Performance analysis of classifiers in social media text for Credibility Assessment

P.SuthanthiraDevi[1], S.Karthika[2]
[1, 2]SSN College of Engineering, Department of Information Technology,
Kalavakkam, Chennai-603 110, Tamilnadu, India
[1]devisaran2004@gmail.com
[2]skarthika@ssn.edu.in

Abstract **- Social media is an interactional tool to disseminate an individual's perception. This research work involves one such micro-blogging platform namely Twitter. The process of mining Twitter data for identifying the credibility of tweets are currently emerging technology in the online community. Since the twitter data comprises of unknown and incomplete data, credibility checking becomes tedious and uncertain. The main objective of this work is to classify the credibility of the tweet with different machine learning algorithms like SVM, NB, and RF. The performance of the credibility analysis is evaluated using the metrics of accuracy, precision, recall, and F1-score. The authors analyze the PHEME dataset which contains tweets from eight real-time events obtained through the thread of conversation. The experimental results of credibility check show that Random Forest Classifier algorithm is best suitable for six out of eight real-time events with an average accuracy of 87%.**

Keywords: Credibility assessment, Machine learning algorithm, Classification, Social Media, N-gram

I. INTRODUCTION

Social media use continues to grow in the world and according to a report, 71% of the country is currently 'socially active' [15]. At present twitter has 326 million active users out of which nearly 80% of tweet users access it using mobile phones [16]. The tweets adapt a structure comprising of text, created date and time, id, status source, screen name, retweet count, favorite count, and location. During the crisis time, Twitter has become a highly influential media for sharing valuable information to a vast community of people. However, this information contains a different form of spams like advertisements, fake images, fake videos, and rumors. This information was spread in an unverified manner, and it causes a harmful penalty to the community. This raises a strong motive for every tweet user to check the credibility of the tweet before sharing in the social media. Oxford dictionary defines credibility as "the quality of being trusted and believed in". The tweet related to the quality of trusted information about the event is said to be credible.

The authors of [1] proposed a real-time system to assess the integrity of tweet messages and the user-generated content has been checked by assigning a score to each tweet. In [2], the authors present a Support Vector Machine Rank (SVM-Rank) model that computes a score to determine the credibility of a tweet based on the User and Twitter features. They designed the web-based system to compute the credit score of a tweet in real time and this system's performance was evaluated using response time and usability. The authors of [3] discuss an ideal methodology for checking the twitter information that uses the content and lexical features of the tweets. The contextual feature uses two approaches called cosine similarity and adjective based model to detect the contextually offensive content. The lexical classifier uses rule-based and LDA based Naïve Bayes (NB) algorithm to analyze the tweets. In this paper different feature of surveying the validity of the twitter has been depicted, and the framework to evaluate the credibility of tweets has been designed based on the adjective model.

In [4] the authors identify the trending topics from the conversational thread of other events and analyze the credibility of the tweet based on the trending topics. They discuss the fashion of information spread in social media. They classify the tweets automatically using user features like user posting and retweets into credible or non-credible classes. The authors of [5] conducted a post-election survey on US adults in the following levels. Level 1 discusses

fake news from three mainstream media articles related to Trump. The second and third levels deal with most recent pre-election headlines from PolitiFact, Snopes and Big True Stories from Guardians election timeline. The authors proved that a single fake news story was more effective than television news, and they present a system that clarifies the level of fake news exposure.

The spread of rumors has serious and chaotic consequences in the moments of crisis, as they can swiftly convey wrong information to people. The amplification of isolated risks has to be curbed. The major contribution of this research work is to make a crucial analysis of social media text in order to prevent the diffusion of inaccurate information using machine learning algorithms and to identify information that is well verified. This research work has analyzed the PHEME dataset which contains 297 conversational threads from eight rumorous events. The authors propose a methodology of the credibility assessment system to determine the credibility of all the tweets in the dataset. The taxonomy of machine learning techniques in credibility analysis and the evaluation metrics are discussed in section 2. Section 3 describes the dataset and pre-processing techniques and the authors present the methodology in section 4. Section 5 discusses the experimental results and various findings are illustrated in section 6. Finally, section 7 concludes the proposed methodology with the future scope as the extension of the work.

II. RELATED WORK

This section presents a brief outline of the existing research techniques used to validate the social media content for credibility. The authors of [6] had collected rumor information from Sina Weibo micro-blogging service. The authors extract a set of features from data and train a classifier to detect rumors automatically from the combination of true and false information. The client program used feature for micro-blogging and the event location feature are used for text classification to improve the accuracy. In [7] the authors built a classifier that assigns a trust value to identify the veracity of the rumor. The rumor veracity value is estimated using three new features like user's present and past behaviour and linguistic characteristics of the messages. Trustworthiness scores are calculated at various time windows from the rumor information using machine learning classifiers. The authors illustrate the results using visual user interface software.

In [8] the authors adopt the NLP and machine learning algorithm to extract the sentiment of the tweet. They use three approaches namely lexicon-based, machine-learning and sentistegth to analyze tweet content. The authors of [9] characterize the user-followee relationship based on User Affinity Score (UAS) and Content Similarity Score (CSS). Also, they developed a Followee-management nudge plug-in that helps the user to identify the correct followees. The authors of [10] proposed a framework namely Inquiry Comments Detection Model (ICDM) to identify the inquiry comments using the rule-based method. They extract the inquiry comments and apply the threshold value to identify the rumor.

In the paper [11], the authors have designed a Probabilistic Collaborative Filter model to identify the source of the rumor-tweet, retweet and to predict the future retweets. The authors of [12] developed a system to describe how rumors have been spread during crisis time. This system generates a target list to detect topics which are related to rumor events. They also identify the source of the rumored candidate and evaluate their results. The authors of [13] extract the rumor tweets from the general query and analyze the online misinformation based on three features namely content-based, network-based and microblog-specific memes.

