DOI: 10.24996/ijs.2022.63.10.33

ISSN: 0067-2904

Ensemble Approach for Detecting COVID-19 Propaganda on Online Social Networks

Akib Mohi Ud Din Khanday *, Qamar Rayees Khan, Syed Tanzeel Rabani

Department of Computer Sciences, BGSB University, Rajouri, India

Received: 23/10/2020 Accepted: 27/11/2021 Published: 30/10/2022

Abstract

COVID-19 affected the entire world due to the unavailability of the vaccine. The social distancing was a contributing factor that gave rise to the usage of Online Social Networks. It has been seen that people share the information that comes to them without verifying its source . One of the common forms of information that is disseminated that have a radical purpose is propaganda. Propaganda is organized and conscious method of molding conclusions and impacting an individual's contemplations to accomplish the ideal aim of proselytizer. For this paper, different propagandistic tweets were shared in the COVID-19 Era. Data regarding COVID-19 propaganda was extracted from Twitter. Labelling of data was performed manually using different propaganda identification techniques and Hybrid feature engineering was used to select the essential features. Ensemble machine learning classifiers were used for performing the binary classification. Adaboost shows an accuracy of 98.7%, which learns from a weak learning algorithm by updating the weights.

Keywords: COVID-19, Propaganda, Hybrid, Ensemble, Adaboost.

1. Introduction

The communication gap has been reduced with the evolution of Social Networks. Social Networks provide countless features for communicating with each other. With the increase in Online Social Networks (OSN's) usage, sharing of information becomes simple. OSN users use social media platforms for various purposes, including brand advertisements, marketing and education etc.[1].

With these countless features, it has multiple negative impacts on society. Some bad agents have used OSN's for criminal operations, which are hazardous for the general public. Fear mongers use social networking platforms to spread false content, rumors and phony content. The bad information can be categorized into Misinformation, Disinformation and Propaganda. Propagandistic information has the distinction that it can either be authentic or fake [2].

Propaganda is the weapon used by political and religious activists to gain fame in the general public. This area has not gained a lot of scientists' enthusiasm because of the semantical nature of its substance. The propaganda can be spread in different structures that may be textual, image-based, video-based, etc. Data is extracted from Twitter, a Social Networking Site generally utilized by government officials, religious activists, celebrities, and influential actors [3]. As discussed by analysts, the propaganda text is the situation when most of the conversation is about governmental, religious issues, and presenting compelling on-

*Email: akibkhanday@bgsbu.ac.in

screen characters. Twitter permits its clients to compose just 256 characters in a single tweet. This is the major challenge in identifying propagandistic posts. Various occasions that are happening in and around the globe are increasing a lot of consideration for advocate clients to spread false, dread, lies and so on. As in late 2019, the infection has been happening in China, Known as Corona Virus, later officially named COVID-19[4]. This infection has affected around 114 million individuals . Because of the exchange with different nations worldwide, this virus spread in every part of globe by affecting western nations like Italy, England, Spain, and United States of America. As well as spreading to countries to Afghanistan, Nepal, Bangladesh, and so forth. Death rate in the Asian Subcontinent was less than Europe. Great deal of examination was accomplished by building up medication of pandemic infection.

A lot of misinformation's was spread through dread mongers utilizing online social media during the time of the pandemic. Deception about the infection was spread immensely and a large amount of false information about curing or preventing the disease was widely shared. An example of false information were, consumption of alcohol, cow urine, that were not scientifically demonstrated to relieve the ailment, presented as cures for COVID.
The World Health Organization additionally considered COVID-19 a worry. Different Politicians throughout the globe spoke to the everyday citizens to avoid potential risk uncovered by the world wellbeing association. Different propagandistic messages were spread utilizing Online social media. Various Hashtags were utilized on Twitter to spread the messages regarding COVID-19. Hashtags are the keywords that are mostly used in Tweets and these keywords are used for the extraction of data. This paper separated information utilizing Twitter Application Program Interface (API) through different hashtags. This paper comprises of 5 sections, the Literature is described in segment 2. Section 3 gives detailed overview of the proposed algorithm. Results are shown and examined in segment 4, and segment 5 summarizes the proposed work.
The noteworthy Influence of the proposed framework is:

- Novel data set was created manually by annotating 5k tweets.
- Hybrid feature selection approach was performed using TF/IDF, n-grams and Length of Tweet for better classification.
- Novel Algorithm is being proposed for this work.
- Ensemble Learning Approaches are implemented to classify between Propagandist and Non-Propagandist text.

