International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.14, No.1 (Jul-23)

http://dx.doi.org/10.12785/ijcds/140112

Topic Modelling, Classification and Characterization of Critical Information

Anuja Soni[1], Sarabjeet Kaur Kochhar[2*], Shruti Jain[3], Megha Karki[3] and Vibha Gaur[3]

[1]*Deen Dayal Upadhyaya College, University of Delhi, Delhi, India*
[2]*Indraprastha College for Women, Department of Computer Science, University of Delhi, Delhi, India*
[3]*Acharya Narendra Dev College, Department of Computer Science, University of Delhi, Delhi, India*
[*]*Corresponding Author*

Received 14 Jun. 2022, Revised 6 May. 2023, Accepted 7 May. 2023, Published 1 Jul. 2023

Abstract: The misinformation, spread on the social media sites such as Twitter, overshadows the utility of such platforms, especially during times of crisis. Fake content is spread to popularize unauthorized treatments or downgrade the efficacy of preventative measures and treatments, resulting in spread of anxiety, depression and chaos amongst society. It is need of the hour, therefore, to apply the technologies like deep learning, natural language programming, and data mining, to develop automated systems that can discern false information from the real information, characterize it for better understanding, and mine it to derive actionable knowledge, that helps to check the spread of misinformation. This work proposes an automated framework that uses a combination of NLP & descriptive and predictive machine learning techniques. COVID-19 related messages on the social media sites are classified as appropriate or misleading using a deep learning model. With an accuracy of 86% BERT classifier was used to classify 3777 tweets. The model tagged 2350 tweets as real and 1427 tweets as fake. The classified social media information is characterized based on its sentimental valence, sentimental intensity and emotional acceptance in public, for better understanding. It was found by the framework that the polarity and intensity of negative fake tweets is much higher than the intensity of positive real tweets. It was found by the framework that the sentiment polarity and intensity of negative fake tweets is much higher than the intensity of positive real tweets. The emotional analysis re-enforced that the fear and negativity of the fake tweets far surpass the fear and negativity spread by real tweets. In fact, conclusions could be drawn that established that the real tweets generated more positivity, joy, trust, and lead to more anticipation. Critical information is retrieved from the authentic information, analyzed for better comprehension, and put in an actionable form, ready to be leveraged. The popular fake information, such as myths or rumors, also equally important to be identified are retrieved and understood, in order to develop counter-strategies for curbing their spread. Results demonstrate that the framework developed in this paper is able to successfully classify information as fake or real; sentimentally and emotionally characterize it, and churn out novel, actionable and interesting knowledge, crucial for the policymakers, to curb the spread of misinformation.

Keywords: Topic Modelling, Classification, Clustering, Sentimental and Emotional Analysis, COVID-19 misinformation

1. INTRODUCTION

The micro-blogging platform Twitter has been not only one of the most popular but also one of the most influential social media sites [1]. People don't just take to Twitter to air views about their day-to-day life and things that concern them, the amount of traffic on Twitter, in the case of some worldwide event, is unprecedented. Through their short messages, people often give very specific information that helps connected people, sometimes, instantly [2]. Such messages also serve as an indispensable tool to guide the policymakers and designated officials to take stock of the situation severity, its effect on the people, and plan the course of action. However, in case of critical events such as disease outbreaks, natural disasters like hurricanes, flooding, earthquake, or a health pandemic, the utility of such media is shadowed by the spread of misinformation [3, 4, 5].

Misinformation broadly refers to false, inaccurate, deceptive, or malicious information. Unverified rumors, conspiracy theories, and misleading information, etc., all are considered forms of misinformation. Even fake news, which refers to the intentionally fabricated information, that is generally crafted to be sensational or spread to charge public emotions, is considered a part of the misinformation. Information that misleads people about the precautions,

E-mail address: anujasoni@ddu.du.ac.in, skaur@ip.du.ac.in, shrutijain.cs.andc@gmail.com, meghakarki.cs.andc@gmail.com, vibhagaur@andc.du.ac.in

preventative measures, treatments, official restrictions, and health advisories regarding the pandemic, puts people at risk, by preventing them from making informed decisions regarding their health. Anxiety, fear, distress, and societal discord are also common consequences of misleading information [6]. Misinformation can also pose a threat to our global economies by affecting stock markets and promoting hoarding etc. Epidemics, such as the recent Ebola, Zika, and yellow fever are case studies that exemplify the aforementioned payoffs of misinformation.

The COVID-19 pandemic has been the most defining global health emergency of the last century. In fact, due to a massive, spontaneous eruption of the misinformation on the internet about COVID-19, the pandemic has also been referred to as an 'infodemic'. Quoting Sylvie Briand, the architect of WHO's strategy to counter the infodemic risk, "We know that every outbreak will be accompanied by a kind of tsunami of information, but also within this information you always have misinformation, rumors, etc." [7].

Understanding information propagation on social media is therefore a pertinent task to develop capabilities to identify and control the spread of misinformation. However, the sheer volume of information on social media sites makes it very difficult to discern critical information, that can be leveraged during the time of crisis, from the unsubstantiated or potentially false information circulating on the web [8, 9]. In this paper, we present an approach towards the resolution of this problem, by developing automated mechanisms, that can address the following tasks:

• Classification of the information on the social media sites as appropriate or misleading. Though the literature includes some manual approaches to perform this task, it is easy to see that a shift towards semi-automation and/or full automation of such classification approaches is highly desirable [9, 10]. Deep learning techniques, a subset of machine learning, usually work on structures modelled on the human brain and have been successfully employed in a vast number of successful applications in heterogeneous big data analytics [11]. Deep neural networks have been shown to achieve human-like performance in identifying, tagging, and sorting image and video data. We use BERT, a deep, a neural network-based technique, for distinguishing real information from the fake information on Twitter [12] in Section 4.

• Characterization of the classified social media information is necessary to understand its sentimental valence, sentimental intensity, and its emotional acceptance in public. The task of characterization of real and fake tweets is taken up through classification in Section 5.

• Critical information must be retrieved from the authentic information and analyzed, to understand its characteristics and finally put in a form, in which it can be leveraged and put to action [2]. To achieve this, we attempt to retrieve the important topics of discussion on Twitter by observing the collective characteristics of the authentic tweets in Section 6.

• The immediate proliferation of misinformation on the social media sites like Twitter play an important part in the formation of public opinion, apart from the other perilous consequences, noted in the introductory discussion [6, 13]. The popular fake information, circulating on social media sites, therefore, requires to be retrieved and understood. Understanding misleading information such as myths or rumors is essential for developing counter-strategies, to help curb their spread. For instance, the nodal medical agencies need to issue myth-busters and formal guidelines against rumors from time to time. An approach towards the realization of this task is also presented in Section 6.

The organization of the paper is as follows: Related Work is reviewed in Section 2. Section 3 presents the general template of the framework proposed in this paper, to classify, characterize and retrieve critical information from the social media platform, Twitter. It also lays down the details of the implementation of the framework and the data used. Section 4 presents the classification framework to discern authentic information from misleading information. Section 5 details the characterization framework that performs a comparative study of the sentimental and emotional characteristics. Section 6 elaborates the clustering framework for retrieval of the critical classified information. An approach for analyzing the retrieved information for greater understanding and control is also presented in this section. Section 7 concludes the paper.

2. LITERATURE REVIEW

We review the literature related to all aspects of our work, classified according to their relevance, in Section 2A (Misinformation detection on social media) and Section 2B (Topic modelling). The works closest to our work are discussed and compared in Section 2C (Misinformation Detection and Modelling). Section 2D discusses the works lying at the crossroads of the studies outlined in Sections 2A-2C and sentimental and emotional analysis.