TABLE 1: TAXONOMY OF MACHINE LEARNING BASED ALGORITHMS FOR CREDIBILITY ANALYSIS IN TWITTER

References	Type of acquired social media data		Types of the machine learning algorithm		Dataset	Inference
	Specified	Unspecified	Supervised	Unsupervised		
Fan Yang et.al.	✓		✓		Sina Weibo micro-blogging service	Rumour detection on Sina Weibo tweets
Giasemidis et.al		✓	✓		Collected all event public tweets, including User and Twitter features	Message classifier detects the veracity of the information
Chen et.al	✓		✓		Twitter data from 20 city government twitter accounts. It included both tweets and re-tweets made as responses to the government accounts	Analyze the tweets from the context of government uses of social media
Anjali Verma et.al		✓	✓		Collect data from randomly identified 100 Twitter users and their followers, followees	Characterize the user-followee relationship to identify correct followee
Zahoor-ur-Rehman et.al.	✓		✓		Collect twitter data from two events like PIA flight crash PK-661 in Havelian and Pakistan Elections 2013	Design a framework to identify the rumors based on the inquiry comments during crisis – Developed model namely ICDM
Zaman et.al.		✓	✓		Collect all the public tweet, retweets in every one hour time period from June 20th,2010 to July 29th,2010.	Predict the information spreading using novel retweet prediction technique
Takahashi et.al.	✓		✓		Collect Japanese language tweet using the keywords “server room or Geek house” and “Cosmo Oil”	To identify the spreading of multiple rumors during disaster time
Qazvinian et.al.		✓	✓		Collect popular rumor related data between 2009 and 2010.	Designed a learning framework that identifies the rumors in the microblog.

TABLE 2: SUMMARY OF MACHINE LEARNING TECHNIQUES AND EVALUATION METRICS

References	Detection Techniques	General Features	Evaluation Metrics
Fan Yang et.al.	SVM classifier and RBF kernel functions	Content and User features	Precision, Recall, F-score
Giasemidis et.al	Supervised binary classification problem, Logistic Regression, Random Forest and Decision Tree	Tweet features and User features	Accuracy, F1-score, Area Under ROC Curve (AUC), Cohen's kappa, Feature importance measure
Chen et.al	Lexicon-Based Techniques, Machine Learning-Based Techniques, Hybrid Techniques	City, date joined, number of days presence, no. of posts, no. of followers, no. of citizen responses	Sentiment score, Sentiment polarity, True positive rate, False positive rate, ANOVA test
Anjali Verma et.al	User-Followee behavioral characterization	The entire source tweet, reply and retweet user behavior.	User Affinity Score (UAS) and Content Similarity Score (CSS), Followee management nudge
Zahoor-ur-Rehman et.al.	Inquiry Comments Detection Model (ICDM)	All posts and comments with related metadata.	Precision, Recall, F-measure, Accuracy
Zaman et.al.	Probabilistic collaborative filtering models	Tweet features and Retweeter's features	Negative log-score, Correlation, Empirical retweet probability
Takahashi et.al.	Entity extraction techniques	Tweet features and Retweeter's features	Burst, Retweet ratio, Difference of Word Distribution, Accuracy
Qazvinian et.al.	Naïve Bayes classifier	Tweet features and Retweeter's features	Kappa coefficient, Likelihood ratio, Average Precision, Recall, KL-divergence, F-Score

The above Table 1 presents the taxonomy of credibility techniques according to the types of data available and detection methods. The events subjected for analysis can have two types of data namely specified and unspecified. The specified class of events have data collected specifically for a particular event or crisis. The unspecified class of events has the data collected for a period of time and the results are derived from this generic data. It is not particular to any event. Table 2 presents various machine learning techniques employed for classification and the metrics used to evaluate the performance.

III. PHEME DATASET

This research work analyzes the PHEME dataset which contains data for eight different events namely Putin-missing, Michael Essien contracted Ebola, Prince to play in Toronto, Germanwings plane crash, Charlie-Hebdo shooting, Ferguson unrest, Ottawa Shooting, and Sydney siege. All these events are related to disseminating a huge volume of rumorous news on Twitter [14]. The following table 3 presents the sample of tweets collected for the above eight events.

TABLE 3: SAMPLE TWEETS FROM PHEME DATASET

S.No	Event	Tweet
1.	Putin missing	*In March 2015, Russsian President Vladimir Putin did not appear for 10 days in public.*
2.	Michael Essien contracted Ebola	*Michael Essien a football player had contracted Ebola on October 12, 2014.*
3.	Prince to play in Toronto	*Musician Prince conducts a secret show in Toronto in November 2014.*
4.	Germanwings plane crash	*The plane crashed in the French Alps on March 24, 2015.*
5.	Charlie Hebdo shooting	*The shooting happened on Charlie-Hedbo in Paris were 11 people killed.*
6.	Ferguson unrest	*Michael Brown shot by a police officer on August 9, 2014, at Ferguson, USA*
7.	Ottawa shooting	*Soldier shot at National War Memorial in Ottawa in Canada*
8.	Sydney siege	*Hostages are being held and a siege is taking place at Sydney's Lindt Chocolate Cafe in Martin Place*

The dataset contains 297 conversational threads and every thread is composed of source tweet, retweets, and replies. The structure of the conversation thread is shown in the following figure 1. The source tweet is the initial tweet that starts or initiates the spread of rumor, a retweet is the reposting of the source tweet and reply is a response to the source tweet.

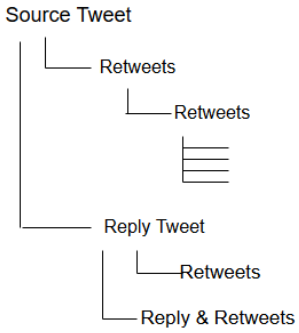

Fig. 1. Structure of Conversation Thread

Table 4 describes the structure of the conversational thread involved in the data related to the real-time events.

