Chapter 11
COVID-19 Fake News on Twitter:
A Statistical and Sentimental Analysis

Yuk Ho Lu
The University of Hong Kong, Hong Kong

Chin Fung Chiu
The University of Hong Kong, Hong Kong

Dickson K. W. Chiu
https://orcid.org/0000-0002-7926-9568
The University of Hong Kong, Hong Kong

Kevin K. W. Ho
https://orcid.org/0000-0003-1304-0573
University of Tsukuba, Japan

ABSTRACT

There have been increasing concerns about spreading COVID-19 fake news and misinformation from social media sites (SNSs) such as Facebook and Twitter, as they facilitate connection and communication on a large scale. Because of the massive amount of information transmitted through SNSs, manual verification of such information is impossible, prompting the development and implementation of automated methods for fake news identification, aka automatic fact-checking. Fake news creators employ a variety of aesthetic tactics to increase their success rates, one of which is to excite the readers' sentiment. Therefore, this research uses sentiment analysis to analyze whether sentimental and emotional words in SNSs content could explain the situations between the spreading of true and fake news. In this way, governments and platform providers could take action to help the general public identify fake news and misinformation and curb them at their source. This research also offers insights to the public on the importance and impacts of sentiment words in SNS content.

DOI: 10.4018/978-1-6684-5959-1.ch011

INTRODUCTION

Nowadays, people receive information through various means. Particularly, young people rely on social and electronic media, instead of traditional media, to know what is happening worldwide (Yu et al., 2021; Ding et al., 2021; Au et al., 2022). This is further facilitated by anytime, anywhere, and ubiquitous mobile Internet access (Lam et al., 2022; Chan et al., 2022). As traditional media usually have the freedom to monitor their content and is edited by professional journalists to ensure professional and ethical standards, traditional media also serve as gatekeepers to provide accurate information to the public (McElroy, 2013; Skovsgaard & van Dalen, 2013). Nevertheless, traditional media are no longer the primary source of information, especially for the younger generation.

Today, we spend a large portion of our time on the Internet, even anywhere, anytime with mobile devices, through social networking sites (SNS), ranging from working and socializing to entertaining (Wang et al., 2021; Dong et al., 2021; Wong et al., 2022). As a result, we are exposed mainly to online information rather than physical ones (Yu et al., 2021; Wang et al., 2016). Despite the popularity of SNS, information accuracy is not guaranteed (Ho et al., 2022). Yet, there are more serious SNS applications, such as communities of practice (Lei et al., 2021) and academic SNS (Yang et al., 2022). Before release, such information does not have a review mechanism and requires SNS to formulate policies on monitoring inappropriate information, including fake news, misinformation, and hateful speech. They often rely on user reports for suspected content investigation after posting, which may be too late to stop spreading fake news and misinformation. Therefore, the public needs tools, like fact-checking mechanisms, for identifying the characteristics of fake news and misinformation fact-checking (Au et al., 2021a; 2021b; 2021c; Guo et al., 2022).

This paper uses 'misinformation' and 'fake news' interchangeably. Fake news is "the deliberate presentation of false or misleading claims as news, where the claims are misleading by design" (Gelfert, 2018, p. 108), while misinformation is released by mistake without bad intention (Ho et al., 2022; Au et al., 2021a). It aims to create a social, political, and economic bias in people's minds for others' benefit, affecting and exploiting people by making misleading information that sounds authorized (Shu et al.,2017). It may lead to a public health crisis or even a riot on the extreme (Ho et al., 2022; Au et al., 2021a; 2021c). The unexpected COVID-19 pandemic has caused an outburst of misinformation about the disease, and SNSs are essential channels for receiving, sharing, and posting information regarding the pandemic (Naeem et al., 2020; Ho et al., 2022).

There are many examples of fake news and misinformation about healthcare issues. Examples like fake cures of gargling with salt water and injecting with bleach and 5G cellular network can cause the pandemic can be easily found online (Bavel et al.,2020). Another example is former US President Trump embracing hydroxychloroquine, an anti-malarial drug against the new COVID-19, with unproven medical evidence on Twitter (Paz, 2020). Besides, specific words in headlines like "die," "death," "disaster," and "end of the world" may arouse the spreading of news as those words are life-related and induce fear in the readers (Paz, 2020). Those topics and words may lead to fake news spreading on a broader scope.

Sentiment to the true and fake news in SNSs may involve the news polarity, whether positive, negative, or neutral. The categories of news sentiment can influence human judgment and decision-making and strongly affect how the news could grab people's attention (Xie et al., 2021; Fang et al., 2021). Moreover, sentiment and emotion could trigger cognitive processing to affect online behavior, like sharing and commenting, which drives the online diffusion of fake news and promotes the public to believe in misleading information (Alonso et al., 2021). This observation motivates the investigation of the differences between true and fake news spreading through sentiment analysis (Dong et al., 2018).

For governments and society, identifying the spread of fake news could make policymakers target the adverse effects and educate the public through contemporary information literacy training (Au et al., 2021a). Otherwise, repeated exposure to fake news and misinformation may lead more readers to believe in them. For SNSs, investigating the most common topics of fake news enables the design of an early warning system for automatically detecting and stopping the misinformation from spreading. Thus, this study explores fake news characteristics to help identify possible misinformation to screen suspicious news for manual fact-checking. More importantly, this study emphasizes words and sentiment with machine learning techniques for such automation.

