

THE DIFFERENT FACETS OF COVID-19 IN ACADEMIA AND PUBLIC: A COMPARISON STUDY BETWEEN COVID-19 LITERATURE AND SOCIAL MEDIA TEXT

Jeevan Kumar Tirumalagiri
University of North Texas, Denton, Texas, USA

Guru Prakash Yanamala
University of North Texas, Denton, Texas, USA

Dakshayani rayini
University of North Texas, Denton, Texas, USA

Jagan Mohan Reddy Nasipireddygar
University of North Texas, Denton, Texas, USA

Srikanth Nampally
University of North Texas, Denton, Texas, USA

Nacaesa Pink
University of North Texas, Denton, Texas, USA

ABSTRACT

This research will be fundamental in enabling individuals to understand the various information dissemination techniques, the role of technology in information dissemination and the factors involved in determining a suitable information transmission tool. It will help understand the role of transmission mediums during periods of uncertainties, the advantages and disadvantages.

GitHub Link: <https://github.com/jeevankumar20/INFO-5810-Project->

KEYWORDS

Algorithms, Models, Word cloud, Misconceptions, Natural Language Processing, Labels.

1 INTRODUCTION

1.1 BACKGROUND

The onset of Covid-19 global pandemic presented a number of challenges and uncertainties across all nations. From paralysis of global economies, business activities, human interactions and ultimately mandatory in-house restrictions. With majority of the world's population locked up in their homes, mobile phones through social media were the only efficient tool that allowed communication and through which information regarding the pandemic would be shared

from different sources to individuals and from person to person. There were several instances where fictitious news and information was disseminated and shared to the public through various social media platforms leading to confusion and misinformation among the public. The aim of this paper is to investigate the various misconceptions propagated during spread of Covid-19 information and build a model that correctly predicts if a tweet propagates misconception, is informative by contradicting the misconception or whether it is neither mis-informative nor informative.

1.2 SIGNIFICANCE OF THE STUDY

This study will be crucial in enabling us to understand the advantages and disadvantages of the various information platforms and more importantly identify the most reliable source of information for future events. Very little research had been done prior to the 2019 Covid-19 global pandemic. Such an unprecedented pandemic happened decades ago before the development and launching of social media. The various academic research papers published give little details on the effects of social media during Covid-19 with little to no evidence on comparative studies on the Covid-19 literature versus social media.

1.3 RESEARCH QUESTION

Some of the various research questions formulated in this study are listed as follows;

Frequent words used during the spread of Covid-19 information?

Frequent words contained in tweets labelled as propagating misconception?

Frequent words contained in tweets labelled as informative by contradicting the misconceptions?

Frequent words contained in tweets labelled as neither mis-informative nor informative?

1.4 RESEARCH PURPOSE

The goal of this research is to provide the most efficient and reliable transmission media after a thorough comparison between social media as a tool for information transmission and academic literature. We will specifically work on a COVIDLIES dataset for misconception detection

on Twitter, to identify the spread of misinformation such as, rumors, hoaxes, fake news, propaganda, spear phishing, and conspiracy theories and annotate whether any of the known misconceptions, curated by researchers, are expressed by these tweets.

1.5 RESEARCH METHODS

One of the methods for data collection and analysis that will be suitable in answering the research questions formulated will include data collection from internet and academic sources. This coupled by data analysis and visualization will help us correctly answer our research questions

2 LITERATURE REVIEW

Several literature sources that investigate the role of social media as an information transmission tool have dwelled much on the misinformation aspect of it. There is a limitation on the number of publications that actually indicate the characteristics of information transmitted through social media. As indicated by (Xing, et al.,2021), this limitation is mainly due to the unicity of the social media platforms and the challenges of expanding data sets.

(Chen,et al.,2021) indicated social media is one of the fast and efficient information transmission channel that is reliable in transmitting covid-19 related news. The author further indicated some of the downsides of social media since misinformation can easily be transmitted to large number of individuals through the various social media platforms.

(Shahi,et al.,2021) indicated that social media is one of the tools that is actively involved in the dissemination of misinformation through fake news. Basing their research on Twitter; one of the most popular social media platforms that is involved in the dissemination of misinformation, the author investigated the various Covid 19 misinformation topics, their authors and the various ways that influence transmission of misinformation. The author further indicated that fake news on twitter had the ability to spread faster than information which is partially faux. This detail is further backed by (Vosoughi et al.,2018) who indicated that in some instances, faux news has the ability to spread faster than the actual news.

(Lovari,2020) based their research on the widespread of misinformation through social media platforms in Italy. The author stated that there was a public mistrust on the information issued by the science and health entities leading to efforts by the government that allowed the government to use its social media platforms to mitigate the public through policies on the awareness of misinformation being spread. (Tidy, 2020) illustrated the efforts by the

various governments towards mitigating the spread of misinformation through social media platforms.

(Velasquez et al.,2021) investigated the negative impacts of malicious covid-19 information and misinformation in spreading hate across a number of social media platforms. The author further indicated the various ways through which this malicious covid-19 information is transmitted. Consequently, mathematical models were suggested in an attempt to control the spread and the transmission of malicious covid 19 information on social media platforms.

(Shahi, & Nandini,2020) constructed a fake news detection data model. The authors further developed a machine learning classifier to identify and expose misinformation regarding Covid 19 .

