (Ruchansky & Seo, 2017).

Tweet Likes

The like symbol when clicked shows that the post is aligned with user's thoughts (Yang & Shu, 2019). It shows the interest and the trustworthiness of users on a piece of information on social media. It can be used to assess the popularity of news by counting the number of likes (Yanagi & Orihara, 2020). It can also be used to strengthen the reliability of a company (Viswanath & Bashir2014). So, the companies always try to increase the likes on social media to get popularity (Nadzir & Harun, 2019).

Retweet

The term Retweet (RT) is actually re-share functionality with purpose to share Tweet with one's followers keeping up the original context. Mostly the followers access a tweet and express their interest about the information to others through Retweets (Gupta et al. 2013). Retweets have not contained one's own reply/comments and are normally used while conveying with interest or eagerness for the main Tweets. Generally, the users don't Retweet or share a Tweet that they have no interest in or not enthusiastic about (Metaxaset al. 2017). The information is made and propagated through retweets. In this way, a Retweet plays a much greater part in spreading information than some other component and because of this sometimes it can create serious similarly as irrelevant, social issues by rehearsing extraordinary information especially in political field, during electoral activities (Shao et al. 2017). In political field, the rumors are spread during electoral activities on Twitter to gain political advantages. Accordingly, Retweet has become an assessment point for various researchers for fake news detection on Twitter.

Comments/Reply

Comments on social media post is very important to get the clear opinions of the users. The users react on the post though writing text in comment box (Guo et al. 2018). Social media users clearly mention their thoughts on the content of post. This would be a gold mine for researchers. The participants react positively and negatively in both ways in comments section (Garimella et al. 2016). They mostly use positive words or phrase which show the participants do agree with the statement which is written in the social media post. Positive comments mostly hold positive thoughts and wording like “Great”, Good, “Agree”, “Congratulation”, “Great Work” etc. The negative comments mostly give negative sense and thoughts which are against the social media post (Anitha et al. 2019).

If a user doesn’t like a post’s content, he/she uses negative words which are not in support of the post. Negative comments mostly hold negative thoughts and wording like “Very Bad”, “Disagree”, “Alarming”, “Angry”, “Harmful”, “Hate” etc. Some users react on post with multimedia like Videos, GIF image and memes in comments box. These are most important to judge the reaction of the people on the post and important for various research point of views. The comments are much important for fake news detection on social media.

Quoted Retweets

The Quoted Retweet (QRT) is an additional element of the Twitter that was presented back in 2015. The traditional Retweet just posts a unique Tweet as an adherent read to his/her own followers without composing any additional remarks. Conversely, the Quoted Retweet lets an adherent present a unique Tweet on his/her own and simultaneously compose his/her remark with respect to the first Tweet. The retweet feature only helps in sharing the data with other users, however, the Quote Retweet lets the adherents share a post by adding their remarks with respect to the information they are keen on. A published QRT is indicated like a Tweet and have similar highlights as that of a Tweet, such as Retweet, Reply and Like.

Moreover, a Quoted Retweet can likewise be cited by different adherents. The Twitter users can utilize QRT feature to communicate their official answers for a specific Tweet to their adherents by remarking or responding (contradict, concur or impartial) on the first Tweet, which wasn’t possible through traditional Retweets (Garimella et al. 2016). The pattern of utilizing Quoted Retweet is ceaselessly expanding and as per an exploration, the explanation is on the grounds that the Retweets are changing to QRT. Since Quoted Retweets assume the function of spreading information as well as contain the cited Tweet and the user’s response on it. It is expected that QRT will be substantially more helpful than the regular Retweets. Quoted Retweet is much important to judge the credibility of information on Twitter. This is flash point and important feature to detect fake news (Jang & Park, 2019).