LITERATURE REVIEW

The Emergence of Fake News

As users consume and share news through SNSs like Facebook, Twitter, and Instagram, the general public is no longer the news receivers but also the news distributors and creators (Yu et al., 2021). Accessible to a broader range of users and audiences, SNSs provide a straightforward way for people in public to participate in discussions of social issues and express themselves freely with a large group of people (Au et al., 2021b). Moreover, entry to posting and delivering news online is much cheaper than traditional print media like newspapers, regardless of its trueness and monitoring (Lazer et al.,2018). Therefore, this trend has favored many small news publishers to publish their content on the Internet. To attract more online readers and viewers and generate more advertising revenue, they may frame some fake news to draw the public's attention. Thus, SNSs have provided opportunities and loopholes for spreading fake news to massive groups of people (Au et al., 2021a; 2021b).

Believing in Fake News

An interesting finding from past studies has shown that fake news and misinformation usually spread faster than some hot topics released by the mainstream. Moreover, most people love to read and tend to believe fake news. Elgerly and Vraga (2020) stated the content and reporting style of the news would largely affect the public to decide whether it is a piece of news or not, and hence their corresponding all-around behaviors. They have outlined four factors that shape the assessments of news: (i) what is communicated, (ii) how it is communicated, (iii) who communicates it, and (iv) where it is communicated. It reveals that the content of the fake news plays an essential role for the reader in determining whether to believe it, which is also a critical identifier for the categories of sentiment analysis. Moreover, the "continued influence effect" would make people believe fake news continuously even if the fake news has been disconfirmed (Ecker et al.,2010).

Responses and Discussion of Fake News

Regarding the news content, Naeem et al. (2020) stated the common types of misinformation regarding COVID-19, including false claims, conspiracy theories, and pseudoscientific health therapies, on how these three types of misinformation determine the responses and actions toward fake news. On the other hand, sharing acts as a formal function on SNS and has made fake news and information more widespread (Au et al., 2021b). Some key opinion leaders (KOLs) and Internet celebrities on Twitter often retweet news from mainstream news channels and add their personal opinions and emotions. Because of their popularity, users often share those postings with their connections, resulting in a more positive sentiment regarding the news. From the sharing, those opinion leaders are in the information flow (Katz & Lazarsfeld, 1995). Those people may easily gain social status and are perceived as reliable news sources. As a result, those leaders can easily drive public opinion, especially during a pandemic, when information flow is messy on SNSs (Ho et al., 2022; Au et al., 2021b).

People have started to use SNS functions to filter, access, and react to the news. For example, about 40% of Internet users love the function that they can customize their favorite news from the site. However, this influencing trend may lead to the self-selective nature of online news and accelerate audience fragmentation and polarization among users (Au et al., 2021b). The customization function in SNSs may result in one-sided information and discussion on the page. The above situation could be explained by the theory of confirmation bias raised by Nickerson (1998), stating that people tend to believe what is consistent with their beliefs and ignore other things inconsistent with their beliefs. As the pandemic continues, people will only continue suffering from fake COVID-19 news, which could induce a big public health crisis (Ho et al., 2022).

Fake News Detection

Bondielli and Marcelloni (2019) presented the qualities that have been evaluated in fake news and rumor detection approaches, offered an overview of the many methodologies used to conduct these tasks, and highlighted how gathering appropriate data to complete them is troublesome. They hypothesized that sentiment analysis techniques might be utilized to extract one of the most important semantic aspects of fake news pieces. Besides, Meel and Vishwakarma (2020) investigated how fake news content destroys web dynamics. They investigated the false information ecosystem, from classifying false information to the incentives for spreading it and the social impact and user perception. They also commented on the current fact-checking performance, including source detection, propagation dynamics, detection approaches, and containment and intervention strategies, and saw sentiment analysis as an effective way of detecting misinformation.

Zhang and Ghorbani (2020) evaluated the negative impact of fake news on the Internet and detection strategies. They pointed out the difficulties in accurate fake news identification due to SNSs' dynamic nature and complexity, diversity of online communication data, and the scarcity of high-quality training data. They also believed sentiment analysis is suitable for illustrating the emotions, attitudes, and ideas expressed on online SNSs, which is crucial for identifying suspicious accounts. Au et al. (2021c) and Ho et al. (2022) also mentioned the difficulty of conducting health fact-checks (as people need to know the domain knowledge in medical science and statistics), and sometimes, people may believe in or even create fake news due to their misinterpretations of the correct data available in the Internet. In addition, a low level of trust in governments and authorities in many countries in current society may make people more likely to believe in fake news as they believe in conspiracy theories (Ho et al., 2023).

