

TECHNISCHE UNIVERSITÄT ILMENAU

Institut für Praktische Informatik und Medieninformatik

Fakultät für Informatik und Automatisierung

Fachgebiet Datenbanken und Informationssysteme

Term Paper Research Seminar

**Data Science in social media: Graph Analysis and Text Mining in Practice**

**Government Communication in the USA regarding COVID-19**

“**Content analysis of US Government Official Tweets regarding COVID-19”.**

Submitted by

Ahmad Raza 64546,

Shaoni Mitra, 63774,

Samira Rezaei Toroghi 64693,

Thanusree Sridhara 64229

Supervisor:

Prof. Emese Domahidi

Dr. Nadine Steinmetz

Ilmenau, 30/09/22

Table of Contents

Abstract

Introduction

1.1 Relevance

1.2 Selection of Social Media Platform

1.3 Research Questions

**2. Literature review**

3. Approach

3.1Data Extraction and Classification

3.2Data Collection

3.3Data Filtering

4.Conclusion and Outlook

**5.Limitations and Challenges**

6.References

# Abstract

Governments took several measures across the world to stop the spread of the COVID-19 pandemic. Among those measures, wearing of masks, using hand sanitizers, social distancing is considered key to winning against the deadly COVID-19 viruses by healthcare professionals. In this study, we focused on obtaining a dataset that could be used to identify what were the major restrictions imposed by the government based on the different waves of covid 19. We obtained data from Twitter based on accounts, timeline, and keywords and cleaned it through several processing steps.

# Introduction

Coronavirus disease 2019 or COVID-19 is an infectious disease that has globally spread since

its first identification in December 2019 from the metropolis Wuhan, China. The COVID-19 has become a problem for the whole world with millions infected and hundreds

of thousands of dead. COVID-19 outbreak was officially classified as a pandemic by the World Health Organization (WHO) on the 11th of March, 2020 because of the ‘alarming levels’ of spread and severity.

The impact of social networks on people's lives is huge, many do not even fully realize the scale of this phenomenon. With cutting-edge technology, the role of social media in people’s lives is crystal clear (Almori. e al., 2021) and it has become an integral part of life, especially among young people. Moreover, during crisis, the increasingly usage of social media platforms offer an opportunity for the government and health institutions to quickly disseminate relevant information among the public.

The COVID-19 disease has attracted the attention on social media which is ideal for quick updating of information for business organisations in crisis, as well as for government institutions. So government institutions, sharing information online, do it not only for those who read them directly but spread this message to broader masses when the media collects and disseminates it further.

At the moment, social networks are essentially a huge database with digital footprints containing a wide variety of information about hundreds of millions of people around the world, which is also well structured.

# Selection of Social Media Platform

There are multiple reasons to analyse this medium for the purpose of this specific research question, instead of other platforms. Even though for example Facebook or YouTube does have a higher number of users and high frequency of use than Twitter in the USA, it is still a significant social media network from a global perspective.

Twitter, with 353 million monthly active users worldwide has met a significant rise in the daily traffic among COVID-19 agenda topics. So, it provides a great possibility to track the dynamics of the Coronavirus (Aguilar-Gallegos et al. 2020). Twitter represents a powerful public health tool for world leaders to communicate information rapidly and directly on Covid-19 to citizens as compared to the more traditional media such as newspapers, radio, television, surveys which is time consuming and sometimes not feasible to obtain large-scale data (Kaeami. A e al.,2021).

There is hardly any other network that spreads news-relevant information as rapidly as Twitter does. Even official news channels are usually slower in spreading the latest news. This attribute contributes to the fact that Twitter is considered one of the most remarkable news platforms. On Twitter, it is not possible to write long texts which leads to short messages with concentration on the most necessary information. The format is beneficial for quick information gathering and data extraction about the latest governmental measures which contributes to the purpose of this research paper. Lastly, most of the USA authorities or political parties responsible for introducing these measures have a Twitter account. This allows a largely complete analysis of governmental tweets about the coronavirus pandemic.

To understand the impact of COVID-19, it is important to collect data which helps to identify the important regulations taken by the government as a protective measure of COVID-19(R. Tang,2021). This paper serves to examine the governmental communication in the USA and their coverage of restrictions due to the COVID-19 pandemic on social media.

# Research Question

What measures did the us government take to stop spreading of covid-19 and communicate the regulations through social media?

Additionally, sub-research questions have also been formulated in order to narrow down the focus of this particular research topic:

How did the gov communicate through the social media to make wearing mask and using hand sanitizers mandatory?