A number of authors have established the role of social media in the spread of misinformation. (Muzy, et al.,2021) indicated that social media was one of the entities that was associated with the spread of fake news regarding Covid-19 and was a major source of mental health issues such as anxiety and depression.

(Al-Zaman,2020) discussed and analyzed the various forms of fake news and the various formats through which they are transmitted through social media. The author described infodemic as a term used to describe the various types of information; fake information and true information in the event of a pandemic. Further, the author stated that some of the most developed countries in the world had already established the various ways through which fake news can be identified and discarded while less developed countries still struggled with the dissemination of fake news.

(Rocha, et al.,2021) conducted research that investigated the various effects of fake news and misinformation during the pandemic. The author highlighted that the spread of fake news and various fake news during the pandemic introduced enormous amounts of fear, panic and anxiety among the public and highlighted that this effect has influences individuals of specific ages. In the study, the author also stated that fake news traversing public health has negative effects such as skepticism toward entities such as the health researchers, physiologists and the government.

A number literature sources that investigate the role of social media as an information transmission tool have dwelled much on the misinformation aspect of it. There is a limitation on the number of publications that actually indicate the characteristics of information transmitted through social media. As indicated by (Xing, et al.,2021), this limitation is mainly due to the unicity of the social media platforms and the challenges of expanding data sets.

(Chen,et al.,2021) indicated social media is one of the fast and efficient information transmission channel that is reliable in transmitting covid-19 related news. The author further indicated some of the downsides of social media since misinformation can easily be transmitted to large

number of individuals through the various social media platforms.

(Shahi, et al., 2021) indicated that social media is one of the tools that is actively involved in the dissemination of misinformation through fake news. Basing their research on Twitter; one of the most popular social media platforms that is involved in the dissemination of misinformation, the author investigated the various Covid 19 misinformation topics, their authors and the various ways that influence transmission of misinformation. The author further indicated that fake news on twitter had the ability to spread faster than information which is partially faux. This detail is further backed by (Vosoughi et al., 2018) who indicated that in some instances, faux news has the ability to spread faster than the actual news.

(Lovari, 2020) based their research on the widespread of misinformation through social media platforms in Italy. The author stated that there was a public mistrust on the information issued by the science and health entities leading to efforts by the government that allowed the government to use its social media platforms to mitigate the public through policies on the awareness of misinformation being spread. (Tidy, 2020) illustrated the efforts by the various governments towards mitigating the spread of misinformation through social media platforms.

(Velasquez et al., 2021) investigated the negative impacts of malicious covid 19 information and misinformation in spreading hate across a number of social media platforms. The author further indicated the various ways through which this malicious covid 19 information is transmitted. Consequently, mathematical models were suggested in an attempt to control the spread and the transmission of malicious covid 19 information on social media platforms.

(Shahi, & Nandini, 2020) constructed a fake news detection data model. The authors further developed a machine learning classifier to identify and expose misinformation regarding Covid 19.

A number of authors have established the role of social media in the spread of misinformation. (Muzi, et al., 2021) indicated that social media was one of the entities that was associated with the spread of fake news regarding Covid-19 and was a major source of mental health issues such as anxiety and depression.

(Al-Zaman, 2020) discussed and analyzed the various forms of fake news and the various formats through which they are transmitted through social media. The author described infodemic as a term used to describe the various types of information; fake information and true information in the event of a pandemic. Further, the author stated that some of the most developed countries in the world had already established the various ways through which fake news can be identified and discarded while less developed countries still struggled with the dissemination of fake news.

(Rocha, et al., 2021) conducted research that investigated the various effects of fake news and

misinformation during the pandemic. The author highlighted that the spread of fake news and various fake news during the pandemic introduced enormous amounts of fear, panic and anxiety among the public and highlighted that this effect has influences individuals of specific ages. In the study, the author also stated that fake news traversing public health has negative effects such as skepticism toward entities such as the health researchers, physiologists and the government.

(Melki, et al., 2021) investigated and analyzed the negative effects of misinformation and the spread of fake news through social media during the pandemic. The author indicated some of the areas of misinformation that include Covid-19 vaccines, conspiracy theories about covid-19 virus and the covid-19 control measures.

(Apuke & Omar, 2021) investigated some of the factors that contribute towards the spread of fake news through social media. Some of these factors highlighted include the easy availability of news on social media, over trust in social media news sources and effective communication exchange through social media.

(van Der Linden, et al., 2021) indicated that a lot of fake news was widespread during the onset of covid -19 pandemic through social media. In the study, the author suggested some of the human behavioral characteristics that can be effectively used in an attempt to mitigate the spread of misinformation.

In a study by (Larson, 2020), the author describes the challenge of misinformation arising from social media usage during the pandemic. The author further highlighted other effects arising from fake news leading to uncertainties, incomplete information and discordant information.

Additionally, (Cinelli, et al., 2020) investigated the various social media platforms and determined the various activities of the users of these platforms during the covid-19 pandemic period. The data obtained by the author illustrated that the user activities across all social media platforms varies on the basis of the topic to be discussed. The author investigated the rate of interaction across the topic of covid 19 during the pandemic period and concluded that the rate of information dissemination in a particular platform is independent of whether the information is true or not.