Lastly, Sharma et al. (2019) revealed the importance of sentiment analysis in detecting fake news. They relied on computational methods to produce a list of accessible data sets for fake news identification and offer a list of obstacles and unresolved issues. They discovered that sentiment analysis was a good signal for detecting fake news, as positive sentiment terms in favorable false reviews tended to be exaggerated relative to their actual counterparts, whereas responses to fake news on SNS tended to be negative. Furthermore, Perdana and Pinandito (2018) found that non-textual characteristics, namely counts of likes and retweets, and textual characteristics were statistically significant to sentiment analysis. However, other scholars, such as Glockner et al. (2022), reported that current NLP fact-checking had its limitations for combating real-world misinformation as it depends on unrealistic assumptions about counter-evidence in the data and fact-checking uncertainty.

Research Gap and Questions

Today, there are many active users on different SNSs. For instance, as of January 2022, Twitter reported an average of 436 million active users (Statista, 2022). Many recent studies have provided insights into how SNSs have changed the ways and forms of our news consumption during the outbreak of COVID-19, which induces the spread of fake news and misinformation (Ho et al., 2022). Besides, previous studies have touched upon the state-of-the-art detection of fake news. However, regarding the research on COVID-19 news, there have been few studies on the most common type of fake news and the spreading dynamics between true and fake news, which are related to sentiment analysis as the primary focus. The following two research questions are established from our literature review and the research gap.

RQ1: What are the most common topics and types of words for identifying fake news about COVID-19 on Twitter?
RQ2: What is the major sentiment in true and fake news about COVID-19 on Twitter?

METHODS AND ANALYTICAL PROCEDURES

Data Preprocessing

Before fitting into machine learning algorithms, the data needs to be preprocessed. This research used open-access data (https://github.com/merry555/FibVID) from Kim et al. (2021). The news claims were collected from two fact-checking sites, Politifact and Snopes, from January 2020 to December 2020.

Data were labeled 0 as COVID true claims and 1 as fake. After that, claim-related tweets and retweets were extracted from Twitter. This research omitted the missing values during data cleansing, stemmed the words, removed the stopwords, converted them to lowercase, and removed all digits and punctuations. Next, two new variables were created: *nchar* to count the number of characters in the clean tweets and *sentiment* to record the sentiment of tweets predicted by deep learning.

The clean dataset yielded 157,882 news claims, including 26,831 true and 131,051 false claims. As the imbalanced dataset could affect the prediction performance of the Naive Bayes classifier, the same number of both claim types, each 5,000, were extracted randomly to facilitate the machine learning model from the clean dataset (Sample 1). Moreover, to investigate the major sentiment of true and fake news about COVID-19, 30,000 random samples were extracted from the clean dataset (Sample 2). Table 1 shows t the variables in the sample of interest.

Table 1. List of variables of interest

	Name of Variables	Explanation
Dependent variable	Group	Fake news label, either 0 (true) or 1 (fake)
Independent variables	Like_count	The number of likes of a tweet
	Retweet_count	The number of retweets of a tweet
	nchar	The count of characters of a tweet
	post_text	The text of a tweet

TF-IDF

Next, TF-IDF (Term Frequency, Inverse Document Frequency) was used to score the term importance in a document based on how frequently they appear across multiple documents. It combines two concepts, Term Frequency (TF) and Document Frequency (DF) (Abdeen et al.,2021). TF measures how frequently a word occurs in a document, which indicates the generality of a word. We divided the frequency by the total number of words in the document to normalize the value in the range of [0 to 1]. Term frequency is calculated as follows:

tf(t,d) = frequency that term t appears in document d

tf(t,d) = log(1+raw counts that term t appears in document d)

Document Frequency (DF) is similar to TF in that it gauges the relevance of a document in the context of the entire corpus. The only difference is that TF is the frequency counter for a term t in document d, whereas DF is the count of occurrences of term t in the document set D. We count a term as one occurrence if it appears in the document at least once; we don't need to know how many times it appears.

In Inverse Document Frequency (IDF), IDF is the inverse of document frequency, which indicates how informative a term *t* is. The most often occurring words, such as stop words, will have a very low value for IDF computation. There are a few other issues with the IDF. For example, if the corpus is large enough, say 10,000, the IDF value explodes, requiring logarithms to overcome the problem. There may be no documents with zero occurrences in the worst-case scenario. To avoid division by 0, we can add 1 to the denominator to smooth out the effect (Abdeen et al.,2021). Finally, the TF-IDF score can take a multiplicative value of TF and IDF.

Classification Techniques

Two machine learning algorithms were used to identify the true and fake news from the datasets. TF-IDF vectorizer was used to convert data into vectors. After feature selection, the result was passed into the Multinomial Naive Bayes model to evaluate feature importance and apply the random forest method as an evaluation benchmark. These techniques could analyze important textual characteristics of identifying true versus fake news.