What rules did the government implement for the non-medical measures and communicated the restrictions using social media?

# Literature Review

The research proposal and data retrieval in this paper bases on a literature review. For this purpose, the authors conducted a literature search on *Google Scholar*. Government communicators in particular should pay attention to the opinions of the public before a problem turns into a disaster since social media in the digital era allows citizens to communicate with others and with organizations. Many academics have been drawn to the two-way communication capabilities of social media to learn how these platforms may be utilized to manage crises effectively and how crisis communication tactics can be used to use social media effectively (Coombs,2019). In terms of issues and crisis management, there has not been much research in public relations on how a government hears the opinions of its people and keeps track of concerns using social media analytics (Chon & Kim, 2022).

This paper's data retrieval and study proposal are based on a literature review. The authors used Google Scholar to search the literature based on the keywords "government" AND "social media" AND "covid" in the papers' titles. By considering of this, 79 results have been founded for this purpose. Those articles were sorted by year which showed 24 articles were published in 2022. Eleven of them were about Twitter, and six of them were most relevant.

Chon and Kim (2022) investigated how a prospective issue becomes a government crisis using a theory-based approach for employing social media analytics. They examined how the federal government dealt with the public health catastrophe as it evolved into a government crisis by using social media data collected during the early phases of the COVID-19 epidemic in the U.S. To comprehend dynamic communication between a government and its digital public before and during a crisis, they suggested the social-mediated crisis, communication model. Their study demonstrated that the COVID-19 scenario changed from a problem to a federal government crisis because Americans blamed the federal government's inadequate crisis management for the COVID-19 virus's spread. The findings also demonstrated that the federal government's incapacity and lack of action hasten the spread of Twitter content. Finally, amid the public health crisis, people's tweeting and retweeting were also influenced by their unpleasant feelings.

Kada et al. (2022) during the first wave of the epidemic in Canada, investigated the account data of 93 Canadian public health and government personnel (December 31, 2019, August 31, 2020). They took all of the English- and French-language tweets, retweets, and related metadata from the study period. They showed that it is possible to examine massive social media datasets using machine learning and natural language processing to gauge public opinion during a public health emergency. According to their findings, the people showed more interest in the Twitter accounts of federal politicians than those of provincial or territorial officials. They assumed that the general populace would have more faith in higher echelons of government. Hashtag trend studies demonstrated how the topic of the Canadian public conversation shifted from COVID-19 mitigation techniques to more recent concerns like COVID-19 mental health consequences. An engagement analysis, a hashtag and trend analysis, and sentiment analysis were used in a triangulation of different data sources and methods to examine public involvement with health and government agencies. In addition, they discovered 11 emotions people had when responding to tweets from Canadian authorities.

The sentiment categories created by Chew & Eysenbach were modified and two additional categories were added. They used the Support Vector Machine technique, which is the most used algorithm for sentiment analysis, in contrast to the authors' analysis approach, which relied on automatic sentiment classification based on search queries of keywords and phrases. They served as an example of how social media can be used to analyze public opinion during a pandemic. They contend that routine examinations of such data by governments can give them and public health professionals access to real-time information on public opinion during a crisis involving public health. They thought the public might get important messages more effectively through social media.

During the COVID-19 pandemic, Mohamad Saleh Hasan (et al, 2022) in their article looked at how risk perception affected self-efficacy and confidence in the government. The researchers collected information from 512 people (students and professors) based in Malaysia during the COVID-19 epidemic to test this idea. The researchers distributed a Google Forms-designed online survey to University of Malaya students and faculty members (UM). They employed partial least squares (PLS) structural equation modeling (SEM) using Smart PLS 3.2.8 to analyze the offered hypotheses (Ringle, Wende, & Becker, 2015). During a public crisis, those who considered the possibility of public health risks were more inclined to have greater trust in the government (Vaughan & Tinker, 2009). Additionally, they discovered that people showed a significant level of adherence to the policies put in place by the Malaysian government to address the COVID-19 threat to the general populace. This finding added to the current understanding of how people might trust their government during a pandemic, underlining that people's perception of danger will boost their trust in the government during any health crisis, such as the COVID-19 pandemic (Slovic, 2000). They suggested that increased use of social media to learn about COVID-19 will strengthen the beneficial association between risk perception and trust in government. Furthermore, these associations were more pronounced when social media was employed to obtain data on COVID-19. In certain instances, it even assisted users in preventing virus exposure. Their study evaluated the link between social media use during the epidemic and increased awareness of risk factors, and also highlighted how social media use affects self-efficacy and trust in the government.