A. *Misinformation Detection*

Fake information detection, by detecting the malicious bots that spread them, has been proposed in [1]. To detect the fake information spreaders, the work employs user-based features and content-based features from Twitter to compute statistics such as TF-IDF, Bag of Words, and average mean time to tweet. Sentiment analysis using the VADER lexicon to compute the average sentiment score of the users is also performed to gauge the overall sentiment polarity of a particular user's tweets. The proposed algorithm is compared with other machine learning classifiers such as decision tree, MLP and random forest, and shown to be more accurate.

A model based on fusing social context along with the news content for categorizing fake news has been proposed in [5]. For an effective classification, the model also uses tensor factorization with a deep neural network. It is shown that a combination of news content and social context gives better performance than using the news content or social context alone. The performance of the model is evaluated on BuzzFeed and PolitiFact real-world fake news datasets.

The authenticity of the tweets about the protests staged in Hong Kong, in 2020, is evaluated by [10]. A dataset of English and Chinese tweets, related to the incident, published by Twitter are processed for the extraction of the top 10 most significant features, according to the mutual information metric, such as the tweet length, TTR, number of punctuation marks, number of periods, average number of characters per sentence, number of adverbs, number of "to", number of verbs, number of entities and tweet entropy. The extracted linguistic features along with semantic polarity are used to ascertain linguistic patterns to distinguish tweets spreading fake from the tweets spreading real news. Four different algorithms are used for the training and evaluation of classification models, namely the Naïve Bayes, SVM, C4.5, and Random Forests of C4.5, using the Scikit-Learn Python module. This work deals with determining the authenticity of tweets, majorly using linguistic features, and formally enlists the utilization of deep neural network algorithms as a possible future direction, a goal that has been taken up in our work.

An adaptation of the BERT model, for fake news detection, has been proposed in [8]. The proposed model builds three blocks of 1d-CNN, with different kernel-sized convolutional layers and different filters to demonstrate better learning on the U.S. General Presidential Elections dataset. Textual fake news detection using deep neural networks has been taken up in [9]. The paper uses Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) RNN methods. The paper compares 'glove', an open-source project at Stanford University, supported by the Flair library, a Natural Language Processing library designed for word embeddings, with the 'twitter', 'news', and 'crawl' word embeddings. Word embeddings are the numerical representations of the text being investigated. An association rule-based framework to classify real and fake news has been proposed in [6]. The tweet corpus is transformed into a binary transactional database. The Apriori algorithm is used to generate association rules, which are used as the basis for the classification of news as real or fake, and visualized for convenient interpretation.

B. Topic Modelling

Some studies focused on the topic modelling and did not involve work on the detection or modelling of misinformation spread during the COVID-19 pandemic. For instance, the discovery of differences in the topics of conversation among the different genders, namely the men, women, and other gender minorities, on Twitter, has been studied in [14]. Identification and comparative analysis of the topics, sentiments, and emotions, using unsupervised learning approaches, from the tweets of two news agencies from Iran and Turkey have been undertaken in [15]. Monitoring the weekly evolution of the most discussed topics regarding COVID-19 on Twitter is taken up in [16]. In this work, two matrix decomposition algorithms, namely, Rolling-ONMF and Sliding-ONMF, are proposed and compared, for identifying the topics and studying their evolution over time, from the collected tweets. Other related works can be found in [17, 18].

C. Misinformation Detection and Topic Modelling

Tweets from the accounts of the six fact-checking agencies in Latin America have been clustered to identify topics of concern and patterns in COVID-related misinformation across the region [19]. In contrast to our work, where a deep neural network is first used to identify authentic information and misinformation, the work by Ceron et al. relies on fact-checking agencies for the identification of the misinformation. Also, our work models, both, the topics of discussion in real tweets, and, myths in fake tweets, as opposed to the aforementioned work, which focuses on topic modelling of misinformation in Latin America.

A text-matching-based misinformation detection approach has been proposed in [4], wherein, the entire tweet corpus is matched with small data set, already labelled as misleading. In contrast to our work, where we use semantic textual clustering to learn the topics of discussion, the misinformation tweets in [4] are put in the form of a matrix tensor model. Associations among the modes term, location, and time, is preserved and used to retrieve the topics, according to the Spatio-temporal characteristics.

A framework for the prediction of potential topics for fake news dissemination is outlined in [20]. The knowledge of a topic's susceptibility to misinformation is proposed to be used for recognizing fake news. The paper also serves to identify, the potential spreaders of unsubstantiated news. The data about the official and fake news is collected from different websites. Topic extraction and sentiment analysis are then performed, before extracting a set of features to determine how information is perceived. The classification of the fake news is then performed the task using machine learning algorithms.

A study exploring the coherence of extracted topics to identify deceptive information is presented in [21]. The study contributes by developing an understanding of the fake news by analyzing thematic deviations between the opening and remainder parts of the news. Experiments are performed on seven datasets to show that fake news depicts less coherence between its opening sentences and its body text.

D. Sentimental and Emotional Analysis

This subsection indicates the works at the intersection of the tasks of misinformation detection, topic modelling, and sentiment analysis reviewed in Section 2A, Section 2B, and Section 2C. These works are discussed here to bring about the facets of sentimental analysis and/or emotional analysis work, undertaken by them.

The following works related to the identification of misinformation, already discussed in subsection 2A, also perform sentiment analysis. The work done by Monica et. al. [1], in the direction of spam detection on Twitter, involves an analysis of the opinions and the emotions of users. For the users who recently posted on Twitter, sentimental analysis is performed on the tweets on their topic of interest. An overall sentiment score for each user is computed, to enable the detection of spam Tweets. Zervopoulos et. al. use sentiment valence derived from tweet text and emojis as one of the factors that contribute to the basis of the classification of fake news on Twitter [10]. The work by Kula et al. also claims to use a sentiment analysis-based framework for fake news detection [9].

The sentimental study done in the works on topic modeling, discussed in section 2B are enlisted as follows. A study to find the differences in the gendered discourses around COVID-19 was taken up in [14]. The work involved sentiment analysis using a Python 3 package called VADER, to calculate the mean sentiment score and categorize it as highly negative, highly positive, or neutral. The study in [15], analyzed tweets of two official news agencies of Iran and Turkey, to identify the emotions and sentiment behind the topics of discussion in tweets and compared them. The results were psychologically and sociologically interpreted.

Sentiment polarity of content associated with the potential topics for fake news, retrieved by the framework, presented by Viacrio et. Al. [20], is discussed in subsection 2C. To the best of our knowledge, none of the works detailed above, undertake to measure and compare the sentiment valence as well as intensity and the emotional response of the people to the real vs. fake information, circulating on the web, as is accomplished in our work.

3. FRAMEWORK FOR INFORMATION RETRIEVAL, CLASSIFICATION, AND CHARACTERIZATION

In this section, we present the outline of the framework, used in this work, to extract, classify and develop understanding and leverage the tweets related to the covid-19 pandemic, posted by people worldwide (Section 3B). The details of the data extraction and its pre-processing are presented in Section 3A. Details of the programming environment and the underlying implementation are laid down in Section 3C.

A. Data Extraction and Preprocessing

Twitter, a popular social media microblogging site, allows users to post short messages, typically restricted to 140 characters in length, to be shared with other users in real-time. The popularity of Twitter, owing to a very large base of its subscribers, has made this social media platform, an indispensable resource for researchers, wishing to study the opinions, sentiments, and emotions of people, on a wide variety of topics.