Fake News on Social Media

Fake News refers to the news which contains fabricated, misleading information which is actually groundless. The term fake news is defined as false material intentionally generated for malicious intentions or unfair political polarization (Shu & Wang, 2017). The Fake News reporting was existing in every era. At the point when online media was progressing quickly back in 2010s, it was abused to disseminate totally wrong information, which was masked in the type of news-casting. Now with the advancement of social media and utilization of counterfeit, news has likewise strongly expanded (Kraski, 2018). Counterfeit news and false reporting have a few likenesses; both use fabricated news reports to spread the information for public trust. The false reporting is formally associated with the press organizations, while the Fake News spreads with counterfeit or fabricated information which is manufactured by an individual or organization on social media (Jang & Park, 2019). Mostly the materials identified with fake news vigorously spread on social media have the accompanying shared characteristics “parody”, “spoof”, “instigate” and “mis-interpretation” (Klunder et al. 2017). Different forms of fake news have been identified in the literature i.e. clickbait, propaganda, parody, sloppy journalism, and misleading headline (Khan et al. 2021). Each is a form of deceiving information that captures the attention and misleads the audience. Moreover, it has gotten progressively harder to distinguish fake news because the methods of spreading fake news are changing each day. Accordingly, there is a need of concrete research which concentrates on the identification of guidelines for fake news detection.

Related Work

Twitter has more than 322 million active users (statista.com) around the world, which is likely to rise up to a figure of 340 million by the year 2024. It is mostly used to share important news about celebrities, political figures and superstars (Garg & Rani, 2017). It is also used as a communication tool for chatter and gossip between notable personalities and their followers (Acharya & Sharma, 2018). Besides these, the use of Twitter has also been seen in emergency situations like pandemics, floods, and earthquakes. Journalists have used Twitter to float important events for fast dissemination as breaking news. Due to the rising scale of Twitter usage, the need for credibility estimation emerged to determine the trustworthiness of the information. These challenges attracted the researchers to explore the influencing factors for the detection of fake news. The existing research on fake news has been divided into four major areas: Knowledge-base, network propagation, features-based, and hybrid approach (Khan et al. 2021).

The knowledge base analyzes the visual features of news content i.e. text, graphics, and multimedia content by using classification, regression, and clustering techniques. Some studies have carried out the lexical analysis by considering the syntactic features of the content. The syntax analysis and user profiles were used to classify the fake or original nature of tweets generated by bots (Wickramarathna & Ganegoda, 2019). The semantic analysis was carried out by investigating sentiment using deep learning technique (Kula et al., 2020). A study (Collins et al. 2020), provides an overview of Text Mining and Natural Language Processing (NLP) for the identification of social bots through posted facts. Visual artifacts were also used to distinguish fake images that were artificially created with certain characteristics (Gupta et al., 2013). The grouping of metadata with relevant text was demonstrated to identify rumors and categorize stances using the linguistic features (Wang, 2017). Network base propagation shares a post or news with the connected community of followers comprising of friends and relatives etc. It has been reported that more than 45% of the posts are further shared by followers. One study utilized the time series technique to study the propagation status at different timestamps and correlated it with the news correctness (Liu et al. 2018). They achieved an accuracy of 85% for Twitter data and established the importance of propagation intensity. Another study analyzed the flagging feature introduced by Facebook allowing users to mark the forged news (Tschiatschholek et al., 2018). A technique "Detective" was developed using the Bayesian Network to evaluate the accuracy of the model.

The feature-based is further classified on the basis of user profiles, social networks, and post contents analysis. The popularity of Twitter due to 200 billion tweets a year and 500 million a day (Aslam, 2021) has made it an ideal platform to be targeted by bots and mechanized programs. Therefore many studies have examined the user accounts from the perspective of bots nature (Giber et al. 2020 ; Grinberg et al. 2019). The other aspect is the correlation between personality traits and fake news (Shu et al., 2018). The experienced and naïve users have been segregated to study the impact of their personalities on fake and real news. The social network was identified as an important criterion, as news is always shared with friends, relatives, and colleagues (Kong et al., 2019). It has been used to associate the propagation and cooccurrence with time series analysis (Liu and Wu, 2018). The post content is dependent upon the user's response in terms of his/her social engagement. A user implicitly expresses his/her point of view through social engagement in the shape of like, sharing, and comments which may be considered to decide about the faith in the news (Yang & Shu, 2019). One study considered the usage of quotes and retweets and determined the propagation intensity to establish the truthfulness of the news (Jang et al., 2019). The research on fake news is in progress by considering different aspects of actions & engagements and applying the relevant techniques and algorithms. The growing amount of data is resulting in new patterns and trends and therefore, further research is required to discover the impact of engagement. There is still room for analyzing the spreading trends linked with user reaction to study its impact on the classification of the news.