Despite (Goel & Gupta, 2020) demonstrating a number of advantages and applications of social media in information dissemination during covid 19, the authors additionally presented some of the limitations of social media during information transmission. Some of these disadvantages include misinformation, misinterpretation of information transmitted, cyber bullying, patient privacy breach among other limitations.

(Venegas-Vera, 2020) illustrated the various ways through which the different social media platforms enabled the widespread of misinformation. The author highlighted and indicated that the fast transmission of fake news has a very high likelihood to deteriorate situations during crises.

In an attempt to investigate issues of misinformation during covid-19 through social media platforms, (Neely, et al.,2021) reported that quite a number of individuals heavily relied on social media for dissemination of information during the pandemic. The author highlights the rapid increase on reliance of social media channels so as to transmit public health information. social media platforms such as Facebook and Twitter have been mentioned to be some of the social media platforms that continue to pose a risk in the credibility of the news generated and transmitted through them. The author concluded by stating that health care providers and professionals need to formulate effective strategies that will ensure the usage of social media effectively to minimize risks of misinformation through fake news.

Authors such as (Haghani & Bliemer,2020) in their covid 19 research papers reported on the significant number of literatures published within a short time. Compared to the previous pandemics, the authors reported that the uncertainties surrounding covid 19 was a key factor that allowed many researchers to investigate this area. Further, this paper investigates the various efforts by scientists and researchers around the globe in understanding the pandemic. The author also raised the concern on the future direction of the research publication on the pandemic and predicts that more literature on covid 19 will continue to be published.

(Gianola,et al.,2020) described the various types of publications on covid 19 and their features. The author indicated that more than 3600 publications and more than 3800 preprints were retrieved from the various registers and databases. The author indicated that the bulk of the literature published within the first few months of the pandemic comprised of mainly brief reports based on opinions of the authors and was mostly described as the letter to the editor. These brief descriptions focused on secondary data reports and not primary data sources. The author further stated on the importance of publishing credible and reliable scientific literature on the pandemic highlighting that extravagance of numerous publications within a short period of time can lead to public mistrust of scientific facts.

(Sotiropoulou, et al.,2022) analyzes the debate on whether or not sufficient and accurate information regarding Covid-19 is indicated in the recently published scientific publications regarding Covid-19. The author tried to uncover some of the underlying criticisms and further encourages health professionals to gather relevant and high-quality data on Covid-19 in order to generate sufficient information on Covid-19. This will ensure that the quality of article published by health professionals is of high quality.

(Liu,2020) stated that some of the earliest medical publications of Covid-19 were published by Asian authors. The author indicated the various research gaps in the publications of Covid-19 literature and suggested the numerous advantages that would emanate from such

research publications. Some of these research gaps mentioned a d need to be included in future studies included the impact of covid-19 on the mental health of the public, health workers and the covid-19 victims. The author also addressed on the need for health researchers to focus on the role of Artificial intelligence in covid-19 spread prediction and analysis.

(Picciariello, et al.,2020) highlighted Covid-19 pandemic as one of the most interesting areas for research. Additionally, (WHO,2022) has demonstrated a number of efforts towards research on the various measures useful in the mitigation of the spread of the virus. The organization further highlighted on its joint efforts with healthcare professionals and scientific researchers across the world towards ensuring rapid development process. From the organizational website, the Covid-19 database is included which contains approved publications consisting of accurate and relevant information.

(Lou, et al.,2020) identified and investigated a number of research publication of covid-19 virus gathered from the various researchers worldwide. More than 180 research publications were identified and collected from the various institutions such as universities, hospitals and from the research organizations.

(Tran,2020) additionally identified and analyzed a number of research publications on Covid-19. The author extracted the data from a number of databases from which more than 5000 publications were identified. A majority of these publications were from countries such as the United States and China. The author also indicated that the economy of a country determined its efforts towards curbing the spread of the virus.

In a study by (Farooq, et al.,2021) the author indicated that an approximate of more than 6000 research publications were extracted from the Web of Science database. These were papers published between the years 2019-2020. One of the key highlights mentioned in this paper indicated that countries such as China and the United States were among the leaders in in Covid-19 pandemic literature publications.

(Liu,2021) additionally analyzed the various publications between 2019-2020. The author concluded by indicating that the various papers were associated to each other directly and indirectly and highlighted that referencing and citation of these papers from one paper to the other gave room for future research work to be conducted on the same.

(Wang,et al 2020) mentioned that the United States was the leading country in conducting Covid-19 researches and collaborations. The author further investigated some of the research topics that have been analyzed over time and additionally determines the trends in the research of the Covid-19 pandemic.

Additionally, China was also leading on the forefront towards ensuring that sufficient research and publications on public health and general viral information on Covid 19 was developed.

(Aristovnik, et al., 2020) attributed the rapid research strategies on the novel covid-19 virus and the effects of the pandemic to the lack of existing literature. Most of these publications would have been developed in 2020 and a number of research gaps have been noted. The author in his research, investigated the various literature published from the Scopus database and used the various bibliometric approaches.