Splitting the training and testing datasets is required to run machine learning models. 80% of the data was used for training the models to optimize the parameters, and the remaining 20% was used to predict the labels and test their accuracy. Naive Bayes methods are a set of supervised learning algorithms. The basis of these methods is Bayes' Theorem, the probability of an event based on past knowledge of factors that may be associated with the event (Puga et al., 2015). A prior for the probability of each class was counted by the relative frequency of class labels in the dataset. This research adopted one of the Naive Bayes methods, the Multinomial Naive Bayes, a multinomially distributed data implementation commonly used in text classification, in which the data is typically represented as word vector counts and TF-IDF vectors. The distribution is parametrized by vectors $\theta_y = \left(\theta_{y1}, \cdots, \theta_{yn}\right)$ for each class y, where n is the number of features (in text classification, the size of the vocabulary) and θ_{yi} and is the probability $P(x_i \mid y)$ of feature i appearing in a sample belonging to the class y. A smoothed variant of maximum likelihood, i.e., relative frequency counting, is used to estimate the parameters θ_y:

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n}$$

where N_{yi} is the number of times feature i appears in a sample of class y in the training set, and N_y is the total count of all features for the class. Smoothing priors α accommodate for features that aren't present in the learning samples and prevent zero probabilities in subsequent calculations. The logarithm of the probability of the features can be utilized to evaluate the informative features, which is the feature importance.

Random Forest

Our data is in the context of high-dimensional, and thus the random forest model can serve as a benchmark and a standard model for classification (Couronné et al., 2018). Firstly, for the model fitting part, Random Forest would first do a resampling called bootstrapping (Wang et al., 2020). It refers to the process of taking a random sample of the training data, but the sample size is the same as that of the training set but allows duplication. The random forest will have more than one decision tree for the number of trees. Users define the number of trees. When considering each split, we select the square root of the total number of features after bootstrapping. First, the Gina impurity scores of all variables can be calculated to find the best split. If the node has the lowest scores, it becomes a leaf node. If separating the data improves the result, the separation with the lowest impurity value is picked.

For the prediction, the observation follows the split till reaching the terminal node to know which label it belongs to. The random Forest method allows the observation to pass through each tree and get a series of labels. In the series, which label is the majority becomes the label of that observation.

Sentiment Analysis

Sentiment Analysis (SA) is a branch of Natural Language Processing (NLP) responsible for the development and implementation of models and techniques for determining whether a text contains objective or subjective information and, in the latter case, whether that information is expressed in a positive, neutral, or negative manner, as well as whether it is expressed strongly or weakly (Iwendi et al., 2022)

Deep learning is about learning the knowledge chunk in multiple levels of representations and abstraction to make up higher-level information from lower-level information, such as sound and images (Shin et al.,2020). The *Transformers* package generates the variable, which employs a deep learning approach to predict the sentiment of the text (Vaswani, 2017) for generating the sentiment variable.

This part has several expected outcomes from the analytical procedure. First, Naive Bayes's accuracy should be similar to a random forest. Second, the precision score is around 80%. Thirdly, the Naive Bayes score is around 80%. Lastly, false news claims are mainly negative in sentiment, while positive news claims are mainly positive. The performance of machine learning models was evaluated using the following criteria.

Datasets were initially classified into groups. Machine learning algorithms first learned the parameter of features to predict the group. The four classification prediction results were true positive, true negative, false positive, and false negative. The following equation was used to compute the precision measure.

$$\text{Precision} = \text{TruePositive} / (\text{TruePositives} + \text{FalsePositives})$$

The precision score was computed on the prediction result, valued between 0.0 for no precision and 1.0 for perfect precision.

Despite the rigorous approach of deep learning, the sentiment of tweets is just a prediction; thus, the performance requires evaluation. According to the literature review, *like_count* and *retweet_count* are statistically significant for fake news classification. Therefore, p-value and logistic regression can evaluate the significance. The sentiment is accurate enough if non-textual characteristics are statistically significant in analyzing sentiments.

RESULTS

Descriptive Statistics and Basic

Table 2. Descriptive Statistics for all dependent variables expect post_text in Sample 1

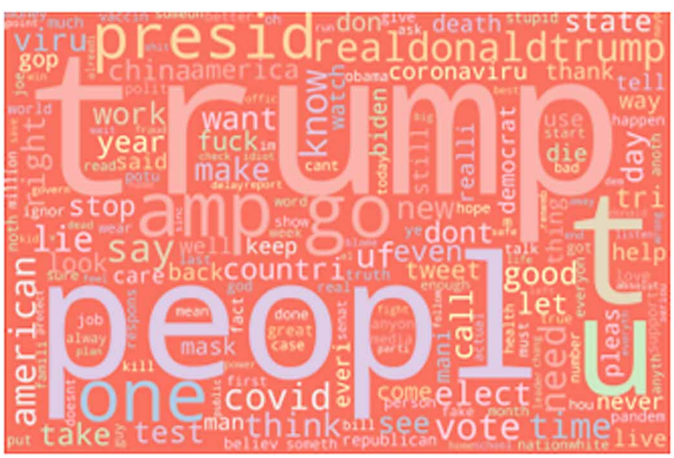
Variables	n	Minimum	Maximum	Mean	Standard Deviation
Like_count	10,000	0	293,731	302.261	5474.752
Retweet_count	10,000	0	54,472	73.99	1226.822
nchar	10,000	5	280	74.459	53.570

Table 3. Descriptive Statistics for all dependent variables expect post_text in Sample 2

Variables	n	Minimum	Maximum	Mean	Standard Deviation
Like_count	30,000	0	993,167	412.730	9168.278
Retweet_count	30,000	0	234,387	101.735	2202.435

Tables 2 and 3 show the descriptive statistics. Figure 1 is a word cloud as a visual representation of the word prominence from the fake news samples. The greater prominence of words, the greater their size is shown in the word cloud.