Objectives

The following are the objectives of the study:

Objective 1: To learn the user's behavior on fake news posts on Twitter.

- H1: There are differences in users' behaviors on real and fake information on social media

Objective 2: To identify and analyze the key parameters for detecting fake news.

- H2: There are different parameters to detect fake news; the analysis power of social engagement parameters on Twitter would help to detect fake news.

Data Collection and Pre-Processing

The model of data collection and preprocessing is depicted in figure 2. It shows the steps of data collection, preprocessing, visualization of featured values, and data analysis.

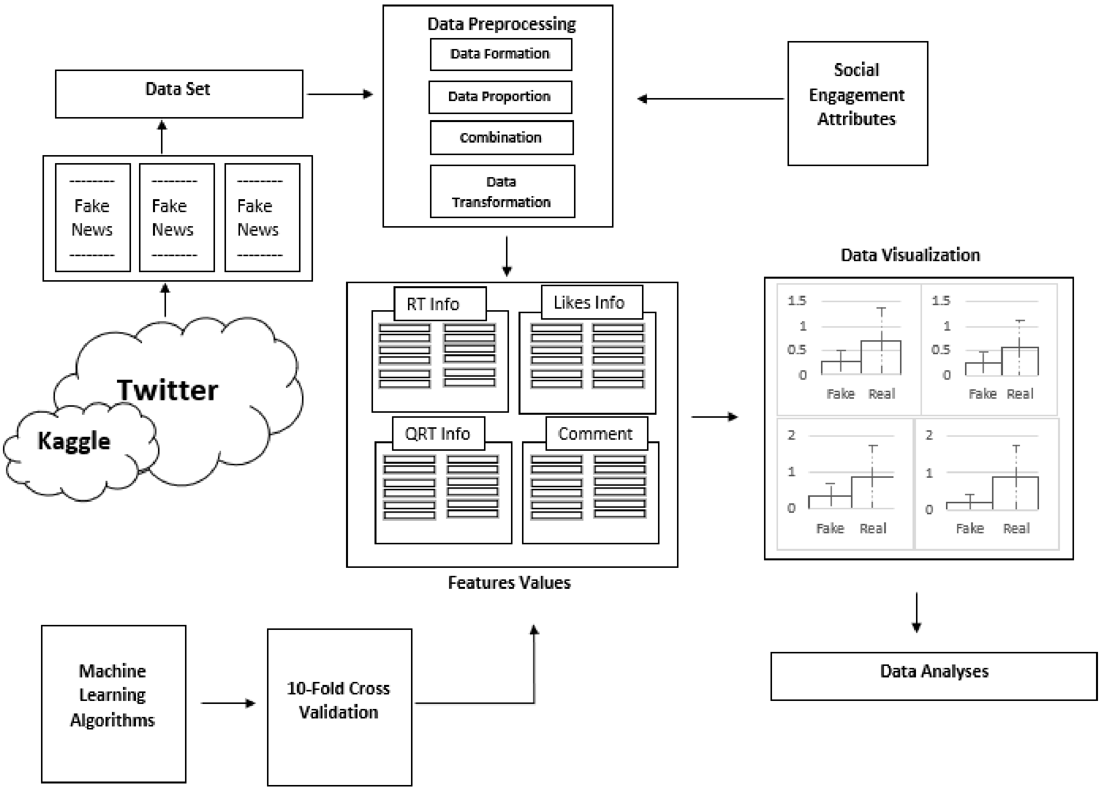

Figure 2: A Complete Data Analysis Model

A sample of fake and real news comprising 300 records was collected from the Kaggle open data source. Kaggle data was comprised of multi-source data related to the news. The Twitter-related data were filtered using the Twitter Advanced Search filter (ASF). The second step was data preprocessing, where data was cleaned by removing the duplicate values and converted into MS Excel format with the features mentioned in table 1. The filtered data was further extracted and reformed using Twitter API in Laravel 5.7 as a web scraping tool (Jang & Aslam, 2019). The API enabled to execution of the commands for the extraction of the desired data. The function shows the code segment for collecting the news data