Bambang Irwan (2022) researched the East Kalimantan Provincial Government's Covid-19 pandemic control strategy through social media communication. They used the content, network, and word cloud analysis software QDA Miner, also known as qualitative data analysis miner Nvivo 12 Plus. They obtained their information from the public records of the Provincial Government and the Office of Communication and Information Province East Kalimantan. He suggested that this Twitter account serves as a tool for crisis management and mitigation. Nvivo 12 Plus study of the information content revealed that the policies for handling and expanding Covid-19 cases in East Kalimantan were directly tied to the information provided during the corona outbreak.

Meihua Su (2021) looked into the speech actions and pragmatic strategies of persuasion used by the official Twitter account of the US government agency to promote widespread vaccination for the Covid outbreak. To do this, information from 92 microblogs on the social media platform Twitter between February 2021 was collated.

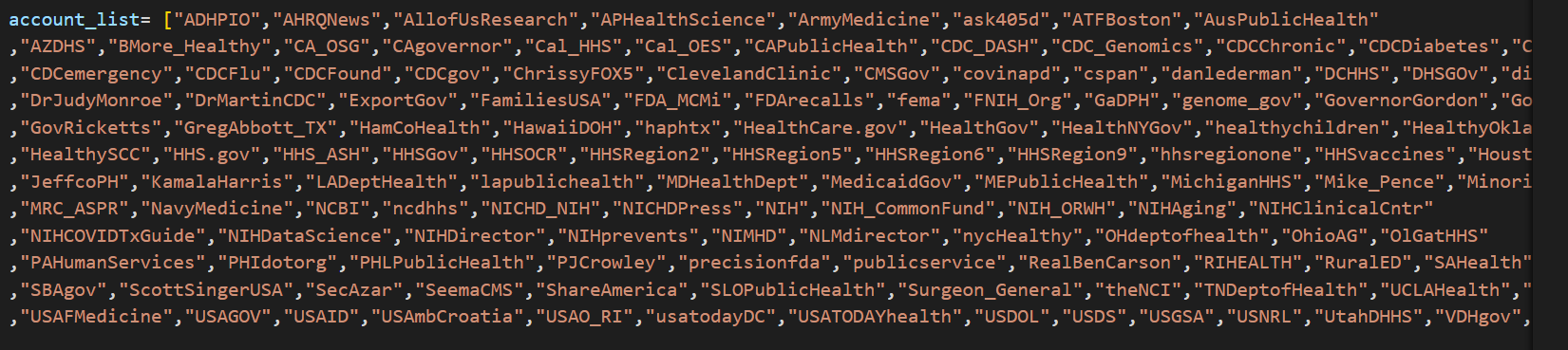
The study used Aristotle's rhetorical techniques for further debate and Searle's speech acts theory to categorize all microblogs. NVivo was used to examine the 92 microblogs (version 12). They employed a qualitative study of the US government's communication tactics taken from the White House Covid-19 Response Team's official Twitter account. Microblogs were initially sorted according to origins, and the strategies with similar purposes were then placed under the same category. The strategies were discovered by observation of the full dataset using the Searle's taxonomy of speech acts.

The findings showed that the official account prefers information from third parties and that it used two of the five speech acts—the forceful speech act and the commissive speech act—to convince the public to be immunized. The official account primarily used the assertive strategy in the current dataset, which appeared to include primarily quotes from health experts rather than original content. The commissive speech act, which is the second-most popular strategy, guaranteed a greater immunization rate throughout the long and short terms.

El Mouloudi (2022) set out to investigate how Health Canada (HC) and the Public Health Agency of Canada (PHAC) interact with Canadians on Twitter and other social media. He sought to determine the types of social media strategies used to carry out crisis communication during the COVID-19 pandemic based on insights from social media practices drawn from the work of Wendling et al. (2013); Lin et al. (2016) and from social-mediated crisis communication theory of Austin et al. (2012). He chose a combination of quantitative and qualitative analysis for this study to uncover meaning categories and their trend through thematic analysis.

He discovered that while PHAC and HC's social media strategies share many characteristics with the theoretical framework. He also offered numerous subtleties in actual practices, which are mostly influenced by the length of the crisis and the uncertainty it has created. This thesis put theory into reality by investigating a current situation and comparing social media usage patterns to communications approaches

## In order to extract all the required information, a total of 138 Twitter accounts was selected and analyzed. The accounts of political parties, federal states, as well as health institutions and spokesmen chosen for the data extraction.