TABLE 4: DATASET WITH CONVERSATIONAL THREADS AND ASSOCIATED TWEETS

Event name	Conversational Threads	Tweets
Putin missing	9	73
Michael Essien contracted Ebola	2	41
Prince to play in Toronto	12	111
Germanwings plane crash	74	1237
Charlie Hebdo shooting	46	1207
Ferguson unrest	58	844
Ottawa shooting	25	311
Sydney siege	71	1232

IV. METHODOLOGY FOR ACCESSING THE CREDIBILITY OF SOCIAL TEXT

The architecture of the proposed system consists of five major modules namely Datastore, Pre-processor, Data annotation, Feature selection and Classifier for analyzing the content of tweets and its credibility. In figure 2 the authors present the proposed architecture of credibility assessing system for classifying tweets as Credible or Non-Credible. The modules involved in credibility assessment system are described as below:

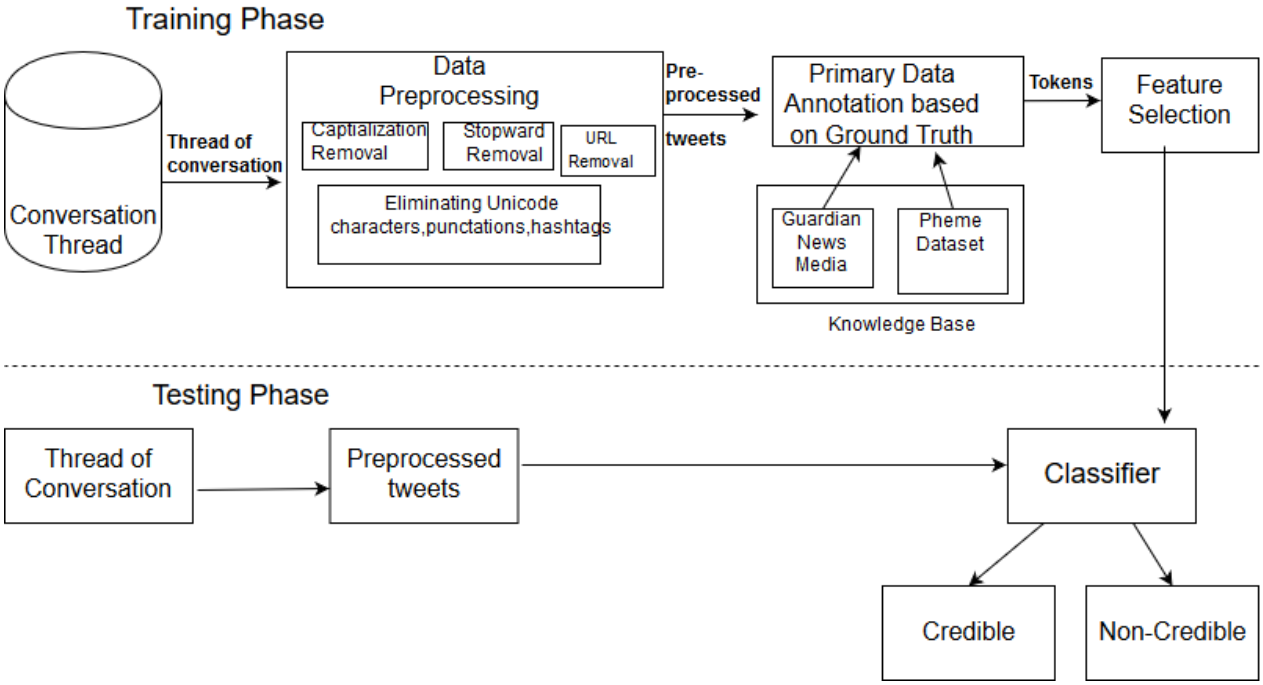

Fig. 2.Architecture for a credible assessment system

A. Data Store

The authors have used the PHEME dataset to build the repository. The repository consists of a conversation thread of tweet with the user feature and Twitter features. Each data contains the tweet features namely text, mentions, hashtags, URLs and retweets. The user features are like verified or not, number of friends, followers, the age for a user account. These features are used for feature extraction and generation.

B. Data pre-processing

The eight real-time events contain 297 conversational threads with 5,043 tweets. Twitter data may be incomplete and contains noisy which can produce misclassification results. Data pre-processing is used to transforming unstructured tweet data into an understandable format. These tweets are pre-processed using rules namely conversion of uppercase to lowercase, removal of RT, URLs, whitespaces, Unicode characters, punctuations, hashtags and stop words are applied.

C. Data Annotation

The tweet retrieved from the pre-processing module is classified as credible or non-credible tweet based on the information provided by the direct observation of news media (Ground truth). The quality of being trusted and believed tweets are classified as credible tweets. The labeling of the tweets is done based on the following definitions:

Credible Class (CC): The tweet related to the quality of trusted news about the event [2].
Non-Credible Class (NC): The tweet contains irrelevant information

D. Feature Selection

The pre-processed tweet text is converted into a weight matrix resulting in the tf-idf matrix. The feature selection is used to remove irrelevant features and this reduces the volume of the features. The feature selection using tf-idf is explained in algorithm 1. The highly related term vectors are generated based on the N-gram model which influences the identification of credible and non-credible classification.

Algorithm 1: Feature Selection using TF-IDF

Input: [Text Document-T_d, Feature Set – Fs, u_i-Unique terms in the text, Weight Matrix-TF, i(tdfs)-Inverse term frequency]
Output: [TF-IDF weights for selected feature set]
Create an array with a selected feature set
for each term u_i in F_s do
    for each text document $(td)_j$ T_d
        if $(TF)_{ij} \neq 0$ then (tdfs) ++
    end for of text document
        $i(tdfs)_i = log(T_d / (tdfs)_i)$
end for of term
for each term u_i in F_s do
    for each text document $(td)_j \in T_d$
        $(TF\text{-}IDF)_{ij} = TF_{ij} * i(tdfs)_i$
    end for text of text document
end for of term

E. Classifier

The tweets are identified as credible or non-credible based on the selected feature set using the state-of-art machine learning algorithms like Naïve Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF). NB algorithm is a probabilistic classifier that makes a simplified assumption about how the features interact with each other [17]. SVM algorithm is suitable for finding the category of the text by using the hyperplane and it splits the text documents into two categories [18]. RF is the most versatile algorithm for determining the category of the text. This model has trained data several times randomly to achieve good

accuracy [19]. The following algorithm 2 formally represents the generation of the n-gram model for the proposed credibility assessment system.

Algorithm 2: Generation of n-grams

Input : [T_d –Text Document, C – {c_{id}} ,where c – category, c_{id} – Total number of category), T_u – Unigram,T_b –Bigram]
Output: [Top five correlated Unigrams and Bigrams]
Input N, where N is top 5 most correlated unigram, bigrams
for each category $c_{id} \in C$ do
Generate chi-square scores for each term in T_d
Calculate array of indices
Arrange feature set in an array
Computation of unigram and bigram terms
end for

The selected features and n-gram models are used for assessing the credibility of social text. This system is trained and tested with different features, to determine the suitable credibility assessment as explained in the following algorithm.

