Int. j. inf. tecnol. (February 2021) 13(1):115–122
https://doi.org/10.1007/s41870-020-00550-5

ORIGINAL RESEARCH

Identifying propaganda from online social networks during COVID-19 using machine learning techniques

Akib Mohi Ud Din Khanday[1] · **Qamar Rayees Khan**[1] · **Syed Tanzeel Rabani**[1]

Received: 20 July 2020 / Accepted: 12 October 2020 / Published online: 29 October 2020

Abstract COVID-19, affected the entire world because of its non-availability of vaccine. Due to social distancing online social networks are massively used in pandemic times. Information is being shared enormously without knowing the authenticity of the source. Propaganda is one of the type of information that is shared deliberately for gaining political and religious influence. It is the systematic and deliberate way of shaping opinion and influencing thoughts of a person for achieving the desired intention of a propagandist. Various propagandistic messages are being shared during COVID-19 about the deadly virus. We extracted data from twitter using its application program interface (API), Annotation is being performed manually. Hybrid feature engineering is performed for choosing the most relevant features.The binary classification of tweets is being performed with the help of machine learning algorithms. Decision tree gives better results among all other algorithms. For better results feature engineering may be improved and deep learning can be used for classification task.

Keywords COVID-19 · Online social networks · Propaganda · Machine learning · Decision tree

1 Introduction

Social networks have bridged the gap of communication by providing a vast number of features for transferring the data from one client to other. With the advancement of online social networks, information sharing has become easy. People use online social networks for various purposes like for brand advertisements, marketing, education, business and for other purposes [1]. However, with these benefits it has some limitation/side effects, various Filthy users use this platform for various illegal activities that are very dangerous for the society. Various hate mongers have used this platform for spreading false content, rumors and fake information. The information that is being shared can be misinformation, disinformation and propaganda. Political and religious activists mostly use propaganda for gaining influence, propaganda can either be true or false [2]. The propaganda is spread in various forms that may be based on text, image and video. Since the twitter has a much influence on peoples behavior and is mostly used by politicians, religious activists, celebrities and influential actors [3], the spike in the graph of propaganda increases exponentially. The study done by previous researchers indicated that propaganda text is mostly related to sectarian and political discussions. Twitter allows its users to write only 280 characters at a time in a single tweet, here is the challenge of how to detect propaganda posts. Various events that are trending in and around the world are gaining much attention for propagandist users to spread hate, fear, hoaxes etc. In late 2019, a virus occurred in Wuhan China Known as COVID-19 [4]. This virus affected almost 10 million people in the world. Due to the trade with other countries around the globe, the virus has spread in every corner of the world by effecting mostly the European countries like Italy, UK, Spain and USA. This virus has

✉ Akib Mohi Ud Din Khanday
akibkhanday@bgsbu.ac.in

[1] Department of Computer Sciences, Baba Ghulam Shah Badshah University, Rajouri 185234, Jammu and Kashmir, India

also spread to Iran, India, Pakistan etc. till now there is less mortality rate in the Asian sub continent. A lot of research is being done for developing a drug for this pandemic virus. Various misinformation's are being spread by fear mongers using social networks. Misinformation about curing this virus is spread enormously some of the misinformation that were claimed to cure this deadly virus are, drinking alcohol, drinking cow urine etc. which has not medically proven for curing this disease. The politicians have also considered COVID-19 a concern. Various politicians around the world appealed to the common people to take precautions revealed by the world health organization. Various propagandistic messages are being spread using online social networks. Various hashtags are being used on twitter for spreading the messages regarding COVID-19. In this paper we extracted data using twitter application program interface (API) by giving various hashtags. Our work consists of five sections, the background is being described in Sect. 2. Section 3 gives the detailed methodology for the proposed system, results are being shown and discussed in Sects. 4 and 5 concludes our work.

The significant contribution of this paper is as follows:

- Novel Data set of 5 K tweets is being generated.
- Enhanced Feature Engineering has been done for achieving better accuracy.