## 

Twitter Accounts

# Approach

# Creating the dataset and methods

The process of building a dataset is restricted to identifying relevant twitter accounts and keywords, to gather the data and of course to clean the data ([Rakshana B. S](https://www.semanticscholar.org/author/Rakshana-B.-S/2114838362), 2021; A. Kumar, 2021). Our goal with the research project is to prepare a cleaned dataset that could be used further.

Twitter offers its API services to collect data. However, it only allowed us to extract data for the last seven days wherein it was expected to analyse data for about 6 months. Therefore, we used web scraping techniques to collect the data for the required duration.

To reach a better understanding of the USA government’s role in COVID-19 on Twitter and to answer our inquiries in this regard, we need to parse meaningful data from this social media. In this section, the procedure of collecting, cleaning, and structuring efficient datasets are elaborated.

**Table 1** demonstrates the list of packages used during this project in various stages of creating the target datasets. These packages are used in Python 3 programming language and executed on Miscrosoft Visual Studio Code

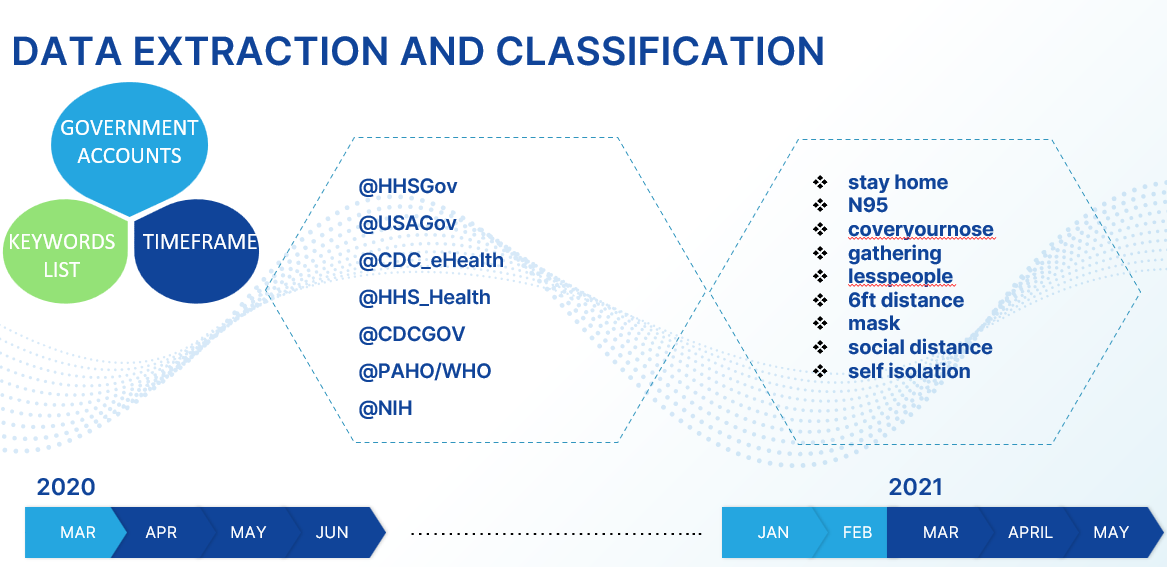
|  |  |
| --- | --- |
| **Package Name** | **Usage** |
| Re | Built in package for Python used to work with Regular Expressions |
| Emoji | It is a module in the Python Package used to create and remove emojis |
| Pandas | Database management and cleaning |
| Matplotlib | Plotting graphs and illustration of Data |
| Nltk | Natural language toolkit is a python pakage that can be used for Natural Language Processing |
| snscrape | Python library for scraping the data. snscrape is a scraper for social networking services (SNS). It scrapes things like user profiles, hashtags, or searches and returns the discovered items |

**Table 1. List of packages used during the project**

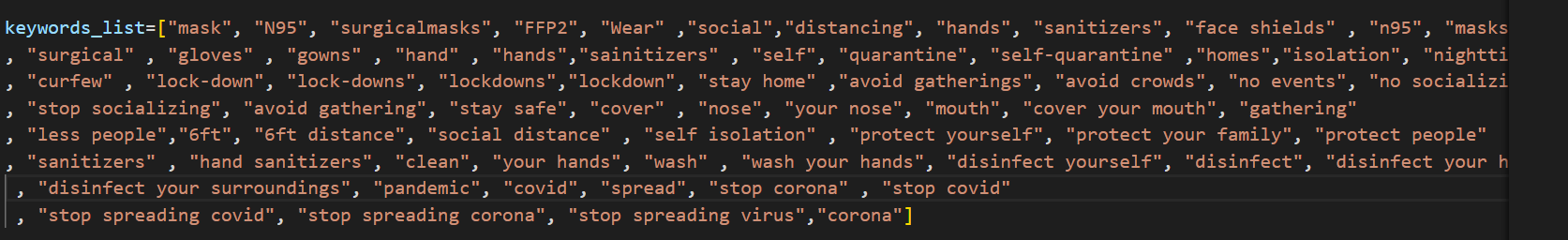
Data Extraction & Classification

For data extraction includes information about the:

* Social media platform (Twitter)
* Analysed accounts (138 official governmental accounts, refer Table 1)
* Time period of analysis (01.03.2021 to 30.06.2021)



To extract data relevant to COVID 19 protective measures, we developed a list of keywords



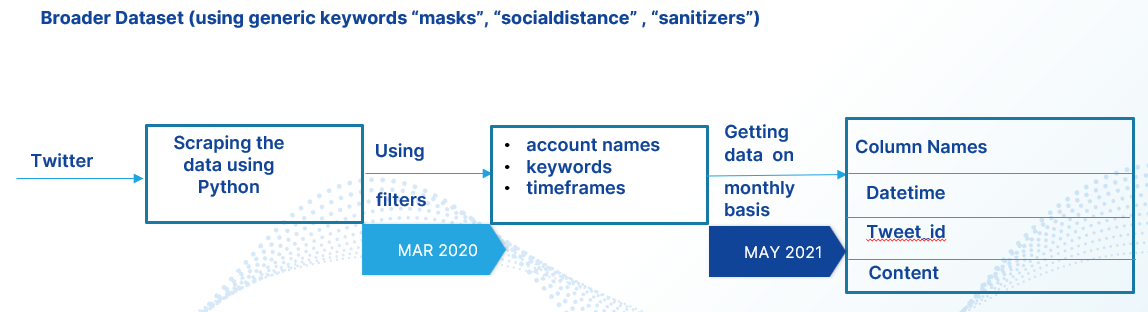
After compiling the code, it was executed on the Visual Studio Code. The data extraction attributes are as follows:

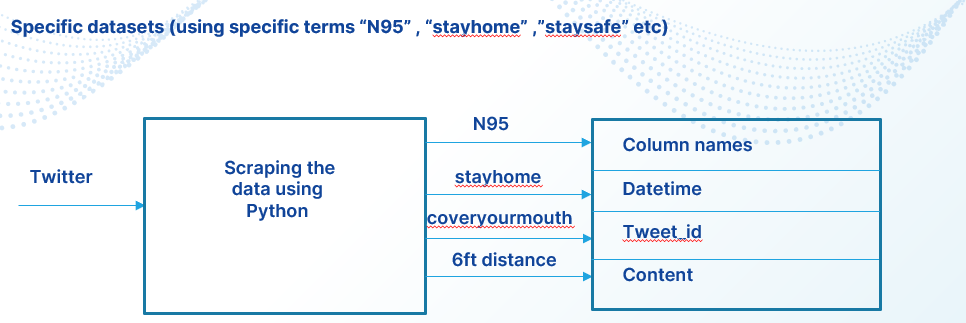
• Date

• Time

• Tweet\_id

• Content





Data Collection

We used an open-source Python Library named Snscrape for gathering data from selected Twitter accounts. Snscrape is an advance tool for Twitter scraping. It allows data scraping from Twitter accounts and profiles. Snscrape utilizes Twitter’s search operators to allow us for scraping Tweets from specific users, scrape Tweets referring to certain topics, hashtags & trends, or sort out sensitive information from Tweets like e-mail and phone numbers.

Snscrape has some major advantages:

Twitter API has restrictions to scrape only the last 3200 Tweets. But Snscrape can fetch most Tweets.

Setting up is quick as there is no hassle of setting up Twitter API.

It can be used anonymously without sign-up in Twitter.

It is free so no pricing limitations.

## 

Data Filtering

The collected data from Twitter is the primary source of data. However, the data in its raw

form cannot be used for information retrieval. Data pre-processing is an important step in

analysing the data from tweets by reshaping the unstructured data into structured data

which helps to visualize in a better way. We have focused on data cleaning process in our project.

Pre-processing

Tweets contain URLs, usernames, hashtags, long words, emojis, slang, misspelled words, etc.

if such features are removed, better insight can be drawn from the data.

*Username*

The presence of username in text-based information provides no information about

sentiments, except for simply expanding dimensionality, ensuring discredited execution of

classifier.

*URL*

Existence of URLs till classification might prompt bogus expectations. For instance, the

extremity of expression "Shaoni signed on to www.good.com" ought to be characterized as

neutral, yet it is countered as positive due to the "good" word. Hence, it is obligatory to

dispose of such events at the beginning phase.

