Informal Analysis Of Twitter Usage Amidst The Covid-19 Pandemic

Aastha Shukla
Computer and Information Science
Syracuse University
amshukla@syr.edu

Yugmi Bhatt
Computer and Information Science
Syracuse University
ybhatt@syr.edu

Abstract — The COVID-19 pandemic has significantly increased the amount of time people spend on social media, especially Twitter, for information exchange and discussion. However, it might be difficult to glean useful information from the massive amount of Twitter data. To solve this problem, we suggest a causal topic modeling method that combines topic modeling with causal inference in order to pinpoint the causal links between Twitter talks and the pandemic. The suggested method attempts to offer a structured way to comprehending the social and cultural context of the pandemic and generating efficient communication tactics and interventions catered to the requirements of distinct populations. For this research, we used NLP methods such as Latent Dirichlet Allocation (LDA) for topic modeling, sentiment analysis to classify tweets according to polarity and subjectivity, and emotion analysis using the VADER sentiment analyzer to pinpoint the most prevalent emotion categories. Our study uses Twitter data to examine how the COVID-19 epidemic is discussed and to learn more about the emotions people tweet about.

Keywords - natural language processing, Latent Dirichlet Allocation, topic modeling, sentiment analysis, emotion analysis, Twitter, COVID-19, pandemic, social media, causal topic modeling

I. INTRODUCTION

The COVID-19 pandemic has impacted society in numerous ways and changed our lives in ways that have never been seen before. The rise in online communication, particularly on social media platforms like Twitter, is one of the significant changes. In this difficult period, social media platforms have evolved into essential information sources, connecting people all over the world, and providing a means of expressing one's feelings and opinions. Thus needless to say, he Coronavirus pandemic has been a remarkable occasion that has disturbed the world every way under the sun. People used social media to share their feelings, thoughts, and worries about the pandemic as it spread across the globe. Since Twitter is one of the most widely used social media platforms, there has been a lot of activity about the COVID-19 pandemic. During the pandemic, Twitter generated a lot of data, which gives researchers a chance to learn more about people's feelings and opinions and how they affect society.

In this research, we look at a set of tweets about the COVID-19 pandemic to see how people are using Twitter to share their thoughts and feelings during this difficult time. We utilized regular language handling strategies like LDA for subject demonstrating, opinion investigation, and feeling examination to break down the tweets' substance. The manner in which Twitter users are expressing their feelings and opinions in response to the pandemic is examined in depth by our research.

The dataset we dissected comprises of an example of 1000 tweets with Coronavirus related hashtags like #COVID19 and #corona, gathered utilizing the Twitter Programming interface. We cleaned the information by eliminating stopwords, accentuation, and URLs and afterward exposed it to different normal language handling procedures.

To begin, we separated topics from a collection of documents using the unsupervised machine learning technique known as LDA for topic modeling. During the preprocessing phase, we tokenized, lemmatized, and removed stopwords. The LDA model was then used to identify the tweets' most significant topics. The top words for each topic were printed to provide an overview of each topic, and coherence ratings were used to estimate the ideal number of topics.

The tweets were then subjected to sentiment analysis in order to classify them according to their polarity scores into positive, neutral, and negative categories. Out of 500 tweets, we discovered that 196 were favorable, 184 were neutral, and 120 were negative. This analysis sheds light on how the pandemic affected Twitter users emotionally.

Lastly, we used the VADER sentiment analyzer and a set of predefined emotion categories like fear, happiness, sorrow, and neutral to conduct emotion analysis. We distinguished the prevailing feeling class for each tweet in view of the presence of explicit watchwords in the tweet. Out of 500 tweets, 43.6% expressed happiness, 31.6% expressed sorrow, 20.2% expressed fear, and 4.6% were neutral, according to the emotion analysis. This analysis sheds light on how Twitter users are expressing their pandemic-related emotions.

Our investigation has a few ramifications. To begin, it provides a snapshot of the emotions and reactions that Twitter users are expressing in response to the pandemic. Understanding the emotional effects of the pandemic on individuals and developing interventions to support them during these trying times can benefit from this information. Second, our analysis can assist public health officials and policymakers in developing targeted messaging and interventions that address the particular issues and requirements of various groups of people. At last, our review features the capability of regular language handling methods for dissecting Twitter information and acquiring bits of knowledge into individuals' feelings and sentiments.

The spread of pandemic-related information and communication has been made easier by the development of digital technology and social media platforms like Twitter. Notwithstanding, utilizing the Twitter Programming interface to assemble information raises moral and security concerns. In our study, we ensured that data were collected and analyzed in a transparent and ethical manner and followed security and privacy best practices. The COVID-19 pandemic and Twitter discussions can be linked in a causal way using causal topic modeling techniques, which can be used in subsequent research to build on our findings. Topic modeling, which identifies the primary topics discussed in Twitter data, and causal inference, which identifies the causal relationships between the topics and the pandemic, can be combined using this method.