2 Related work

The ever-growing attractiveness and beauty of using social networks directly or indirectly effects our daily life. It is not surprising that social media has become a weapon for manipulating sentiments by spreading disinformation as per the trend. The adversal use of these platforms are mostly used for spreading unreliable or ambiguous information which is a communal, financial, and political threat [5]. Gupta et al. [6] Analysed fake content on twitter during Boston attack the results showed that fear mongers effectively use social media for triggering mass hysteria and panic. Arts et al. [7] discussed about three types of attacks that took place in cyber network operations—physical, syntactic and semantic attacks. Physical attacks are attacks that affect the hardware of the system. Syntactic attacks occur due to the technologies, and there is no human hand in this attack. Semantic attacks are the most dangerous attacks which change the information content or the meaning of information [7]. Semantic attacks diverse from the other two forms of cyber-attacks. Semantic attacks attack the human–computer interface, and its effect is not visible as physical or the syntactic attacks. Semantic attacks are divided into many categories viz overt attack (include phishing, spam, etc.) and covert attack. Cybenko et al. [8] focused on covert attacks, i.e. misinformation, disinformation and propaganda. Kumar et al., Sarwar et al. [9, 10] analysed textual data for predicting various diseases. They showed that text classification showed better results in detecting the disease as well as any type of fraud from the text. Babcock et al. [1] suggested to use the social calculating characteristics of the consumers on online social media for determining the credibility of the information. The information on social networks can be shared deliberately or un-deliberately and are categorized in misinformation and disinformation. Mis-information is that information where the user does not know the truthfulness of information that is being spread. In contrast, Kumar et al. [11] described dis-information as the information in which the user deliberately gives false/accurate information for sharing [11]. Dis-Information usually occurs in politics, health, finance, technology etc. Howard et al. [12] studied Orchestrated Astroturf which is used for manipulating political conversations, even during election times. Esposito [13] proposed a semantic graph-based approach for radicalization detection in social media. They showed that pro-ISIS users tend to discuss about religion, historical events and ethnicity while anti-ISIS users focus more on politics, geographical locations and interventions against ISIS. Varol et al. [14] detected early promoted campaigns on social media. The results showed that compromised accounts are being used for spreading disinformation, and these accounts may also be used for spreading propaganda. According to O'Donnell et al. [15] propaganda comes under the type of disinformation which is defined as the systematic and deliberate process to shape opinions, influence thoughts, and direct behaviour of a person for achieving the desired intention of a propagandist. Paul et al. [16] showed that propaganda is mainly used for gaining the people's faith in some person or some community or party and plays a significant role in politics. Lightfoot [17] studied the effect of social bots on politics (political propaganda through social bots). The study found that social bots play a vital role in spreading fake news and accounts that continuously spread misinformation are significantly more likely to be Bots [18] showed that In USA presidential election 2016, political propaganda has a significant role in the winning of Donald Trump. Badawy et al. [19] analysed jihadist propaganda they showed that radical propaganda can be shared by posting four types of messages, religious and sacred topics, violence, sectarian discussion, and dominant celebrities and events.

3 Methodology

The proposed system for identifying propaganda during COVID-19 consists (i) data collection (ii) data preprocessing (iii) feature engineering and (iv) classification. The graphical representation of the proposed system is depicted in Fig. 1.

3.1 Data collection

Data is being extracted using twitter API [20], with the help of python tweepy by mentioning trending hashtags during COVID-19. About 5.1 million tweets are extracted using hashtags COVIDINDIA, CORONAVIRUS, CORONAJIHAD, CHINESEVIRUS, CORONAMUSLIM, etc. But after analyzing we got 3 hashtags that are related for spreading misinformation and propaganda, these tweets were #CoronaJihad, #CoronaMuslim and #Chinesevirus.

3.1.1 Manual annotation

We performed manual annotation to these tweets based on the content and semantics with the help of 18 different techniques of propaganda. We hire two journalists and one computer expert to perform labelling of the data.

3.1.2 Corpus collection

In the annotation about 5 K tweets were labelled into binary class as propaganda and non-propaganda. Based on various propaganda identification techniques. Figure 2 depicts the labelled dataset with their length in characters.

3.2 Data preprocessing

The textual data in the corpus consists of many missing values, URL's, hyperlinks, digits, stop words. For refining the data, various preprocessing tasks were performed, some of the tasks are as follows:

3.2.1 Tokenization

Tokenization splits the tweets into tokens. A sentence is being fragmented into the number of tokens, each word is considered as a separate token.

3.2.2 Stop words

Stop words like a, an, the etc. are being removed using English stop word dictionary.

3.2.3 Lemmatization

In this step lemma of the word is determined based on the intended meaning of the particular word.