*Punctuation*

A text draft is loaded up with punctuations however they are of least significance for

assessment of feelings. Henceforth, the dismissal of punctuation causes no damage except

for helping decrease the dimensionality of information.

*Special Characters*

Special characters like "&%$#" utilized alongside text are interpreted as one with words

which make it an obscure word for classifier. For instance, words like "#good" isn't perceived

by the classifier due to # symbol.

*Removed Emojis*

Emojis in text can be identified with the facial expressions of an individual while talking. As

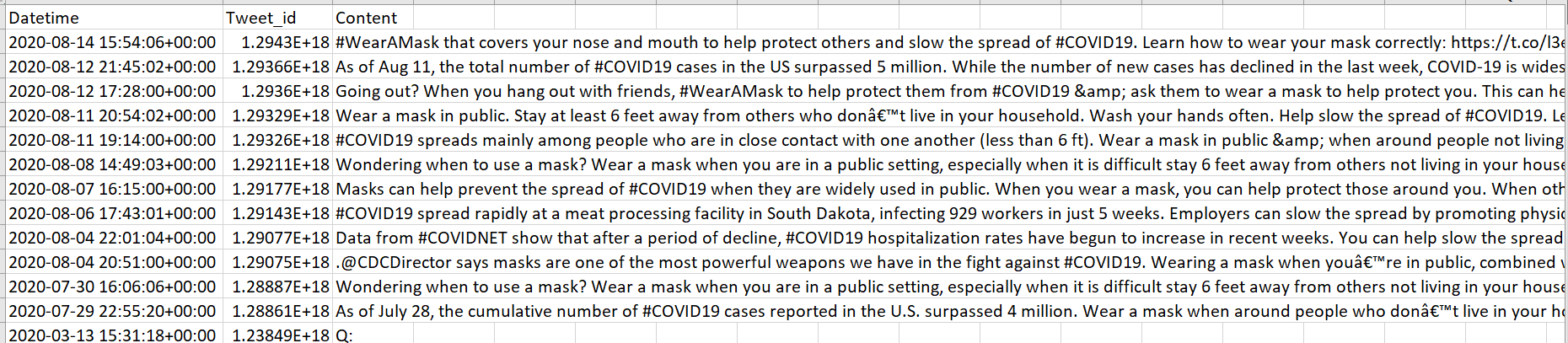
looks clarify best with regards to an individual's sentiments, similarly emojis in messages

assume a critical part in assessment of an individual's perspective.

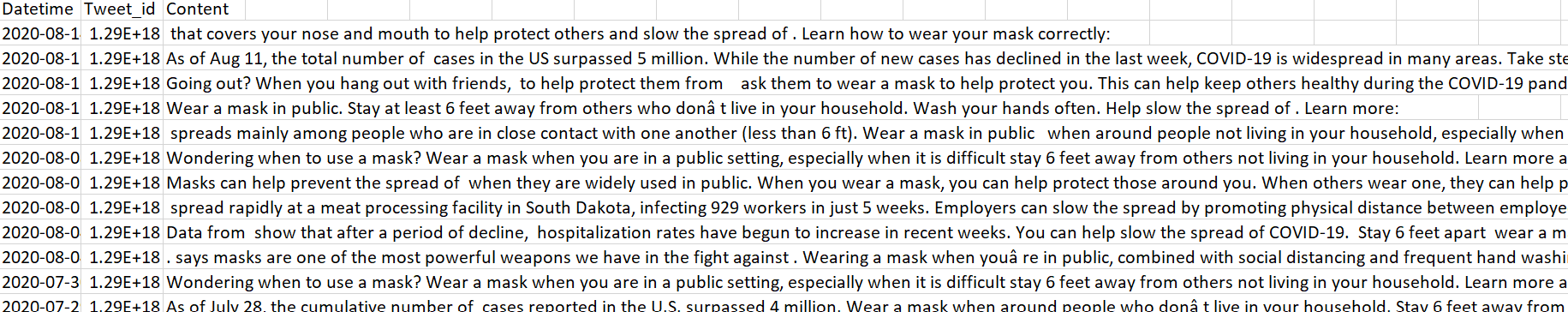
*Remove Duplicates and Null Values*

The data is extracted using a web-scraper; the result may contain duplicates as other users

likely retweet a tweet. In order to avoid having the same tweet, we eliminate the duplicates out ofthe gathered tweets. We have also excluded the columns for which there we null values.