Function 1

```
public function get Twitter Post Detail ($post_id)
 {
 $data = Twitter :: get Tweet ($post_id);
 $retweets = Twitter :: get Rts ($post_id);
 $media = '';
 $type = '';
 $url = '';
```

```
if (count($data->extended_entities->media)==1)
    {
    $media = 'Yes';
    $type = $data->extended_entities->media[0]->type;
    $url = $data->extended_entities->media[0]->media_url_https;
    }
    else
    {
    $media = 'No';
     }
    echo "<h3>Twitter User Detail</h3>";
    echo "<strong>Tweet User Name:</strong> ".$data->user->name."<br>";
    echo "<strong>Tweet User ScreenName:</strong> ".$data->user-
>screen_name."<br>";
    echo "<strong>Tweet User Location:</strong> ".$data->user->location."<br>";
    echo "<strong>Tweet User Description:</strong> ".$data->user-
>description."<br>";
    echo "<strong>Tweet User Followers:</strong> ".$data->user-
>followers_count."<br>";
    echo "<strong>Tweet User Friends:</strong> ".$data->user->friends_count."<br>";
    echo "<strong>Tweet User Since:</strong> ".date('d M, Y', strtotime($data->user
    >created_at))."<br>";
    echo "<strong>Tweet User Profile Picture URL:</strong> ".$data->user
    >profile_background_image_url;
    echo "<h3>Tweet Post Detail</h3>";
    echo "<strong>Tweet:</strong> ".$data->text."<br>";
    echo "<strong>Retweets:</strong> ".$data->retweet_count."<br>";
    echo "<strong>Likes:</strong> ".$data->favorite_count."<br>";
    echo "<strong>Comments:</strong> ".$data->comments_count."<br>";
    echo "<strong>Quoted Retweets:</strong> ".$data->quoted retweets_count."<br>";
    echo "<strong>Media Available or Not ?:</strong> ".$media."<br>";
    echo "<strong>Media Type:</strong> ".$type."<br>";
    echo "<strong>Media URL:</strong> ".$url."<br>";
    echo "<strong>Most Recent Tweets:</strong> ".count($retweets)."<br>";
    }
```

The featured values were comprised of six attributes as shown in table 1.

Table 1: All the Features of Dataset

Sr No.	Attributes	Description
1	**Retweets**	No's sharing posts with others
2	**Likes**	Interest in the post
3	**Comments**	User's opinion on the post through text
4	**Quoted Retweets**	User's retweets with remarks
5	**Multimedia**	Use of videos and animations
-6	**Images**	Use of images

During the next step the data proportion technique was applied to convert the frequency-based data into binomial values (Jang, 2019; Carlos, 2011) by using the following formula:

$$\text{Proportion} = \frac{\text{Value of a Instance}}{\text{Sum of Instances Values}}$$

The researchers then formed the combination of two attributes ((Jang et al., 2019; Lakshamanarao et al., 2019) and proposed five models as shown in table 2.

Table 2: All the Features of Dataset

Models	Attributes Information
Model 1	Retweets, QRT
Model 2	Images, Multimedia
Model 3	Comments, QRT
Model 4	Retweet, Likes
Model 5	Likes, QRT

The data was further transformed and converted into first CSV format and then to arff for executing the machine learning algorithm by using Waikato Environment for Knowledge Analysis (WEKA) tool. The researchers also performed a statistical analysis in order to find the average difference between fake and real news attributes.

Statistical Data Analysis

Statistical analysis was performed to find the average differences between the attributes of fake and real news as shown in table 3.