2. Related Work

The adversarial utilization of web-based social networking spread questionable or vague information that poses a mutual, monetary, and political danger[5]. Dread mongers are successfully utilizing online sites for triggering widespread panic and fear [6]. As indicated by [7] there are three classes of attacks which occur in digital system – physical, syntactic, semantic attacks. Physical assaults effect the apparatus of framework. Syntactic assaults happen because of the advancements, and it does not involve any human intervention. While, semantic attacks are the riskiest type that alter context of data or data impact [8].

These days' semantic attacks are progressively regular in informal online organizations. Semantic assaults differs from the other two types of digital attacks. They focus on the human-PC interface, and their impact is not obvious as that of physical or syntactic attacks. Semantic attacks can be classified into two classes: Overt attack (incorporate Phishing, Spam, and so forth.) and Covert attack. This research will focus on Covert attack, which include deception, disinformation and propaganda[9].

Text characterization shave demonstrated the promising outcomes in identifying the illness just as a misrepresentation from the text[10]. Numerous creators have recommended utilizing the social ascertaining qualities of the buyers on online web-based life to decide the believability of the information. Social groups are immensely used to spread manufactured information. Dubious information can be shared purposely or unintentionally and can be arranged for falsehood and disinformation.

Misinformation is happens when the user does not have the vaguest idea about the righteousness of data which is spread. Whereas, Disinformation is spread when a user intentionally gives fake/false information for sharing[11]. Disinformation typically occurs in legislative topics, health, finance, latest technologies and so forth. With the usage of organized Astroturf, political discussion can be controlled and is mostly used at election times[12] [13]. The undermined accounts are used to spread disinformation. These accounts may likewise spread propaganda. Propaganda falls under the class of disinformation considering that it is an orderly and purposeful practice for shaping feelings, impact musings, and direct the behavior of a group of people to achieve the ultimate expectancy of a propagandist. Propaganda is chiefly utilized for improving people's confidence in some individual or some network or gathering. Thus, it adopts a critical role in legislative issues. Accordingly political propaganda gained much interest from scientists around the world. During presidential political race 2016, in the United Stated of America, political propaganda played a critical role in Donald Trump's triumph [14]. Radical propaganda may be shared through four sorts of messages, devout and consecrated themes, viciousness, partisan conversation, and prevailing big names and events[15].

Sentiment Analysis was performed using Machine Learning techniques and the Bayesian Rough Decision Tree (BRDT) algorithm has showed better accuracy in extracting Social Media Posts sentiments [16]. [17] Identified Propaganda using Traditional Machine Learning Algorithms using Twitter Data. [18] Proposed intrusion detection system based on data stream classification and the algorithm was applied on CICIDS2017 datasets that contain various types of attacks.

3. Methodology

The framework which recognizes propaganda in COVID-19 Era (I) Collection of Data (ii) Data Pre-Processing (iii) Feature Engineering and (iv) Classification. The graphical portrayal of proposed framework presented in Figure 1.