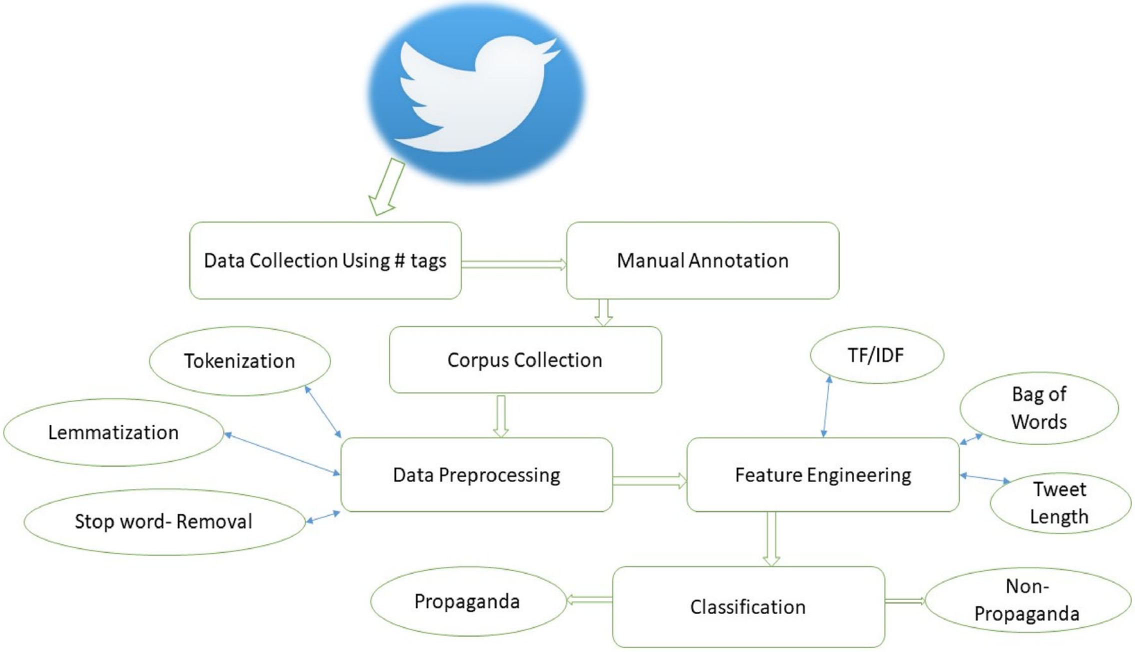


Fig. 1 Proposed system for identifying propaganda on online social networks

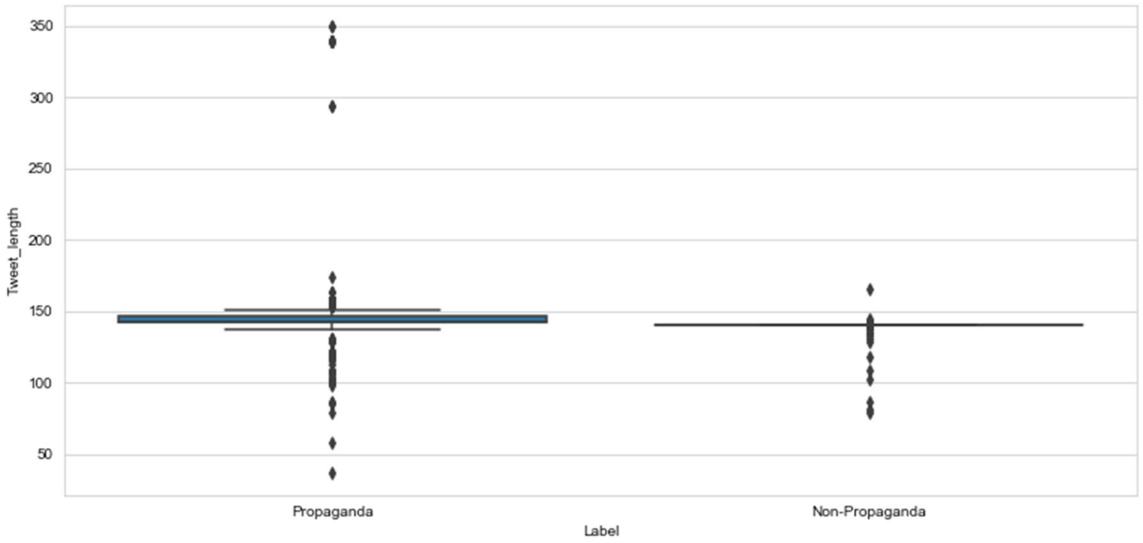


Fig. 2 Annotated corpus with their tweet length in characters

3.3 Feature engineering

For performing classification various features are needed for performing this task. We consider hybrid feature engineering by combining three types of features extracted using three different techniques TF/IDF, bag of words and tweet length.