(Mondolfi, et al., 2020) also conducted a bibliometric analysis to determine whether or not sufficient research on Covid-19 had been carried out. The author gathered data on covid 19 pandemic from the various medical journals that include PubMed, Scopus and Science Citation Index. Some of the findings obtained from this research indicated the various countries that are leading in publishing Covid-19 research papers and generally the trend of research from 1996 January to 2020 January on the same.

Similarly, a bibliometric analysis on the covid-19 pandemic was conducted by (Hossain, 2020) to investigate the various research gaps on the topic. The author highlighted that less was known about the pandemic and encouraged for more research to be carried out. Data mining techniques using the relevant software were used to analyze the various data obtained from the Web of Science. Research collaborations were additionally cited as a strategy through which all global economies can work together towards ensuring that the effects of covid -19 across all countries in the world are thoroughly investigated with sufficient evidence.

In an attempt to understand the various research topics on Covid-19 virus and the pandemic, (Dehghanbanadaki, et al., 2020) carried a scientometric research through which numerous publications were extracted from Scopus and analyzed. The author indicated that the publications analyzed investigated in detail the key areas surrounding the virus such as the structure of the covid-19 virus and how it came into existence, the rate of spread and the severity of the virus, how it is transmitted from one person to another, how it is detected in the human body, how it is treated and the mitigation procedures to prevent the spread of the virus.

(Latif, et al., 2020) conducted research that reviewed the various literature publications on Covid 19. The author further analyzed the various data sets from the various data repositories and analyzed this data to determine the rate of spread of the virus and also compare the covid-19 prevention strategies. The author introduced data science techniques that were aimed towards enabling easy identification of research challenges and issues identified during bibliometric analysis of the existing covid 19 literatures. Moreover, the author highlighted data science as an effective strategy towards ensuring that covid 19 prevention strategies are effective. Furthermore, the author indicated that data science will give room for collaboration between machine learning researchers, data analysts and healthcare experts towards ensuring that the applications of data science in combating challenges of Covid 19. Some of

the various areas of applications described in this paper include remote expert systems for covid 19 diagnosis, contact tracing and development of simulation and modelling tools such as epidemic models so as to understand and predict the behaviors of the virus.

3 DATA COLLECTION

Misconceptions were extracted from a Wikipedia article about misinformation related to the COVID-19 pandemic (Wikipedia, 2020). The extracted statements were manually examined, and statements that are not misinformation were removed. The misconceptions statements were manually to a positive expression of that misinformation.

We collected COVID-19-related tweets identified by Chen et al. (2020). We used tweets from March and April 2020, and filtered out non-English tweets.

BERTSCORE (Zhang et al., 2019) was used to identify tweets related to our list of misconceptions, a similarity metric on tweet-misconception pairs was computed. For each given misconception, the 100 most similar tweets were selected for annotation. Each of these tweet-misconception pairs was manually labeled by researchers in the UCI School of Medicine as either: a tweet propagating the misconception, informative by contradicting the misconception or neither mis-informative nor informative tweet.

4 DATA ANALYSIS PLAN

The data collected will undergo critical evaluation and classifications. We are going to use Python Natural Language Processing packages to analyze the data and for visualizations.

Effective and reliable methodologies and techniques will be essential in answering the various research questions identified and formulated for this research. The various phases of this research are illustrated using the diagram below.

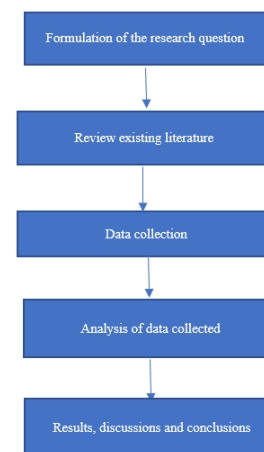


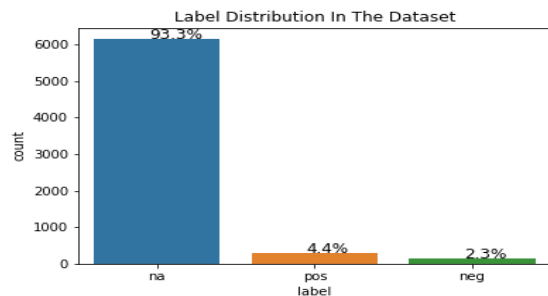
Figure 1: Data Analysis Flow Chart.

4.1 Data Cleaning

We performed the data cleaning in python software. We imported the necessary packages we were going to use.

The dataset has 6591 rows, and 4 columns; misconception id that represents different pieces of COVID-19 related misinformation, misconception variable holds the tweets identifying any of the known misconceptions, tweeter id, and lastly label that represents whether the tweet propagates the misconception (agree/pos), is informative by contradicting it (disagree/neg), or is neither mis informative nor informative (no stance/na). The dataset has two numerical variables and two categorical variables.

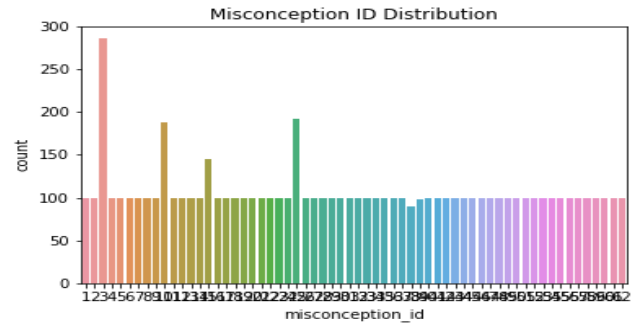
The label variable has 3 unique values; na, pos and neg. The label variables identify 6149 tweets as ones propagating the misconceptions, 288 tweets being informative by contradicting the misconception and 154 tweets as neither mis-informative nor informative. We plot a distribution of each type of label and the percentages. The data is heavily skewed with 93.3 of tweets as na, 4.45% as pos and 2.3% as neg.