Raw data before cleaning



Data after cleaning

# Conclusion and Outlook

The information demonstrates that COVID-19 widespread disease has affected the lives of people in 223 countries and demands prevention and other measures taken by the officials in power. During this pandemic, social networks played a big role in spreading information among the population. Various governments and health institutes took advantage of it and used social networks to reduce disinformation and disseminate the necessary information to fight COVID-19. Governmental communication strategies applied during the pandemic can be investigated by analysing the content of the provided tweets and finding dependencies in it.The provided data set with 153000 tweets from relevant governmental accounts can be a suitable basis for analysis of governmental communication in the USA and may be applied by communication scientists. However, it should be expanded with the data from other social media such as Facebook, Instagram or YouTube that are also used for informing the citizens about new governmental restrictions in the pandemic. Moreover, the effectiveness of the communication can be measured in future research by combining the provided data with other quantitative indicators such as infection rates and number of vaccinated people.

Further research could focus more on other countries, for example, those in which the national health system is even better or worse than in the US and additionally other non-English speaking countries as well. The list of keywords can be used and developed in other studies about the communication in the coronavirus era.

# Limitations

The main challenges were faced in collecting the data.We were facing issue in fetching the data from Twitter API, it was only allowing us to fetch the data for last 7 days using developers account. So, we wrote the code for web scraping.

Tweepy – Initially we used Tweepy for fecthing the tweets but it provided data only for few months or few days but not year-old data.

Getoldtweets3 is a Python library used for fetching data from twitter but it got deprecated as a result no relevant data was found.

Snscrape – We are getting inconsistent data using only keywords without using specific accounts.

Although vaccination is considered one of the most important preventive measures of covid19 but we are not considering it on our study because it is a separate domain itself.