Figure 1. Word cloud of fake news text after cleansing



The precision score of the Naive Bayes classifier is 76.905% (correct to 3 decimal places), higher than that of the random forest of the standard classification model, 72.152%, indicating that the trained parameters of the Naive Bayes classifier are convincing in discussing the feature importance. Likewise, Agarwal et al. (2019) conducted a study on the accuracy of machine learning classifiers in detecting fake news and found that the precision score of the Naive Bayes classifier and random forest classifier was approximately 64% and 60%, respectively. Nevertheless, a more recent study found a much higher precision score of approximately 87% concerning the Naive Bayes classifier in fake news detection (Pandey, 2022). Therefore, a significant difference regarding the precision score can be observed between the existing literature and our research.

One possible explanation is that the length of tweets is likely to negatively correlate with fake news detection accuracy. This is consistent with a study on detecting COVID-19 fake news that when the length of tweets was over 70, the accuracy would be significantly reduced from approximately 91.43% to 57.14% (Samuel et al., 2020). Similarly, our sample's average length of tweets is 74.459, indicating our dataset contains relatively longer tweets (see Tables 2 and 3). Thus, the accuracy discount of the Naive Bayes classifier on COVID-19 tweets is probably reasonable and necessary. Therefore, we can apply the trained parameters of the Naive Bayes classifier to evaluate the feature importance to discuss the textual characteristics of identifying fake news.

DISCUSSION

As mentioned in the Methods section, the logarithmic probability can show the features' importance. We observed that words with disapproving tones are informative in identifying fake news, such as "imposter" and "hypocrisies." The logarithmic probabilities of these two words are at the top, among other features, indicating that the words with disapproving tones are likely to be the common type of words in fake news. One possible explanation is that fake news authors hope to arouse people's awareness with their tweets. Our findings resemble studies about the textual characteristics of fake news, in which fake news mainly contains negative valence to attract more replies and attention (Brummette et al.,2018, Tandoc et al., 2018). It implies that fake news probably spread negative sentiments on SNSs amid the COVID-19 pandemic.

The Naive Bayes classifier shows words like "Trump," " President," "Donald Trump," "COVID," and "virus" are often presented in the headlines or content of the fake news. One possible explanation is that people tend to believe the news from representative figures like President Trump as he represents authority. As the president represents the whole government, most people would not fact-check the news content but believe and spread the news. This is consistent with Wang and Huang's (2021) finding that people tend to believe the news that the government publishes. Also, Murayama et al. (2021) stated that SNS influencers, such as politicians and government officials, may inspire others to use the term "fake news" to question the opposition and support thoughts and viewpoints similar to their ideology.

Moreover, citizens may decrease their belief in the government once they find the government released fake news and decrease their belief in similar government news. As a result, people's overall contentment with the government in the relevant subject area may be badly affected by such an encounter with the government's lack of legitimacy. In other words, the public's awareness of the government's lack of credibility has an immediate effect on the general public and continues to impact people's attitudes toward the government in the future (Wang & Huang, 2021).

Besides, the Naive Bayes classifier result also predicts that fake news is shared and spread quickly on popular Twitter accounts like the former US President Donald Trump with some arousing topics like lies, death figures, and Americans, which induce an intense discussion, likes, and retweets. As a result, tweets from Trump about COVID-19 news spread widely through Twitter, and people easily believe in such fake news. This aligns with a study on fake news on Twitter by Brummette et al. (2018), stating that fake news discussions take place in big clusters (i.e., online communities) dominated by members of the general public, as opposed to the media, non-governmental organizations, or politicians. Indeed, when joining a community on Twitter, a user can tweet directly to the community rather than to all followers. People love to join different big communities to view and retweet posts they like and believe. As a result, people would believe the fake news retweeted largely from online communities, resulting in echo-chamber effects (Au et al., 2021b).

Logistic regression indicated significant results for like_count and retweet_count, with $p < 0.05$, implying that sentiments predicted by the deep learning model are likely to resemble human analysis and are largely accurate. Sample 2 is a random sample of the clean dataset; thus, the major news is also fake. One possible explanation is that negative sentiment is more likely to draw attention and increase the number of counts and retweets. Our findings are similar to Perdana and Pinandito (2018), predicting sentiment by likes, retweet counts, and textual characteristics.

Notably, the result indicated that both true and fake news sentiment scores are negative. A possible explanation is that the negative atmosphere has been aroused because of the seriousness of the COVID-19 pandemic, causing widespread lockdowns and affecting people's life worldwide (Huang et al., 2021; 2022; Chan et al., 2022; Yu, Chiu, & Chan, 2023; Yu, Lam, & Chiu, 2022; Meng et al., 2022).

Besides, the negative sentiment ratio in fake news is slightly higher than in true news, while the ratio of positive sentiment in true news is slightly higher than that of fake news. One of the possible explanations is that fake news sustains more negative and emotional words that exaggerate the seriousness, aiming to attract more people to discuss and share false information on prominent SNSs. The above result can be explained by Ding et al.'s (2020) result that negative sentiment is higher in fake news than true news, and fake news tweets may include more emotive content to attract readers' attention, i.e., more negative content. In contrast to fake news, true news mainly reports facts and actual situations about the pandemic, in which the use of words and headings is slightly milder than fake news.