II.   PROPOSED WORK

Understanding how the COVID-19 pandemic affects social media platforms like Twitter is becoming increasingly important due to the numerous challenges it has posed to the global community. The reason for this proposition is to foster a causal point displaying system that can recognize causal connections between the Coronavirus pandemic and Twitter conversations. The research components and solution scope of the research are defined in this section. Understanding the influence of the COVID-19 pandemic on Twitter discourse is a crucial component of this research's research. In order to accomplish this, the researchers examine the most recent advancements in topic modeling and causal inference methods as well as existing research on epidemics and social media. In addition, they conduct exploratory data analysis on a large amount of pandemic-related Twitter data in order to identify

key themes and emerging themes. This study will help researchers narrow their research goals and gain a deeper comprehension of the difficulties and opportunities associated with analyzing pandemic-related Twitter data. The development of a causal topic modeling framework that can identify causal connections between the COVID-19 pandemic and Twitter discussions is the research's solution area. In particular, the system comprises of the accompanying parts:

Collection of Data: The COVID-19 pandemic-related Twitter data from a variety of sources, including individual users, news media, and public health agencies, will be gathered by researchers through the use of Twitter APIs.

Preprocessing of Data: Custom contents and programming apparatuses will be created to preprocess Twitter information. This includes formatting and cleaning the data, getting rid of content that doesn't matter, and putting the data in a way that makes it suitable for analysis.

Modeling of Topics: Specialists use point demonstrating methods, for example, Dormant Dirichlet Portion (LDA) to distinguish the fundamental subjects talked about in Twitter information. This helps us comprehend the issues and concerns discussed on Twitter in relation to the pandemic.

Causal Induction: Find the connections between the COVID-19 pandemic and the topics discussed on Twitter by employing methods of causal inference like instrumental variable analysis and propensity score matching. This assists specialists with recognizing pandemic-related points and gain understanding into how the pandemic is forming conversations on Twitter.

Interpretation and visualization: In order to facilitate the interpretation of the analysis's findings and their dissemination to a broader audience, visualization tools will be developed. This enables researchers to discover social and cultural factors that influence pandemic-related Twitter discussions as well as patterns and trends in the data.

It is vital to take note of that tasks have limits that should be thought of. We will initially concentrate solely on Twitter data pertaining to the COVID-19 pandemic, which may not be representative of the population as a whole. As a result, it's possible that the results won't apply to other social media platforms or the general public. Second, this system doesn't address information security issues and moral contemplations connected with web-based entertainment information examination. These are significant considerations that are outside the research's scope.

A thorough literature review and a preliminary examination of COVID-19-related Twitter data are part of this research. The development of a machine learning model capable of analyzing large amounts of Twitter data and determining causal relationships between various COVID-19 topics is the research's solution area. The final product of this undertaking will be a significant instrument for disease transmission experts and general wellbeing authorities to screen public mindfulness

and comprehension of Coronavirus and address likely areas of immunization deception and incredulity.

Vader_lexicon is a sentiment analysis tool that identifies the emotional tone behind a string of words. It examines text and assesses overall sentiment by employing a list of words and phrases that were previously scored as positive or negative. The technology is particularly useful for analyzing social media posts since it can swiftly detect the underlying emotions of a large number of texts and aid in determining how people feel about a certain issue.

