



## Research Article

Md Enamul Kabir\*

# Topic and sentiment analysis of responses to Muslim clerics' misinformation correction about COVID-19 vaccine: Comparison of three machine learning models

<https://doi.org/10.1515/omgc-2022-0042>

Received May 16, 2022; accepted August 27, 2022

### Abstract

**Purpose:** The purpose of this research was to use develop a sentiment model using machine learning algorithms for discerning public response about the misinformation correction practices of Muslim clerics on YouTube.

**Method:** This study employed three machine learning algorithms, Naïve Bayes, SVM, and a Balanced Random Forest to build a sentiment model that can detect Muslim sentiment about Muslim clerics' anti-misinformation campaign on YouTube. Overall, 9701 comments were collected. An LDA-based topic model was also employed to understand the most expressed topics in the YouTube comments.

**Results:** The confusion matrix and accuracy score assessment revealed that the balanced random forest-based model demonstrated the best performance. Overall, the sentiment analysis discovered that 74 percent of the comments were negative, and 26 percent were positive. An LDA-based topic model also revealed the eight most discussed topics associated with ten keywords in those YouTube comments.

**Practical implications:** The sentiment and topic model from this study will particularly help public health professionals and researchers to better understand the nature of vaccine misinformation and hesitancy in the Muslim communities.

**Social implications:** This study offers the joint task force of Muslim clerics and medical professionals, and the future misinformation campaigns a sentiment detection model to understand public attitude to such practices on social media.

---

**Article Note:** This article underwent double-blind peer review.

---

\*Corresponding author: **Md Enamul Kabir**, Bowling Green State University, Bowling Green, OH 43402, USA, E-mail: [mkabir@bgsu.edu](mailto:mkabir@bgsu.edu). <https://orcid.org/0000-0003-3810-4000>

**Originality:** While the impact of misinformation on public sentiment and opinion on social media has been researched extensively, Muslim perspectives on combating misinformation have received less attention. This research is the first to evaluate responses towards Muslim clerics correcting religious vaccine misinformation using machine learning models.

**Keywords:** Balanced Random Forest; computational method; COVID-19 vaccine misinformation; machine learning; misinformation correction; muslim clerics; sentiment analysis; topic model

## 1 Introduction

The study of public response to misinformation correction and debunking is a relatively new development in the field of vaccine misinformation research. This issue is embedded in the ongoing discussion on strategies for correcting misinformation. Traditionally, medical professionals and researchers held an assumption that debunking faith-based misinformation about the COVID vaccination can be readily accomplished by sharing accurate information. However, recent studies showed that this strategy often leads to public responses that are rather negative and adamant about the original misinformation (Ecker et al. 2022; Larson and Broniatowski 2021; Walter and Riva 2020).

The case of faith-based rhetoric has been a particularly significant contributor to people's persistence with misinformation (Alimardani and Elswah 2020). To such a length that recent studies recognized faith-based misinformation as a global concern on social media such as Twitter, Facebook, Instagram, etc (Wardle and Singerman 2021). Despite the various misinformation correction campaigns on social media, misinformation reportedly persisted and led to vaccine hesitancy and resistance (Kait 2020; Stecula et al. 2020). Such hesitancy and resistance appeared in the expression of anti-vaxxer groups as well as ordinary people on social media platforms (Kanozia and Arya 2021). In the United States, for example, millions of white evangelical Christians believed that the COVID-19 vaccine was manufactured using the tissues of aborted cells (Pew Research Center 2021). Some Muslim communities feared that the COVID-19 vaccine was merely a conspiracy against Muslims around the world, which could cause miscarriage, infertility, and DNA damage (Abbas et al. 2021; Khan et al. 2020). Thus, some Muslim clerics and institutions joined the practice of correcting misinformation using religious counterspeech and social media engagement (Syed and Wajid 2021). By interpreting Islamic texts, for instance, these clerics attempted to clarify the Islamic law on vaccines. Some of these clergymen even released videos of themselves receiving

the COVID vaccine and detailing their experiences. Numerous religious leaders and Muslim health professionals have even established a “National Task Force on COVID-19” for this reason. Along the line, studies showed that, in addition to medical professionals, Muslim clerics' engagement was a rare and crucial step in correcting vaccine misinformation as they wield a great deal of influence within the Muslim community (Arief and Karlina 2019). In addition, Muslim clerics have a deeper understanding of vaccination intake from an Islamic standpoint. The efficacy of Muslim clergies' scholarly and influential misinformation correction campaign can be called into question in light of recent evidence showing that similar methods typically fail to elicit favorable responses from the public in other situations (Walter and Riva 2020). Due to this, the current work is highly pertinent to the field of misinformation correction studies.

This study explored Muslim sentiment and response to Muslim Clerics' misinformation correction messages and performed a sentiment analysis. Muslim sentiment is extremely relevant to vaccine misinformation research since this religion has the most rigorous rulings on the consumption of food and medicine. For example, pork meat and alcohol are strictly prohibited in Muslim culture. While finding how people's sentiment toward misinformation correction messages vary is necessary, it is quite impossible to do so without a large-scale experiment or survey. Even in that case, the survey would require quite a large sample and spontaneous participation. As a result, it would be quite difficult to examine the public response and sentiment towards the misinformation correction campaign about the COVID vaccine. However, the thousands of comments on YouTube videos of misinformation correction are a useful resource to analyze those sentiments. To carry out the sentiment analysis, this study employed three machine learning algorithms (i.e., Naive Bayes, Support Vector Machine and Random Forest) in conjunction with a Balanced Random Forest model, which is particularly tailored for imbalanced datasets. The study's goals were to (1) examine people's sentiments and responses to the misinformation correction campaign led by Muslim clerics and (2) compare the accuracy of different machine learning models in detecting the sentiments in responses to misinformation correction campaigns.

## 2 Literature review

### 2.1 Continued influence effect (CIE) of misinformation theory

The persistence of misinformation is influenced by several cognitive, social, and emotional factors in addition to a lack of access to high-quality information. A common belief is that correcting misinformation simply requires supplying the

appropriate information. Misinformation, however, may often continue to persist in people's minds even after they receive a correction (Lewandowsky et al. 2012). This persistence with misinformation despite correcting efforts is the core concept of the continued influence effect of misinformation theory (CIE) (Ecker et al. 2022). The socio-affective factor is an important aspect of CIE theory which suggests that persistence will continue to triumph if the misinformation correction threatens a person's belief system (e.g., religious beliefs) (Ecker and Chang 2019). This indicates the intensity of faith-based misinformation, especially because emotional factors are also associated with CIE effects. It is possible for misinformation corrections to cause people psychological anguish, prompting the public to ignore the correction in an effort to alleviate the emotional distress they experience. Misinformation that incites negative emotions, such as fear or anger, may be more effective at causing a continued influence (Susmann and Duane 2021). In this study, the design and theoretical approach of the continued influence effect is adapted to examine how the general public responds to Muslim clergies' efforts to correct misinformation.