***Algorithm 3: Credibility assessment* of social text**

Input : [Training data -T_r,Testing data -T_s)]
Output : [Predicted Credible Class]
Split training and testing data as 75%,25%
Generate Matrix of Token counts from text documents
Create a dictionary with a list of tokens
Convert count matrix to a normalized tf-idf
Learn a list of tokens and idf
Fit the model according to the T_r and T_s

V. RESULTS AND DISCUSSION

The tweets repository of 5,043 tweets was manually annotated as either Credible Class or Non-Credible class with respect to the ground truth. The tweets are pre-processed with the rules as discussed in data pre-processing. The following table 5 shows a sample pre-processing.

TABLE 5.PRE-PROCESSING OF SAMPLE TWEET FOR CHARLIE HEDBO EVENT

Sample Tweet	***"RT@France: 10^ people dead after shooting at HQ of satirical weekly newspaper #CharlieHebdo, according to witnesses & http://t.co//FkYxGmuS58"***	
Rule No.	**Rule Name**	**Pre-Processed Tweet**
1.	Convert to lowercase	rt@france: 10^ people dead after shooting at hq of satirical weekly newspaper #charliehebdo, according to witnesses & http://t.co//fkyxgmus58
2.	Removal of RT	@france: 10^ people dead after shooting at hq of satirical weekly newspaper #charliehebdo, according to witnesses & http://t.co//fkyxgmus58

3.	Replacement of amp with and	@france: 10^ people dead after shooting at hq of satirical weekly newspaper #charliehebdo, according to witnesses and http://t.co//fkyxgmus58
4.	Removal of hashtags	@france: 10^ people dead after shooting at hq of satirical weekly newspaper charliehebdo, according to witnesses and http://t.co//fkyxgmus58
5.	Removal of Punctuations and Symbols	france 10 people dead after shooting at hq of satirical weekly newspaper charliehebdo according to witnesses and httptcofkyxgmus58
6.	Removal of URL	france 10 people dead shooting hq satirical weekly newspaper charliehebdo according witnesses and
7.	Unwanted Whitespaces Removal	france 10 people dead shooting hq satirical weekly newspaper charliehebdo according witnesses
Pre-Processed Tweet	france 10 people dead shooting hq satirical weekly newspaper charliehebdo according witnesses	

The pre-processed tweets are further represented as tf-idf matrix and computed scores are used for feature selection. Chi-square technique is applied to determine the terms and its correlation with the corresponding category of the class. Vector space model is constructed using the best features, and these features are used to compute unigram, bigram, and tri-gram terms. The following table 6 shows the total number of the thread of conversations, number of pre-processed data and number of training and testing tweets, number of unigrams, bigram and trigram generated for all the eight events. The study of the table presents that the generated uni-grams are predominantly more than bi-grams and tri-grams.

TABLE 6.NUMBER OF PRE-PROCESSED DATA, TRAINING AND TESTING INSTANCES, N-GRAM TERMS

Tweet Category	**Putin missing**	**Michael Essien contracted Ebola**	**Prince to play in Toronto**	**German-wings plane crash**	**Charlie Hebdo shooting**	**Fergus on unrest**	**Ottawa shooting**	**Sydney siege**
No of Conversations	**9**	**2**	**12**	**74**	**46**	**58**	**25**	**71**
Preprocessed Data	64	32	102	1046	1108	600	298	1014
Training Data	**48**	**24**	**77**	**785**	**831**	**450**	**224**	**761**
Testing Data	16	8	26	262	277	150	75	254
Unigrams	**5**	**7**	**14**	**272**	**327**	**173**	**62**	**276**
Bigrams	0	6	7	110	54	58	19	120
Trigrams	0	5	4	78	23	49	13	96

The authors have calculated the tf-idf vector using unigrams (1, 1) and a combination of unigram and bigram (1,2). It is inferred that, using only unigrams in this computation results in classifying most of the credible tweets as non-credible. The combination of unigram and bigram lowers the misclassification rate compared to the previous method. The following

figure 3 and 4 depict the Correctly Predicted (CP) classes' vs. InCorrectly Predicted (ICP) classes for eight events using the three state-of-art machine learning algorithms.