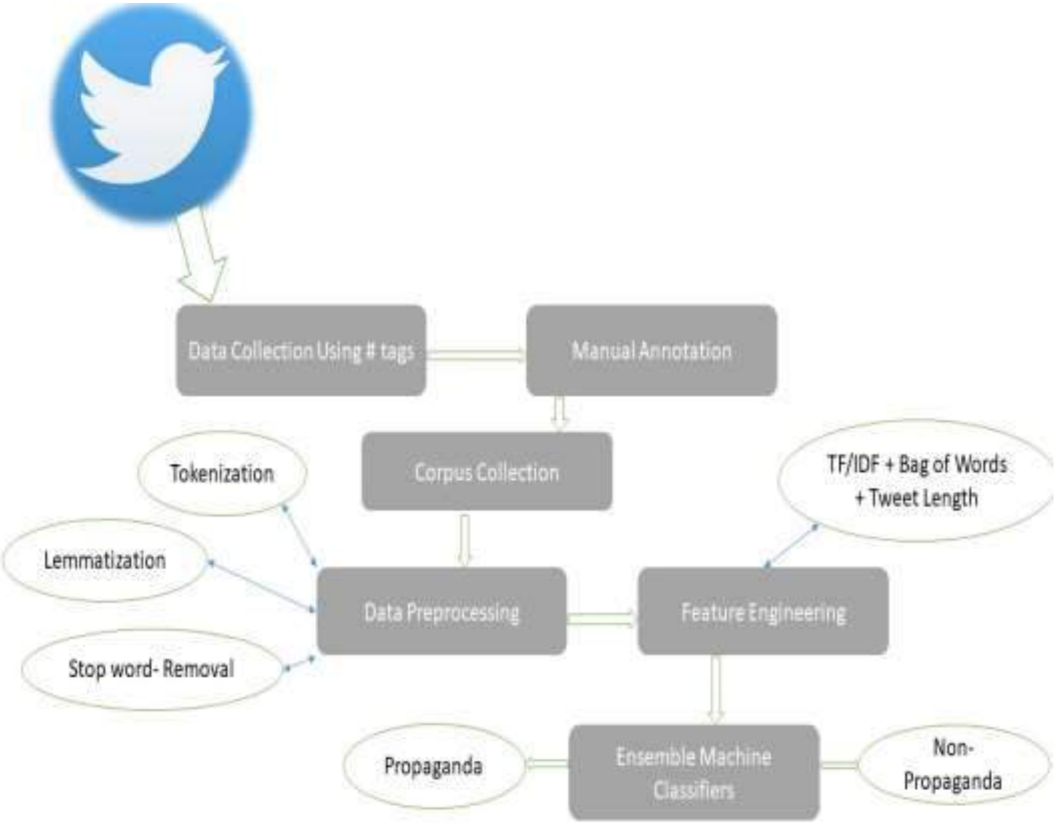

Figure 1: Proposed Framework for classifying tweets into binary class.

The Various steps of the Proposed Algorithm (Prop) are as follows:

```
Prop Algorithm for classification of tweets in Propaganda and Non-Propaganda
Input: Filtered Tweets from Twitter ($T_{input}$)
Output: Propagandistic Tweets ($T_P$) and Non-Propagandistic Tweets ($T_N$)
1: for i from 1 to k (total number of tweets)
Corpus[i] = $T_{input}$ [i] + Label //Manual Annotation
Tweet[i] = Tweetlength(Corpus[i])
2: for i from 1 to k
Processed[i] = Tokenize (Corpus[i])
Processed[i] = RemoveStopWord (Processed[i])
Processed[i] = Lamitize (Processed[i])
3: for i from 1 to k
Feng[i] = BagofWords( TfIdf(Processed[i]))
Feng[i] = Feng[i] + Tweet[i]
4: Classify(Feng[i])
```

3.1 Data Collection

3.1.1 *Twitter information extraction:* Data was extracted from social media platform Twitter. It was extracted through Application Program Interface (API) [19], with the assistance of python tweepy library by using various keywords about COVID-19. Almost 5.1 million tweets were extracted using Hashtags . Some of the Hashtags that were used are #COVIDINDIA, #CORONAVIRUS, #CORONAJIHAD, #CHINESEVIRUS, #CORONAMUSLIM, and so on. Yet, after examining that data, three hashtags were connected to spreading deception and propaganda. The ambiguous hashtags were CoronaJihad, CoronaMuslim and Chinesevirus.

3.1.2 *Manual Annotation:* Labelling of these tweets was done manually depending on substance and semantic of the tweet. About 18 of the unique methods of propaganda were used during labeling. Two journalists and computer master graduate were hired to do annotation of tweets.

3.1.3 *Corpus Collection:* Corpus of about 5K tweets , marked into binary label Propaganda and Non-Propaganda depending on different propaganda recognition strategies. Figure 2 delineates named dataset with their corresponding length in characters. After analyzing the labeled dataset, it can be seen that the propaganda class tweets is larger than non-propagandistic posts.

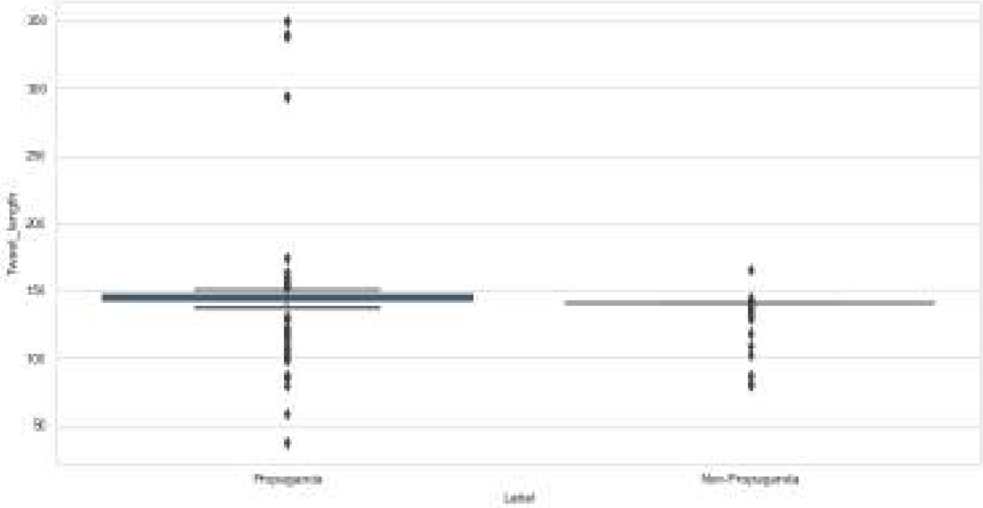

Figure 2: Annotated Dataset with corresponding Tweet length.