## 2.2 Correcting faith-based misinformation

Misinformation is defined as incorrect or misleading information presented as fact, either intentionally or unintentionally (Merriam-Webster Dictionary 2022). Faith-based misinformation is a special case of misinformation practice where subjects believe the misinformation message they received and are concerned that it could be detrimental and offensive to their religious culture and practice. According to Alimardani and Elswah (2020), faith-based misinformation is characterized by a lack of discernment, a misinterpretation of religious texts or records, or a claim of divine knowledge or power. Fear, emotion, or religious authority are generally used to persuade the recipients of religious misinformation. Faith-based misinformation, in contrast to more standard forms of misinformation research, presents unique challenges and demands an in-depth familiarity with the targeted religion and its sociopolitical setting. As an example, some Hindus were outraged as they thought cow meat was used in the production of the COVID-19 vaccine, whereas other Muslims were hesitant to get vaccinated because they suspected pork was used (Abbas et al. 2021; Etutu and Goodman 2021). The Muslim cultural position on pork is unique from the Hindu cultural perspective on cow meat, and this difference needs to be highlighted. Muslims were outraged because, according to their faith's teachings, consuming pork is a grievous sin. Alternatively, Hindus do not eat cow meat because cows are held with high regard in their community. Abbas et al. (2021) also discovered that one of the major pieces of misinformation regarding the COVID-19 vaccine that shocked Muslim communities was that the vaccine was designed to make Muslims

impotent. Considering CIE theory's emotional and faith-related factors, the fact that these misinformation beliefs or conspiracy theories are grounded in people's faith and values made the correction process sensitive and potentially threatening (Susmann and Duane 2021). The most detrimental effect of this entails ultimately resulting in vaccine denials or vaccine hesitancy (Wardle and Singerman 2021). Studies showed that people who received vaccine-related information on social media were more likely to be exposed to vaccine misinformation and, in turn, become vaccine-hesitant (Stecula et al. 2020). Given the inefficiency of the social media companies in halting the spread of vaccine misinformation on their platforms, Wardle and Singerman (2021) suggested that social media users now hold a crucial role in the battling front of misinformation correction, especially because the users are at the epicenter of the continued influence effect of misinformation to a great extent (Walter and Riva 2020).

Various practices of misinformation correction were observed throughout the pandemic. Previous studies extended several novel approaches that are more suited for correcting faith-based misinformation. For example, Bavel et al. (2020) claimed that an understanding of cultural differences and the social context of the misinformation source is necessary to correct vaccine misinformation. Similarly, Arief and Karlinah (2019) proposed a communication approach to offset vaccine misinformation in the Muslim community, which would involve utilizing both trustworthy and influential members of the religious community as well as medical experts to educate the general population. This requires deep knowledge of Islamic perspectives to establish credibility before ordinary Muslims. Along the line, Muslim clerics' use of social media for misinformation correction has turned quite effective. Muslim clerics who are quite influential on social media platforms like YouTube attempted to correct such misinformation and suggested that Muslims can take the vaccine without a doubt (Syed and Wajid 2021). These clerics mostly clarified the ruling on the vaccines using the interpretation of Islamic texts. Some of these clerics even shared their experiences and posted videos of themselves taking the COVID vaccine. Such practices presumably worked well to misinformation correction, as social media sites like Twitter, Facebook, and YouTube have been found to be the most common means of disseminating false information about vaccines (Clamor et al. 2022; Melton et al. 2021).

In short, early misinformation correction studies mostly shed light on the correcting or debunking techniques and approaches. However, recent studies showed that the misinformation correction practice by only professionals misses out the voice of the ordinary public or crowds which could add further insights to the issue (Micallef et al. 2020). According to Micallef et al. (2020) "Compared to professional fact-checkers, concerned citizens, who are users of the platform where misinformation appears, have the ability to directly engage with people who propagate false claims either because of ignorance or for a malicious purpose"

(p. 1). Besides, regardless of the strategies, misinformation correction messages often fail to yield a positive response from the public. The following review gives way to understanding the dynamics of public response to misinformation correction.

### 2.3 Response to misinformation correction messages

The reaction from the public to misinformation correction or debunking depends on the properties of the corrective message, message source, etc. For instance, in vaccine-related misinformation debunking, a corrective message may not be successful even if contained detailed factual information (Chan et al. 2017; Larson and Broniatowski 2021). Rather, directly addressing the concern of the public is key for effectively correcting misinformation and fostering trust. Along the line, the credibility of corrective message sources is a significant factor in misinformation correction practice, but so is the source of misinformation. Corrective messages fail to succeed if people believe the original misinformation to be more credible (Walter and Riva 2020). As a result, efforts to retract misinformation often backfire and increase the persistence with misinformation. Tenney et al. (2009) suggested that an alternative causal explanation while correcting misinformation can mitigate such backfire issues.

Recent research examined how people respond to misinformation correction or debunking in a range of situations. When confronted with misinformation on Twitter, according to Wang and Zhuang (2018), users either (1) delete the tweet(s) or (2) post a new tweet to further clarify the situation. This indicates a reinforced reaction to misinformation debunking. Nyhan and Reifler (2015) also observed a similar pattern of reinforced support after corrective efforts. The study by Chan et al. (2017) demonstrated that the persistence of misinformation was stronger when audiences developed justifications in support of the initial misinformation. On the other hand, the debunking effect of misinformation correction was shown to be weaker in general. Surprisingly, a detailed debunking message was also positively connected with the misinformation-persistence impact. This suggests that, regardless of the messaging tactics employed for misinformation correction (e.g., thorough messages), anti-misinformation attitudes are not readily fostered in common people. Instead, it can contribute to increased persistence to misinformation. Furthermore, depending on the properties of the corrective message, it may just cause a negative response to the message's presenter (Larson and Broniatowski 2021). Therefore, the objective of this study was to analyze how the public responded to the misinformation correction campaign launched by Muslim clerics. Since the Muslim community was emotionally triggered originally by the

widespread belief that alcohol and pork were used in the manufacture of the COVID vaccine (Abbas et al. 2021), comments posted by Muslim users addressing the corrective efforts of Muslim clergies can be a good source of insights about their emotion and sentiment.

While much research has been done on misinformation correction strategies, misinformation persistence and public response in the context of faith-based misinformation remained unknown. To bridge that chasm, this study aimed to examine public sentiment and responses towards the misinformation correction messages of Muslim Clergies (Chan et al. 2017).

This leads to the first research question,

**RQ1:** What was the Muslim sentiment towards Muslim clerics' misinformation correction campaign for the Covid vaccine?

Nonetheless, Muslim clerics' misinformation correction for the COVID-19 vaccine was conducted on YouTube where the comments posted by common people are a useful source for understanding the sentiment expressed towards the practice. In addition to sentiment analysis, the comments are also an excellent resource for obtaining an in-depth grasp of the discussion. It would be intriguing to explore the underlying trends and issues in Muslims' conversations on anti-Misinformation campaign videos. This gave way to the following research question:

**RQ2:** What were the most expressed topics in the comments?

## 2.4 Machine learning techniques for sentiment analysis

Machine learning is an emerging field of Artificial Intelligence that design computational algorithms to imitate human intelligence by gaining knowledge about their surroundings. They are especially well-tailored to handle big data analysis. In today's era of data abundance, machine learning-based methods have been successfully implemented in much social scientific research (Mason et al. 2014). Especially, sentiment analysis based on machine learning has been frequently applied to comprehending public opinion and misinformation around COVID-19 (Gradoń et al. 2021). It is a natural language processing (NLP) technique of subclassifying textual materials into multiple categories: positive and negative polarity, or neutrality and positivity, to extrapolate the sentiment of a subject, idea, event, or phenomenon (Liu 2012). A traditional technique of sentiment analysis is lexical-based classification, which is similar to the classification of labeled historical data in machine learning. Melton et al. (2021) indicated in their lexical-based classification that Reddit groups in the vaccine-related discussions exhibited more positive than negative emotions, which has continued to stabilize over