We use this platform to gather tweets, posted by people, worldwide, related to the COVID-19 pandemic. Approximately 21k tweets were scraped from the Twitter API for a period extending two weeks from October 2020 and two weeks from January 2021. Two different periods were chosen for study, to enable the study of time-variant patterns. The extracted tweets were subjected to horizontal and vertical data selection. During the horizontal data selection, filters were set to retrieve only the tweets corresponding to the COVID-19 pandemic. This resulted in the extraction of 6985 pandemic-related tweets. During the vertical selection, only the attributes pertinent for the study at hand were retained, thus reducing the dimensionality of data to be studied. The selected tweets were pre-processed using the Natural Language Toolkit (NLTK) package, an open-source Python library for natural language processing. The tweets were tokenized and cleaned. This included removal of noise i.e., removal of Twitter handles, special characters, URLs, punctuations (not representing emotions), Twitter handles and stop words, etc. Then the text of tweets was converted to lowercase and stemming was performed.

B. Research Framework

The schematic diagram of the general framework is depicted in Figure 1. The first component of the framework was data extraction and preprocessing, the details of which were presented in the previous subsection (Section 3A).

The next component focused on the classification of tweets, wherein, the tweets, scraped from the social media, were classified as real tweets and fake tweets. The classifier was first trained on a sample of tweets, manually annotated by multiple human experts. The classified tweets were then subjected to comparative analysis to assess their sentimental and emotional characteristics. Sentimental valence analysis, Sentimental intensity analysis, and emotional analysis were performed on the classified tweets to ascertain the polarity of the opinion expressed in the tweets, and how people emotionally responded to them. The details of classification and the subsequent characterization of tweets are presented in Sections 4 and 5 respectively.

The real and fake tweets were separately clustered to observe, characterize and compare the clusters of classified tweets. Significant information from the clusters of tweets containing authentic tweets was retrieved. The clusters of misleading information were also analyzed to successfully detect myths and rumors regarding COVID-19 circulating on Twitter. The formation of clusters and their analysis are detailed in Section 6.

C. Implementation Details

The research framework was implemented using Python 3.9.1. For deployment of the BERT (Bidirectional Encoder

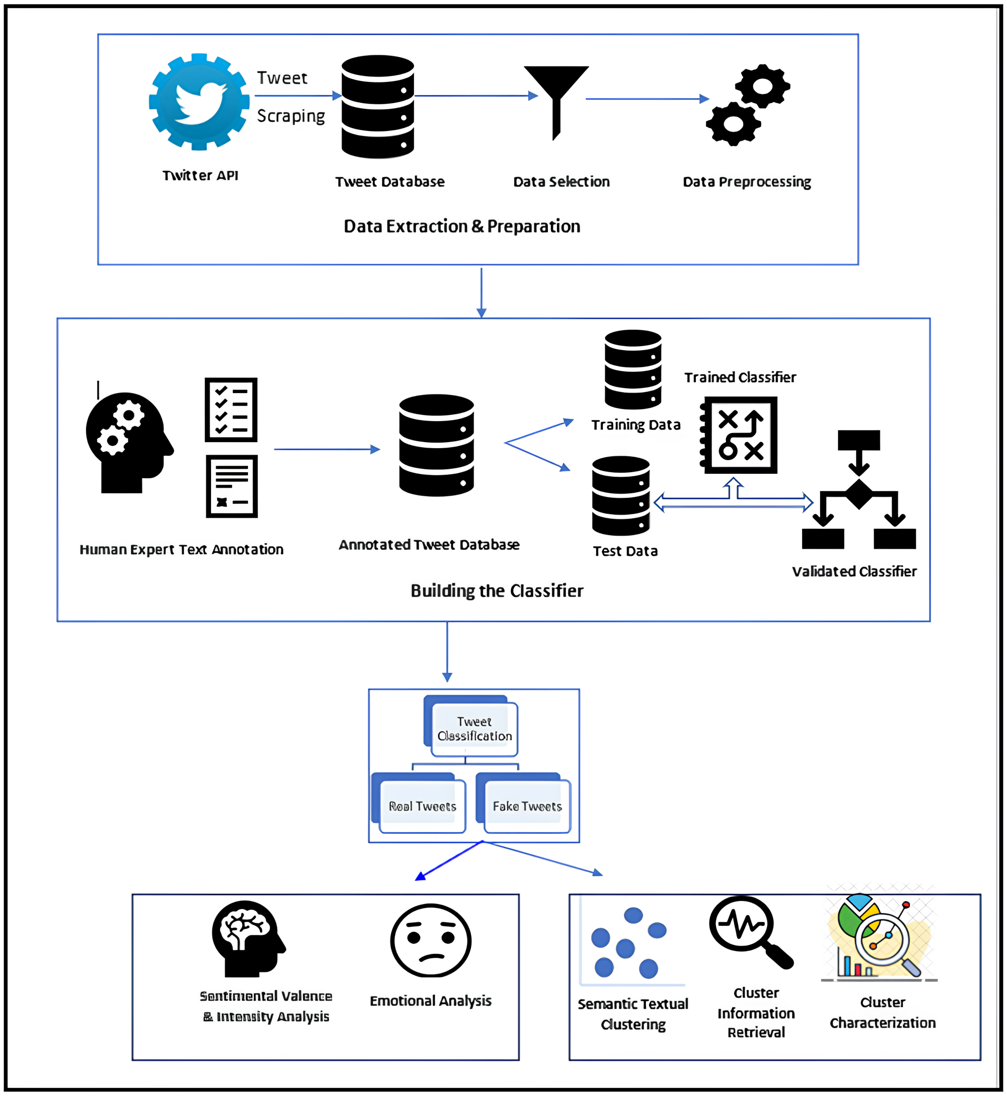


Figure 1. Research framework to classify, characterize and leverage the knowledge from COVID-19 tweets.

Representations from Transformers) classifier, the following python libraries and packages were used: PyTorch, an open-source machine learning library, was used for developing and training the neural network-based deep learning model BERT. Transformers were used with the BERT, for natural language processing. TensorFlow, another open-source library for machine learning, was used for the training and fitting of data for the BERT model. PyCaret, a Python equivalent of the Caret machine learning package in R, was used for tuning and evaluating the BERT model. Matplotlib, a plotting library, was used for plotting the graphs and its numerical mathematics extension, NumPy was used for performing statistical, and algebraic functions on the arrays. Seaborn, another visualization library was used for statistical graphics.

For Sentimental and emotional analysis, the re library was used to clean the noisy tweets. pandas was used for merging, shaping, data cleaning, and data wrangling. NRCLex was used for measuring the emotional effect of tweets. Plotly 4.14.3, NumPy1.20, and Matplotlib 3.3.4 were used for visualizing the data.

For the implementation of the k-means clustering algorithm, the sklearn library was extensively used for data pre-processing, tokenization, filtering the stop-words. For instance, the sklearn TfidfTransformer function was used to get the frequencies. nltk package was used for text pre-processing. The string module was used for manipulating strings. Collections, for example, counters, lists, dicts, sets, and tuples, etc., were used extensively in the code. Kneed, an implementation of the Kneedle algorithm, was used for identifying the knee or elbow point, to determine the optimal number of clusters. pandas, matplotlib, NumPy, seaborn, and re, used for classification and/or sentimental and emotional analysis were also used for clustering.

4. TWEET CLASSIFICATION

The BERT model was employed in this work to classify the tweets as real tweets or fake tweets. BERT (Bidirectional Encoder Representations from Transformers), is an open-source machine learning framework for natural language processing, known for its deep contextual understanding of the ambiguous text due to its bidirectional learning capabilities.

Training of the supervised learning models, such as classifiers, is essential for building predictive models and ensuring their efficacy. Annotation is the process of assigning labels, manually or through technological interventions, to the textual, audio, image, or video data for compiling a tailored training sample. Human expert annotation has been successfully applied to implement and improve a variety of machine learning and artificial applications such as speech recognition, chatbots, improving the relevance of search engine results, product recommendations, etc.