3.2 Data Preprocessing

Corpus's data comprised several omitted qualities, “URL's”, “Hyperlinks”, “digits”, “Stop words”. For cleaning the information, different pre-processing methods were conducted, a portion of the undertakings are as per the following:

3.2.1 *Tokenization*: Tokenization is the process of splitting tweets into tokens. A textual tweet is fragmented in various tokens and each word is representation of a different token.

3.2.2 *Stop words:* Stop words do not have impact on the content of tweets. Stop words were removed using English stop word dictionary.

3.2.3 *Lemmatization:* A lemma of a word is determined depending on specific word's proposed significance.

3.3 Hybrid Feature Engineering

For training and testing, a machine learning model various features are required. Hybrid Feature building was conducted by combining three features separated utilizing three distinct procedures TF/IDF, n-grams and Length.

3.3.1 *TF/IDF:* Term Frequency (TF)/Inverse Document Frequency (IDF) mirrors a word's significance of tweets/entire corpus through its arithmetical insights. It was determined utilizing the accompanying condition.

$$TFIDF(t,w,D) = TF(t,w) * IDF(t,D)$$

$$TF(t,w) = f_{t,w} / \sum_{t\prime \in w} f_{t\prime,w}$$

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

Here t = term , w = textual tweets in corpus and D = total number of tweets.

3.3.2 *Bag of Words*: Contains words and lemma that can be Uni, bi and trigrams. In this work Unigrams, bigrams and trigrams were used.

3.3.3 *Tweet Length*: on Twitter, only 256 characters can be used in a single Tweet . In this work length of the tweets was also considered as a feature. During calculations results uncovered that the Propagandistic tweets are having more noteworthy length as compared with no propagandistic tweets. This feature is combined with TF/IDF and Bag of Words to accomplish better testing outcomes in our work.

In the wake of performing highlight building the most connected bigrams were “risky muslim”, “ascent coronajihad”, “coronavirus report”, “rt billyperrigo”, “coronajihad nar”, “india come”, “come coronavirus”, “billyperrigo as of now”, “effectively perilous”, “muslim india”, “rt rose_k01” and “hashtag coronajihad”.

3.4 Classification

Binary classification of tweets was performed using Ensemble Machine Learning algorithms. This work comprises four types of ensemble learning algorithms. The algorithms include Random Forest, Bagging, Adaboost and Stochastic Gradient Boosting. These algorithms are fine tuned to improve their accuracy.

3.4.1 Random Forest

The best hyper parameters of Random Forest were: “bootstrap”=False, “maximum_depth”=30, “maximum_features”= “sqrt”, “minimum_samples_leaf”=1, “minimum_samples_split”=5, “n_estimators”=800. class_weight=None, criterion='gini', “maximum_leaf_nodes”=None, “minimum_impurity_decrease”=0.0, “minimum_impurity_split”=None, “minimum_weight_fraction_leaf”=0.0, “n_jobs”=None, “oob_score”=False, “random_state”=8, “verbose”=0, “warm_start”=False

3.4.2 Bagging

The hyper parameters of Bagging were: “base_estimator”=None, “bootstrap”=True, “bootstrap_features”=False, “maximum_features”=1.0, “maximum_samples”=1.0, “n_estimators”=10, “n_jobs”=None, “oob_score”=False, “random_state”=8, “verbose”=0, “warm_start”=False

3.4.3 Adaboost

The hyperparameters of Adaboost were:
“Algorithm”=”SAMME.R”, “base_estimator”=None, “learning_rate”=1.0, “n_estimators”=50, “random_state”=8.

3.4.4 Stochastic Gradient Boosting

The hyper parameters of Stochastic Gradient boosting were:
“criterion”=”friedman_mse”,”init”=None,”learning_rate”=0.1,”loss”=”deviance”,”maximum_depth”=3,”maximum_features”=None,”maximum_leaf_nodes”=None,”minimum_impurity_decrease”=0.0,”minimum_impurity_split”=None,”minimum_samples_leaf”=1,”minimum_samples_split”=2,”minimum_weight_fraction_leaf”=0.0,”n_estimators”=100,”n_iter_no_chang

e”=None, “presort”='auto', “random_state”=8, “subsample”=1.0, “tol”=0.0001, “validation_fraction”=0.1, “verbose”=0, “warm_start”=False.