time. Recent development in the use of machine learning for sentiment analysis include algorithms such as support vector machine (SVM), random forest (RF), decision tree (DT), logistic regression (LR), and k-nearest neighbors (KNN) etc. (Demircan et al. 2021). Furthermore, a Latent Dirichlet Allocation (LDA) based topic modeling analysis showed that group members were more concerned about side effects than conspiracy ideas. While the sentiment analysis was able to perform the sentiment classification, the LDA-based topic modeling uncovered the discussion further in-depth. The LDA (Latent Dirichlet Allocation, or LDA) is a generative probabilistic technique for creating explicit representations of subjects based on text input and has been widely used in generating topic models (Nwankwo et al. 2020). In a more recent covid-misinformation study, Clamor et al. (2022) employed a pointwise KL divergence in scoring “informativeness” and “phraseness” to remove misleading tweets, as well as BTM for topic modeling, to ensure an additional level of meticulousness. This new classifier model proved to be more effective than previous analyses in that it was able to identify 3533 misleading tweets with an accuracy of 74.25 percent, which was higher than many previous models. However, in understanding the diverse array of misinformation in the sub-Saharan region, Nwankwo et al. (2020) study claimed that the Bidirectional Encoder Representations from Transformers (BERT) model performs more effectively in addition to the LDA-based topic modeling. BERT made it possible to train in both directions. Its transformer is an encoder-decoder that is used to understand the contextual connections between words and to develop language models for speech recognition (Devlin et al. 2018). Sentiment analysis with machine learning has developed quite rapidly in the past years to the extent that automatic analytic software has been developed using certain lexicon databases. Such a software, developed by W3A.PL, was employed in Klimiuk et al. (2021) investigation into Polish vaccine-deniers’ misinformation on Facebook.

The unstructured nature of social media text data has made it quite challenging to train and test them using traditional supervised learning techniques. As a result, many computational studies in sentiment analysis research leaned more towards deep learning approaches (Song et al. 2021). Islam et al. (2020) provided an extensive review of misinformation research where deep learning (DL) can be used to automatically analyze text data and identify patterns that not only extract universal characteristics but also accomplish superior outcomes. Islam and colleagues demonstrated that text data-driven deep learning (DL) is a powerful and adaptable approach. Song et al. (2021) recent work proposed a classification-aware neural topic model (CANTM), which made text classification and topic identification much easier to accomplish. The overall performance accuracy of the CANTM model was 63.34 percent.

Deep learning methods offer some novel creative models and themes in misinformation research as well. For instance, Green et al. (2021) interrupted time series analysis observed the misleading information about COVID in the UK, specifically before and after the first lockdown announcement. This was an unprecedented and ingenious approach considering the panic and resistance observed during the lockdown announcement. Although the number of post-announcement tweets increased, no evidence of post-announcement misinformation tweets was found in this study. The machine-learning approach in medical research especially demands extreme rigor and meticulous evaluation. Thus, multiple approaches were observed in the previous research on misinformation. For instance, in addition to convolutional neural networks, Du et al. (2021) employed four more machine learning algorithms, such as support vector machines, logistic regression, extremely randomized trees, and a recurrent neural network to classify misinformation about human papillomavirus (HPV). Out of the five approaches, the convolutional neural network produced the best results. Similarly, Klimiuk et al. (2021) study on Polish vaccine-deniers' misinformation rendered a fair score of accuracy (70%) using the BiLSTM neural network.

To recap, prior misinformation-centered research relied heavily on lexical-based classification for sentiment analysis and software development for automatic sentiment detection. A variety of deep learning strategies were also observed in recent studies for detecting disinformation and misinformation that offered unique and unprecedented machine learning models. Now, this study aimed at building a machine learning model for detecting Muslim sentiment in addition to analyzing the sentiment in this context. For such, Random Forest, Support Vector Machine, and Naïve Bayes models were used as methods in this study. This leads to the third research question,

**RQ3:** Which ML algorithm (i.e., Balanced Random Forest, SVM, and Naïve Bayes) perform more accurately in detecting sentiment towards misinformation correction?

## 3 Method

### 3.1 Dataset

The data was collected from YouTube using a social media data extractor and analyzer application called Netlytic. Netlytic is a cloud-based text analyzer and social network visualization tool that is available on the web. It is useful to trace and visualize social networks from conversations on sites like Twitter,

YouTube, blog comments, and online forums and chats using Netlytic's text summarization technology (Gruzd 2016). The app, initially developed by Dr. Gruzd at Ryerson University, is now widely used in human behavior research. Many such research using Netlytic has been published in top journals including *American Behavioral Scientist* and the *Journal of Education for Library and Information Science* (Gruzd 2009; Hampton 2010). Using the YouTube API, Netlytic allows access to historic posts on Youtube. One of the several advantages of Netlytic is that it is also a community-supported text analyzer and social network analyzer that is designed to assist social media academics and educators in analyzing public debate on social media platforms (Gruzd 2016).

Initially, all the YouTube videos related to Muslim clerics' misinformation correction about COVID-19 vaccine were assembled using search keywords such as, 'Covid vaccine halal or haram', 'Is Covid vaccine permissible for Muslims?', 'Islamic ruling about Covid vaccine', 'Is the COVID vaccine halal?' etc. 20 videos were found in the initial search. However, most of the videos found had their comment option turned off, which is not surprising given the fact that comment sections in such videos were turning chaotic. Many comments were even accusing the cleric of being sold out to western media and government. Only 13 videos remained with comment options open which were posted from Muslim Clerics of various parts of the world including the United States, United Kingdom, Malaysia, Zimbabwe, etc. The videos featured the most popular Muslim scholars and clerics in such regions including Dr. Jakir Naik, Dr. Yasir Qadhi, Mufti Menk, and others. A couple of videos among them were telecasted on news-based tv channels such as, CNN and Newsmax, etc. Some of these videos have a far higher audience than those that have disabled the comment feature. For instance, out of the 13, the video that featured Dr. Jakir Naik garnered 1,470,922 views, and the one that features Mufti Menk received 282,217 views whereas Dr. Farzana's video without comment option received only 14,807. In short, the 13 videos selected for data collection were among the most viewed pieces.

Overall, 10,000 comments were collected. After cleaning and preprocessing the texts, the dataset remained with 9701 comments. Considering the large dataset, this study was conducted using a computer-assisted text analysis method. Unlike manual content analysis, this method offers an opportunity to analyze a larger dataset. It also removes the traditional concern about the effectiveness of sample data to represent the population.

## 3.2 Procedure

In building a machine learning model, a training set of documents that is already labeled is used to train the algorithm, and a test set of documents is used to check

the model. In this procedure, lexical methods are used to find possible key phrases, and a set of train-document that have been labeled is created (Onan et al. 2016). As per the procedure, this study was conducted in two stages, first, a train-dataset (the dataset used for the algorithm to learn) was created using 'dictionary-based text analysis' for the positive and negative sentiment using and create a machine learning model that can detect the negative and positive sentiments in YouTube comments. Second, to answer the research question **RQ2**, what were the most expressed topics in the comments this study also employed the LDA based topic model which is an unsupervised machine learning algorithm.

The 'dictionary-based text analysis' has been a long tradition to label datasets. This method is widely used in computer-assisted content analysis (Guo et al. 2016). The dictionary-based approach is generally useful because it can "automatically classify text of any kind into groups of any kind" (Guo et al. 2016 p. 335). However, recent studies criticized that dictionary-based labeling alone can be quite shallow, time-consuming, and labor-intensive technique (Onan, 2022). Also, the dictionary-based approach functions more appropriately in text data that are structured (e.g., newspaper contents) whereas the datasets containing YouTube comments are unstructured and grammatically ill-constructed. The linguistic nuances in such a text set are best understood by human coders (Cunha Lassance et al. 2019). Especially, in social studies, manual labeling can be essentially helpful to validate the dictionary-based labeling, so the analysis is grounded in qualitative reasoning before training the algorithm and furthering into machine learning procedure (Chen et al. 2018). Thus, randomly pulled 500 tweets were manually labeled for negative and positive. In the coding procedure, each comment that is relevant and a compliment to the misinformation correction videos is given a 'positive' label. For example, the post "Good job Imam! It is highly important to be scientific literate" is labeled as a positive comment. On the other hand, the comments were labeled as negative whenever the remarks expressed criticism and disapproval of the video, for example, "you are paid off, or you are intentionally lying. It contains" dead babies". The agreement between manual labeling and dictionary-based labeling was 85%.