The BERT-base-uncased model was used to do the classification job. TensorFlow 2.0 and Python's Transformers package were used to build the BERT model. The pre-trained BERT model was fine-tuned for the binary classification job at hand. To avoid overfitting, a dropout layer with a dropout rate of 10% was added to the BERT layer's output. Following that, a Dense layer of 768 neurons with ReLU activation function was added, and a Dense layer of 512 neurons was used with Softmax activation function. The tweets were divided into two categories: real and fake. To maximise the performance of the BERT model, the Negative Log Likelihood Loss (NLLLoss) function was utilised as the loss function. During the training phase, different values of hyper-parameters such as Optimizer function, Epochs, number of layers, and Learning Rate were used to execute experiments numerous times. The model was eventually fine-tuned using the best parameters, as shown in Table 1.

The python packages and libraries used for implementation of BERT have already been detailed in Section 3C. For incorporating contextual information about the real and fake tweets in the BERT framework, a sample of tweets was annotated by human experts. Multiple domain experts were involved in the annotation process to rule out bias and incorporate multiple subjective perspectives. Before application, the BERT model was trained on the annotated tweet data set. A total of 3208 tweets were tagged as real tweets or fake tweets by manual inspection of the tweet content by multiple annotators. A balanced training set was constructed by randomly merging the annotated tweets. This annotated set of tweets was partitioned into a training set, using the 'train - validation split' option, which consisted of 2642 tweets. Using the 'validation-test split' option, a test set was created, that consisted of 566 tweets. This training set of annotated tweets were used to train the classifier. When validated on the annotated tweets test set, the classifier gave an 86.6 accuracy, F1-score of 83.66, and precision and recall as 81.57 and 85.59 respectively. The validated BERT model was then used to classify 3777 tweets. The model tagged 2350 tweets as real and 1427 tweets as fake.

TABLE I. Parameters used for tuning BERT

Hyper-parameters	**Optimal Value**
Max. Sequence Length	15
Optimizer	Adam
Loss Function	Negative Log Likelihood Loss (NLLLoss)
No. of Epochs	10
No. of Layers	12
Batch Size	32
Learning Rate	1e-5
Activation Function	ReLU (hidden layer), Sigmoid (output layer)
Train, test, and validation split	70%, 15% and 15% respectively
Dropout	0.1

Identification of misinformation by weak supervision has been taken up in [22]. The study attempts classification of real and fake news on the social media, compares various learners on the basis of performance metrics and chooses XGBoost classifier that gives an F1 score of 0.9. This is comparable to a F1-score of 83.66 yielded by the BERT model implemented in our framework. Another work to classify reliable and unreliable news and tweets on the social media has been presented in [23]. Again, the results of the framework (71%<=F1 score<= 83% for unreliable news and 43%<=F1 score<= 67% for reliable news) are comparable with our results. In a more recent work, a combination of Bi-LSTM model along with the BERT classifier is used for misinformation detection with an accuracy of 87.02% in [24], also comparable to the accuracy of our framework at 86.6%. A BERT based model to identify fake tweets in two Indian languages viz. Hindi and

Bengali has been presented in [25], that achieves 81% F-Score in Hindi and 78% F-Score for Bengali Tweets, which is lower than the F1 score 83.66 of the model presented in our framework for English tweets.

5. SENTIMENTAL AND EMOTIONAL CHARACTERIZATION OF CLASSIFIED TWEETS

Sentiment analysis, also sometimes referred to as subjectivity analysis, or opinion mining is a branch of natural language processing, used to ascertain the attitude or sentiments of the speaker, behind a piece of text. The text, to be analyzed for the sentiments it embodies, maybe highly unstructured or heterogeneous. It may be a small piece of text such as a tweet, a complex sentence, a paragraph, or a complete document.

The last decade has seen a lot of work in the field of sentiment analysis, especially, on Twitter data. However, most of the work has focused on the detection of sentimental valence i.e., positivity, negatively, or neutrality. It is important to note that a holistic assessment of the sentiment should not only take into consideration its polarity but also its intensity, i.e., the strength with which a particular sentiment has been expressed. In this work, we work to ascertain both the polarity and the intensity of the sentiment expressed by the tweets.

The real and fake tweets, classified by the BERT model, were subjected to sentimental and emotional analysis. The libraries and packages used for the analysis have been detailed under the subsection implementation details (Section 3C). The findings of the same are depicted from Figure 2 to Figure 5.

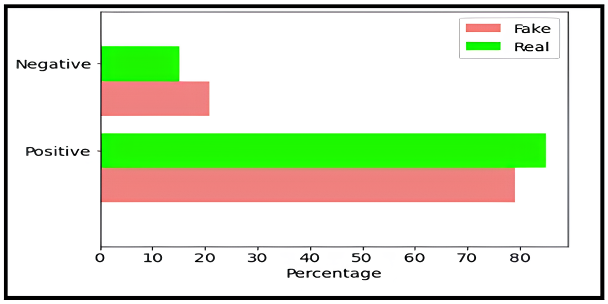


Figure 2. Sentimental polarity of Real vs. Fake Tweets

Figure 2 shows the overall sentiment polarity, in percentage, of the positivity and negativity of real and fake tweets. It can be interpreted from Figure 2, that the fake tweets related to the COVID-19 pandemic, worldwide, lagged in positivity as compared to the real tweets related to the pandemic. The fake tweets were at least 5% more negative than the real tweets.

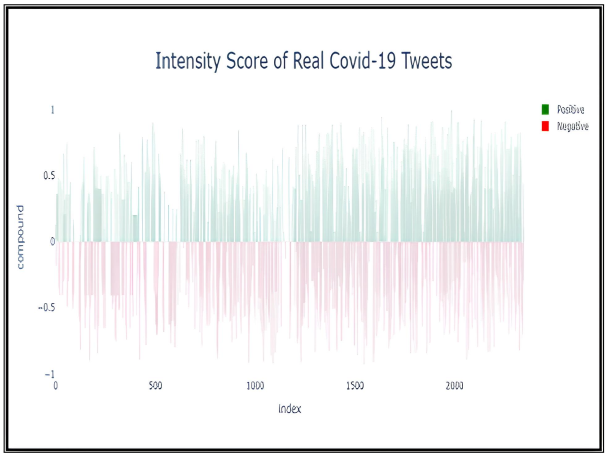


Figure 3. Intensity Score of Real Tweets related to COVID-19.

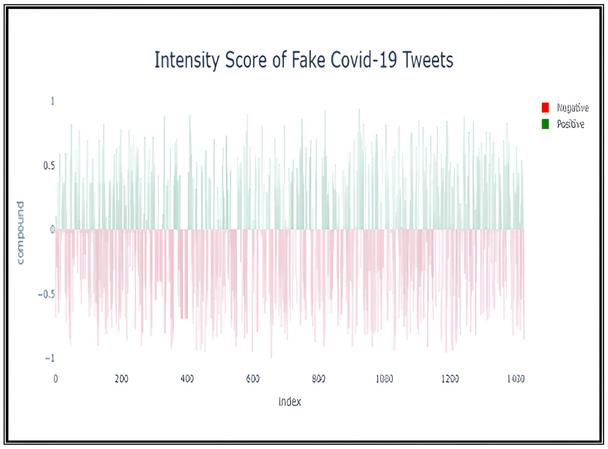


Figure 4. Intensity Score of Fake Tweets related to COVID-19.

The intensity of each individual real tweet and fake tweet is plotted in Figures 3 and 4 respectively. The figures reinforce the conclusions drawn from Figure 2. It can be seen from Figure 4 that the intensity of negative fake tweets is much higher than the intensity of negative real tweets. Thus, it can be concluded that the tweeters of fake news intentionally spread a strong negative sentiment on social media. The impetus for doing so may stem from intrinsic fear or anxiety due to the pandemic, or, the tweeter may just be acting like a rumor monger, wishing to spread fear. The emotional analysis of the classified tweets (Figure 5), strengthens the conclusions drawn above. The fear and negativity of the fake tweets far surpass the fear and negativity spread by the real tweets. It can also be seen that the real tweets generate more positivity, joy, trust, and lead to more anticipation.