4. Results and Discussions

The calculations were performed on a workstation having 8 GB RAM and 6 inbuilt processors. The Novel dataset was produced during this work. Different preprocessing steps were conducted to refine the dataset and make it appropriate for performing parallel characterization. Hybrid Feature selection was done, as the three distinct features (TF/IDF, Bag of Words and Tweet Length) were joined. Around 100 features were selected for playing out the parallel characterization yet because of the computational intricacy data gain was utilized for choosing the most compelling highlights. In this work 70:30 ratio was used, 70% was utilized for preparing the ensemble ML models and 30% were utilized for testing the models. The classification report was evaluated based on Precision, Recall and F1-Score.

$P=Tp/(Tp+Fp)$

$R=Tp(Tp+Fn)$

$F1\text{-}Score = (2*P*R)/(P+R)$

Where *P=Precision,*
R= Recall,
Tp= True Positive
Fp=False Positive
Fn=False Negative

The outcomes indicated that Adaboost outperforms all other Ensemble learning classifiers by accomplishing 98.7% Accuracy with 0.98 precision, 0.98 recall and 0.99 F1-Score. Table 1 gives a classification report and correlation of all Ensemble Machine Learning Classifiers.

Table I: Classification Report of Ensemble Machine Learning Algorithms.

Ensemble Technique	Precision	Recall	F1-Score	Accuracy
Random Forest	99%	98%	98%	98.61%
Bagging	98%	98%	98%	98.55%
Adaboost	98%	98%	99%	98.70%
Stochastic Gradient Boosting	98%	97%	98%	98.50%

The confusion matrix of all the Ensemble Machine Learning classifiers is shown in Figures 3 to 6.