Following the labeling, the training dataset appeared to be leaning heavily toward the negative sentiment. Consequently, the data set turned out to be imbalanced. Such an imbalance occurs in machine learning applications when one class has an exceedingly small number of instances, and the other class has an exceptionally large number of instances. There are several approaches recommended to address that issue, including consensus cluster-based under-sampling, Balanced Random Forest model, and others (Chen et al. 2004; Onan 2019). The Balanced Random Forest (BRF) approach was employed in this study.

### 3.3 Preprocessing and lemmatization

A text categorization technique called TF-IDF was used for classification. The TF-IDF is computed using the weight factor of words. On the other hand, it quantifies the importance of Count Vectorization. The more instances of a term in a document, the less valuable it gets. For example, commonly occurring words in phrases such as “and”, “or”, and “the” might be omitted when analyzing the meaning of a statement. The fewer times a word is repeated, the higher the IDF value. The TF-IDF equation is shown below, where  $i$  is the number of occurrences of  $I$  in  $j$ ,  $d_j$  is the number of documents that include  $I$  and  $N$  is the total number of documents (Demircan et al. 2021).

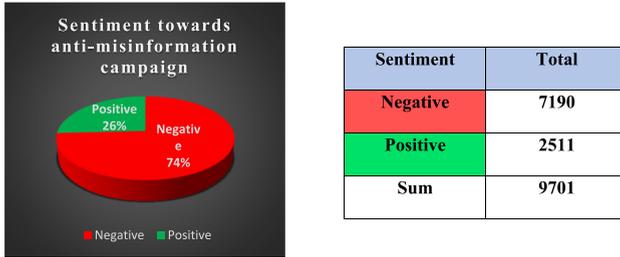
$$w_{i,j} = tf_{i,j} \times \log \left( \frac{N}{d_j} \right)$$

The NLTK library in python programming language was used to preprocess and lemmatize the dataset. During the preprocessing procedure, it was determined whether to include comments as feature data. Sentences of less than 1000 characters were included in trained data, since lengthy articles may result in the loss of textual feeling. To get an algorithm to understand any text, it is necessary to break the word down into smaller units that our machine can comprehend. That is the purpose of stemming and tokenization in natural language processing. First, stemming was employed for stripping the suffixes (“ing”, “ly”, “es”, “s” etc.) from a word. For example, – “play”, “player”, “played”, “plays”, and “playing” are the different variations of the word – “play”.

The next step is the tokenization technique which divides a large body of text into smaller components called tokens (Figure 1). Tokenization may be categorized roughly into three types: word tokenization, character tokenization, and sub-word tokenization. It is considered as the most critical stage in text data modeling. As seen in Figure 2, tokenization is applied to transform a full sentence into a set

	Description	Label	text_final
0	[what, a, load, of, cobblers, ., a, robot, cou...	0	['load', 'cobbler', 'robot', 'could', 'do', 'g...
1	[blind, sheep]	0	['blind', 'sheep']
2	[what, a, clueless, guy]	0	['clueless', 'guy']
3	[I, 'd, rather, drink, alcohol, and, pork, th...	0	['rather', 'drink', 'alcohol', 'pork', 'take',...
4	[confused, individual, ., do, not, listen, to,...	0	['confuse', 'individual', 'listen', 'sha', 'al...

Figure 1: Word tokenization.



**Figure 2:** Sentiment towards Muslim Clerics' anti-misinformation campaign.

of tokens. The tokens are then utilized to construct a vocabulary that is further used as a feature in traditional NLP algorithms such as the Count Vectorizer and the TF-IDF. Each word in the vocabulary is considered to have a distinct property. Following is the tokenized example in the data. The column `text_final` in Figure 2 contains the tokenized version of the text.

### 3.4 Topic modeling

To answer the first research question, what were the most expressed topics in the comments? Multiple approaches were employed to compare the results of dictionary-based text analysis and LDA based topic models. For the dictionary-based analysis, Netlytic (Gruzd 2016) and `text2data`'s exploratory analysis was performed. In addition to that, a topic modeling was performed using the latent Dirichlet allocation (LDA) approach. Using the Python library "Gensim", LDA was trained across each batch of tweets. The number of topics to train is entirely up to the researcher in LDA modeling (Guo et al. 2016). For this study, eight topics were chosen.

### 3.5 Classification models

In the second phase, using the trained dataset, overall, three supervised machine learning methods such as Naïve Bayes, Support Vector Machine (SVM), and Random Forest model were employed to generate an algorithm that can detect negative and positive sentiments in the YouTube comments.

First, the **Naïve Bayes** classifier is a well-known technique for supervised classification. It is a probabilistic classifier based on the Bayes' theorem that considers the Naïve (Strong) independence condition. It was introduced into the

text retrieval community under a different name and has remained a popular (baseline) approach for text categorization, the issue of classifying documents using word frequencies as a characteristic. A benefit of Naïve Bayes is that it uses a modest quantity of training data to estimate the classification parameters. Naïve Bayes' is a conditional probability model in its simplest form. Despite its simplicity and reliance on strong assumptions, the naïve Bayes' classifier has been shown to perform well in a wide variety of domains, especially in sentiment analysis (Dey et al. 2016).

Second, the **Support Vector Machine (SVM)** algorithm has been used to a broad variety of classification issues. SVM can be employed as a text classifier that uses hyper-plane to split the data. The hyperplane is a straight line in two-dimensional space that optimizes the margin between two classes. SVM's core notion is to discover line separators in the search space to divide diverse groups. The mathematical formulation of the SVM approach makes use of the cost (C), epsilon ( $\epsilon$ ), gamma ( $\gamma$ ), and three kinds of kernel functions, such as linear, RBF, and polynomial. To enhance the performance of SVM, a grid search was also used to find the best parameters. Two separate performance criteria were used to assess the models' performance: the accuracy score and the confusion matrix.

Third, the **Random Forest** is an ensemble of unpruned classification or regression trees that is generated by randomly selecting features from bootstrap samples of training data. A forecast is formed by aggregating the ensemble's predictions (a majority vote for classification or an average for regression). To keep the overall error rate as low as possible, this model prioritizes the accuracy of the predictions for the majority group, which often results in low accuracy for the minority group. Considering the imbalanced training data in the current study, a Balanced Random Forest (BRF) was employed to achieve better performance (Chen et al. 2004).

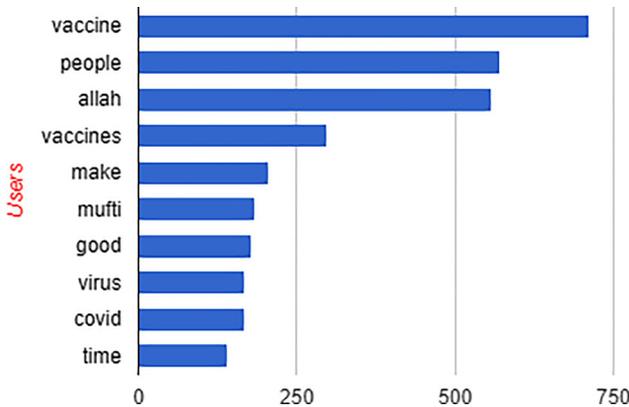
## 4 Results

To answer the research question of what the Muslim sentiments towards Muslim clerics' anti-misinformation campaign were, a dictionary-based analysis of text2-data revealed that 74 percent of the comments were negative, and 26 percent of the comments were positive (see Figure 2). Some examples of negative and positive comments are presented in Table 1.