3.3.1 TF/IDF

Term frequency/inverse document frequency reflects the importance of a word in a tweet or in a whole corpus by giving its numerical statistics. It is calculated using the following equation.

$$\mathrm{TFIDF}(t,w,D) = \mathrm{TF}(t,w) * \mathrm{IDF}(t,D)$$

$$\mathrm{TF}(t,w) = f_{t,w} \Bigg/ \sum_{t' \in w} f_{t',w}$$

$$\mathrm{IDF}(t,D) = \log \frac{|D|}{1 + |\{w \in D : t \in w\}|}$$

where t is the term as a feature, w denotes each tweet in the corpus and D is the total number of tweets in the dataset (document space).

3.3.2 Bag of words

Consists of words and lemma uni, bi and trigrams. We included bigrams, trigram words such that more information can be extracted from the text.

3.3.3 Tweet length

Since twitter allows only 280 characters in a single tweet, we considered the length of the tweet also. While performing computations it was revealed that the propagandistic tweets are having greater length than non-propagandistic tweets. In our work, we used this feature with TF/IDF & bag of words for achieving better testing results.

After performing feature engineering the most correlated bigrams were ‘dangerous muslim’, ‘rise coronajihad’, ‘coronavirus report’, ‘rt billyperrigo’, ‘coronajihad nar’, ‘india come’, ‘come coronavirus’, ‘billyperrigo already’, ‘already dangerous’, ‘muslim india’, ‘rt rose_k01’, ‘hashtag coronajihad’.

3.4 Classification

The main motive of work is to build a classifier which will classify a tweet into propaganda and non-propaganda class. Supervised machine learning algorithms are used as our corpus is labelled. Various traditional machine learning classifiers are trained and tested for this task.

3.4.1 Logistic Regression

Based on class relationship with the label it predicts the numerical class value. Logistic regression is fine-tuned as:

C = 1.0, classweight = None, dual = False, fit-intercept = True, intercept-scaling = 1, max-iter = 100, multi-class = ‘warn’, n_jobs = None, penalty = ‘l2’, random_state = 8, solver = ‘warn’, tol = 0.0001, verbose = 0, warm_start = False.

3.4.2 *Multinomial Naïve Bayes*

Multinomial Naïve Bayes (MNB) uses a classical Bayes algorithm for text classification. Multinomial Naïve Bayes is fine-tuned as:

alpha = 1.0, class-prior = None, fit-prior = True.

3.4.3 *Support vector machine*

Supervised machine learning approach used for classification tasks as well as for regression problems. It takes 'n' number of features for the particular text with the given label. Support vector machine (SVM) is fine-tuned as:

C = 0.1, cache-size = 200, class-weight = None, coef0 = 0.0, decision-function-shape = 'ovr', degree = 3, gamma = 'auto_deprecated', kernel = 'linear', max-iter = − 1, probability = True, random-state = 8, shrinking = True, tol = 0.001, verbose = False.

3.4.4 *Decision tree*

In this approach input space is broken down into regions. Every region is classified independently. Decision tree classifier is fine-tuned as:

Class-weight = None, Criterion = 'gini', max-depth = None, max-features = None, max-leaf-nodes = None, min-impurity-decrease = 0.0, min-impurity-split = None, min-samples-leaf = 1, min-samples-split = 2, min-weight-fraction-leaf = 0.0, presort = False, random-state = 0, splitter = 'best'.