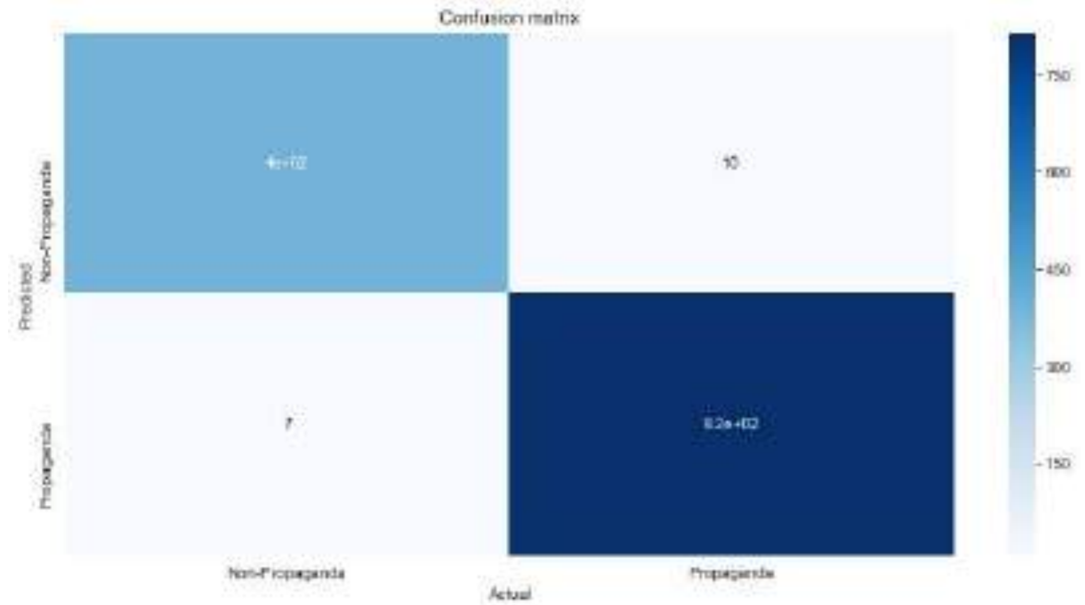

Figure 3: Confusion Matrix of Random Forest

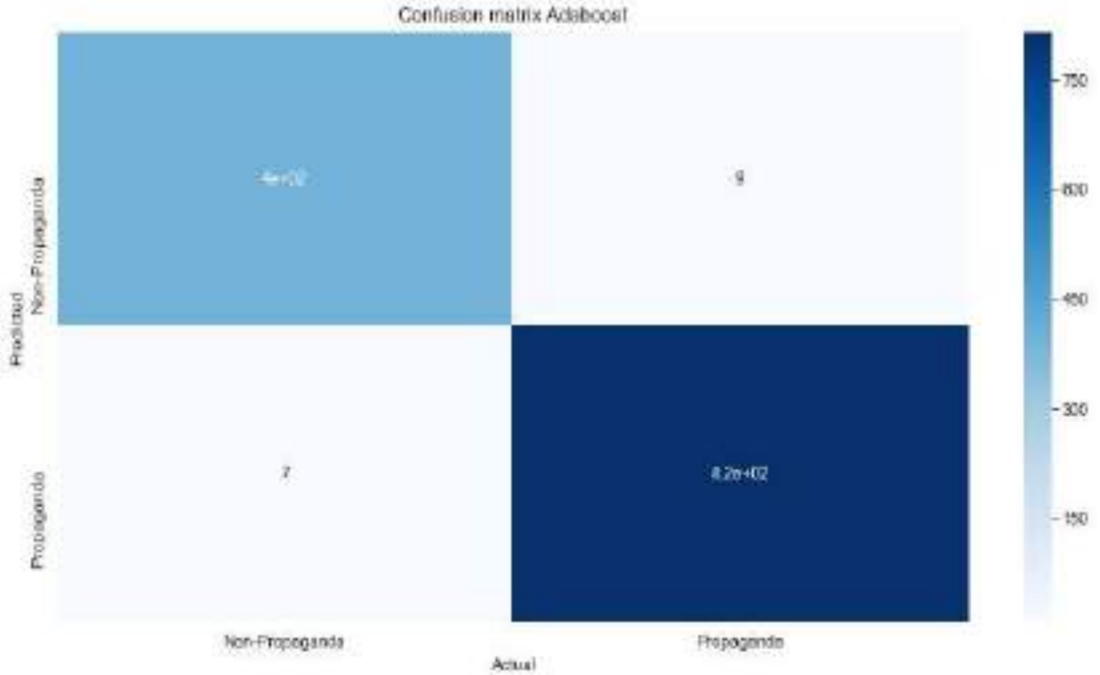

Figure 4: Confusion Matrix of Bagging.

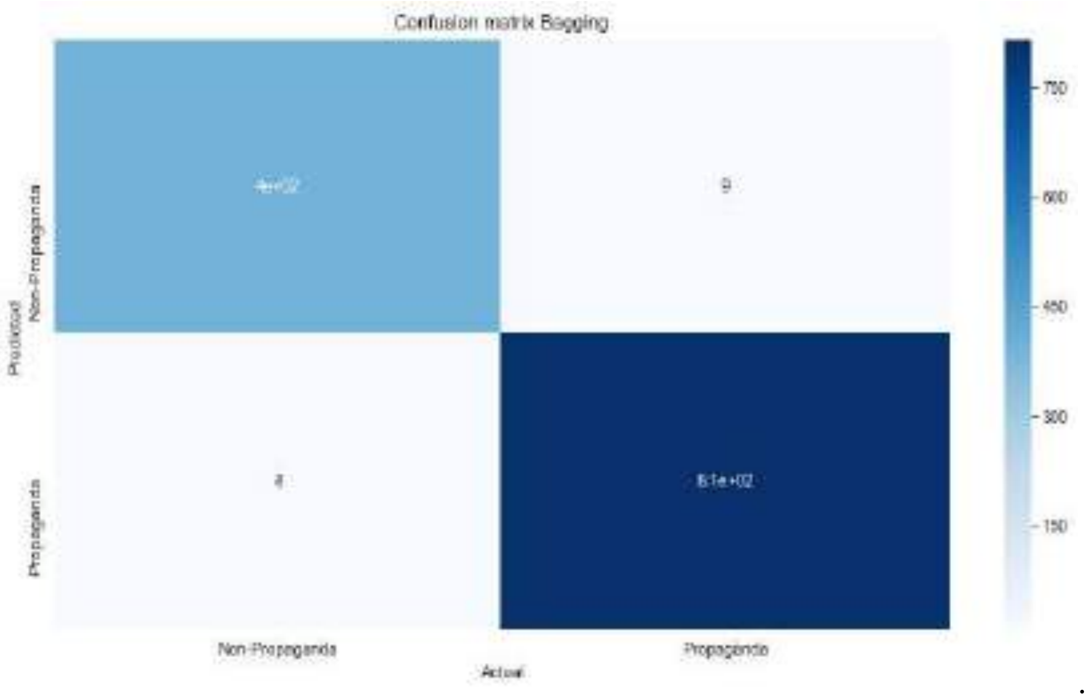

Figure 5: Confusion Matrix of Adaboost.