To answer the second research question, what the most expressed topics in the comments were, exploratory analysis from two different applications 'Netlytic' (see Figure 3) and 'text2data' (see Figure 4) showed respectively top ten topics and top twenty-six topics. A substantial similarity in the discussed keywords appeared

**Table 1:** Example of negative and positive comments.

Negative Comments	Positive comments
1 You are paid off, or you are intentionally lying. The blood of innocent are now in your hands you are 100% accountable for your actions against humanity.	Bcoz of such people who are commenting below this particular community us always a target ... The guy spoke truth .. vaccination is necessary for us.
2 Bro. It contain dead babies. Speech without knowledge is dangerous.	Good job imam! it is highly important to be scientific literate.
3 MayAllah forgive you for spreading falsehood.	Most scholars say it is permissible, so why making things more complicated!
4 Please confirm. Is it permissible for us as muslims? Even though there are cells taken from abortions from decades ago.	And most of the vaccines are halal just listened and comprehend well what the imam is talking about



**Figure 3:** Top ten keywords (Netlytic).

from two applications. Most expressed keywords in the discussion found from both application’s analysis were as follows, vaccine, Allah, people, mufti, covid etc. However, the LDA based topic model revealed 8 distinct topics and 10 associated words throughout the corpus which shows a better cluster of themes discussed in the comments (see Table 2).

The exploratory text analysis also revealed that the 10 most vocal commenters contributed to a significant portion of the total comments: 18.3 percent of comments came from one single commenter. Respectively second, third, and fourth top commenters posted 11.5 percent, 11.5 percent, and 10.7 percent of all the comments (see Figure 5).

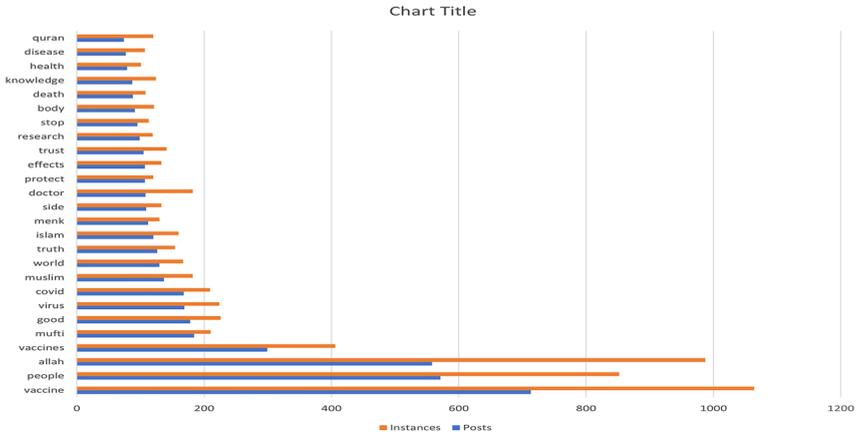


Figure 4: Most expressed keywords (text2data).

It is fascinating to see that two of the top commenters were either not Muslims or did not use names associated with the Muslim faith. The social network analysis carried out with Netlytic (Gruzd 2016), on the other hand, reveals that these two commentators were not among the most influential clusters (see Figures 6 and 7). These clusters were constructed to assess the influence of commenters in a social network based on the number of times a commenter was mentioned and reacted to

Table 2: LDA based topic model.

Topic	Top 10 words
Topic #1:	['Just', 'need', 'use', 'say', 'doctor', 'make', 'people', 'said', 'vaccine', 'Allah']
Topic #2:	['Life', 'man', 'tell', 'heaven', 'john', 'jesus', 'shall', 'matthew', 'father', 'god']
Topic #3:	['Doctor', 'speak', 'don', 'right', 'allah', 'understand', 'video', 'good', 'let', 'trust']
Topic #4:	['Time', 'world', 'don', 'body', 'covid', 'vaccines', 'people', 'virus', 'just', 'vaccine']
Topic #5:	['Just', 'know', 'did', 'brother', 'don', 'menk', 'allah', 'mufti', 'vaccine', 'people']
Topic #6:	['Old', 'praise', 'la', 'abu', 'prophet', 'times', 'ameen', 'wa', 'al', 'allah']
Topic #7:	['Like', 'long', 'effects', 'muslims', 'covid', 'don', 'know', 'vaccines', 'people', 'vaccine']
Topic #8:	['Thought', 'giving', 'muslims', 'scholars', 'lost', 'seeker', 'advice', 'truth', 'muslim', 'sheikh']

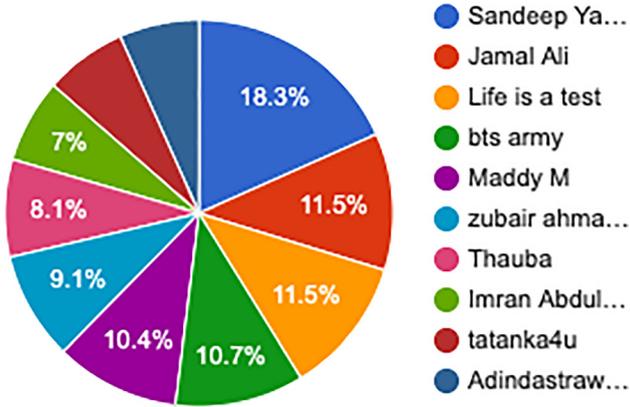


Figure 5: Top ten commenters.

by other commentators (Gruzd 2016). This indicates that the top two non-Muslim commentators have made a substantial number of comments and responses to others' remarks, but they have not received as many responses from other commentators nor been cited as often. To put it another way, they were not among the most influential commentators, even though they were the most active commenters

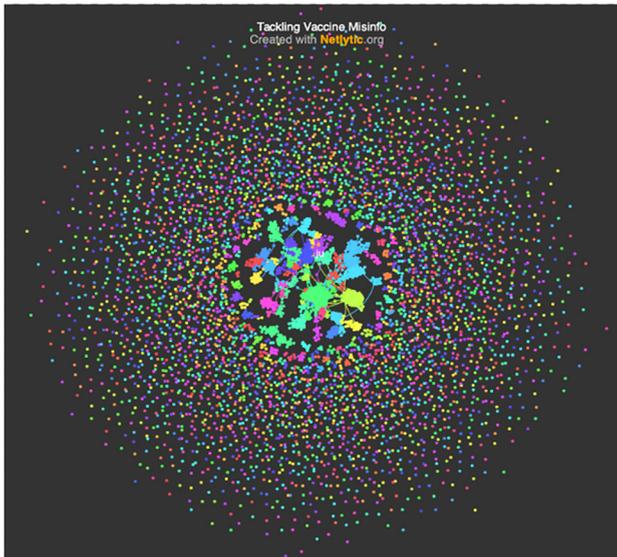


Figure 6: Network analysis of commenters.



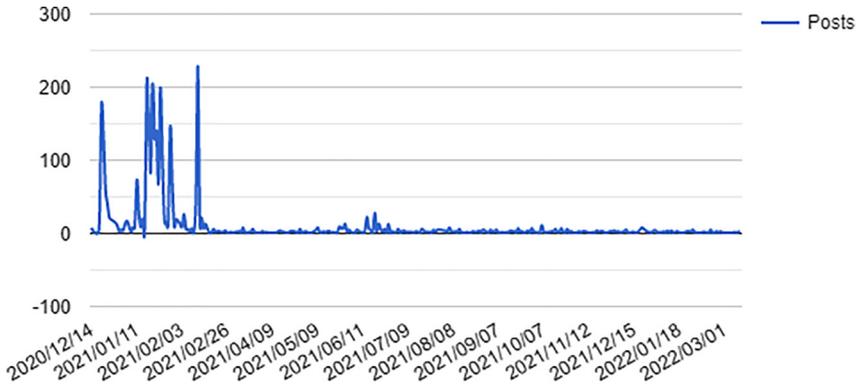


Figure 8: Posting timeline of comments.