The dataset has 62 unique misconception ids, we visualize the count of these ids across the whole dataset. Most of the ids have equal number of frequencies in the dataset that is 100 counts. But there are few ids that have higher frequencies.

The tweets having a count of more than 100 are; 'corona is genetically engineered' having a count of 286, 'corona is caused by 5G' having a count of 182, 'corona virus can only survive in cold temperatures' having a count of 150 and 'Holy communion cannot be the cause of spread of corona virus' having a count of 185. Three of the most frequent tweets have been marked as neither mis-informative nor informative and one marked as a tweeting propagating misconception.

The tweet with a count of less than 100 is 'shaking hands cannot spread corona virus' that has a count of 90 and is a tweet marked as neither mis-informative nor informative.



4.2 Data Preprocessing

The label variable data type is object, this is a categorical variable; we convert it to numerical variable by changing the datatype.

We convert all the text held in the misconception variable holding our tweets to lower case, so that all tweets can follow one format, this will be important later as we feed this variable into our models.

Another important part in data pre-processing is removing the punctuation marks that can reduce the quality of the analysis. We pre-process the misconception variable to get rid of the punctuation marks.

We also remove stop words in the tweets. Stop words are words that have less meaning and could also lessen the quality of the analysis. Stop words include prepositions, articles, conjunctions among other English words.

We tokenize our tweets, this is splitting the sentences into specific words and then assign them to vectors referred to as tokens. This is important since the data analysis models do not accept plain words, they recognize integers.

We assign words that have the same meaning together by stemming; this is grouping words that have similar meaning together for instance words like; fighting, fighter and fight will be stemmed to fight. After stemming we combine the different words it one sentence once again and remove the commas separating the words. We assign the new processed tweets to a new variable named clean_tweet.

5 METHODOLOGY

We train the model on our cleaned tweets, we split the dataset into two portions 75% training and 25% testing set. We vectorize our tweets before we feed them into the model. This is meant to convert the string objects into machine and model readable format for analysis and prediction. The words are set to vectors.

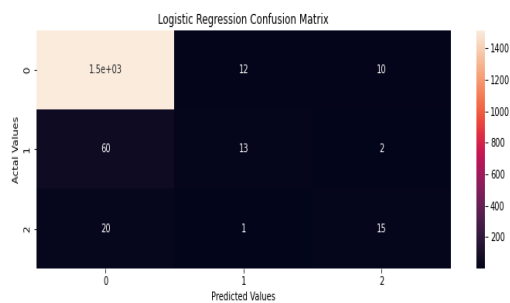
We train our data on a logistic regression model, a K-nearest neighbor classifier and a Multinomial NB classifier. We evaluate our models based on confusion matrix, F1-score and accuracy scores. The confusion matrix

visualizes how the model performed in predicting our tweets in the correct labels.

6 MODEL TRAINING

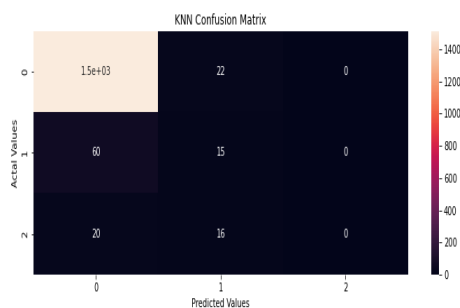
6.1 Logistic Regression

We fit and train a logistic regression model on the dataset. It has an overall accuracy score of 93.62% and an f1 score of 56.7%. The model was able to correctly predict 1515 na tweets, 13 pos tweets and 15 neg tweets in a test set of 1648 tweets. This was picked as the best model as it had the best score and correctly predicted many tweets in the correct label.



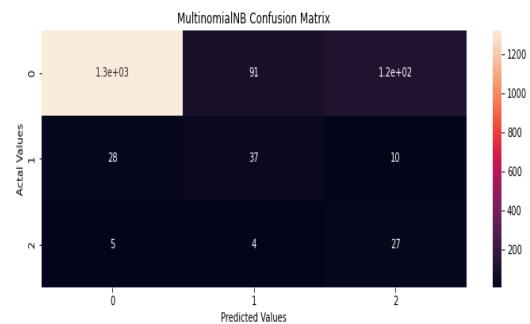
6.2 KNN-classifier

We fit and train a K-nearest neighbor classifier on the dataset. It has an overall accuracy score of 92.83% and an f1 score of 40.06%. The model was able to correctly predict 1515 na tweets, 15 pos tweets and 0 neg tweets, in a test set of 1648 tweets. This model had slightly lower accuracy scores when compared to the logistic model.