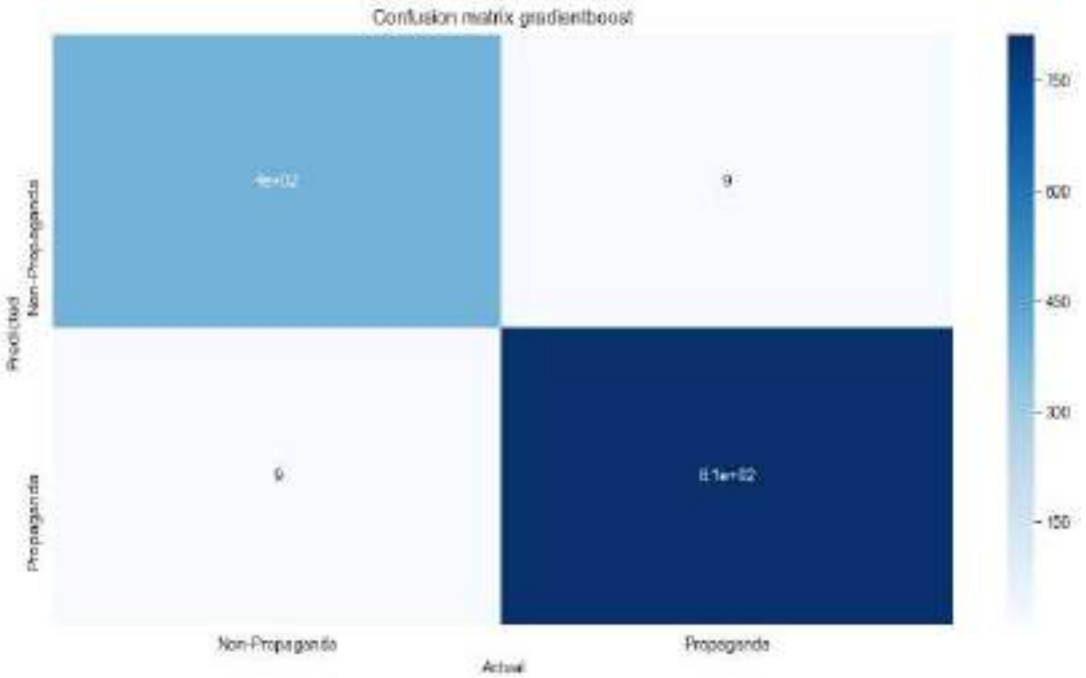

Figure 6: Confusion Matrix of Stochastic Gradient Boosting.

No previous literature was found in which Ensemble Machine Learning was used for identifying propaganda. Accordingly for validating this work 10-Fold cross-validation was performed and it was seen that there is no under skewness issue. Likewise, it was observed that no Under-fitting/Overfitting occurred during preparing and testing of the proposed model. Under-fitting occurs when a model is unable to capture the underlying trend of the data and it effects the accuracy of the model. At the same time, overfitting occurs when a model trains from massive amounts of data and starts learning from noise. The Comparison of Ensemble Machine Learning Classifiers is shown in Figure 7.

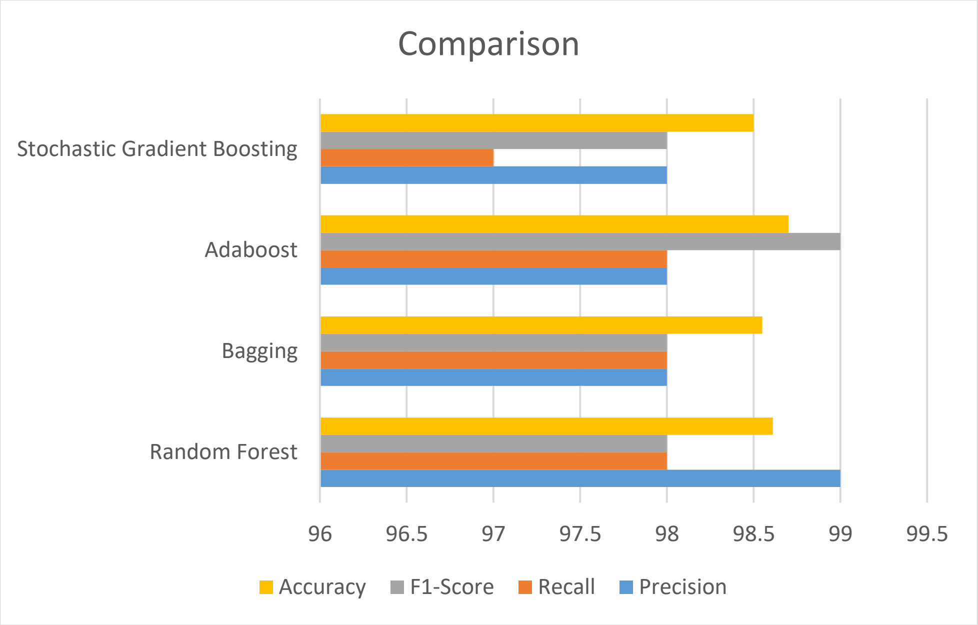

Figure 7: Comparison of Ensemble Machine Learning Classifier.