4 Results and discussion

In our experiment, we have used logistic regression, multinomial Naïve Bayesian, support vector machine and decision tree algorithms for performing the task of classifying propagandist text from non-propagandist text. The proposed hybrid feature engineering technique is used to extract the useful features that are supplied to the fine tuned machine learning models. About 100 features are chosen

Table 1 Classification report and comparison of machine learning algorithms

Algorithm	Precision	Recall	F1-Score	Accuracy (%)
Logistic regression	0.98	0.98	0.98	98.3
Multinomial Naïve Bayes	0.97	0.97	0.97	97.23
Support vector machine	0.98	0.98	0.98	98.2
Decision tree	0.99	0.99	0.99	98.53

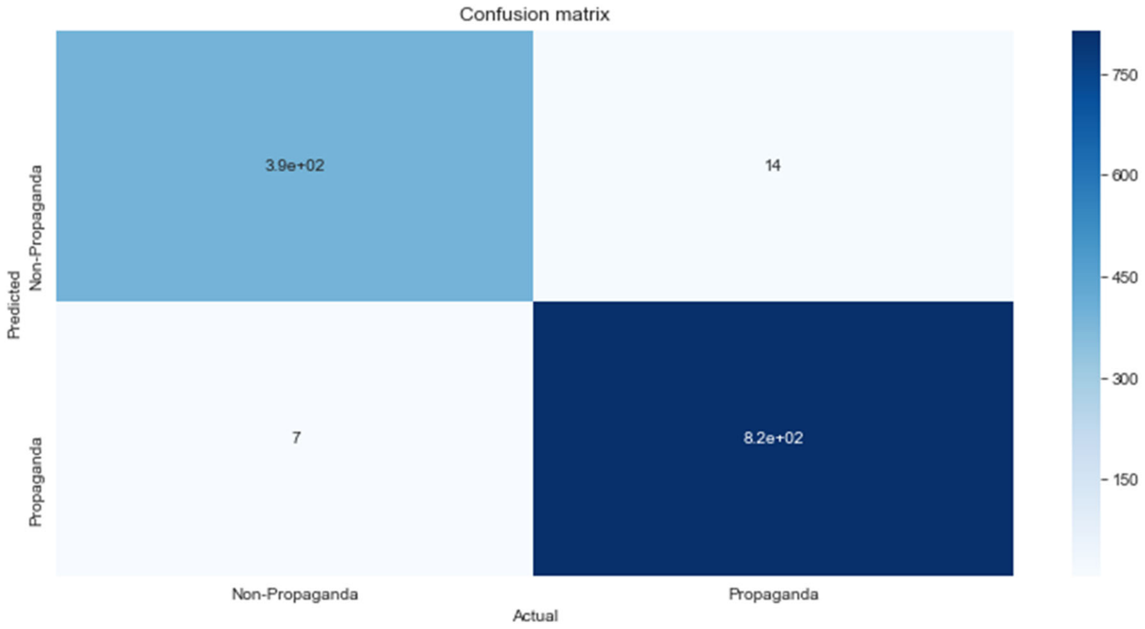


Fig. 3 Confusion matrix of logistic regression

for performing the binary classification but due to the computational complexity information gain is used for selecting the most influential features. The dataset is being split into 70 by 30 ratio, 70% is used for training the machine learning models and 30% are used for testing the models. Machine learning algorithms are finetuned in such a way that they give better results. The algorithms are tested by giving them different parameters. In support vector machine we used three kernel RBF, poly and linear. The linear kernal showed the better results as compared to other two kernals. Similarly other machine learning algorithms showed better results by finetunning their particular parameters. Multinomial Naïve Bayes showed better results when alpha was set to 1.0, Logistic regression showed good results when C was assigned value of 1.0 and maximum iteration of 100 were taken. In decision tree gini coefficient was used for information gain and it showed promising results. The comparision of all machine learning