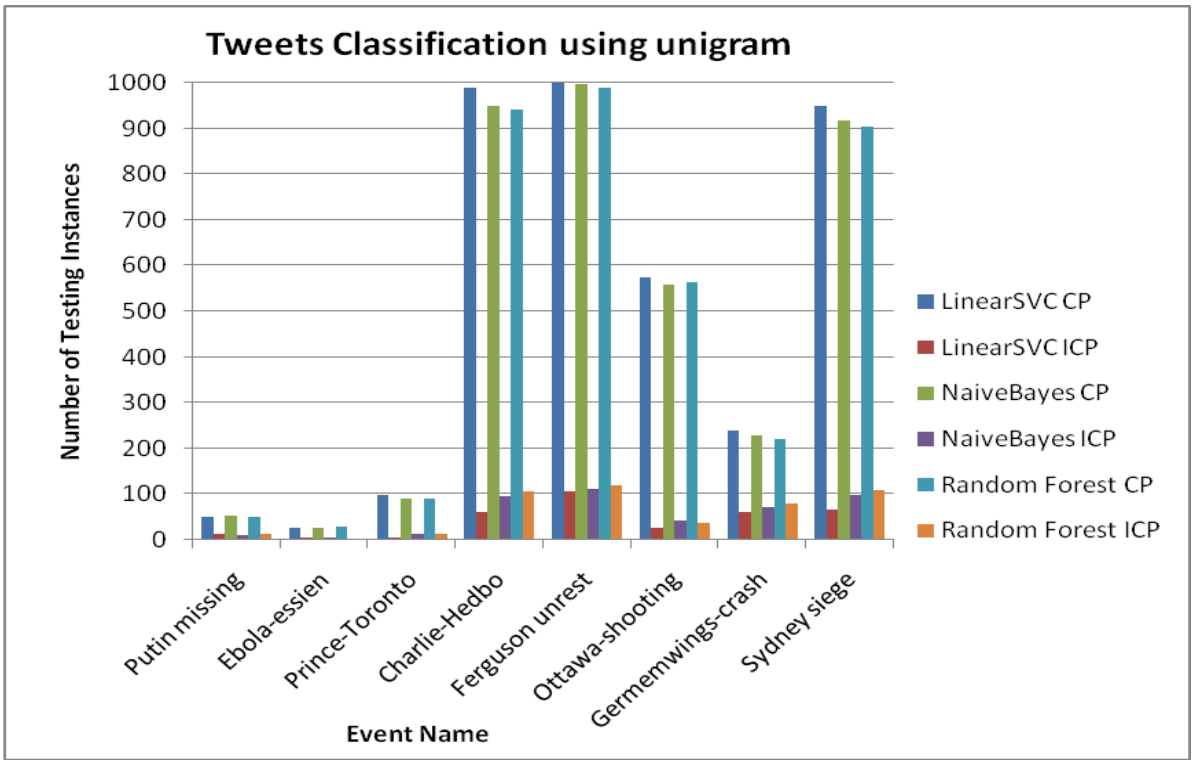

Fig. 3.Classification of credible tweets using unigram

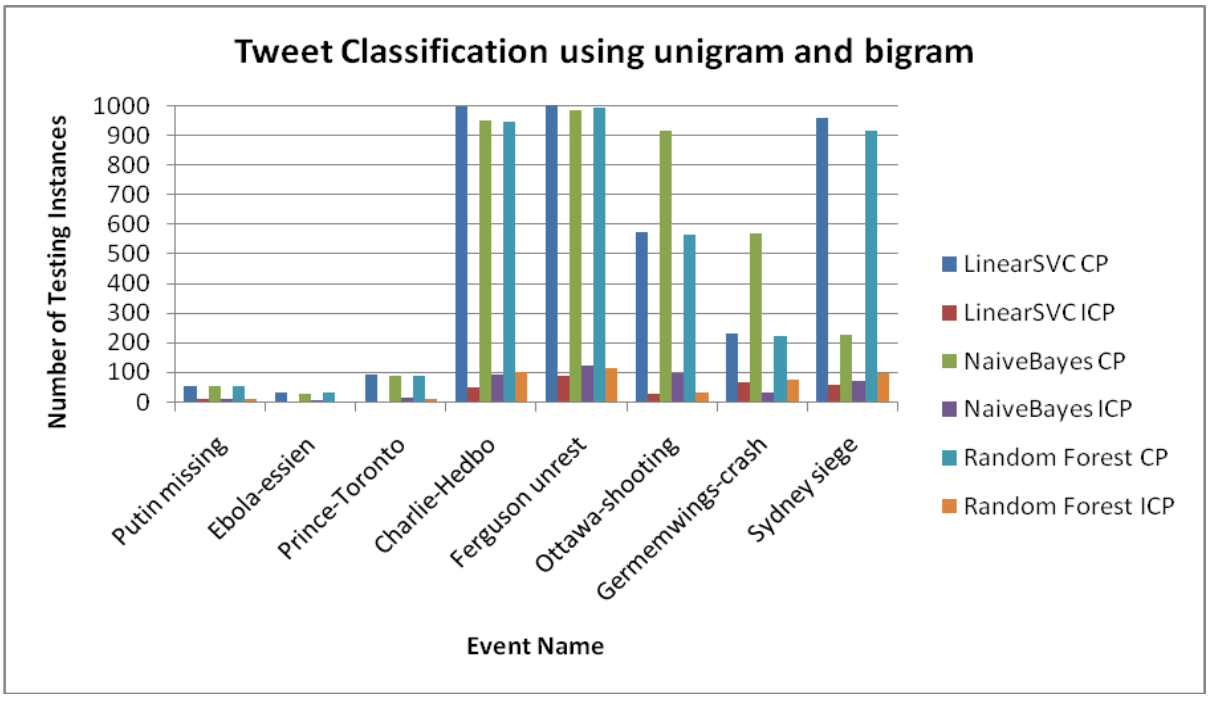

Fig. 4.Classification of tweets using unigram and bigram

Based on the n-gram analysis, the classification algorithms SVM, NB, and RF have been applied for credibility assessing. The prediction of credible instances using only unigram terms performs well for Putin-missing, Ebola Essien events. It is because of the terms like ***birth, Russian*** for Putin-missing event and ***contracted, unconfirmed*** for Ebola Essien. The authors have inferred that unigram terms perform well only on low volume dataset. The

prediction of credible instances using unigram and bigram combination is high for the events like Prince-Toronto, Ferguson unrest, Ottawa-shooting, and Germanwings-crash, Charlie-Hedbo, Sydney-siege. It is because of the bigram terms like ***held inside, isis flag, flag held*** instead of using unigram terms like ***inside, isis, flag*** for the event Sydney siege. The bigram terms ***terrorist attack, police storm*** are influenced for Charlie-Hedbo events instead of unigram terms ***Islam,Muslims***.

TABLE 7. THE TEST RESULTS FOR CORRECTLY PREDICTED INSTANCES (CP) AND INCORRECTLY PREDICTED INSTANCES (ICP) USING UNIGRAM AND BIGRAM (1, 2) CLASSIFICATION FOR EIGHT EVENTS