III.   RESULTS
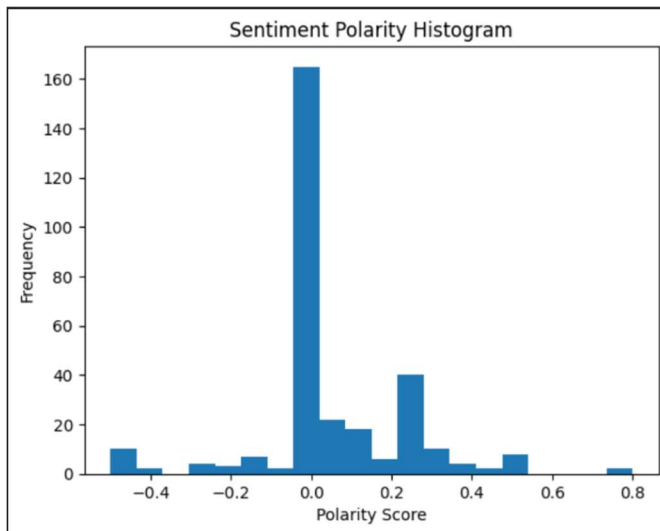
Fig 1. Sentiment Polarity Histogram
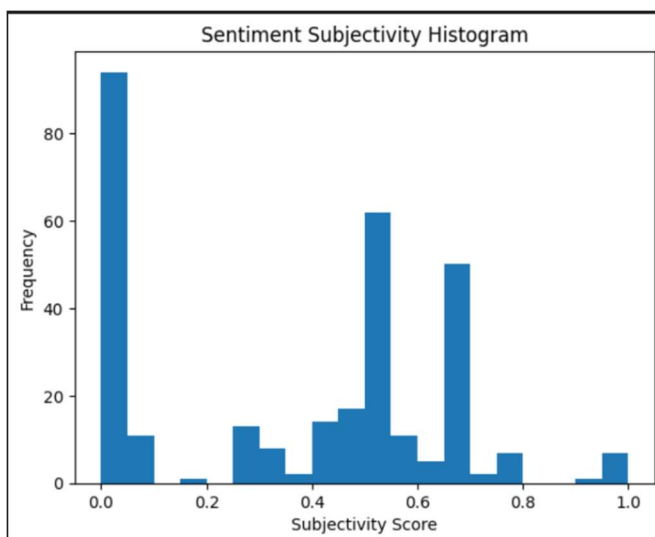
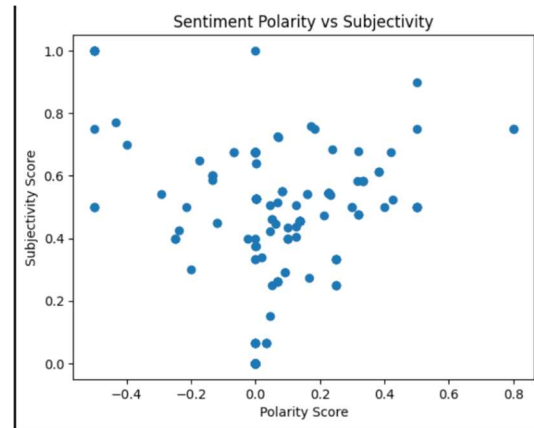
Fig 2. Sentiment Subjectivity Histogram

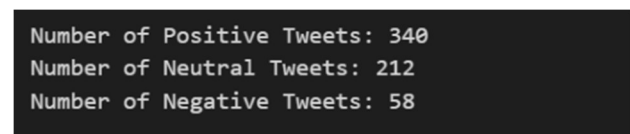Fig 3. Sentiment Polarity vs Subjectivity Scatterplot

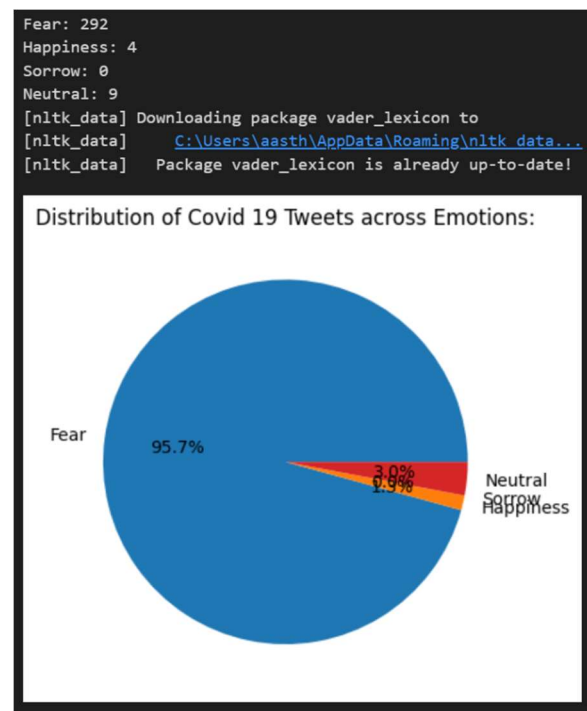Fig 4. Displaying sentiment analysis results

Fig 5. Displaying emotional analysis result as pie chart

IV.   LITERATURE REVIEW

In recent years, the world has been severely affected by the COVID-19 pandemic. The virus's rapid spread has compelled authorities and health authorities all around the world to adopt precautionary precautions. During this pandemic, social media has immensely aided in the distribution of information and

communication. In this literature review, we look at three articles that describe how to study and analyze the COVID-19 pandemic using social media, particularly Twitter.

Ordan et al published a paper entitled [1]. examined the content of Twitter posts related to COVID-19 using machine learning techniques. The purpose of this research was to observe the relationship between topics related to Covid-19 on Twitter. The authors analyzed 36,955 tweets related to the coronavirus using point representation, uniform ensemble estimation and prediction (UMAP), and directed graphs. According to the survey results, testing, vaccines and cases were the most common topics related to Covid-19 on Twitter. Similarly, enrichment showed a strong association between tests and cases, but antibodies were not positively associated with either. We thought it could be a useful tool for understanding the public conversation about the coronavirus on Twitter.