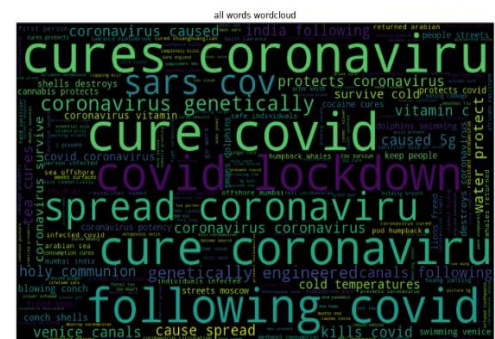
6.3 Multinomial_NB classifier

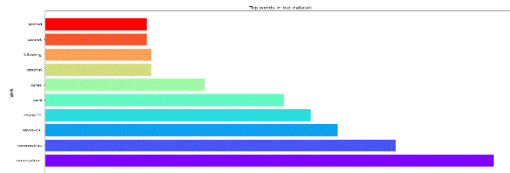
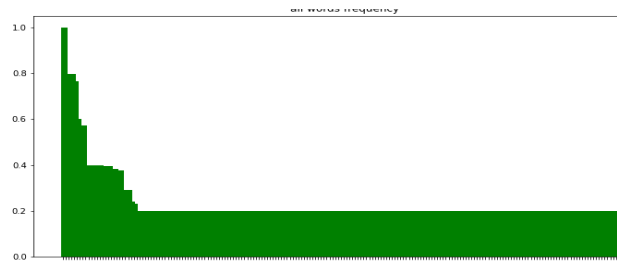
We fit and train a K-nearest neighbor classifier on the dataset. It has an overall accuracy score of 85.19 % and an f1 score of 52.04%.. The model was able to correctly predict 1343 na tweets, 37 pos tweets and 24 neg tweets, in a test set of 1648 tweets. A comparison made between Multinomial NB classifier and logistic regression confusion matrix shows the latter is poor at predicting the correct labels though the former has a better overall accuracy.



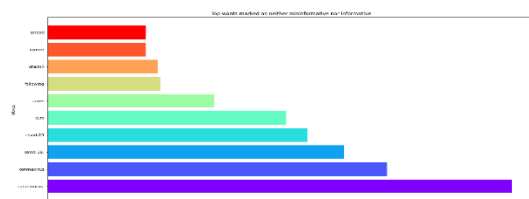
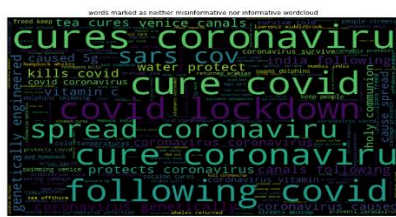
7 RESULTS AND DISCUSSION

We visualize the top repetitive words in the whole dataset using a word cloud. We visualize the top 10 repetitive words. Another graph is also plotted to show the frequencies of each word it reveals most of the words appear equally frequent, only for a few that are more frequent than others.

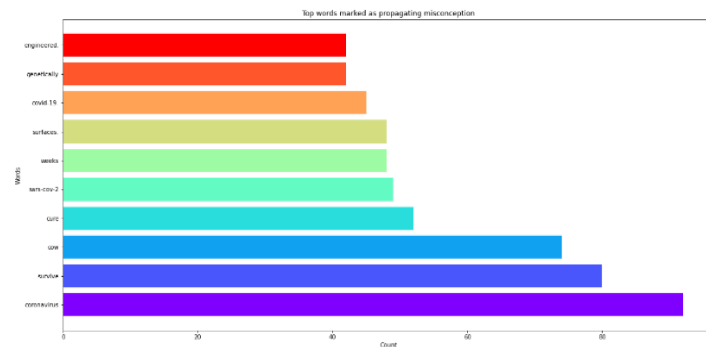




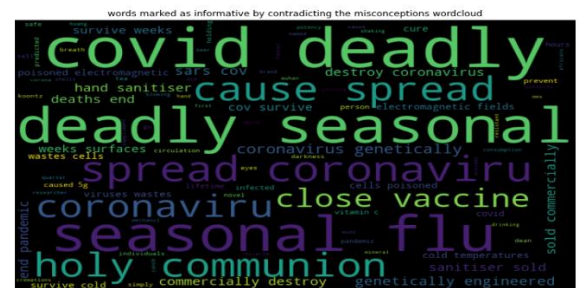
We visualize the top words for tweets labelled as neither mis-informative nor informative in a word cloud and bar plot showing the top words in that label.

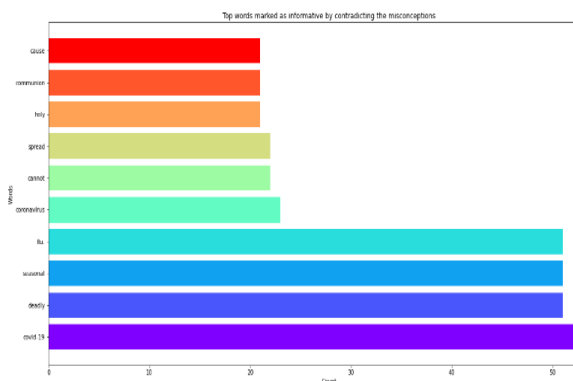


We visualize the top words for tweets labelled as ones propagating the misconceptions as a word cloud and bar plot showing the top words for that label.