Table 3: Accuracy score of Naïve Bayes, SVM, and Balanced RF.

Accuracy score	Naïve Bayes	SVM (linear)	SVM (rbf)	SVM (polynomial)	Random forest	Balanced random forest
	0.74	0.77	0.72	0.72	0.74	0.62

at ‘C’: 50, ‘kernel’: ‘rbf’ by tuning hyperparameters, as shown in Figure 9. While the best score found in GridSearch CV was 0.78. the GridSearch CV score on the test set was 0.97 which is quite significant.

Next, a confusion matrix was employed to assess the performance of the sentiment detection model. The confusion matrix result of the RF-based model is

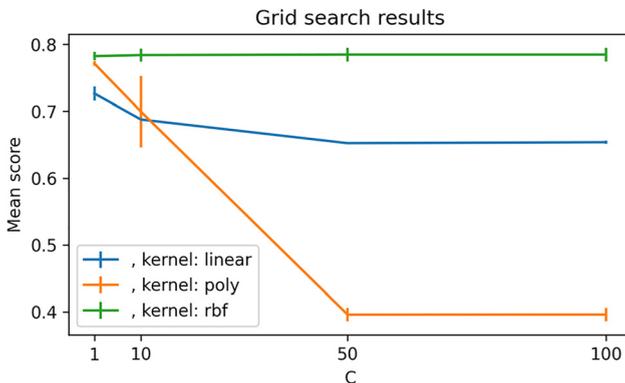
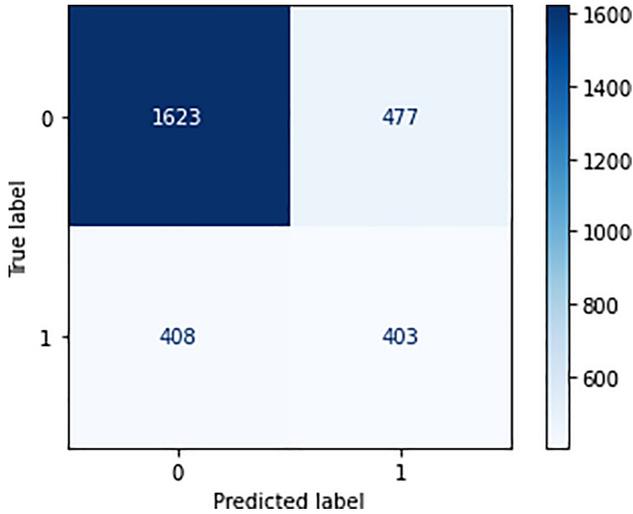


Figure 9: GridSearch score of linear, RBF, polynomial kernel..



**Figure 10:** Confusion matrix for Balanced Random Forest.

shown in Figure 10. Out of the three models, Naïve Bayes, SVM, and Balanced Random Forest (BRF), the BRF's model performed the best in the assessment of the confusion matrix. It was observed that overall, out of 2000 negatively labeled comments, the classifier correctly predicted 1623 comments as negative, whereas 477 comments were misclassified as positive. Out of 811 positively labeled comments, the classifier correctly predicted 403 comments as positive, whereas 408 comments were misclassified as negative. The imbalanced nature of the dataset in this study caused the other models' low performance. Chen et al. (2004) suggested that, instead of the traditional Random Forest model, a Balanced Random Forest can yield better performance as assessed in the confusion matrix in Figure 10.

## 5 Discussion

Based on the empirical results of the comparison of the three-machine learning models, several insights are revealed about the response to misinformation correction by Muslim clerics. First, the result showed a high negative sentiment towards the misinformation correction of Muslim clerics. It is not surprising that the negative sentiment would persist based on the CIE theory. However, the heavy (almost three-fourth) negative sentiment in the comments shows quite strong resistance to the acceptance of the vaccine. Many comments that were directed

particularly towards the clerics were excessively harsh and accusatory. Thus, it can be assumed that people with negative sentiments might have been more vocal in the comment section than others.

Second, the results also indicate that the Balanced Random Forest model yields promising results in analyzing the unstructured YouTube comments dataset. The accuracy of the models that were presented in this study was reported to be 77%, which is rather large when compared to the accuracy scores of prior sentiment models constructed for vaccine-deniers' misinformation, which were 70% (Klimiuk et al. 2021). The unbalanced nature of the dataset employed in this investigation was the main reason for the deficient performance of the other models. In place of the conventional Random Forest model, Chen et al. (2004) suggested using a Balanced Random Forest since it has the potential to produce better results. In light of this, the significance of this 'sentiment' detection model becomes extremely relevant in the field of public health research and study about social media. Particularly because sentiment analysis is gaining increased popularity in the field of misinformation research as a whole, future studies on misinformation can make use of this sentiment model (the balanced Random Forest model) to better comprehend the nature of misinformation discourse on social media. Also, given the greater number of comments with a negative sentiment than those with a positive sentiment, the negative prediction rate turned out to be always greater when the dataset was fed to a machine learning model. The models used for sentiment analysis in this study will particularly help public health sector workers or researchers to better understand the nature of vaccine hesitancy in Muslim population. Additionally, the Muslim practice of correcting disinformation might benefit from the meaningful understanding that this study provides. The purpose of the joint task force that was established by Muslim doctors and clerics was just to illustrate the practice of refuting misinformation; nonetheless, the impact of their campaign has been uncovered through the course of this investigation. In other words, the Muslim clerics and medical professionals who joined the task force to correct vaccine misinformation can observe how Muslim users responded towards the correction message. This study offers the joint task force, and the future misinformation campaigns a detection model to understand the impact of such practices on social media. Despite producing unprecedentedly insightful results, this study is not without its drawbacks. Most importantly, it's possible that the views and feelings expressed in the comments collected from YouTube users do not represent the views and feelings of the public at large or of all Muslims. The dataset only contains data on the people who impulsively posted and expressed their concerns. Another limitation that this study has is that the dataset was relatively small. However, that was all the data available to researchers for this particular topic. In future, should the

misinformation correction campaigns continue to post more videos, the newer comments that may come along with them can be observed and analyzed using the model proposed in this study. In subsequent research, it may be possible to investigate the ways in which misinformation correction tactics are operating in various geographical regions and in relation to other religions.

## 6 Conclusion

The purpose of this study was to understand the sentiment of Muslim users towards the anti-misinformation campaign of Muslims clerics on YouTube and develop an algorithm to detect sentiment using supervised machine learning models. As faith-based misinformation caused propagation of COVID-19 vaccine misinformation across the world, many Muslim clerics have attempted to debunk the misinformation using religious counterspeech and interaction on social media. In this study, three machine learning algorithms called Naive Bayes, SVM, and a Balanced Random Forest were used to build a sentiment model that can detect Muslim sentiment regarding Muslim clerics' anti-misinformation campaign on YouTube. The model can detect positive and negative sentiment in the response about the campaign. Confusion matrix and accuracy score were employed to evaluate the model. According to the confusion matrix and the accuracy score analysis, the Balanced Random Forest-based model displayed the best performance. This study was designed to specifically analyze the YouTube comments in faith-based misinformation correction videos. However, this method and models can be applied for sentiment analysis of other YouTube comments. This is one of the very few research in the discipline of communication that has attempted to address the challenge of imbalanced learning in the context of YouTube comment datasets. The use of machine learning in the field of communication is gaining traction, although this is mostly due to multidisciplinary collaboration and consultation with academics from the field of computer science. In fact, despite producing more publications, misinformation research in the communication field receives less attention and citations than computer science (Ha et al. 2021). While interdisciplinary efforts do make important contributions, it would be beneficial for communication academics to have some degree of self-reliance in the computational methods they use. The use of the Balanced Random Forest model, as well as other machine learning applications in this research, will inspire more inquiries into similar social media campaigns especially while handling an imbalanced dataset.