Table 3: Averages Values of Features

	Retweet	Likes	Comments	Quoted RT
Fake	4571.77	8930.11	795.53	570.78
Real	26242.3	152869.3	9053.5	3551.4

Table 3 shows that the average frequency of social engagement features of fake news is far less than that of real news. The average values of social engagement features were visualized using the boxplot as shown in figures 3 to 6. The news data was divided into two groups i.e. group 1 representing the fake and group 2 denoting the real news. The graph displays similar trends as given in table 3 showing less visualization and proportions for fake news. The fake news has been re-tweeted less (figure 3a), It has fewer likes (3b), commented smaller (3c), and lesser QRT (3d) than real news.

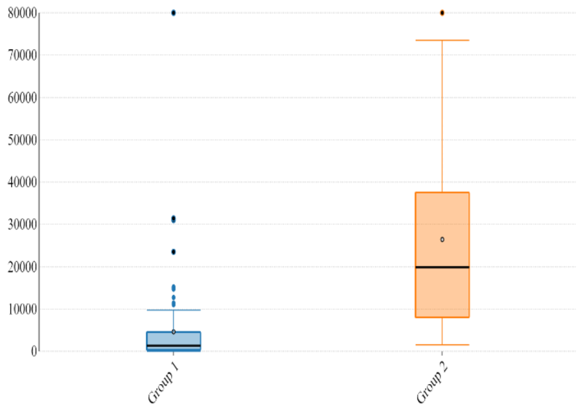

Figure 3: Average Retweets Fake and Real News

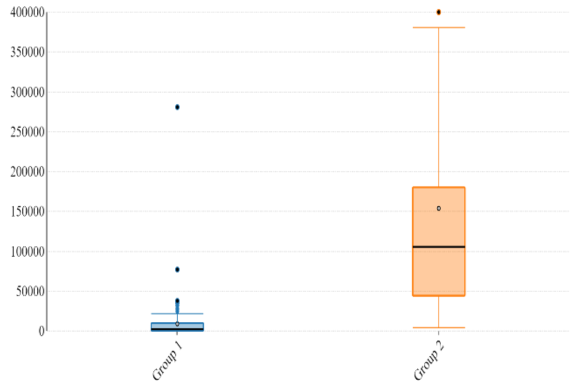

Figure 4: Average Likes Fake and Real News

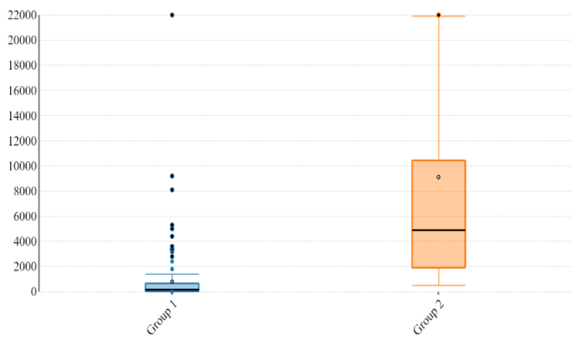

Figure 5: Average Comments Fake and Real News

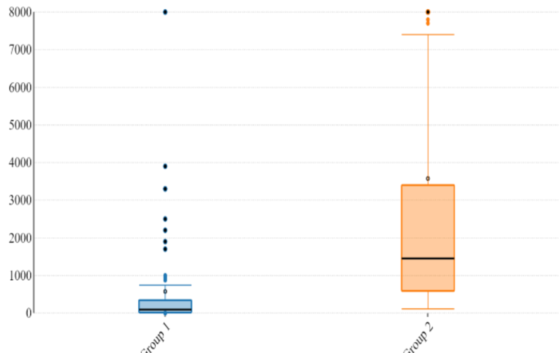

Figure 6: Average Quoted RT Fake and Real News

Results of Machine Learning algorithms on Models

The researchers measured the performance of each model by employing fifteen machine learning algorithms. The study performed a 2-fold cross-validation with 30 % validation and 70 % training for each model. Some of the algorithms were neglected due to low performance. Model 1 used the Retweet and Quoted Retweets features of social engagement. The Bayes Network classified the instances with 62.85% of accuracy. The same trend was continued as the Decision Stump and K Star classified the instances with 60.95% and 62.85 % accuracy respectively.