Tweets	Actual Class	Predicted Class	Credibility Check	Event
Mr Putin building the nation is great Good country man and best president	CC	CC	CP	Putin-missing
haha all he wanted was a week off putindead	NC	CC	ICP	
In response to inquiries we can confirm that Prince will not be performing at tonight	CC	CC	CP	Prince -Toronto
Clearly prince is having a show in Toronto prince concert	NC	CC	ICP	
its a Sunni flag ISIS are Sunni So are Al Qaida BlackSunni GreenShia	NC	NC	CP	Sydney siege
Flag in window of Sydney Lindt cafe not an ISIS flag Reads There is no God but Allah and Muhammad is the messenger of God	CC	NC	ICP	
Unconfirmed reports claim that Michael Essien has contracted Ebola	CC	CC	CP	Ebola-essien
AC Milan midfielder Michael Essien has been diagnosed with Ebola Get well soon Michael Daily Times	CC	NC	ICP	
Several hostages freed at Jewish supermarket in Paris Photo Thomas Samson AFP	CC	CC	CP	Charlie-Hedbo
This is a fair revenge French planes with impunity in Afghanistan Syria Libya and Iraq	NC	CC	ICP	
Shocker In Ferguson police beat a man and charged him for bleeding on them	CC	CC	CP	Ferguson-unrest
Sov Citizen in Dallas this week PLANNED TO KILL COPS and was brought in alive how come a Shoplifter is denied same courtesy	NC	CC	ICP	
ISIS Media account posts picture claiming to be Michael ZehafBibeau dead OttawaShooting suspect Canada	NC	CC	ICP	Ottawa-shooting
Very reliable source on Parliament Hill tells me the assailant has been killed cdnpoli ottnews	CC	CC	CP	
Accident aircraft looks to be Germanwings Airline code U or GWI flight Barcelona to Dusseldorf U	CC	CC	CP	Germanwings-crash
The plan would be to level out Id be going with hijacking	NC	CC	ICP	

The above table 7 test results show that the prediction of Credible Classes (CC) and Non-credible Classes (NC) for eight real-time events using unigram and bigram combination. The classes are predicted based on the rules discussed in data annotation. The authors conclude that the prediction of the classes mainly depends on the n-gram terms. The following figure 5 presents the proportion of the classified true positives and false positives. True positive is a result of the model, which correctly predicts the credible class, whereas false positive incorrectly predicts the credible class. These values are being used in evaluation metrics like precision, recall, and F1-score.

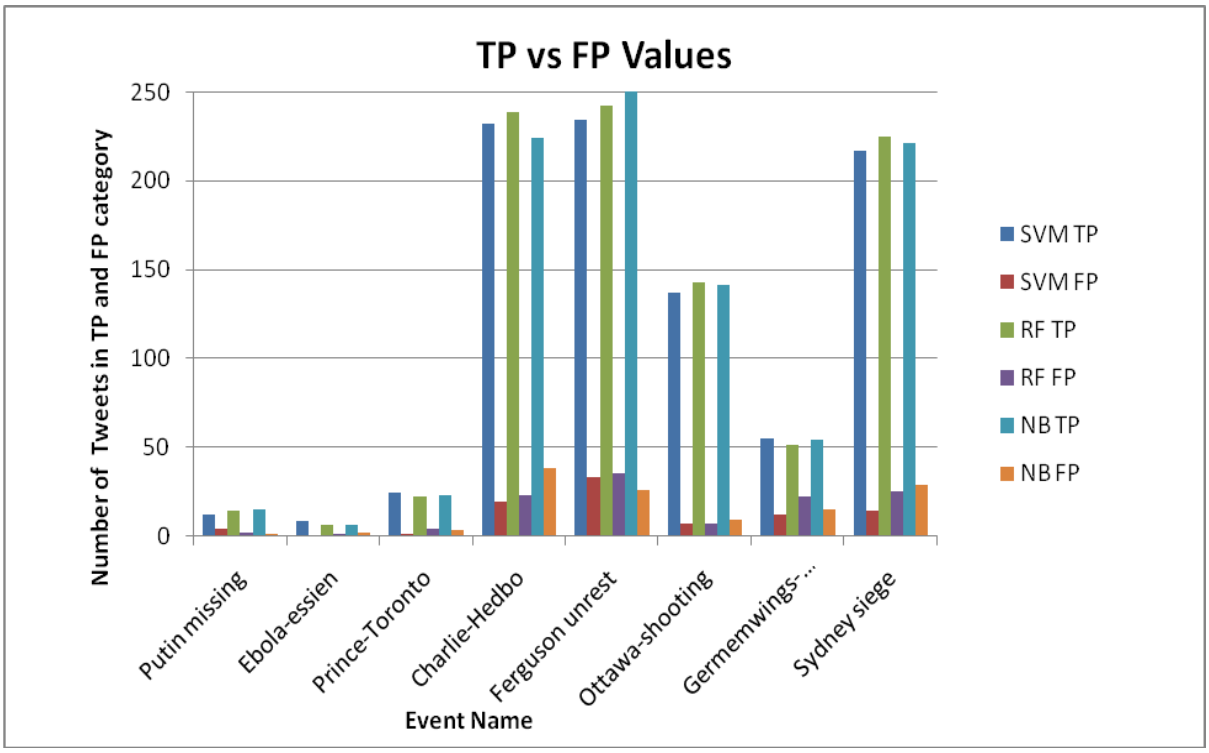

Fig. 5.The test results for eight real-time Tweets with the evaluation parameter of TP and FP

The classification results for dataset with respect to the accuracy are shown in figure 6. The RF algorithm performs better due to the fact that a lot of decision trees are repeatedly generated and hence more probabilities are taken to compute accuracy [23]. The performance of the SVM algorithm is better for some of the events namely Ebola-Essien, Charlie-Hedbo, and Ottawa-shooting because it has a regularisation parameter and uses kernel trick to classify the tweet text [21]. The performance of the NB algorithm is poor in most of the events like Charlie-Hedbo, Ottawa-shooting, and Sydney-siege. Since it makes a very strong hypothesis on the dataset related to the selected features. This leads to potentially low accuracy for classification of tweet texts [22].