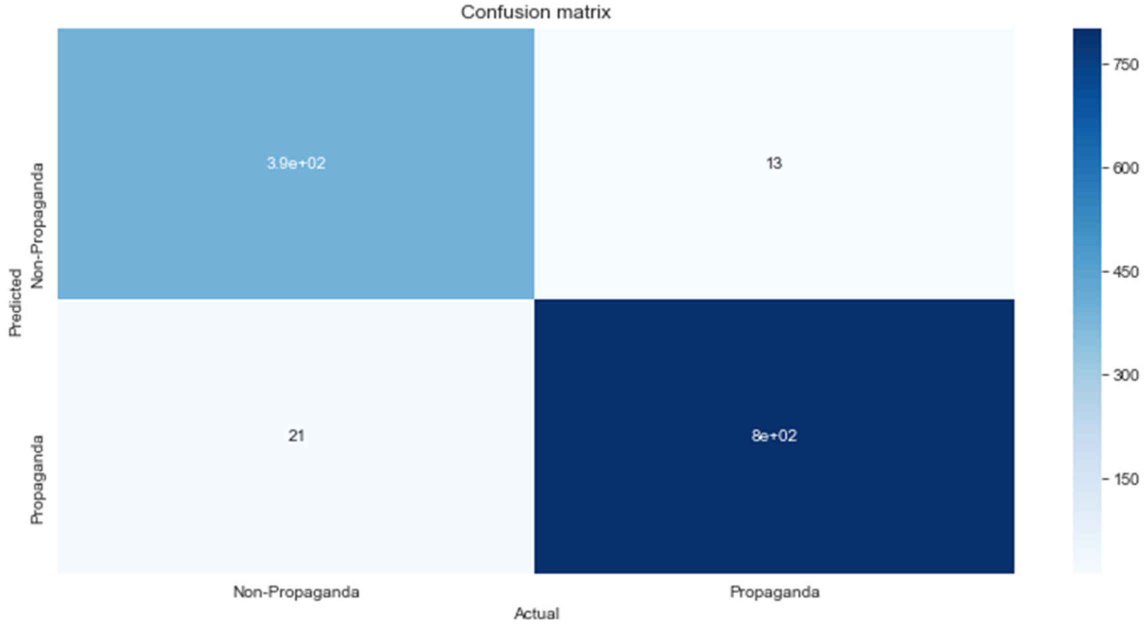


Fig. 4 Confusion matrix of multinomial Naïve Bayes

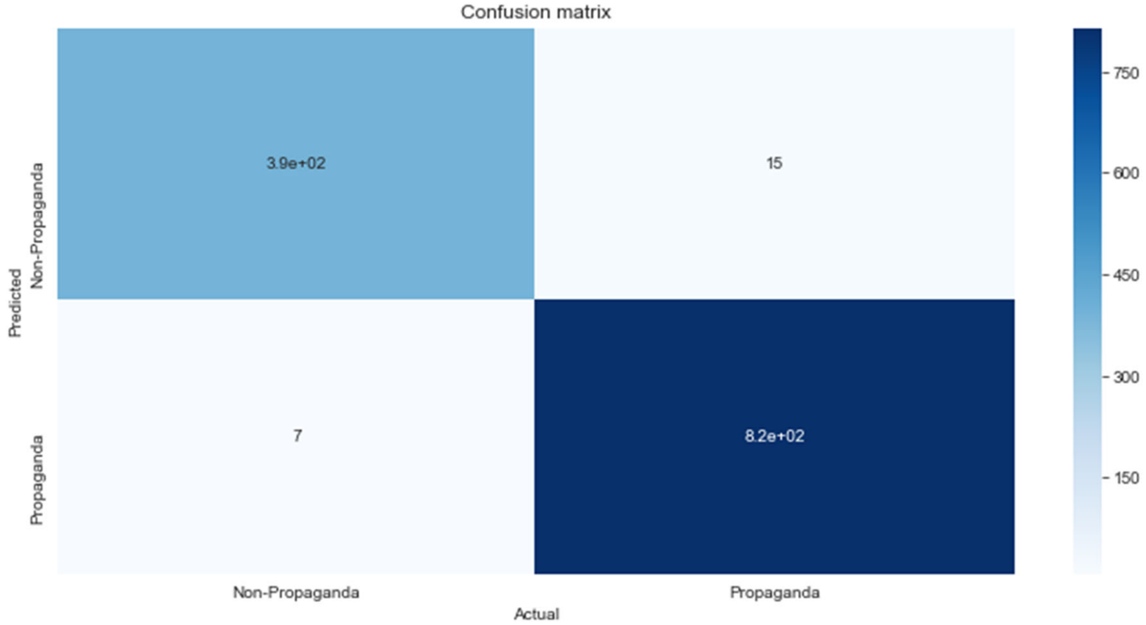


Fig. 5 Confusion matrix of support vector machine

algorithms are performed and the results showed that decision tree outperforms all other traditional machine learning algorithms by achieving 98.5% accuracy with 0.99 precision, 0.99 recall and 0.99 F1-Score. Support vector machine and logistic regression showed also good results by achieving 0.98, 0.98, 0.98 precision recall and F1-score, respectively. Table 1 gives a detailed classification and comparison of all machine learning algorithms. The confusion matrices of all the machine learning algorithms are shown in Figs. 3, 4, 5, 6. For validating our work we performed tenfold cross-validation. It is also seen that there is neither Under-fitting nor overfitting during training and testing of the propsed model. After performing analysis it was revealed that propagandistic tweets have greater length than non-propagandistic tweets. The majority of data used in our research was related to COVID-19. More data that range numerous fields should be gathered for better analysis of propaganda. We need more human exertion to play out the labelling of the tweets into different classes. As the data increases, manual annotation gets intense, therefore, requiring an automatic nnotation program that will learn from the semantics of provided text. More feature engineering is required for accomplishing better text classification results.The comparative analysis of machine learning algorithms used in our work is shown in Fig. 7.