## Appendix

### YouTube videos for data collection

1. Imam Kashif Ahmad on the Covid-19 vaccines and saving lives <https://www.youtube.com/watch?v=xdFM3Mcq1jE&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=1&t=25s>
2. Is the COVID Vaccine HALAL? | Dr. Shabir Ally <https://www.youtube.com/watch?v=YLbRjkIOyw4&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=2>
3. It is YOUR decision—Mufti Menk #Vaccine <https://www.youtube.com/watch?v=IJquyaiVkJXA&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=3>
4. Muslims in Singapore allowed to receive COVID-19 vaccine after Mufti's decision <https://www.youtube.com/watch?v=tXn8An2E5K4&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=4>
5. COVID-19: Ethnic minority communities being 'targeted' by anti-vaxxers <https://www.youtube.com/watch?v=kTBeF6441w&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=5>
6. To Vax or not to Vax—The Truth! - Covid Vaccine <https://www.youtube.com/watch?v=Pd1IoxYZbZA&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=6>
7. "I took the vaccine and the effects of it are felt by others" - Mufti Menk <https://www.youtube.com/watch?v=MMXQbC4Swf4&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=7>
8. Covid Vaccines | Are Covid Vaccines Halal or Haram? | Dr Mashhood Qazi <https://www.youtube.com/watch?v=Az8Xim4Upko&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=8>
9. Dr Zakir Naik on Corona Virus <https://www.youtube.com/watch?v=QDu9RgAGJL4&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=9>
10. Dr Amir Khan: Muslim doctor says COVID vaccines are halal <https://www.youtube.com/watch?v=MILoTG5ARU&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=10>
11. COVID vaccine between Fiqh and Medicine |Q&A | Shaykh Dr. Yasir Qadhi with Dr. Sulaiman Abawi <https://www.youtube.com/watch?v=Qqe4ZKE6Yaw&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=11>

12. what is your opinion on taking covid 19 vaccine Dr Zakir Naik #new #Ramadan #HUDATV <https://www.youtube.com/watch?v=qU-4b9ICQQ0&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=12>
13. CNNi: British Muslims and Hindus fight vaccine hesitancy <https://www.youtube.com/watch?v=ZcXIU-fiKUY&list=PLrpweZPVrRgmNQUU94sr43UHacCeUhe3L&index=13>

## References

- Abbas, Qamar, Fatima Mangrio & Sunil Kumar. 2021. Myths, beliefs, and conspiracies about COVID-19 vaccines in Sindh, Pakistan: An online cross-sectional survey. *Authorea Preprints* 1–7. <https://doi.org/10.22541/au.161519250.03425961/v1>.
- Alimardani, Mahsa & Mona Elswah. 2020. Online temptations: COVID-19 and religious misinformation in the MENA region. *Social Media + Society*.
- Arief, Nurlaela N & Siti Karlinah. 2019. The role of Ulama (Islamic religious leaders) in correcting anti-vaccination rhetoric in Indonesia. *Journal of Asian Pacific Communication* 32(2). 254–271.
- Bavel, Jay J. Van, Katherine Baicker, Paulo S. Boggio, Valerio Capraro, Aleksandra Cichocka, Mina Cikara, Molly J. Crockett, Alia J. Crum, Karen M. Douglas, James N. Druckman, John Drury, Oeindrila Dube, Naomi Ellemers, Eli J. Finkel, James H. Fowler, Michele Gelfand, Shihui Han, S Alexander Haslam, Jolanda Jetten, Shinobu Kitayama, Dean Mobbs, Lucy E. Napper, Dominic J. Packer, Gordon Pennycook, Ellen Peters, Richard E. Petty, David G. Rand, Stephen D. Reicher, Simone Schnall, Azim Shariff, Linda J. Skitka, Sandra Susan Smith, Cass R. Sunstein, Nassim Tabri, Joshua A. Tucker, Sander van der Linden, Paul van Lange, Kim A. Weeden, Michael J. A. Wohl, Jamil Zaki, Sean R. Zion & Robb Willer. 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour* 4(5). 460–471. PMID:32355299.
- Chan Man-pui Sally, Christopher R Jones, Kathleen Hall Jamieson & Dolores Albarracín, 2017. Debunking: A meta-analysis of the psychological efficacy of messages correcting misinformation. *Psychological Science* 28(11). 1531–1546.
- Chen, Chen, Andy Liaw & Leo Breiman. 2004. *Using random forest to learn imbalanced data*, vol. 110, 1–12. Berkeley: University of California.
- Chen, Nan-Chen, Margaret. Drouhard, Rafal Kocielnik, Jina Suh & Ceceilia Aragon. 2018. Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity. *ACM Transactions on Interactive Intelligent Systems* 8(2). 1–20.
- Clamor, S. Thomas Daniel, Geoffrey A. Solano, Nathaniel Oco, Jasper Kyle Catapang & Jerome Cleofas & Iris Thiele Isip-Tan. 2022. Identification and analysis of COVID-19-related misinformation tweets via kullback-leibler divergence for informativeness and phraseness and biterm topic modeling. In *2022 international conference on artificial intelligence in information and communication (ICAIIIC)*, 451–456.
- Cunha Lassance Alexandre Ashade, Melissa Carvalho Costa & Marco Aurélio C. Pacheco. 2019. Sentiment analysis of YouTube video comments using deep neural networks. In *Artificial intelligence and soft computing*, 561–570: Springer International Publishing.

- Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Toutanova Kristina. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Demircan, Murat, Adem Sella, Fatih Abut & Mehmet Fatih Akay. 2021. Developing Turkish sentiment analysis models using machine learning and e-commerce data. *International Journal of Cognitive Computing in Engineering* 2. 202–207. ISSN 2666-3074.
- Dey, Lopamudra, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose & Sweta Tiwari. 2016. Sentiment analysis of review datasets using naïve Bayes' and K-nn classifier. *International Journal of Information Engineering and Electronic Business* 8. 54–62.
- Du, Jingcheng, Sharice Preston, Hanxiao Sun, Shegog Ross, Rachel Cunningham, Julie Boom, Lara Savas, Muhammad Amith & Tao Cui. 2021. Using machine learning-based approaches for the detection and classification of human papillomavirus vaccine misinformation: Infodemiology study of reddit discussions. *Journal of Medical Internet Research* 23(8). e26478.
- Ecker, Ullrich KH & LiAng Chang. 2019. Political attitudes and the processing of misinformation corrections. *Political Psychology* 40. 241–260.
- Ecker, Ullrich KH, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K. Fazio, Nadia Brashier, Panayiota Kendeou, Emily K. Vraga & Michelle A. Amazeen. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology* 1(1). 13–29.
- Etutu, Joice & Jack Goodman. 2021. *Misleading claims targeting ethnic minorities*. BBC News. <https://www.bbc.com/news/55747544>.
- Gruzd, Anatoliy. 2016. *Netlytic: Software for automated text and social network analysis*. <http://Netlytic.org>.
- Gruzd, Anatoliy. 2009. Studying collaborative learning using name networks. *Journal of Education for Library & Information Science* 50(4). 243–253.
- Gradoń, Kacper T., Janusz A. Hołyst, Wesley R. Moy, Sienkiewicz Julian & Krzysztof Suchecki. 2021. Countering misinformation: A multidisciplinary approach. *Big Data & Society* 8(1). <https://doi.org/10.1177/20539517211013848>.
- Green, Mark, Elena Musi, Francisco Rowe, Darren Charles, Frances Darlington Pollock, Chris Kypridemos, Andrew Morse, Patricia Rossini, John Tulloch, Andrew Davies, Emily Dearden, Hendramoorthy Maheswaran, Alex Singleton, Roberto Vivancos & Sally Sheard. 2021. Identifying how COVID-19-related misinformation reacts to the announcement of the UK national lockdown: An interrupted time-series study. *Big Data & Society* 8(1). 205395172110138.
- Guo, Lei, Chris J. Vargo, Zixuan Pan, Weicong Ding & Ishwar Prakash. 2016. Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Probation Journal* 93(2). 398–415.
- Ha, Louisa, Loarre Andreu Perez & Rik Ray. 2021. Mapping recent development in scholarship on fake news and misinformation, 2008 to 2017: Disciplinary contribution, topics, and impact. *American Behavioral Scientist* 65(2). 290–315.
- Hampton, Keith N. 2010. Internet use and the concentration of disadvantage: Glocalization and the urban underclass. *American Behavioral Scientist* 53(8). 1111–1132.
- Islam, Md Rafiqul, Shaowu Liu, Xianzhi Wang & Guandong Xu. 2020. Deep learning for misinformation detection on online social networks: A survey and new perspectives. *Social Network Analysis and Mining* 10(1). 82.
- Kanozia, Rubal & Ritu Arya. 2021. “Fake news”, religion, and COVID-19 vaccine hesitancy in India, Pakistan, and Bangladesh. *Media Asia* 48(4). 313–321.