We visualize the top words for tweets labelled as informative by contradicting the misconceptions using a word cloud and a bar plot showing the top words for that label.





8 CONCLUSION AND LIMITATIONS

We answer our research questions from the concluded analysis.

Frequent words used during the spread of Covid-19 information?

These are; Covid-19, Corona virus, cure, vitamin, lockdown, spread, prevent and Sars-Cov2.

Frequent words contained in tweets labelled as propagating misconception?

These were found to be words highly associated with propagating misconceptions; corona virus, survive, cow, Covid-19, cure, Sars-Cov2, weeks, surface and genetically.

Frequent words contained in tweets labelled as informative by contradicting the misconceptions?

These words were found to be informative in contradicting the misconception; Covid-19, deadly, seasonal, flu, corona virus, spread, holy, communion and cause.

Frequent words contained in tweets labelled as neither mis-informative nor informative?

These words were found to be neutral, they neither mis-informed or informed the general public about Covid-19 misconceptions; Covid-19, Corona virus, cure, vitamin, lockdown, spread, protects and tea.

The analysis was limited to only one social media platform due to the available dataset and time factor, analysis of datasets from different social media platforms would help us provide extensive insights about our research and study.

9 AUTHOR CONTRIBUTIONS

Jeevan, Jagan, Srikanth, Dakshayani, Guruprakash and Nacaesa altogether have planned and presented the ideas. Research for suitable data for the project is performed by

Dakshaynai and Jagan. Srikanth has carried out Data Cleaning on the selected dataset. Various research papers have been reviewed by Jagan, Jeevan and Prakash. Nacesa, Prakash and Jeevan implemented the model Training. Visualisations are performed collectively by Prakash and Nacaesa. Srikanth, Jagan and Dakshayani aided in interpreting the results. Jeevan wrote the paper with input of all the authors. All the authors have provided critical feedback on the analysis and during discussion in evaluation of the results.

10 REFERENCES

- Al-Zaman, M. S. (2020). COVID-19-related fake news in social media. medRxiv.
- Apuke, O. D., & Omar, B. (2021). Social media affordances and information abundance: Enabling fake news sharing during the COVID-19 health crisis. *Health informatics journal*, 27(3), 14604582211021470.
- Aristovnik, A., Ravšelj, D., & Umek, L. (2020). A bibliometric analysis of COVID-19 across science and social science research landscape. *Sustainability*, 12(21), 9132.
- Bonilla-Aldana, D. K., Quintero-Rada, K., Montoya-Posada, J. P., Ramírez-Ocampo, S., Paniz-Mondolfi, A., Rabaan, A. A., ... & Rodríguez-Morales, A. J. (2020). SARS-CoV, MERS-CoV and now the 2019-novel CoV: Have we investigated enough about coronaviruses?—A bibliometric analysis. *Travel medicine and infectious disease*, 33, 101566.
- Chan, A. K., Nickson, C. P., Rudolph, J. W., Lee, A., & Joynt, G. M. (2020). Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. *Anaesthesia*.
- Chen, K., Luo, Y., Hu, A., Zhao, J., & Zhang, L. (2021). Characteristics of misinformation spreading on social media during the COVID-19 outbreak in China: a descriptive analysis. *Risk Management and Healthcare Policy*, 14, 1869.
- Cinelli, M., Quattrocchi, W., Galeazzi, A., Valensise, C. M., Brugnoti, E., Schmidt, A. L., ... & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific reports*, 10(1), 1-10.
- Dehghanbanadaki, H., Seif, F., Vahidi, Y., Razi, F., Hashemi, E., Khoshmirsafa, M., & Aazami, H. (2020).
- Bibliometric analysis of global scientific research on Coronavirus (COVID-19). *Medical journal of the Islamic Republic of Iran*, 34, 51.
- Deng, W., & Yang, Y. (2021). Cross-Platform Comparative Study of Public Concern on Social Media during the COVID-19 Pandemic: An Empirical Study Based on Twitter and Weibo. *International journal of environmental research and public health*, 18(12), 6487.
- Farooq, R. K., Rehman, S. U., Ashiq, M., Siddique, N., & Ahmad, S. (2021). Bibliometric analysis of coronavirus disease (COVID-19) literature published in Web of Science 2019–2020. *Journal of family & community medicine*, 28(1), 1.
- Garfin, D. R. (2020). Technology as a coping tool during the COVID-19 pandemic: Implications and recommendations. *Stress and Health*.
- Gianola, S., Jesus, TS, Barger, S., & Castellini, G. (2020). Characteristics of academic publications, preprints, and registered clinical trials on the COVID-19 pandemic. *PloS one*, 15 (10), e0240123.
- Goel, A., & Gupta, L. (2020). Social media in the times of COVID-19. *Journal of clinical rheumatology*.
- Haghani, M., & Bliemer, MC (2020). Covid-19 pandemic and the unprecedented mobilization of scholarly efforts prompted by a health crisis: Scientometric comparisons across SARS, MERS and 2019-nCoV literature. *Scientometrics*, 125 (3), 2695-2726.