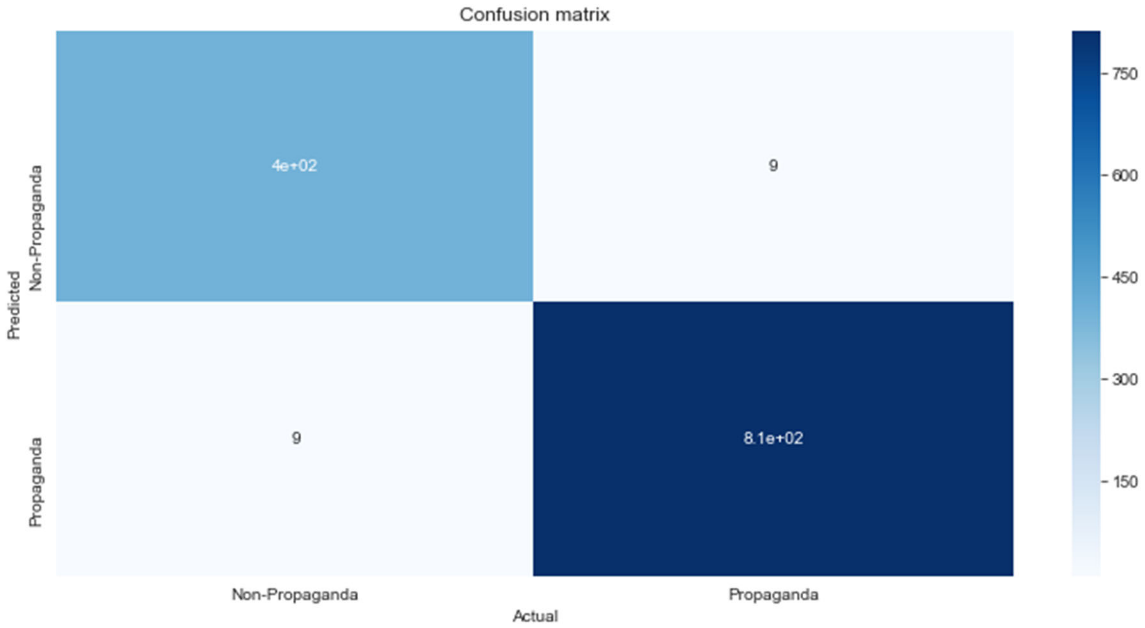
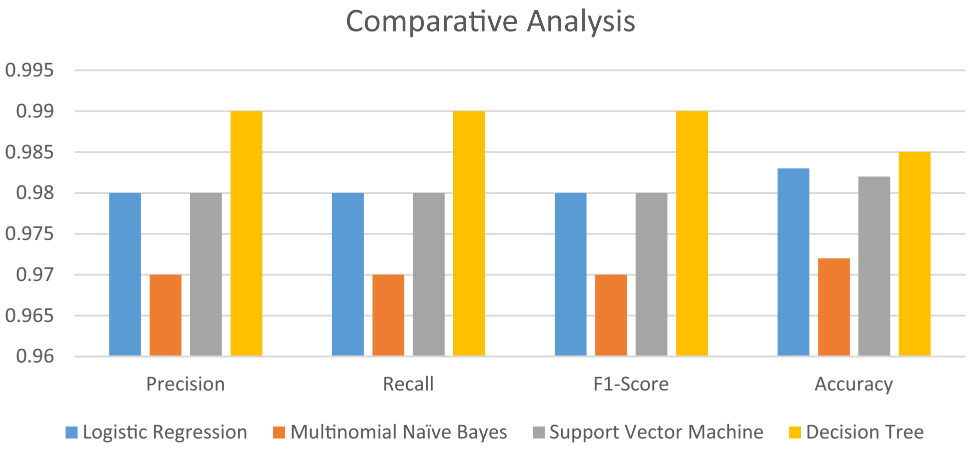