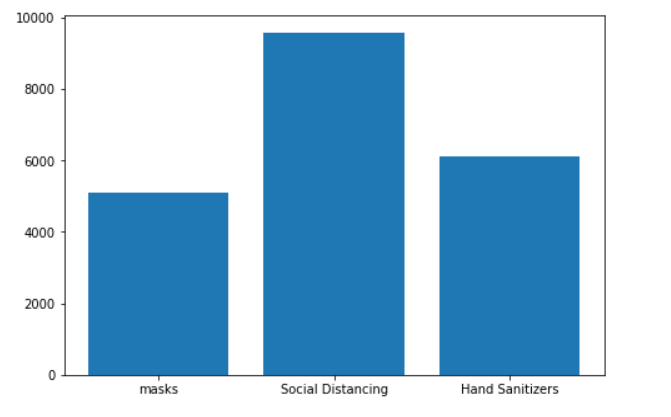


Figure 1. Contains the data related to covid restrictions during the first wave

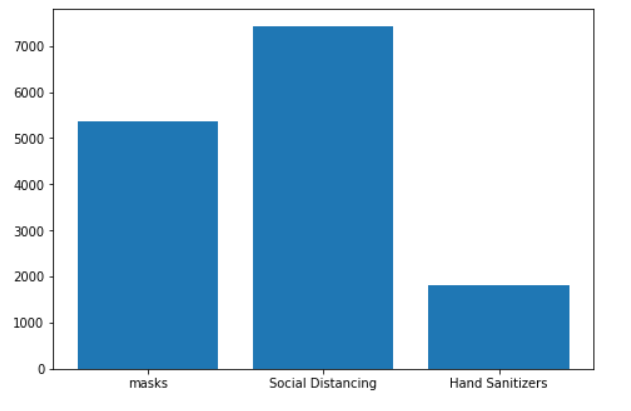


Figure 2. Contains the data related to covid restrictions during the second wave

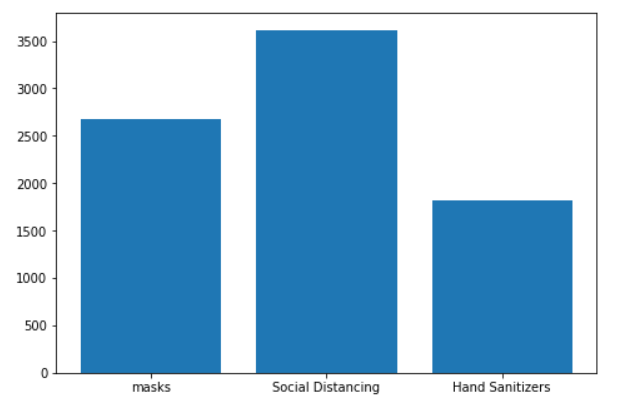


Figure 3. Contains the data related to covid restrictions during the third wave

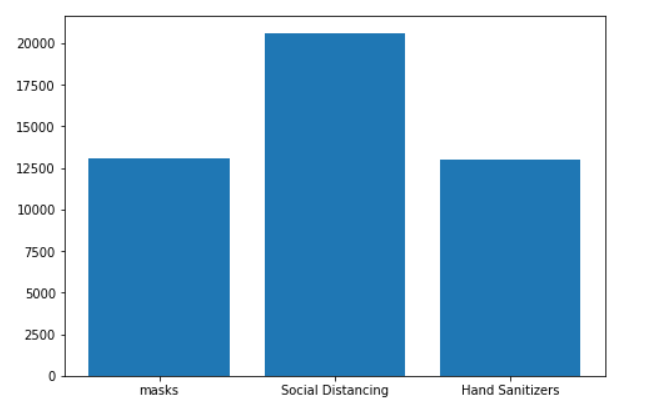


Figure 4. Contains the data related to covid restrictions overall

References

1. https://www.sciencedirect.com/science/article/pii/S036381112200056X

Chon, M. G., & Kim, S. (2022). Dealing with the COVID-19 crisis: Theoretical application of social media analytics in government crisis management. Public Relations Review, 48(3), 102201.

2.https://arxiv.org/abs/2010.04877

Saire, J. E. C., & Zuñiga, E. W. V. (2020). Analysis of Users Reaction around Impeachment in Peru using Twitter. arXiv preprint arXiv:2010.04877.

3.https://ieeexplore.ieee.org/abstract/document/9408930

Tang, R., Zhang, L., Zhang, G., & Wang, J. (2021, April). Analysis of COVID-19 Rebound Based on Natural Language Processing. In 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP) (pp. 333-336). IEEE.

4.https://www.mdpi.com/2076-393X/9/10/1059

Karami, A., Zhu, M., Goldschmidt, B., Boyajieff, H. R., & Najafabadi, M. M. (2021). COVID-19 vaccine and social media in the US: Exploring emotions and discussions on Twitter. Vaccines, 9(10), 1059.

5.https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0273153

Kada, A., Chouikh, A., Mellouli, S., Prashad, A. J., Straus, S. E., & Fahim, C. (2022). An exploration of Canadian government officials’ COVID-19 messages and the public’s reaction using social media data. Plos one, 17(9), e0273153.

6.https://assets.researchsquare.com/files/rs-43836/v3/07043a72-f8d2-4a0f-a6f1-69dd4895dcf9.pdf?c=1656439949

Hassan, M. S., Al Halbusi, H., Najem, A., Razali, A., & Williams, K. A. (2022). Impact of risk perception on trust in government and self-efficiency during covid-19 pandemic: Does social media content help users adopt preventative measures?.

7.https://pubs2.ascee.org/index.php/ijcs/article/view/459

Irawan, B. (2022). Policies for controlling the covid-19 pandemic through social media communications by the East Kalimantan provincial government. International Journal of Communication and Society, 4(1), 125-136.

8.http://download.bcpub.org/proceedings/2021/ECSS2021/ECSS023.pdf

Su, M. (2022). US Government’s Engagement in Social Media: A Case Study of How Government Promote COVID-19 Vaccination. URL: http://download. bcpub. org/proceedings/2021/ECSS2021/ECSS023. pdf [accessed 2022-06-22].

9.https://ruor.uottawa.ca/handle/10393/43180

El Mouloudi, M. (2022). Use of Social Media in Crisis Communication in the Federal Government During COVID-19 Pandemic: Analysis of Responses Strategies (Doctoral dissertation, Université d'Ottawa/University of Ottawa).

10. Bird, S., Klein, E. and Loper, E., 2011. Natural language processing with Python.

Beijing [etc.]: O'Reilly.

11. Han, C., Yang, M. and Piterou, A., 2021. Do news media and citizens have the same

agenda on COVID-19? an empirical comparison of twitter posts. Technological

Forecasting and Social Change, 169, p.120849.

12. Atkins, M. S., Frazier, S. L., Leathers, S. J., Graczyk, P. A., Talbott, E., Jakobsons, L., Adil,

J. A., Marinez-Lora, A., Demirtas, H., Gibbons, R. B., & Bell, C. C. (2008). Teacher key

opinion leaders and mental health consultation in low-income urban schools. Journal of

Consulting and Clinical Psychology, 76 (5), 905–908. https://doi.org/10.1037/a0013036

13. Chen, E., Lerman, K., & Ferrara, E. (2020). Tracking Social Media Discourse About the

COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public

Health and Surveillance, 6 (2), e19273. https://doi.org/10.2196/19273