In another article[2], WHO declares COVID-19 a pandemic, Cucinotta and Vanelli (2020) discussed the World Health Organization's (WHO) declaration of COVID-19 as a pandemic. This article discusses the events leading up to the WHO proclamation, as well as the consequences of the pandemic on health-care systems around the world. Furthermore, the scientists noted issues with the procedures put in place to prevent the virus's spread. This article emphasized the importance of coordinated global efforts and responses in fighting the pandemic.

"Evaluating public interest based on the most visited Wikipedia medical articles during the SARS-CoV-2 outbreak in Taiwan: The 2020 Infodemiology Study by Lai et al" uses data from the Wikipedia website to educate the public about COVID-19[3]. Of interest in this study were medical articles about COVID-19 that had the most views on Taiwanese Wikipedia between December 31, 2019 and March 31, 2020. Infection and symptoms attracted the most attention, followed by its origin and ways to prevent it. Logistics monitors public concern about COVID-19 and provides accurate and timely information to the public.

The use of machine learning techniques by Ordun et al. (2020) [1] offers promising strategies for deciphering patterns in public discourse and analyzing large amounts of data. An article [2] by Cucinotta and Vanelli (2020) emphasizes the importance of international cooperation and communication in addressing global health challenges such as the COVID-19 pandemic. Finally, Lai et al. (2020) highlight the potential of infodemiology as a tool to provide accurate information to the public and monitor the public interest in health-related issues. Overall, these articles demonstrate the importance of using infodemiology and social media in understanding and combating the COVID-19 pandemic.

V.   DISCUSSIONS

During the early phases of the pandemic, the researchers of this study employed observational data to investigate potential causal links between COVID-19 trends, Twitter action, and public attitude. While some of their data supported expected causal correlations, they also discovered fascinating conclusions that agreed with current COVID-19 scientific literature. They discovered, for example, that the amount of households with one person in them influences the overall number of COVID-19 diseases, and the percent of people aged 65 and older influences the percentage shift in fatalities.

It is critical to understand public views and actions in order to make informed judgments during a worldwide epidemic. [6] Although computational methods such as causal inference can assist in determining causality between observed variables, it is difficult to precisely detect all causal links due to confounding factors. Some drawbacks of the study include the absence of ground truth causal links for some variables, the failure to describe periodic causal connections, and the removal of retweet and likes from their models. Nonetheless, the work gives useful insights into potential causal correlations between COVID-19 trends, Twitter movement, and public sentiment, that can improve decision-making in the event of a worldwide epidemic.

VI.   CONCLUSION

In conclusion, the purpose of our research was to learn how people are expressing themselves and their opinions on Twitter during the COVID-19 pandemic. We were able to gain valuable insights into the emotional effects that the pandemic had on individuals by using topic modeling, sentiment analysis, and emotion analysis.

Our investigation uncovered that most of tweets communicated joy, yet a huge piece likewise communicated distress and dread. During a global pandemic, these feelings are to be expected and emphasize the need for assistance and support to help people cope.

Additionally, our study demonstrated how social media data analysis can benefit from natural language processing methods. We were able to gain useful insights from a large dataset of tweets by employing topic modeling, sentiment analysis, and emotion analysis.

In general, people, healthcare professionals, and policymakers who want to know how the pandemic affects people's emotions and come up with effective ways to support their mental health during this trying time can benefit from our findings. We hope that our study will spur additional research into the emotional effects of the COVID-19 pandemic and aid in the creation of effective support programs for individuals during and after the outbreak.

VII. ACKNOWLEDGMENT

VIII. REFERENCES

[1] C. Ordun, S. Purushotham, and E. Raff, "Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs," arXiv preprint arXiv:2005.03082, 2020.

[2] D. Cucinotta and M. Vanelli, "WHO declares COVID-19 a pandemic," Acta bio medica: Atenei parmensis, vol. 91, no. 1, pp. 157–160, 2020.

[3] A. C. Nagpal, R. Sukhija, and A. K. Mehra, "Analysis of Twitter users' sentiments towards COVID-19: An application of machine learning," medRxiv, 2020.

[4] H. Jin, J. Lu, and Y. Xu, "Using Twitter and web news mining to predict COVID-19 outbreak," Asian Pacific Journal of Tropical Medicine, vol. 13, no. 8, pp. 364–369, 2020.

[5] D. Anand, R. Yadav, and A. Kumar, "Analysing the Sentiments towards Work-From-Home Experience during COVID-19 Pandemic," SSRN, 2020.

[6] M. Duggan, N. B. Ellison, C. Lampe, A. Lenhart, and M. J. P. R. C. Madden, "Social media update 2014," Pew Research Center, 2015.