Fig. 6 Confusion matrix of decision tree

Fig. 7 Comparative analysis of used machine learning algorithms



5 Conclusion

Machine learning has grew a lot of attentiveness, due to its better and robust results in every field. During the COVID-19 various misinformation and propaganda is being shared. In this paper, data is extracted from the online social network platform “twitter” using its API. The extracted data is being manually labelled into two classes’ propaganda and non-propaganda. Hybrid feature engineering is being performed by combining three different textual features (TF/IDF, bag of words and tweet length). The results revealed that propagandistic text gave greater length than non-propagandistic text. Machine learning algorithms are used for classifying tweets into propaganda and non-propaganda class. Decision tree classifier showed better results among all other machine learning algorithms by having 98.5% accuracy, 0.99 precision, 0.99 recall and 0.99 F1-Score. In future, more features may be used for getting better results also Deep learning can be used for performing this task.

References

1. Babcock M, Beskow DM, Carley KM (2019) Different faces of false: the spread and curtailment of false information in the black Panther Twitter discussion. J Data Inf Qual. https://doi.org/10.1145/3339468
2. Zhou Y (2017) Pro-ISIS fanboys network analysis and attack detection through Twitter data. In: 2017 IEEE 2nd Int. Conf. Big Data Anal. ICBDA 2017, pp 386–390
3. Kietzmann JH, Hermkens K, McCarthy IP, Silvestre BS (2011) Social media? Get serious! Understanding the functional building blocks of social media. Bus Horiz 54(3):241–251
4. Khanday AMUD, Rabani ST, Khan QR, Rouf N, Mohi ud Din M, (2020) Machine learning based approaches for detecting COVID-19 using clinical text data. Int J Inf Technol 12:731–739. https://doi.org/10.1007/s41870-020-00495-9
5. World Economic Forum (2017) The Global Risks Report 2017, 12th edn. Glob. Compet. Risks Team, p 103
6. Gupta A, Lamba H, Kumaraguru P (2013) $1.00 per RT #Boston Marathon #Pray For Boston: Analyzing fake content on Twitter. APWG eCrime Researchers Summit, San Francisco, CA 2013:1–12. https://doi.org/10.1109/eCRS.2013.6805772
7. Libicki CM (2009) Cyberdeterrence and cyberwar. ISBN 978-0-8330-4734-2, Rand, p. 240
8. Cybenko G, Giani A, Thompson P (2002) Cognitive hacking: a battle for the mind. Computer 35(8):50–56
9. Kumar A, Dabas V, Hooda P (2018) Text classification algorithms for mining unstructured data: a SWOT analysis. Int J Inf Tecnol 12:1159–1169. https://doi.org/10.1007/s41870-017-0072-1
10. Sarwar A, Ali M, Manhas J, Sharma V (2018) Diagnosis of diabetes type-II using hybrid machine learning based ensemble model. Int J Inf Tecnol 12:419–428. https://doi.org/10.1007/s41870-018-0270-5
11. Kumar KPK, Srivastava A, Geethakumari G (2016) A psychometric analysis of information propagation in online social networks using latent trait theory. Computing 98(6):583–607
12. Howard PN, Kollanyi B (2016) Bots,# StrongerIn, and# Brexit: computational propaganda during the UK-EU referendum. Available at SSRN 2798311.
13. Esposito A (2006) The Semantic Web. In: Tarricone L, Esposito A (eds) Advances in Information Technologies for Electromagnetics. Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-4749-5_3
14. Varol O, Ferrara E, Menczer F, Flammini A (2017) Early detection of promoted campaigns on social media. EPJ Data Sci. 6:13. https://doi.org/10.1140/epjds/s13688-017-0111-y
15. Jowett GS, O’donnell V (2018) Propaganda & persuasion. Sage publications
16. Paul C, Matthews M (2016) The Russian
17. Lightfoot S, Jacobs S (2017) Political propaganda spread through social bots. Media, Culture, & Global Politics, pp. 1–22
18. Bessi A., Ferrara E (2016) Social bots distort the 2016 US Presidential election online discussion. First Monday, 21(11-7)
19. Badawy A, Ferrara E (2018) The rise of jihadist propaganda on social networks. Journal of Computational Social Science 1(2):453–470
20. Verma P, Khanday AMUD, Rabani ST, Mir MH, Jamwal S (2019) Twitter sentiment analysis on Indian government project using R. Int J Recent Technol Eng 8(3):8338–8341