The K Star obtained the highest balance in accuracy tradeoff. Overall, the fake news achieved low values of accuracy measures as compared to real news. Model 2 (Image & Multimedia) was dropped due to the repetition of the same values. Model 3 considered the data with Comments and Quoted RT features.

Table 4: Performance of ML Algorithm for Model 1

Model 1							
Classifiers	Class	Correctly Classified	TP	FP	Precision	Recall	F-Score
Bayes Net	Real	62.85%	0.918	0.724	0.575	0.981	0.725
	Fake		0.276	0.019	0.935	0.276	0.426
Decision Stump	Real	60.95%	0.981	0.762	0.563	0.981	0.715
	Fake		0.238	0.019	0.926	0.238	0.379
K Star	Real	62.85%	0.724	0.467	0.608	0.724	0.661
	Fake		0.533	0.276	0.659	0.533	0.589

Random tree correctly classified the instances with 68.09% accuracy, followed by random forest (67.62%) and IBK (65.71 %) accuracies. Model 3 showed a good balance of accuracy measures both for fake as well as real news. IBK algorithm showed the best performance with comments and QRT features that were proved to have significant contributions to fake and real news. Model 4 evaluated Retweets and Likes features on the sample data.

Decision Table showed an accuracy of 60 % followed by Random Tree with an accuracy of 62.38%. The model showed a good balance of accuracy both for fake and real news. Model 5 was comprised of Likes and QRT data where the random forest showed an accuracy of 61.9 % followed by random tree with an accuracy of 62.85 %. It showed higher accuracy for real news as compared to fake news.

Table 5: Performance of ML Algorithm for Model 2

Model 3							
Classifiers	Class	Correctly Classified	TP	FP	Precision	Recall	F-Score
Random Forest	Real	67.62%	0.743	0.390	0.655	0.743	0.696
	Fake		0.610	0.257	0.703	0.610	0.659
Random Tree	Real	68.09%	0.733	0.371	0.664	0.733	0.697
	Fake		0.629	0.267	0.702	0.629	0.663
IBK	Real	65.71%	0.657	0.343	0.657	0.657	0.657
	Fake		0.657	0.343	0.657	0.657	0.657

Table 6: Performance of ML Algorithm for Model 3

Model 4							
Classifiers	Class	Correctly Classified	TP	FP	Precision	Recall	F-Score
Random Tree	Real	62.38%	0.619	0.371	0.625	0.619	0.622
	Fake		0.629	0.381	0.623	0.629	0.626
Decision Table	Real	60.00%	0.952	0.752	0.559	0.952	0.704
	Fake		0.248	0.048	0.839	0.248	0.382

Table 7: Performance of ML Algorithm for Model 4.

Model 5							
Classifiers	Class	Correctly Classified	TP	FP	Precision	Recall	F-Score
Random Forest	Real	61.90%	0.657	0.419	0.611	0.657	0.633
	Fake		0.581	0.343	0.629	0.581	0.604
Random Tree	Real	62.85%	0.648	0.390	0.624	0.648	0.636
	Fake		0.610	0.352	0.634	0.610	0.621

Conclusion

In the current era of digital society, a huge amount of information is passed through social media applications. This information is comprised of both real and fake news related to a particular aspect of society. Many studies have been conducted to analyze the social engagement features on Twitter, however, the growing amount of data is providing an opportunity to continue the analysis of social engagement features. This research has been conducted with the same spirit. It presents the analysis of social engagement features and their impact on fake news detection on Twitter. The data was collected from the Kaggle and LIAR open source which was preprocessed for the analysis. The researchers considered six features of social engagement and divided them into four models i.e. Model 1 (Retweets and QRT), Model 2 (Comments, QRT), Model 3 (Retweet, Likes) and Model 4 (Likes, QRT). The study also performed the statistical analysis of frequencies on fake and real news data.