6. TOPIC MODELLING AND MYTH RETRIEVAL

As discussed in Section 1, Twitter and other social media platforms have become indispensable sources of information, especially, during the time of a crisis. The rapid diffusion of information on social media platforms has helped evoke a global response from the netizens, to help out others, with whatever information they have. Sometimes, the time at which it is shared, such

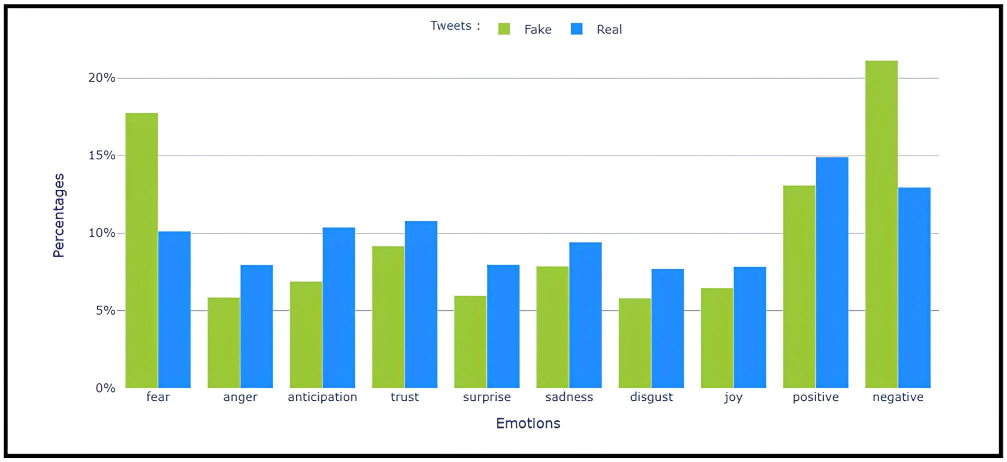


Figure 5. Emotional Analysis of Fake and Real Tweets related to COVID-19

information is not even available with the administrators and policymakers. They keep tuned in to the social networking sites, to know the latest that is happening at a particular geographical location. However, on the flip side, this swift information propagation system also leads to the hasty dissemination of volumes of unsubstantiated information. Apart from the other disadvantages of the propagation of misinformation, discussed in the introduction section, it becomes very difficult to identify and retrieve critical information from the mounds of information present on the web.

To surmount the aforementioned challenge, the information on social media sites needs to be classified as appropriate or misleading. The classification of tweets as real or fake tweets, by using the BERT model, has already been discussed in Section 4. As a next step, critical information must be retrieved from the clusters and analyzed, to understand its characteristics and finally put in a form, in which it can be leveraged and put to action. This is discussed in Section 6A.

A. *Information Retrieval and Analysis from Clusters*

To accomplish the goal of critical information retrieval and analysis, we propose to study the collective characteristics of the real information vs. the fake information circulating on Twitter. For this, we employ the k-means algorithm to perform semantic textual clustering, separately, of the tweets containing real information and fake information. From these clusters, we attempt to draw critical information and leverage it to promote its understanding. Developing capabilities to promote understanding is crucial to facilitate effective decision-making.

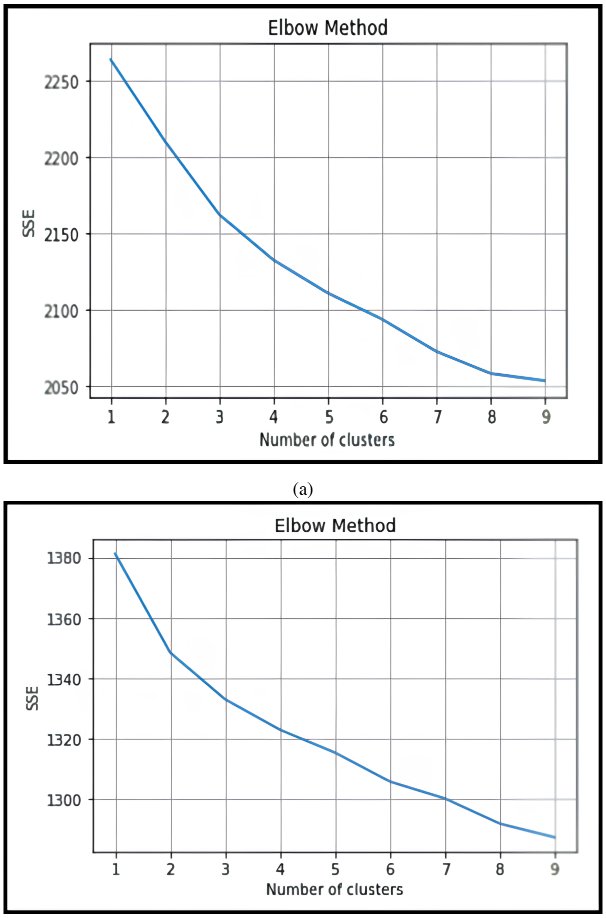
The basic premise behind the working of the k-means algorithm is as follows. Given k, the number of clusters to be formed, the initial data points are selected to form centroids of each of the k clusters. The distance of each data point in the dataset is computed from the centroids of the k clusters. The distance metrics for comparison may be chosen, depending upon the type of attributes and application domain. The data point is assigned to the cluster with the nearest centroid. After repeating this procedure for all the data points in the dataset, the centroid points are recomputed. The next iterations are undertaken in which the points are reassigned to clusters according to the new centroids, till the clusters converge i.e., don't change substantially. The implementation details and the libraries and packages used for implementing the mentioned clustering algorithm have been detailed in Section 3C.

To determine the optimal value of input parameter k, the number of clusters, a clustering visualization method known as the Elbow method was used. The method focuses on finding an 'elbow', which corresponds to the optimal value of k using the 'knee point detection algorithm'. The similarity metric used is the sum of squared distances of each point in the dataset to the centre of the cluster it is assigned.

The algorithm generated the optimal value of k = 4 for the real tweets and k = 3 for the fake tweets. Figures 6 (a) and 6(b) show the graphs of the cluster squared sum of errors vs. the number of clusters for real and fake tweets respectively.

Figures 7(a) and 7(b) shows the scatter plots for visualization of the four real tweet clusters and three fake tweet clusters, respectively, along with their centroids, generated as output by the k-means algorithm. Figures 8 (a) and (b) show a word cloud of the most frequent terms used in the real and fake, COVID-related tweets.

To extract and leverage the critical information regarding the authentic topics of discussion on Twitter, and, to get a quick understanding of the most talked-about fake information and myths circulating on Twitter, we retrieved the top k semantically cohesive frequent words from both



(a)

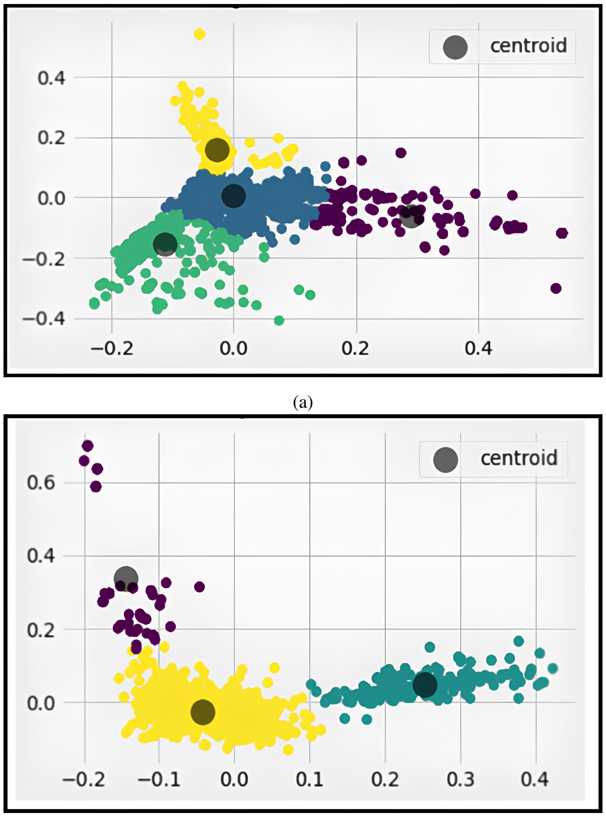
(b)

Figure 6. Optimal number of cluster determination through the Elbow method for (a) real tweets; (b) fake tweets

the real and fake clustering schemes. These words served as pointers to very important, relevant topics of discussion related to COVID-19, that made ripples on social media sites. We were also able to unearth many myths and rumors, circulating on the web, during the period of study.