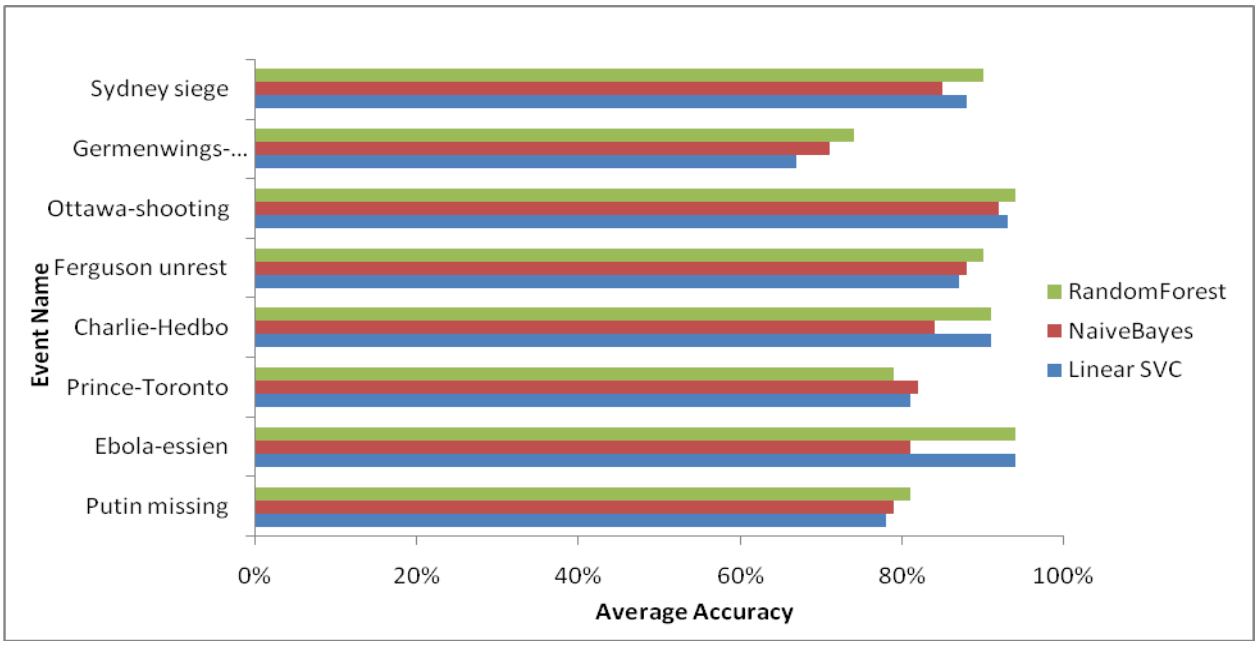

Fig. 6.The accuracy test results for eight real-time events using RF, SVM and NB Algorithms.

VI. EVALUATION METRICS

The credible classification of machine learning algorithms is evaluated based on the metrics of accuracy, precision, recall, and F1-score. Accuracy is defined as a statistical measure of how well a classifier correctly identifies the credible classes of the text. The accuracy results show the difference between a correct result and a true value for credible tweets. The greater value of accuracy implies that the algorithm correctly classifies credible tweets.

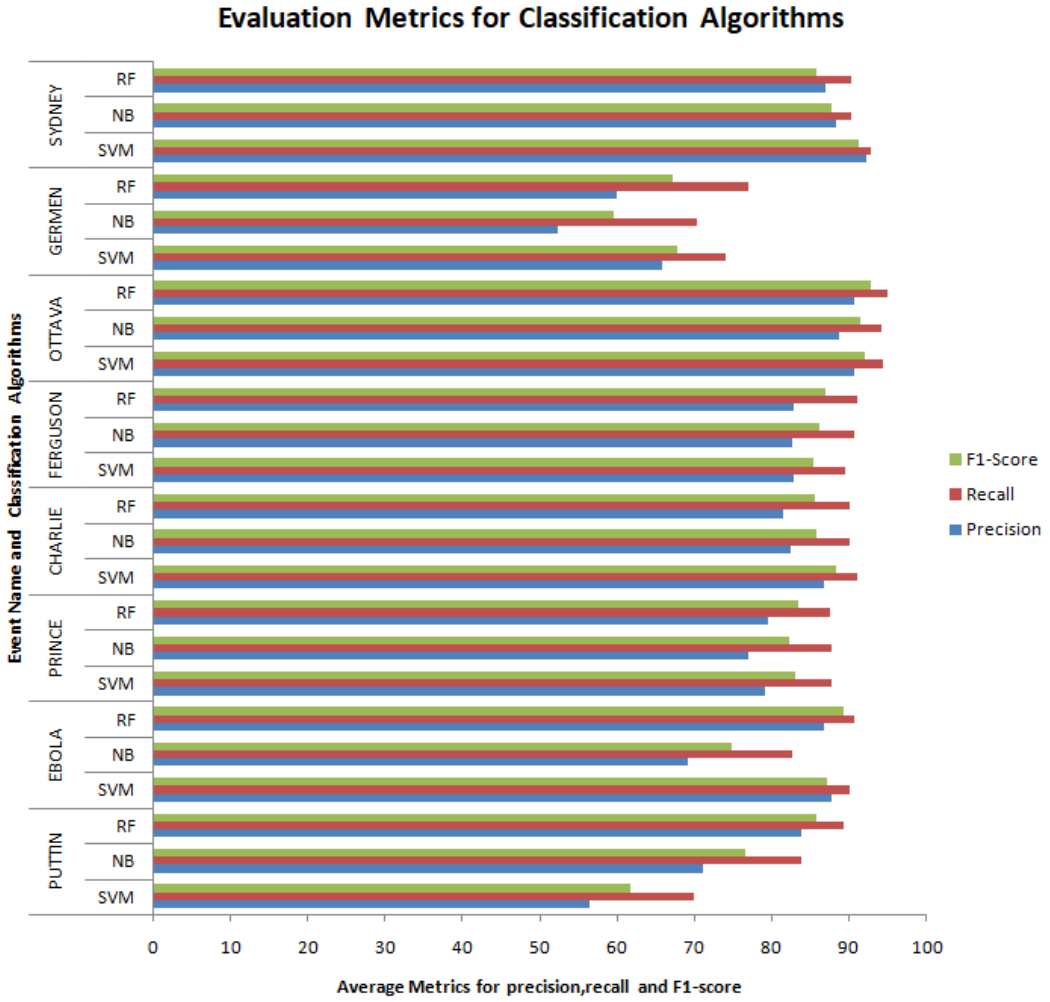

Fig. 7. The test results for eight real-time events with the evaluation parameters precision, recall and F1-score using RF, SVM, and NB Algorithms.