- Klimiuk, Krzysztof, Agnieszka Czoska, Karolina Biernacka & Łukasz Balwicki. 2021. Vaccine misinformation on social media - topic-based content and sentiment analysis of polish vaccine-deniers' comments on facebook. *Human Vaccines & Immunotherapeutics* 17(7). 2026–2035.
- Khan, Yusra Habib, Tauqeer Hussain Mallhi, Nasser Hadal Alotaibi, Abdulaziz Ibrahim Alzarea, Abdullah Salah Alanazi, Nida Tanveer & Furqan Khurshid Hashmi. 2020. Threat of COVID-19 vaccine hesitancy in Pakistan: The need for measures to neutralize misleading narratives. *The American Journal of Tropical Medicine and Hygiene* 103(2). 603–604.
- Kait, Sanchez. 2020. *Facebook will remove COVID-19 vaccine misinformation*: The Verge. <https://www.theverge.com/2020/12/3/22150425/facebook-covid-19-vaccine-coronavirus-misinformation-ban>.
- Liu, Bing. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5(1). 1–167.
- Larson, Heidi J. & David A. Broniatowski. 2021. Why debunking misinformation is not enough to change people's minds about vaccines. *American Journal of Public Health* 111(6). 1058–1060.
- Lewandowsky, Stephan, Ullrich KH Ecker, Colleen M. Seifert, Norbert Schwarz & John Cook. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest* 13. 106–131.
- Mason, Winter, Jennifer Wortman Vaughan & Hanna Wallach. 2014. Computational social science and social computing. *Machine Learning* 95(3). 257–260.
- Melton, Chad A., Olufunto A. Olusanya, Nariman Ammar & Arash Shaban-Nejad. 2021. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the reddit social media platform: A call to action for strengthening vaccine confidence. *Journal of Infection and Public Health* 14(10). 1505–1512.
- Merriam-Webster Dictionary. 2022. *Misinformation* (accessed 19 April 2022).
- Micallef, Nicholas, Bing He, Srijan Kumar, Mustaque Ahamad & Nasir Memon. 2020. The role of the crowd in correcting misinformation: A case study of the COVID-19 infodemic. In *2020 IEEE international Conference on big data (big data)*, p. 748–757.
- Nyhan, Brendan. & Jason Reifler. 2015. Displacing misinformation about events: An experimental test of causal corrections. *Journal of Experimental Political Science* 2. 81–93.
- Nwankwo, Ezinne, Chinasa Okolo & Cynthia Habonimana. 2020. Topic modeling approaches for understanding COVID-19 misinformation spread in s-Saharan Africa. In *Proc. AI social good workshop*.
- Onan, Aytuğ. 2022. Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification. *Journal of King Saud University. Computer and Information Science* 34(5). 2098–2117.
- Onan, Aytuğ, Serdar Korukoğlu & Hasan Bulut. 2016. Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications* 57. 232–247.
- Onan, Aytuğ. 2019. Consensus clustering-based undersampling approach to imbalanced learning. *Scientific Programming* 2019. 1–14.
- Pew Research Center. 2021. *Intent to get vaccinated against COVID-19 varies by religious affiliation in the U.S. 10 Facts about Americans and coronavirus vaccines*. [www.pewresearch.org/fact-tank/2021/03/23/10-facts-about-americans-and-coronavirus-vaccines/ft\\_21-03-18\\_vaccinefacts/](http://www.pewresearch.org/fact-tank/2021/03/23/10-facts-about-americans-and-coronavirus-vaccines/ft_21-03-18_vaccinefacts/).
- Song, Xingyi, Johann Petrak, Ye Jiang, Iknor Singh, Diana Maynard, & Kalina Bontcheva. 2021. Classification aware neural topic model for COVID-19 disinformation categorisation. *Plos One*. 16(2). e0247086.

- Susmann, Mark W. & T. Wegener Duane. 2021. The role of discomfort in the continued influence effect of misinformation. *Memory & Cognition* 50(2). 435–448.
- Syed, Sana & Arshia Wajid. 2021. Muslim community engagement efforts to tackle COVID-19 vaccine misinformation. *Harvard Medical School Primary Care Review* April 21. <https://info.primarycare.hms.harvard.edu/review/muslim-community-engagement-efforts>.
- Stecula, Dominik Andrzej, Ozan Kuru & Kathleen Hall Jamieson. 2020. How trust in experts and media use affect acceptance of common anti-vaccination claims. *The Harvard Kennedy School Misinformation Review* 1(1). <https://misinforeview.hks.harvard.edu/article/users-of-social-media-more-likely-to-be-misinformed-about-vaccines/>.
- Tenney, Elizabeth R., Hayley MD Cleary & Barbara A. Spellman. 2009. Unpacking the doubt in “beyond a reasonable doubt”: Plausible alternative stories increase not guilty verdicts. *Basic and Applied Social Psychology* 31. 1–8.
- Wardle, Claire & Eric Singerman. 2021. Too little, too late: Social media companies' failure to tackle vaccine misinformation poses a real threat. *BMJ* 372. <https://doi.org/10.1136/bmj.n26>.
- Walter, Nathan & Tukachinsky Riva. 2020. A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research* 47(2). 155–177.
- Wang, Bairong & Jun Zhuang. 2018. Rumor response, debunking response, and decision makings of misinformed Twitter users during disasters. *Natural Hazards* 93(3). 1145–1162.

## Bionote

### Md Enamul Kabir

Bowling Green State University, Bowling Green, OH 43402, USA

[mkabir@bgsu.edu](mailto:mkabir@bgsu.edu)

<https://orcid.org/0000-0003-3810-4000>

Md Enamul Kabir is a Doctoral student at Bowling Green State University in the United States. Kabir studies the use of computational methods in communication, with a focus on big data analytics, natural language processing, and machine learning applications. His research centers around communication and racism, social media activism, misinformation, and pedagogy. His recent award-winning research created a novel scale to gauge ‘the degree of racial identity’ in modern American society.