CONCLUSION

Today, we spend much time on SNS and rely on them to receive information. Therefore, a post can be very influential and crucial to our decision-making and daily lives. COVID-19 is a public health issue; misinformation should be effectively resolved, and people should be aware of the characteristics. This research has revealed negative valence, and some controversial figures are probably common topics and words of fake news. Furthermore, negative sentiments are prominent in both true and fake news and should be noted. The public should be more careful when observing such characteristics on SNSs and fact-check to determine information accuracy to combat the COVID-19 pandemic eventually.

One notation limitation in this research is that the precision score is relatively low compared to other existing literature. Further research can combine textual and non-textual features into the training dataset to improve the precision score. Secondly, sentiment analysis may contain inaccurate results using machine learning methods, which may misclassify the sentiments compared to human analysis. Only two features and one machine learning algorithm are applied in this research. Therefore, further research can focus on using predicted sentiment in other machine learning algorithms.

REFERENCES

Abdeen, M. A., Hamed, A. A., & Wu, X. (2021). Fighting the COVID-19 infodemic in news articles and false publications: The NeoNet text classifier, a supervised machine learning algorithm. *Applied Sciences (Basel, Switzerland)*, *11*(16), 7265. doi:10.3390/app11167265

Agarwal, V., Sultana, H. P., Malhotra, S., & Sarkar, A. (2019). Analysis of Classifiers for Fake News Detection. *Procedia Computer Science*, *165*, 377–383. doi:10.1016/j.procs.2020.01.035

Alonso, M.A., Vilares, D., Gómez-Rodríguez, C., & Vilares, J. (2021). Sentiment analysis for fake news detection. *Electronics 10*.

Au, C. H., Ho, K. K., & Chiu, D. K. W. (2022). Managing users' behaviors on open content crowdsourcing platform. *Journal of Computer Information Systems*, *62*(6), 1125–1135. doi:10.1080/08874417.2021.1983487

Au, C. H., Ho, K. K. W., & Chiu, D. K. W. (2021a). Does political extremity harm the ability to identify online information validity? Testing the impact of polarisation. *Government Information Quarterly*, *38*(4), 101602. doi:10.1016/j.giq.2021.101602

Au, C. H., Ho, K. K. W., & Chiu, D. K. W. (2021b). (in press). The role of online misinformation and fake news in ideological polarization: Barriers, catalysts, and implication. *Information Systems Frontiers*. Advance online publication. doi:10.100710796-021-10133-9

Au, C. H., Ho, K. K. W., & Chiu, D. K. W. (2021c). Stopping healthcare misinformation: The effect of financial incentives and legislation. *Health Policy (Amsterdam)*, *125*(5), 627–633. doi:10.1016/j.healthpol.2021.02.010 PMID:33712320

Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., & Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*(5), 460–471. doi:10.103841562-020-0884-z PMID:32355299

Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, *497*, 38–55. doi:10.1016/j.ins.2019.05.035

Brummette, J., DiStaso, M., Vafeiadis, M., & Messner, M. (2018). Read all about it: The politicization of "fake news" on Twitter. *Journalism & Mass Communication Quarterly*, *95*(2), 497–517. doi:10.1177/1077699018769906

Chan, V. H. Y., Ho, K. K. W., & Chiu, D. K. W. (2022). Mediating effects on the relationship between perceived service quality and public library app loyalty during the COVID-19 era. *Journal of Retailing and Consumer Services*, *67*, 102960. doi:10.1016/j.jretconser.2022.102960

Cheng, W. W. H., Lam, E. T. H., & Chiu, D. K. W. (2020). Social media as a platform in academic library marketing: A comparative study. *Journal of Academic Librarianship*, *46*(5), 102188. doi:10.1016/j.acalib.2020.102188

Couronné, R., Probst, P., & Boulesteix, A.L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics, 19*(1).

Ding, L., Ding, L., & Sinnott, R. (2020). Fake News Classification of Social Media Through Sentiment Analysis. In *Big Data – BigData 2020* (pp. 52–67). Lecture Notes in Computer Science. Springer International Publishing. doi:10.1007/978-3-030-59612-5_5

Ding, S. J., Lam, E. T. H., Chiu, D. K. W., Lung, M. M., & Ho, K. K. W. (2021). Changes in reading behavior of periodicals on mobile devices: A comparative study. *Journal of Librarianship and Information Science*, *53*(2), 233–244. doi:10.1177/0961000620938119

Dong, G., Chiu, D. K. W., Huang, P.-S., Lung, M. M., Ho, K. K. W., & Geng, Y. (2021). Relationships between research supervisors and students from coursework-based master's degrees: Information usage under social media. *Information Discovery and Delivery*, *49*(4), 319–327. doi:10.1108/IDD-08-2020-0100

Dong, L. Y., Ji, S. J., Zhang, C. J., Zhang, Q., Chiu, D. W., Qiu, L. Q., & Li, D. (2018). An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews. *Expert Systems with Applications*, *114*, 210–223. doi:10.1016/j.eswa.2018.07.005

Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, *38*(8), 1087–1100. doi:10.3758/MC.38.8.1087 PMID:21156872

Edgerly, S., & Vraga, E. K. (2020). Deciding what's news: News-ness as an audience concept for the hybrid media environment. *Journalism & Mass Communication Quarterly*, *97*(2), 416–434. doi:10.1177/1077699020916808

Fang, J., Chiu, D. K. W., & Ho, K. K. W. (2021). Exploring cryptocurrency sentimental with clustering text mining on social media. In Z. Sun (Ed.), *Handbook of Research on Intelligent Analytics with Multi-Industry Applications* (pp. 157–171). IGI Global. doi:10.4018/978-1-7998-4963-6.ch007

Gelfert, A. (2018). Fake news: A definition. *Informal Logic*, *38*(1), 84–117. doi:10.22329/il.v38i1.5068

Glockner, M., Hou, Y., & Gurevych, I. (2022). Missing counter-evidence renders NLP fact-checking unrealistic for misinformation. Cornell University. doi:10.48550/arXiv.2210.13865

Guo, J., Ji, S., Cao, N., Chiu, D. K. W., Su, N., & Zhang, Z. (2022). (in press). MDG: Fusion learning of the maximal diffusion, deep propagation and global structure features of fake news. *Expert Systems with Applications*.

Ho, K. K. W., Chan, J. Y., & Chiu, D. K. W. (2022). Fake news and misinformation during the pandemic: What we know, and what we don't know. *IT Professional*, *24*(2), 18–24. doi:10.1109/MITP.2022.3142814

Ho, K. K. W., Chiu, D. K. W., & Sayama, K. L. C. (2023). (in press). When privacy, distrust, and misinformation cause worry about using COVID-19 contact-tracing apps. *IEEE Internet Computing*. Advance online publication. doi:10.1109/MIC.2022.3225568

Huang, P. S., Paulino, Y., So, S., Chiu, D. K. W., & Ho, K. K. W. (2021). Special Issue Editorial - COVID-19 Pandemic and Health Informatics (Part 1). *Library Hi Tech*, *39*(3), 693–695. doi:10.1108/LHT-09-2021-324

Huang, P.-S., Paulino, Y. C., So, S., Chiu, D. K. W., & Ho, K. K. W. (2022). Guest editorial: COVID-19 Pandemic and Health Informatics Part 2. *Library Hi Tech*, *40*(2), 281–285. doi:10.1108/LHT-04-2022-447

Iwendi, C., Mohan, S., Khan, S., Ibeke, E., Ahmadian, A., & Ciano, T. (2022). Covid-19 fake news sentiment analysis. *Computers & Electrical Engineering*, *101*, 107967. doi:10.1016/j.compeleceng.2022.107967 PMID:35474674

Katz, E., & Lazarsfeld, P. F. (1995). Between media and mass/the part played by people/the two-step flow of communication. In O. Boyd-Barrett & C. Newbold (Eds.), *Approaches to Media* (pp. 124–134). Arnold.

Kim, J., Aum, J., Lee, S., Jang, Y., Park, E., & Choi, D. (2021). FibVID: Comprehensive fake news diffusion dataset during the COVID-19 period. *Telematics and Informatics*, *64*, 101688. doi:10.1016/j.tele.2021.101688 PMID:36567815

Lam, A. H. C., Ho, K. K. W., & Chiu, D. K. W. (2022). (in press). Instagram for student learning and library promotions? A quantitative study using the 5E Instructional Model. *Aslib Journal of Information Management*. doi:10.1108/AJIM-12-2021-0389

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., & Schudson, M. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. doi:10.1126cience.aao2998 PMID:29590025

Lei, S. Y., Chiu, D. K. W., Lung, M. M., & Chan, C. T. (2021). Exploring the Aids of Social Media for Musical Instrument Education. *International Journal of Music Education*, *39*(2), 187–201. doi:10.1177/0255761420986217

McElroy, K. (2013). Where old (Gatekeepers) meets new (media): Herding reader comments into print. *Journalism Practice*, *7*(6), 755–771. doi:10.1080/17512786.2013.774117

Meel, P., & Vishwakarma, D. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Systems with Applications*, *153*, 112986. doi:10.1016/j.eswa.2019.112986

Meng, Y., Chu, M. Y., & Chiu, D. K. (2022). (in press). The impact of COVID-19 on museums in the digital era: Practices and challenges in Hong Kong. *Library Hi Tech*. doi:10.1108/LHT-05-2022-0273

Murayama, T., Wakamiya, S., Aramaki, E., & Kobayashi, R. (2021). Modeling the spread of fake news on Twitter. *PLoS One*, *16*(4), E0250419. doi:10.1371/journal.pone.0250419 PMID:33886665

Naeem, S., Bhatti, R., & Khan, A. (2021). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information and Libraries Journal*, *38*(2), 143–149. doi:10.1111/hir.12320 PMID:32657000

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. doi:10.1037/1089-2680.2.2.175

Pandey, S., Prabhakaran, S., Reddy, N. S., & Acharya, D. (2022). Fake news detection from online media using machine learning classifiers. []. IOP Publishing.]. *Journal of Physics: Conference Series*, *2161*(1), 012027. doi:10.1088/1742-6596/2161/1/012027

Paz, C. (2020 Oct 2). All the President's Lies about the Coronavirus. *The Altanic*. https://www.theatlantic.com/politics/archive/2020/11/trumps-lies-about-coronavirus/608647/

Perdana, R.S., & Pinandito, A. (2018). Combining likes-retweet analysis and naive bayes classifier within Twitter for sentiment analysis. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10*(1-8), 41-46.