A. Precision, recall, and F1-score

Precision denotes the fraction of correctly classified credible tweets among all sample tweets that were identified to be credible. Recall can be denoted as the number of correctly classified credible tweets among all the actual credible tweets. F1-score is the harmonic mean of the precision and recall values. It is useful to combine both recall and precision values to predict the credibility of the tweet [20]. These metrics are evaluated by the presence of most influential terms and these terms computed by TF-IDF. Top correlated unigram, bigram and trigram terms are generated for evaluation. The highly credible related vectors are computed using n-gram range of (1,1),(2,2),(3,3),(1,2) and (1,3) using SVM, NB, RF algorithms for PHEME dataset. This precision, recall, F1-score values are high for RF algorithm in the events of Putin-missing, Ebola Essien, Ferguson unrest, Ottawa-shooting, Germanwings-plane crash and Sydney-siege events. SVM algorithm has achieved nearly similar score values for the event Charlie-Hedbo. NB and SVM algorithm has achieved a high score for Prince Toronto event. The authors have observed that all the three algorithms perform well

for unigram terms with a minimum volume of the dataset. But as the volume of dataset increases, the performance of NB and SVM algorithms comparatively deteriorates. Figure 7 shows the average of precision, recall and F1-score values for eight real-time events using three machine learning algorithms.

VII. CONCLUSION

While inaccurate and questionable information has always been a reality, the emergence of the Internet and social media has increased this concern due to the ease with which such information can be spread in unverified fashion to large communities of users. The major contribution of this research work is to assess the credibility of social text and facilitate to prevent the dissemination of inaccurate information using machine learning algorithms. This credibility assessment system analyzes various features of the tweet conversation threads from eight real-time events and classifies the tweet text as credible or non-credible based on the quality of trusted news about the events. The RF, SVM, NB algorithms are adapted to determine the credibility of the tweet using the n-gram model. The experiment results advocate RF and SVM algorithm upon the NB algorithm. RF normalizes the overfitting during credibility analysis and results more accurate. The RF classifier is found to be the best algorithm for credibility assessment for social media text.

REFERENCES

[1] Gupta P., Pathak V., Goyal N., Singh J., Varshney V., Kumar S. (2018) Content Credibility Check on Twitter. In: Deka G., Kaiwartya O., Vashisth P., Rathee P. (eds) Applications of Computing and Communication Technologies. ICACCT 2018. Communications in Computer and Information Science, vol 899. Springer, Singapore

[2] Gupta, A., Kumaraguru, P., Castillo, C., Meier, P.: TweetCred: real-time credibility assessment of content on Twitter. In: Aiello, L.M., McFarland, D. (eds.) SocInfo 2014. LNCS, vol. 8851, pp. 228–243. Springer, Cham (2014).

[3] Gupta, P., Kamra, A., Thakral, R., Aggarwal, M., Bhatti, S., Jain, V.: A proposed framework to analyze abusive tweets on social networks. Int. J. Mod. Educ. Comput. Sci. **1**, 46–56 (2018).

[4] Castillo, C., Mendoza, M., Poblete, B.: Information credibility on Twitter. In: Proceedings of the 20th International Conference on World Wide Web, pp. 675–684. ACM, March 2011

[5] Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211-36.

[6] Yang, F., Liu, Y., Yu, X., & Yang, M. (2012, August). Automatic detection of rumor on Sina Weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics* (p. 13). ACM.

[7] Giasemidis, G., Singleton, c., Agrafiotis, I., Nurse, J. R., Pilgrim, A., Willis, C., &Greetham, D. V. (2016, Nov). Determining the veracity of rumors on Twitter. In *International Conference on Social Informatics* (pp. 185-205). Springer, Cham.

[8] Chen, H. M., & Franks, P. C. (2016). Exploring Government Uses of Social Media through Twitter Sentiment Analysis. *Journal of Digital Information Management*, *14*(5).

[9] Verma, A., Wadhwa, A., Singh, N., Beniwal, S., Kaushal, R., & Kumaraguru, P. (2018, August). Followee Management: Helping users follow the right users on Online Social Media. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1286-1290). IEEE.

[10] Moin, R., Zahoor-ur-Rehman, K. M., Alzahrani, M. E., &Saleem, M. Q. (2018). Framework for Rumors Detection in Social Media. *The framework*, *9*(5).

[11] Zaman, T. R., Herbrich, R., Van Gael, J., & Stern, D. (2010, December). Predicting information spreading in twitter. In *Workshop on computational social science and the wisdom of crowds, nips* (Vol. 104, No. 45, pp. 17599-601). Citeseer.

[12] Takahashi, T., &Igata, N. (2012, November). Rumor detection on twitter. In *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on* (pp. 452-457). IEEE.

[13] Qazvinian, V., Rosengren, E., Radev, D. R., & Mei, Q. (2011, July). Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1589-1599). Association for Computational Linguistics.

[14] Wiegand, S., & Middleton, S. E. (2016, April). Veracity and velocity of social media content during breaking news: Analysis of November 2015 Paris shootings. In *Proceedings of the 25th international conference companion on world wide web* (pp. 751-756). International World Wide Web Conferences Steering Committee.

[15] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

[16] https://www.omnicoreagency.com/twitter-statistics/

[17] Narayanan, Vivek, Ishan Arora, and Arjun Bhatia. "Fast and accurate sentiment classification using an enhanced Naive Bayes model." In *International Conference on Intelligent Data Engineering and Automated Learning*, pp. 194-201. Springer, Berlin, Heidelberg, 2013.

[18] https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners/

[19] Jakkula V. Tutorial on support vector machine (SVM). School of EECS, Washington State University. 2006;37.

[20] Du S, Zhang F, Zhang X. Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. ISPRS Journal of photogrammetry and remote sensing. 2015 Jul 1;105:107-19.

[21] Huina Mao, Xin Shuai, Apu Kapadia, "Loose tweets: An analysis of privacy leaks on Twitter," in Proceedings of the 10th Annual ACM workshop on Privacy in the electronic society, WPES '11,Chicago, Illinois, USA, Oct 17-17, 2011, pp. 1-12.

[22] Cawley, Gavin C., and Nicola LC Talbot. "On over-fitting in model selection and subsequent selection bias in performance evaluation." Journal of Machine Learning Research 11, no. Jul (2010): 2079-2107.

[23] Fawagreh, Khaled, Mohamed Medhat Gaber, and Eyad Elyan. "Random forests: from early developments to recent advancements." Systems Science & Control Engineering: An Open Access Journal 2, no. 1 (2014): 602-609.