The statistical analysis of frequencies showed that the intensity of social engagement on fake news is less than that of real news. The fake news was found with the smaller count of retweets, likes, comments, and QRT engagement. The researchers further applied machine learning algorithms on the four proposed models. For Model 1, The highest performance was obtained by IBK algorithm, however, an imbalance was found among the accuracy measures among fake and real news. The Retweat and QRT engagements were found less important for fake news data. Model 2 showed the same performance on feature engagements for both fake and real news and highest accuracy was obtained by IBK algorithm. The comments and QRT were found important for both fake and real news. For Model 3, the performance of the classifier achieved higher accuracy on fake news using the random tree algorithm. In Model 4, the highest accuracy was obtained by random forest but overall, the model showed low performance on fake news data.

The study concludes that social engagement can be used to detect the faith in news on Twitter. The analysis may be extended to a larger dataset with more features. The techniques of Natural Language Processing (NLP) may also be used for the content analysis.

References:

Acharya, M., & Sharma, S. (2018, December). Semantic Analysis of Twitter Posts. In 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS) (pp. 104-109). IEEE.

Alhayan, F., Pennington, D. R., & Ayouni, S. (2018, April). Measuring passive engagement with health information on social media. In 2018 21st Saudi Computer Society National Computer Conference (NCC) (pp. 1-6). IEEE.

Anitha, K. M., Saraswathy, V., Gayathri, K., & Priya, S. S. (2019, February). An Approach to Comment Analysis in Online Social Media. In 2019 3rd International Conference on Computing and Communications Technologies (ICCCT) (pp. 332-335). IEEE.

Aslam, S., 2021. Twitter by the Numbers: Stats, Demographics & Fun Facts. https://www.omnicoreagency.com/twitter-statistics/.

Collins, B., Hoang, D.T., Nguyen, N.T., Hwang, D., 2020. Trends in combating fake news on social media–a survey. Journal of Information and Telecommunication 1–20.

Garimella, K., Weber, I., & De Choudhury, M. (2016, May). Quote RTs on Twitter: usage of the new feature for political discourse. In Proceedings of the 8th ACM Conference on Web Science (pp. 200-204).

Garg, N., & Rani, R. (2017, June). Analysis and visualization of Twitter data using k-means clustering. In 2017 International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 670-675). IEEE.

Gibert, D., Mateu, C., Planes, J., 2020. The rise of machine learning for detection and classification of malware: research developments, trends and challenges. J. Netw. ComputAppl. 153, 102526.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. Science, 363(6425), 374-378.

Guo, H., Cao, J., Zhang, Y., Guo, J., & Li, J. (2018, October). Rumor detection with hierarchical social attention network. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management (pp. 943-951).

Gupta, A., Lamba, H., Kumaraguru, P., & Joshi, A. (2013, May). Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In Proceedings of the 22nd international conference on World Wide Web (pp. 729-736).

Jang, Y., Park, C. H., &Seo, Y. S. (2019). Fake news analysis modeling using quote retweet. Electronics, 8(12), 1377.

Khan, T., Michalas, A., & Akhunzada, A. (2021). Fake news outbreak 2021: Can we stop the viral spread?. Journal of Network and Computer Applications, 103112.

Klunder, J., Schmitt, A., Hohl, P., & Schneider, K. (2017). Fake news: Simply agile. Projekt management und Vorgehensmodelle 2017-Die Spannungzwischen dem Prozess und den Mensch imProjekt.

Kong, X., Shi, Y., Yu, S., Liu, J., Xia, F., 2019. Academic social networks: modeling, analysis, mining and applications. J. Netw. Comput. Appl. 132, 86–103. Kshetri, N., Voas, J., 2017. The economics of “fake news”. IT Professional 19.

Kraski, R. (2017). Combating Fake News in Social Media: US and German Legal Approaches. John's L. Rev., 91, 923.

Kula, S., Choraś, M., Kozik, R., Ksieniewicz, P., & Woźniak, M. (2020, June). Sentiment analysis for fake news detection by means of neural networks. In International Conference on Computational Science (pp. 653-666). Springer, Cham.

Kumar, A. S., &Revathy, S. (2020, June). Review on Social Network Trust With Respect To Big Data Analytics. In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) (pp. 721-727). IEEE.