Puga, J. L., Krzywinski, M., & Altman, N. (2015). Bayes' theorem. *Nature Methods*, *12*(4), 277–278. doi:10.1038/nmeth.3335 PMID:26005726

Samuel, J., Ali, G. G. M. N., Rahman, M. M., Esawi, E., & Samuel, Y. (2020). COVID-19 public sentiment insights and machine learning for Tweets classification. *Information (Basel)*, *11*(6), 314. doi:10.3390/info11060314

Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., & Liu, Y. (2019). Combating fake news. *ACM Transactions on Intelligent Systems and Technology*, *10*(3), 1–42. doi:10.1145/3305260

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K.-C. (2020). Enhancing social media analysis with visual data analytics: A deep learning approach. *Management Information Systems Quarterly*, *44*(4), 1459–1492. doi:10.25300/MISQ/2020/14870

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *SIGKDD Explorations*, *19*(1), 22–36. doi:10.1145/3137597.3137600

Skovsgaard, M., & van Dalen, A. (2013). Dodging the Gatekeepers?: Social media in the campaign mix during the 2011 Danish elections. *Information Communication and Society*, *16*(5), 737–756. doi:10.1080/1369118X.2013.783876

Statistica (2022). Most popular social networks worldwide as of January 2022, ranked by number of monthly active users. *Statistica*. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Stoltzfus, J. C. (2011). Logistic Regression: A Brief Primer. *Academic Emergency Medicine*, *18*(10), 1099–1104. doi:10.1111/j.1553-2712.2011.01185.x PMID:21996075

Tandoc, E.C., Lim, Z.W., & Ling, R. (2018). Defining "Fake News": A typology of scholarly definitions. *Digital journalism, 6*(2), 137-153.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. Cornell University.

Wang, C., & Huang, H. (2021). When "fake news" becomes real: The consequences of false government denials in an authoritarian country. *Comparative Political Studies*, *54*(5), 753–778. doi:10.1177/0010414020957672

Wang, P., Chiu, D. K. W., Ho, K. K. W., & Lo, P. (2016). Why read it on your mobile device? Change in reading habit of electronic magazines for university students. *Journal of Academic Librarianship*, *42*(6), 664–669. doi:10.1016/j.acalib.2016.08.007

Wang, Q., Zhou, Y., Ding, W., Zhang, Z., Muhammad, K., & Cao, Z. (2020). Random forest with self-paced bootstrap learning in lung cancer prognosis. *ACM Transactions on Multimedia Computing Communications and Applications*, *16*(1s), 34. doi:10.1145/3345314

Wang, W., Lam, E. T. H., Chiu, D. K. W., Lung, M. M., & Ho, K. K. W. (2021). Supporting higher education with social networks: Trust and privacy vs. perceived effectiveness. *Online Information Review*, *45*(1), 207–219. doi:10.1108/OIR-02-2020-0042

Wong, J., Chiu, D. K. W., Leung, T. N., & Ho, K. K. W. (2022). (in press). Exploring the associations of addiction as a motive for using Facebook with social capital perceptions. *Online Information Review*. doi:10.1108/OIR-06-2021-0300

Xie, R., Chu, S. K. W., Chiu, D. K. W., & Wang, Y. (2021). Exploring public response to COVID-19 on Weibo with LDA topic modeling and sentiment analysis. *Data and Information Management*, *5*(1), 86–99. doi:10.2478/dim-2020-0023 PMID:35402850

Yang, Z., Zhou, Q., Chiu, D. K. W., & Wang, Y. (2022). Exploring the factors influencing continuance intention to use academic social network sites. *Online Information Review*, *46*(7), 1225–1241. doi:10.1108/OIR-01-2021-0015

Yu, H. H. K., Chiu, D. K. W., & Chan, C. T. (2023). Resilience of symphony orchestras to challenges in the COVID-19 era: Analyzing the Hong Kong Philharmonic Orchestra with Porter's five force model. In W. Aloulou (Ed.), *Handbook of Research on Entrepreneurship and Organizational Resilience During Unprecedented Times* (pp. 586–601). IGI Global.

Yu, H. Y., Tsoi, Y. Y., Rhim, A. H. R., Chiu, D. K. W., & Lung, M. M. W. (2021). (in press). Changes in habits of electronic news usage on mobile devices in university students: A comparative survey. *Library Hi Tech*. doi:10.1108/LHT-03-2021-0085

Yu, P. Y., Lam, E. T. H., & Chiu, D. K. W. (2022). (in press). Operation management of academic libraries in Hong Kong under COVID-19. *Library Hi Tech*. doi:10.1108/LHT-10-2021-0342

Zhang, X., & Ghorbani, A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, *57*(2), 102025. doi:10.1016/j.ipm.2019.03.004