The results of the top 10 words retrieved from each of the clusters in the fake clustering scheme are depicted in Figures 9(a), 9(b), and 9(c). The first cluster of fake news brought to light many popular rumors being spread on social media platforms, for instance, ‘spraying chlorine or alcohol on the skin kills viruses in the body [26]. The rumor was, in fact, so popular that the World Health Organization had to release a myth buster to counter it [27]. We could also unearth a couple of other rumors using this cluster’s findings, namely, ‘drinking alcohol reduces the risk of infection’ and ‘hand dryers kill the coronavirus’ [26]. Again, these myths were so popular, WHO refuted them in its myth busters [27].

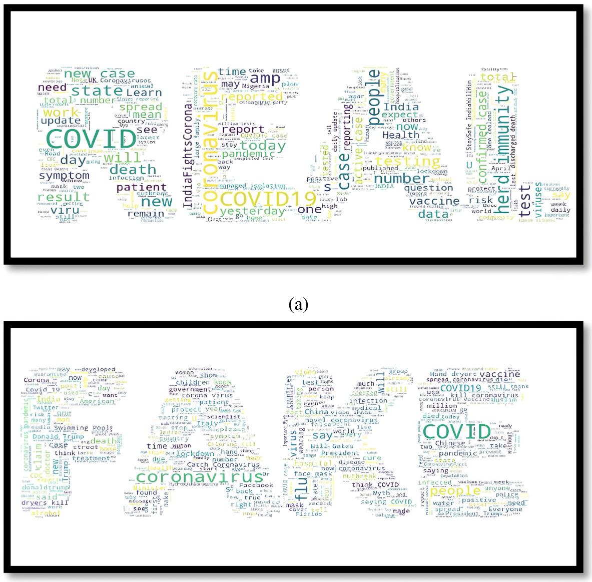
The second cluster of fake tweets was also successful in exposing contradicting misconceptions being tweeted,



(a)

(b)

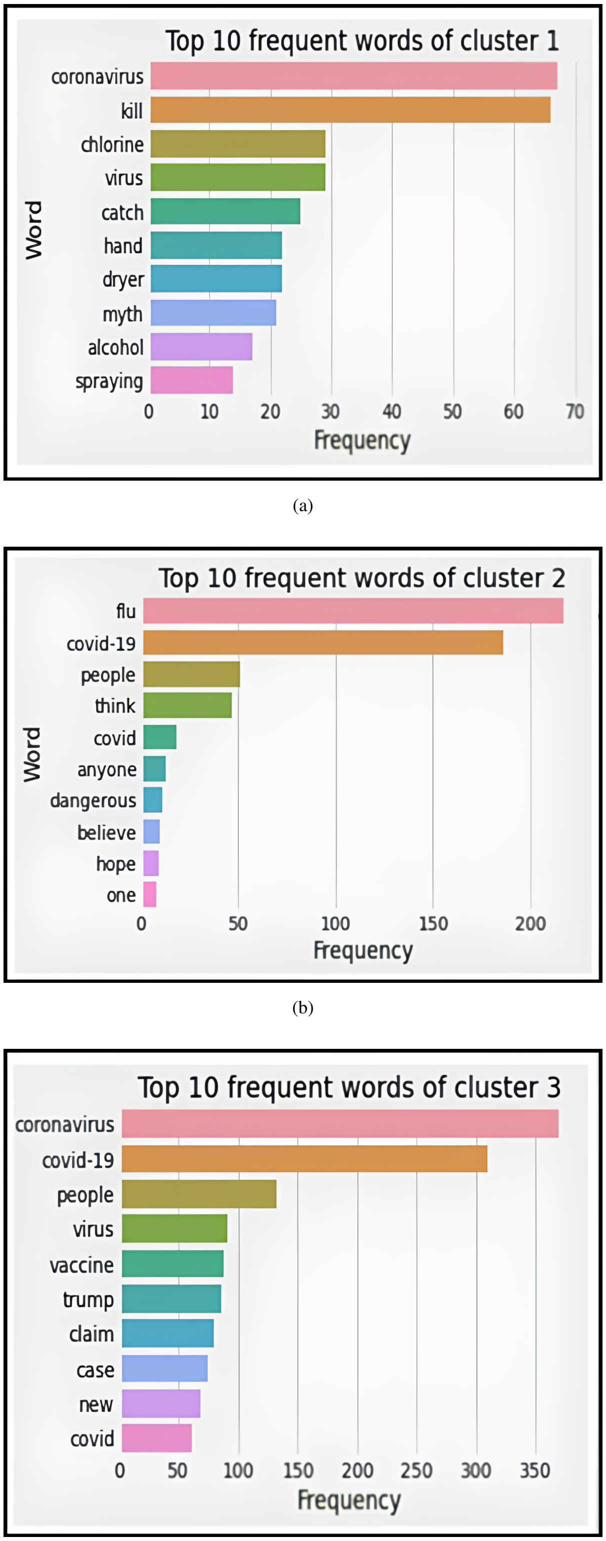
Figure 7. Scatter plots depicting visualization of the (a) real COVID-19 related tweet clusters;(b) fake COVID-19 related tweet clusters.



(a)

(b)

Figure 8. Word Clouds of COVID-19 related tweets (a) real tweets; (b) fake tweets.



(a)

(b)

(c)

Figure 9. Top 10 frequent words drawn from the three clusters of fake tweets, that is, (a) in Cluster 1; (b) in Cluster 2; and (c) in Cluster 3.

such as, ‘COVID-19 is just like the normal flu’, ‘Flu and pneumonia vaccines can protect against COVID-19’, and ‘The coronavirus is one of the most dangerous (deadliest) virus known to humans’ [27].

The third cluster of fake tweets brings forward, that false tweets were being circulated about the United States Ex-President Donald Trump. This finding is validated by the BBC news report [7]. The news article reports, that the Republican supporters tweeted a baseless rumor that the president was somehow deliberately infected with Covid-19 and at the same time, tweets from the Democrats falsely claimed that President Trump was more ill than was acknowledged and that his public appearances were staged with a body double. The third cluster of fake tweets also depicts the frequent use of words articulated in the following myths: ‘The new vaccine will turn people into crocodiles or monkeys.’, ‘The new COVID-19 vaccine has been proven to cause infertility in 97 percent of its recipients.’ [28, 29].