Liu, Y., Wu, Y.-F.B., 2018. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Thirty-Second AAAI Conference on Artificial Intelligence. Liu, Y., Wu, Y.-F.B., 2020. Fned: a deep network for fake news

Maryam, A., & Ali, R. (2019, March). Misinformation source identification in an online social network. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.

Metaxas, P. T., Mustafaraj, E., Wong, K., Zeng, L., O'Keefe, M., & Finn, S. (2015, April). What Do Retweets Indicate? Results from User Survey and Meta-Review of Research. In ICWSM (pp. 658-661).

Nadzir, M. M., Harun, N. H., & Hassan, M. G. (2019, April). Social media engagement on Malaysian government agencies Facebook pages: An empirical analysis. In 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT) (pp. 717-719). IEEE.

Nyow, N. X., & Chua, H. N. (2019, November). Detecting fake news with tweets' properties. In 2019 IEEE Conference on Application, Information and Network Security (AINS) (pp. 24-29). IEEE.

Oh, C., Roumani, Y., Nwankpa, J. K., & Hu, H. F. (2017). Beyond likes and tweets: Consumer engagement behavior and movie box office in social media. Information & Management, 54(1), 25-37.

Ray, S. (2019, February). A quick review of machine learning algorithms. In 2019 International conference on machine learning, big data, cloud and parallel computing (COMIT Con) (pp. 35-39). IEEE.

Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 797-806).

Sayed, H., Dafoulas, G. A., &Saleeb, N. (2019, December). Social Networking Sites (SNS) and Digital Communication Across Nations. In 2019 14th International Conference on Computer Engineering and Systems (ICCES) (pp. 228-233). IEEE.

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques. ACM Transactions on Intelligent Systems and Technology (TIST), 10(3), 1-42.

Shu, K., Zhou, X., Wang, S., Zafarani, R., & Liu, H. (2019, August). The role of user profiles for fake news detection. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 436-439).

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). Fake newsnet: A data repository with news content, social context and spatial temporal information for studying fake news on social media. arXiv preprint arXiv:1809.01286.

Shu, K., Wang, S., Liu, H., 2018. Understanding user profiles on social media for fake news detection. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, pp. 430–435.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., & Menczer, F. (2017). The spread of fake news by social bots. arXiv preprint arXiv:1707.07592, 96, 104.

Song, G., Kim, S., Hwang, H., & Lee, K. (2019, January). Blockchain-based notarization for social media. In 2019 IEEE international conference on consumer electronics (icce) (pp. 1-2). IEEE.

Tschiatschholek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., Krause, A., 2018. Fake news detection in social networks via crowd signals. In: Companion Proceedings of the the Web Conference 2018. International World Wide Web Conferences Steering Committee, pp. 517–524.

Viswanath, B., Bashir, M. A., Crovella, M., Guha, S., Gummadi, K. P., Krishnamurthy, B., & Mislove, A. (2014). Towards detecting anomalous user behavior in online social networks. In 23rd {USENIX} Security Symposium ({USENIX} Security 14) (pp. 223-238).

Vogel, I., & Meghana, M. (2020, October). Detecting Fake News Spreaders on Twitter from a Multilingual Perspective. In 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 599-606). IEEE.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false new online. Science, 359(6380), 1146-1151.

Wang, W.Y., 2017. Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection arXiv preprint arXiv:1705.00648.

Wickramarathna, N. C., & Ganegoda, G. U. (2019, December). Detecting Automatically Generated Tweets Using Lexical Analysis and Profile Credibility. In 2019 4th International Conference on Information Technology Research (ICITR) (pp. 1-6). IEEE.

Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019, July). Unsupervised fake news detection on social media: A generative approach. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 5644-5651).

Yanagi, Y., Orihara, R., Sei, Y., Tahara, Y., & Ohsuga, A. (2020, July). Fake News Detection with Generated Comments for News Articles. In 2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES) (pp. 85-90). IEEE.

Web references: https://www.statista.com/statistics/303681/twitter-users-worldwide, 2021