5. Conclusion

Machine Learning has gained interest nowadays as it has various applications. During the COVID-19 pandemic, false Information and propaganda are constantly shared through Online Social Networks. Data was extracted from Online Social Networking site "Twitter" utilizing its API. The extricated information was manually labelled in binary classes' propaganda and non-propaganda. In this paper, Hybrid feature designing was performed by consolidating three distinctive literary features (TF/IDF, Bag of Words and Tweet Length). Outcomes uncovered that propagandistic content has more prominent length than no propagandistic content. Ensemble ML techniques were utilized for performing classification of tweets into propaganda and non-propaganda category. Adaboost classifier indicated superior outcomes amongst other Ensemble ML techniques with 98.7 % Accuracy, 0.98 precision, 0.98 recall and 0.99 F1-Score. In future, more features engineering may improve accuracy. Additionally, Deep learning can improve the classification task and can be used in place of Ensemble Machine Learning Classifiers.

References:

[1] M. Babcock, D. M. Beskow, and K. M. Carley, "Different faces of false: The spread and curtailment of false information in the black Panther Twitter discussion," *J. Data Inf. Qual.*, vol. 11, no. 4, 2019.

[2] Y. Zhou, "Pro-ISIS fanboys network analysis and attack detection through Twitter data," *2017 IEEE 2nd Int. Conf. Big Data Anal. ICBDA 2017*, pp. 386–390, 2017.

[3] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media," *Bus. Horiz.*, vol. 54, no. 3, pp. 241–251, 2011.

[4] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *Int. J. Inf. Technol.*, 2020.

[5] World Economic Forum, "The Global Risks Report 2017 12th Edition," *Glob. Compet. Risks*

Team, p. 103, 2017.

[6] A. Gupta, H. Lamba, and P. Kumaraguru, "$1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on twitter," *eCrime Res. Summit, eCrime*, 2013.

[7] T. H. E. Arts, C. Policy, C. Justice, N. Security, P. Safety, and H. Security, *Rand_Mg877*. .

[8] A. M. U. D. Khanday, Q. R. Khan, and S. T. Rabani, "Detecting Textual Propaganda Using Machine Learning Techniques," *Baghdad Sci. J.*, no. December, pp. 199–209, 2020.

[9] G. Cybenko, A. Giani, and P. Thompson, "Cognitive hacking: A battle for the mind," *Computer (Long. Beach. Calif).*, vol. 35, no. 8, pp. 50–56, 2002.

[10] S. T. Rabani, Q. R. Khan, and A. M. U. D. Khanday, "Detection of suicidal ideation on Twitter using machine learning & ensemble approaches," *Baghdad Sci. J.*, vol. 17, no. 4, pp. 1328–1339, 2020.

[11] K. P. K. Kumar, A. Srivastava, and G. Geethakumari, "A psychometric analysis of information propagation in online social networks using latent trait theory," *Computing*, vol. 98, no. 6, pp. 583–607, 2016.

[12] P. N. Howard and B. Kollanyi, "Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum," *Ssrn*, 2016.

[13] O. Varol, E. Ferrara, F. Menczer, and A. Flammini, "Early detection of promoted campaigns on social media," *EPJ Data Sci.*, vol. 6, no. 1, 2017.

[14] C. Paul and M. Matthews, "The Russian 'Firehose of Falsehood' Propaganda Model: Why It Might Work and Options to Counter It," *Russ. "Firehose Falsehood" Propag. Model Why It Might Work Options to Count. It*, 2017.

[15] E. Ferrara, “Contagion dynamics of extremist propaganda in social networks,” *Inf. Sci. (Ny).*, vol. 418–419, pp. 1–12, 2017.

[16] H. A. Alatabi and A. R. Abbas, "Sentiment analysis in social media using machine learning techniques," *Iraqi J. Sci.*, vol. 61, no. 1, pp. 193–201, 2020.

[17] A. M. U. D. Khanday, Q. R. Khan, and S. T. Rabani, "Identifying propaganda from online social networks during COVID-19 using machine learning techniques," *Int. J. Inf. Technol.*, 2020.

[18] A. A. Abdualrahman and M. K. Ibrahem, "Intrusion Detection System Using Data Stream Classification," *Iraqi J. Sci.*, vol. 62, no. 1, pp. 319–328, 2021.

[19] P. Verma, A. M. U. D. Khanday, S. T. Rabani, M. H. Mir, and S. Jamwal, "Twitter sentiment analysis on Indian government project using R.," *Int. J. Recent Technol. Eng.*, vol. 8, no. 3, pp. 8338–8341, 2019.