Figure 10 shows the results of the top 10 most semantically cohesive words retrieved from each of the clusters in the real clustering scheme. The first cluster of the real tweets, reveals an important topic being talked, all over the world. ‘Herd Immunity’ saw a storm of tweets. Even WHO posted many tweets regarding the herd immunity, some describing the phenomenon, others warning against achieving it without an effective vaccine. One of the tweets posted by the world health organization in October 2020 states, ‘Herd immunity or population immunity is achieved by protecting people from a virus, not by exposing them to it.’. Another one states, ‘Letting #COVID19 spread through populations, of any age or health status, will lead to unnecessary infections, suffering, and death. Without a widely available vaccine, herd immunity is a dangerous and counter-productive strategy for stopping the virus.’. Whether India had achieved herd immunity, was also a hot topic of discussion on Twitter, in October 2020. Times of India tweets, ‘Has India achieved herd immunity against Covid-19?’. Voicing similar sentiments, chairperson of the Scientific Advisory Committee, National Institute of Epidemiology, Dr. Jayaprakash Muliyil tweets, ’Active cases drop in India as more regions show signs of herd immunity.’

Another surprising viral topic that surfaced from the above findings was that the Indian Union Cabinet Minister for Textiles and Women & Child Development, and a popular celebrity star, Mrs. Smriti Irani, tested positive for corona, in October 2020. Economic Times tweets, ‘Union Cabinet Minister for Textiles and Women & Child Development, @smritiirani tests positive for #Coronavirus.’. Another interesting tweet story that could be derived from the cluster one of the real tweets was regarding the uproar created by then-president Donald Trump’s pitch to women on coronavirus recovery, ’We’re getting your husbands back to work’.

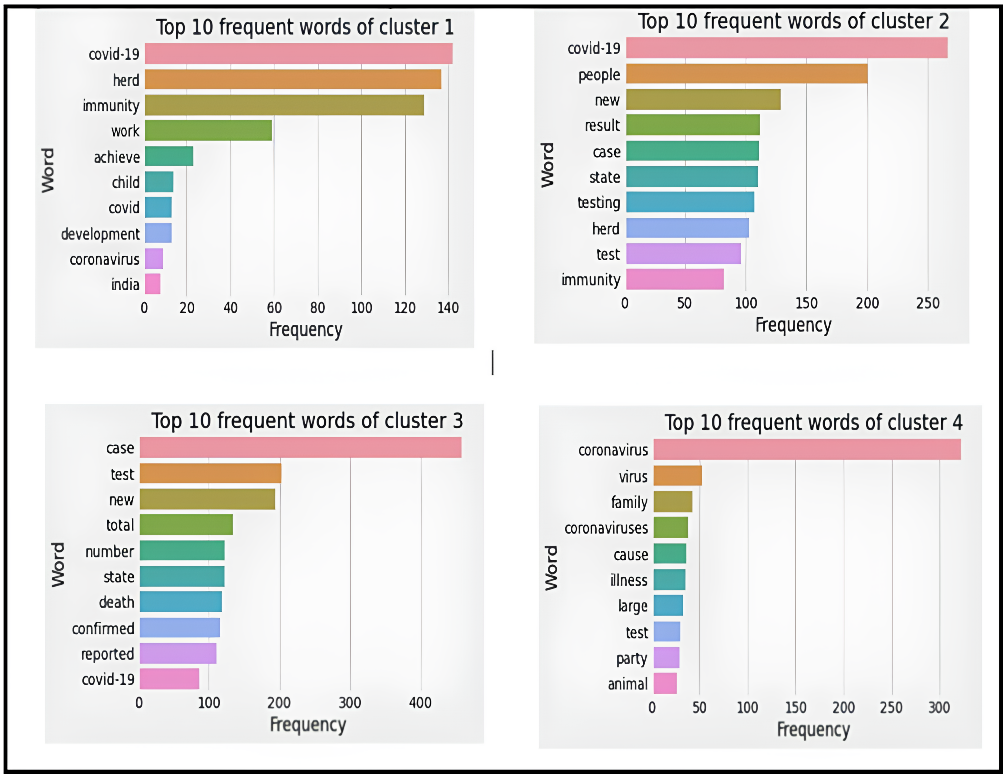


Figure 10. Top 10 frequent words drawn from the four clusters of real tweets

Cluster two of real tweets unearths discussion on Twitter related to a new COVID-19 virus strain, discovered in the United States just before October 2020, and a lot of chatter thereof about the testing. Herd immunity, a hot topic of discussion captured by cluster one, is also featured in cluster two. However, here, the central topic of discussion emerged as the people voiced out their concerns regarding herd immunity in wake of the detection of the new strain of the virus.

Tweets in the third cluster of real tweets revolved around people's concern about the amount of testing done by governments and the number of confirmed new cases and deaths as reported by the government or confirmed from other sources.

The fourth cluster brings forth an interesting topic of worldwide resentment about the restrictions being posed by the governments regarding the use of alcohol, closing of restaurants, pubs, beaches, and the imposition of night curfews, especially, around the new year. For instance, Dr. Tony Holohan, Ireland's CMO, in order to reduce alcohol consumption, recommended to the Ireland government to close gastropubs and restaurants from December 28 and issued a statement, 'This virus loves alcohol, that is a concern for us'". A lot of Irelanders tweeted their resentment. One tweet from Ireland, in this context, goes, 'I really wish they'd stop referring to covid like it's an alcoholic extroverted party animal it doesn't love alcohol and it doesn't love to party it's a virus looking for a host. . . '. Another tweet, resenting against the night curfews, goes, 'no see curfew is a great idea because the CDC has established the virus is a party animal and a night owl who only comes out to play after 8 pm.'. Yet another tweet, resenting against the closure of beaches, goes, 'And beaches are closed, but okay to walk on the pavement next to it. . . . Clever virus. . . . Be home by 21:00. . . The virus comes out after. . . real party animal this virus. . . '.

The fourth cluster of real tweets also uncovered a lot of people posting the status of their family taking ill due to coronavirus, around the new year. Also, specifically, in focus were the tweets regarding the withdrawal and later the re-joining of the fast bowler Mitchell Starc, from the #AUSvIND T20I squad, due to a family illness.

7. Conclusions and Future Directions

Due to the power of instant proliferation of ideas and news, the social media has emerged as an important platform for the people, especially, during a crisis. However, these platforms have been exploited as conduits for spreading fake information, which raises serious concerns, given the possibility of dangerous consequences.

This work proposed a single framework that uses a deep neural network classifier to automatically identify the information related to the COVID-19 pandemic, circulating on Twitter, as authentic or fake. The BERT classifier employed for this purpose achieves a high accuracy of 86.6% and an F1 score of 83.66%, which is, as discussed in the paper, comparable to the results of most of the classifiers proposed for misinformation detection. The classified tweets are subjected to a comparative analysis of the sentimental and emotional aspects, to successfully draw out conclusions, like the fake tweets spread far more negativity and fear than the real tweets. The classified tweets were clustered to retrieve topics of discussion from the authentic tweets and retrieve popular rumors, being spread on Twitter. The results of the framework were very interesting. While real tweet clusters brought to fore the major topics of discussion during the pandemic, the fake tweets were able to uncover some very popular myths being circulated on social media. All the findings of the framework were searched from the web, and corroborated. The framework, such as the one proposed in this paper, should be essential for understanding what kind of information is being circulated on the internet and developing the counter strategies, to limit the misinformation.

It will be both interesting and useful to extend the above model for classification of real and fake news and articles on the other social media platforms and utilize the generalized framework to identify and retrieve critical information from the web, especially in case of an adversity.

References

[1] C. Monica and N. Nagarathna, “Detection of fake tweets using sentiment analysis,” *SN Computer Science*, vol. 1, no. 2, pp. 1–7, 2020.

[2] S. Madichetty *et al.*, “A stacked convolutional neural network for detecting the resource tweets during a disaster,” *Multimedia tools and applications*, vol. 80, no. 3, pp. 3927–3949, 2021.

[3] E. Alothali, K. Hayawi, and H. Alashwal, “Hybrid feature selection approach to identify optimal features of profile metadata to detect social bots in twitter,” *Social Network Analysis and Mining*, vol. 11, no. 1, pp. 1–15, 2021.