- Hamilton, J. L., Nesi, J., & Choukas-Bradley, S. (2020). Teens and social media during the COVID-19 pandemic: Staying socially connected while physically distant.
- Hossain, M. M. (2020). Current status of global research on novel coronavirus disease (Covid-19): A bibliometric analysis and knowledge mapping. Hossain MM. Current status of global research on novel coronavirus disease (COVID-19): a bibliometric analysis and knowledge mapping [version 1].
- Koh, J. X., & Liew, T. M. (2020). How loneliness is talked about in social media during COVID-19 pandemic: Text mining of 4,492 Twitter feeds. *Journal of psychiatric research*.
- Larson, H. J. (2020). A call to arms: helping family, friends and communities navigate the COVID-19 infodemic. *Nature Reviews Immunology*, 20(8), 449-450.
- Latif, S., Usman, M., Manzoor, S., Iqbal, W., Qadir, J., Tyson, G., ... & Crowcroft, J. (2020). Leveraging data science to combat COVID-19: A comprehensive review. *IEEE Transactions on Artificial Intelligence*, 1(1), 85-103.
- Liu, N., Chee, M. L., Niu, C., Pek, P. P., Siddiqui, F. J., Ansah, J. P., ... & Ong, M. E. H. (2020). Coronavirus disease 2019 (COVID-19): an evidence map of medical literature. *BMC medical research methodology*, 20(1), 1-11.
- Liu, N., Xu, Z., & Skare, M. (2021). The research on COVID-19 and economy from 2019 to 2020: analysis from the perspective of bibliometrics. *Oeconomia Copernicana*, 12(2), 217-268.
- Lou, J., Tian, S. J., Niu, S. M., Kang, X. Q., Lian, H. X., Zhang, L. X., & Zhang, J. J. (2020). Coronavirus disease 2019: a bibliometric analysis and review. *Eur Rev Med Pharmacol Sci*, 24(6), 3411-21.
- Lovari, A. (2020). Spreading (dis) trust: Covid-19 misinformation and government intervention in Italy. *Media and Communication*, 8(2), 458-461.
- Melki, J., Tamim, H., Hadid, D., Makki, M., El Amine, J., & Hitti, E. (2021). Mitigating infodemics: The relationship between news exposure and trust and belief in COVID-19 fake news and social media spreading. *Plos one*, 16(6), e0252830.
- Muzi, S., Sansò, A., & Pace, C. S. (2021). What's happened to Italian adolescents during the COVID-19 pandemic? A preliminary study on symptoms, problematic social media usage, and attachment: relationships and differences with pre-pandemic peers. *Frontiers in Psychiatry*, 12, 590543.
- Neely, S., Eldredge, C., & Sanders, R. (2021). Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. *J Med Internet Res*, 23(6), e29802.
- Picciariello, A., Gagliardi, G., & Altomare, D. F. (2020). It's COVID o'clock. *The British Journal of Surgery*.
- Rocha, Y. M., de Moura, G. A., Desidério, G. A., de Oliveira, C. H., Lourenço, F. D., & de Figueiredo Nicolette, L. D. (2021). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: A systematic review. *Journal of Public Health*, 1-10.
- Shahi, G. K., & Nandini, D. (2020). FakeCovid--A multilingual cross-domain fact check news dataset for COVID-19. *arXiv preprint arXiv:2006.11343*.
- Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2021). An exploratory study of COVID-19 misinformation on Twitter. *Online social networks and media*, 22, 100104.
- Sotiropoulou, M., Schizas, D., Mulita, F., Maroulis, I., & Vailas, M. (2022). The impact of "COVID-mania" on medical scientific literature. Are we facing a publication pandemic?. *European Surgery*, 54(2), 126-127.
- Tidy, J. 2020. Coronavirus: Facebook alters virus action after damning misinformation report.
- Tkáčová, H., Pavlíková, M., Jenisová, Z., Maturkanič, P., & Králik, R. (2021). Social media and students' wellbeing: An empirical analysis during the covid-19 pandemic. *Sustainability*, 13(18), 10442.
- Tran, B. X., Ha, G. H., Nguyen, L. H., Vu, G. T., Hoang, M. T., Le, H. T., ... & Ho, R. C. (2020). Studies of novel coronavirus disease 19 (COVID-19) pandemic: a global analysis of literature. *International journal of environmental research and public health*, 17(11), 4095.
- Tsao, S. F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L., & Butt, Z. A. (2021). What social media told us in the time of COVID-19: a scoping review. *The Lancet Digital Health*, 3(3), e175-e194.
- van Der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. *Frontiers in psychology*, 11, 566790.
- Velasquez, N., Leahy, R., Restrepo, N. J., Lupu, Y., Sear, R., Gabriel, N., ... & Johnson, N. F. (2021). Online hate network spreads malicious COVID-19 content outside the control of individual social media platforms. *Scientific reports*, 11(1), 1-8.
- Venegas-Vera, A. V., Colbert, G. B., & Lerma, E. V. (2020). Positive and negative impact of social media in the COVID-19 era. *Reviews in cardiovascular medicine*, 21(4), 561-564.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *science*, 359(6380), 1146-1151.
- Wang, J., & Hong, N. (2020). The COVID-19 research landscape: Measuring topics and collaborations using scientific literature. *Medicine*, 99(43).
- Xing, Y., He, W., Cao, G., & Li, Y. (2021). Using data mining to track the information spreading on social media about the COVID-19 outbreak. *The Electronic Library*.