[4] T. Balasubramaniam, R. Nayak, K. Luong, M. Bashar *et al.*, “Identifying covid-19 misinformation tweets and learning their spatio-temporal topic dynamics using nonnegative coupled matrix tensor factorization,” *Social Network Analysis and Mining*, vol. 11, no. 1, pp. 1–19, 2021.

[5] R. K. Kaliyar, A. Goswami, and P. Narang, “Echofaked: improving fake news detection in social media with an efficient deep neural network,” *Neural computing and applications*, vol. 33, no. 14, pp. 8597–8613, 2021.

[6] J. A. Díaz-García, C. Fernandez-Basso, M. D. Ruiz, and M. J. Martin-Bautista, “Mining text patterns over fake and real tweets,” in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, 2020, pp. 648–660.

[7] “False coronavirus claims and rumours about trump,” Oct 2020. [Online]. Available: https://www.bbc.co.uk/news/blogs-trending-54387438

[8] R. K. Kaliyar, A. Goswami, and P. Narang, “Fakebert: Fake news detection in social media with a bert-based deep learning approach,” *Multimedia tools and applications*, vol. 80, no. 8, pp. 11 765–11 788, 2021.

[9] S. Kula, M. Choraś, R. Kozik, P. Ksieniewicz, and M. Woźniak, “Sentiment analysis for fake news detection by means of neural networks,” in *International conference on computational science*. Springer, 2020, pp. 653–666.

[10] A. Zervopoulos, A. G. Alvanou, K. Bezas, A. Papamichail, M. Maragoudakis, and K. Kermanidis, “Hong kong protests: using natural language processing for fake news detection on twitter,” in *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 2020, pp. 408–419.

[11] C. Huang, “Special issue on deep learning-based neural information processing for big data analytics,” pp. 1513–1515, 2020.

[12] L. Barbaglia, S. Consoli, S. Manzan, D. Reforgiato Recupero, M. Saisana, and L. Tiozzo Pezzoli, “Data science technologies in economics and finance: A gentle walk-in,” in *Data Science for Economics and Finance*. Springer, Cham, 2021, pp. 1–17.

[13] C. Lanius, R. Weber, and W. I. MacKenzie, “Use of bot and content flags to limit the spread of misinformation among social networks: a behavior and attitude survey,” *Social Network Analysis and Mining*, vol. 11, no. 1, pp. 1–15, 2021.

[14] A. Al-Rawi, K. Grepin, X. Li, R. Morgan, C. Wenham, and J. Smith, “Investigating public discourses around gender and covid-19: a social media analysis of twitter data,” *Journal of Healthcare Informatics Research*, vol. 5, no. 3, pp. 249–269, 2021.

[15] W. Ahmad, B. Wang, H. Xu, M. Xu, and Z. Zeng, “Topics, sentiments, and emotions triggered by covid-19-related tweets from iran and turkey official news agencies,” *SN Computer Science*, vol. 2, no. 5, pp. 1–19, 2021.

[16] C.-H. Chang, M. Monselise, and C. C. Yang, “What are people concerned about during the pandemic? detecting evolving topics about covid-19 from twitter,” *Journal of healthcare informatics research*, vol. 5, no. 1, pp. 70–97, 2021.

[17] C. P. Sankar, S. V. Chandra, and K. S. Kumar, “Dynamics of semantic networks of independence day speeches,” in *2018 International CET Conference on Control, Communication, and Computing (IC4)*. IEEE, 2018, pp. 383–387.

[18] C. P. Sankar, D. A. Thumba, T. Ramamohan, S. V. Chandra, and K. S. Kumar, “Agent-based multi-edge network simulation model for knowledge diffusion through board interlocks,” *Expert Systems with Applications*, vol. 141, p. 112962, 2020.

[19] W. Ceron, G. Gruszynski Sanseverino, M.-F. de Lima-Santos, and M. G. Quiles, “Covid-19 fake news diffusion across latin america,” *Social Network Analysis and Mining*, vol. 11, no. 1, pp. 1–20, 2021.

[20] M. D. Vicario, W. Quattrociocchi, A. Scala, and F. Zollo, “Polarization and fake news: Early warning of potential misinformation targets,” *ACM Transactions on the Web (TWEB)*, vol. 13, no. 2, pp. 1–22, 2019.

[21] M. S. Dogo, P. Deepak, and A. Jurek-Loughrey, “Exploring thematic coherence in fake news,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2020, pp. 571–580.

[22] S. Helmstetter and H. Paulheim, “Collecting a large scale dataset for classifying fake news tweets using weak supervision,” *Future Internet*, vol. 13, no. 5, p. 114, 2021.

[23] X. Zhou, A. Mulay, E. Ferrara, and R. Zafarani, “Recovery: A multimodal repository for covid-19 news credibility research,” in *Proceedings of the 29th ACM international conference on information & knowledge management*, 2020, pp. 3205–3212.

[24] D. Raevan Faisal and R. Mahendra, “Two-stage classifier for covid-19 misinformation detection using bert: a study on indonesian tweets,” *arXiv e-prints*, pp. arXiv–2206, 2022.

[25] D. Kar, M. Bhardwaj, S. Samanta, and A. P. Azad, “No rumours please! a multi-indic-lingual approach for covid fake-tweet detection,” in *2021 Grace Hopper Celebration India (GHCI)*. IEEE, 2021, pp. 1–5.

[26] “29 coronavirus myths busted,” 2021. [Online]. Available: https://www.medicalnewstoday.com/articles/coronavirus-myths-explored

[27] “Covid-19 mythbusters.” [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters

[28] “Russia spreads fake news about oxford’s covid-19 vaccine; claims it turns humans into monkeys.” [Online]. Available: https://www.wionews.com/world/russia-spreads-fake-news-about-oxfords-covid-19-vaccine-claims-it-turns-humans-into-monkeys-335768

[29] “Top 10 covid-19 vaccine myths.” [Online]. Available: https://www.understandinganimalresearch.org.uk/news/top-10-covid-19-vaccine-myths

Dr. Anuja Soni is an Associate Professor in Deen Dayal Upadhyaya College, University of Delhi and has about 22 years of teaching experience. She has done her Ph.D. from University of Delhi, Delhi, India. She has published many papers in international journals and conferences. Her research interests include Software Engineering, Fuzzy Logic, Multi-Agent System, Soft Computing, and Machine Learning.

Dr. Sarabjeet Kaur Kochhar (Corresponding Author) was awarded a Ph.D. Degree in Computer Science from Department of Computer Science, University of Delhi, Delhi, India. She is an Associate Professor in the Department of Computer Science, Indraprastha College for Women, University of Delhi, Delhi, India, with over 20 years of teaching experience. She has published extensively in international journals, conferences and books. Her research interests are currently aligned along the fields of Data Mining, Data Analytics, and Natural Language Processing.

Ms. Shruti Jain is a Associate Engineer in Nagarro and has recently completed B.Sc. Hons. in Computer Science from Acharya Narendra Dev College, University of Delhi, Delhi, India. Her research interests span the fields of Data Science, Sentiment Analysis, Emotion Analysis, Deep Learning and Machine Learning.

Ms. Megha Karki is a Associate Engineer in Nagarro and has recently completed Bachelor’s degree in Computer Science from Acharya Narendra Dev College, University of Delhi, Delhi, India. Her research interests span the fields of Data Science, Sentiment Analysis, Emotion Analysis, Deep Learning and Machine Learning.

Dr. Vibha Gaur is a professor in the Department of Computer Science, Acharya Narendra Dev College, University of Delhi. She has published more than 40 papers in international journals and conferences. Her research interests include Requirement Engineering, Software Quality, Fuzzy Logic.