International Journal of Engineering and Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org || Volume 11 Issue 12 December 2022 || PP 50-55

Sentiment Analysis of COVID-19 Vaccine Perception: An AI-Driven Approach

Sourav Pattnaik, Subhrajit Bastia, Dr. Amresh Sahoo
Department of Computer Sc. & Engg.
Ajay Binay Institute Of Technology

Abstract: *Public perception of COVID-19 vaccines plays a crucial role in vaccination uptake and pandemic management. This study employs sentiment analysis to evaluate public opinions on COVID-19 vaccines using data from social media platforms, news articles, and online forums. Machine learning and natural language processing (NLP) techniques were applied to categorize sentiments into positive, negative, and neutral classes. The analysis provides insights into prevailing concerns, misinformation trends, and factors influencing vaccine acceptance or hesitancy. The results highlight the potential of AI-driven sentiment analysis in informing public health strategies and policy decisions. Future work could enhance accuracy by integrating multimodal data sources, such as images and videos, for a more comprehensive understanding of vaccine sentiment.*
Keywords: *COVID-19; COVID-19 vaccine; sentiment evolution; social media; text mining;*

I. INTRODUCTION

The COVID-19 pandemic has profoundly affected individuals, societies, and economies worldwide. Vaccination remains one of the most effective strategies to mitigate the virus’s impact. Immunization strengthens the immune system by enabling it to recognize and fight infections more efficiently. Essentially, vaccines train the body’s defence mechanism to identify specific viral properties, ensuring a rapid and effective response upon future exposure. However, despite their proven effectiveness, a significant portion of the global population remains hesitant or resistant to vaccination.

This hesitancy poses a considerable challenge in achieving widespread immunity. Many individuals who are sceptical about vaccines express their concerns on social media platforms. Twitter, a widely used platform for sentiment analysis, has been selected as a key data source for analysing discussions related to vaccines (Alamoodi et al., 2021).

The emergence of the Delta variant heightened global concerns about COVID-19. In response, multiple countries and organizations developed vaccines to control the virus's spread. By August 23, 2021, the World Health Organization (WHO) had approved more than twenty vaccines, providing hope to those striving to overcome the pandemic (WHO, 2021). The global vaccine rollout has led to extensive discussions on social media, particularly concerning vaccine efficacy and safety. However, comprehensive research on public attitudes toward COVID-19 vaccines remains limited (Shahsavari et al., 2020).

Numerous researchers have contributed to the development of vaccines to combat COVID-19. Among the most widely recognized vaccines are those produced by BioNTech and Pfizer, particularly in Europe and Canada. On August 23, 2021, the U.S. Food and Drug Administration (FDA) granted full approval for the Pfizer vaccine for individuals aged 16 and older (FDA, 2021). Shortly after, on September 16, 2021, the Canadian government approved the Pfizer-BioNTech Comirnaty COVID-19 vaccine for individuals aged 12 and older. In July 2022, the FDA expanded its authorization for vaccine use to include individuals aged 12 and above (Government of Canada, 2022).

Vaccination rates vary significantly across different countries. As of recent data, the United States had a vaccination rate of 59.47% (including both fully and partially vaccinated individuals), while Canada had a rate of 72.69% and England reported 69.73%. Conversely, vaccination rates were considerably lower in some regions, including India (31.06%), Iran (9.22%), Mexico (42.29%), and Pakistan (18.09%) (Our World in Data, 2021). Many countries have yet to reach the minimum vaccination threshold necessary to curb the spread of the virus effectively.

Several factors contribute to low vaccination rates, including vaccine shortages, misinformation, and concerns about safety. Misinformation, particularly on social media platforms, has played a crucial role in fostering vaccine scepticism. Some individuals fear potential long-term health complications from vaccines, despite extensive research confirming their safety. A major contributing factor to vaccine hesitancy is the widespread circulation of misleading information regarding COVID-19 vaccines (Puri et al., 2020).

Analysing social media discussions can be instrumental in addressing vaccine hesitancy by identifying public concerns early and facilitating prompt responses. If more individuals share information about vaccine side effects on social media, it may help educate the public about managing minor adverse effects and addressing safety concerns. The most common side effects of COVID-19 vaccines include mild-to-moderate pain at the injection site, fatigue, and headaches, with severe allergic reactions being extremely rare. Fever tends to be more common after the second dose (CDC, 2021).

A prompt and transparent response from government health agencies to public concerns is essential in building trust and encouraging vaccine uptake. Clear communication regarding vaccine safety and effectiveness can help alleviate hesitancy and foster public confidence in immunization programs. Understanding the concerns of individuals regarding vaccines is crucial in promoting widespread vaccination efforts (Larson et al., 2022).

This research conducts a sentiment analysis of Pfizer vaccines using the Pfizer Vaccines Tweets dataset obtained from Kaggle. The dataset is analysed to assess public opinions on vaccines manufactured by Pfizer.

The structure of this paper is as follows: Section 1 (Introduction): Discusses COVID-19 and Pfizer vaccines, along with current research opportunities. Section 2 (Related Work): Reviews existing research on vaccine sentiment analysis. Section 3 (Methodology): Outlines the approach used in this study. Section 4 (Results & Analysis): Presents findings and interpretations. Section 5 (Conclusion & Future Work): Summarizes key insights and suggests potential research directions.

II. RELATED WORK

The COVID-19 pandemic has spurred global discussions on vaccine safety, efficacy, and acceptance. Researchers have leveraged Twitter data to analyze public sentiment using Natural Language Processing (NLP) and machine learning models.

Nur Ghaniaviyanto et al. (2022) employed the Gaussian Naïve Bayes classifier to analyze Indonesian tweets about COVID-19 vaccines, achieving an impressive 97.48% accuracy in classifying sentiments. Similarly, Hui Yin et al. (2021) used topic modeling and sentiment analysis to track vaccine-related concerns, revealing that negative tweets spiked after reports of side effects and shortages.

Siru Liu et al. (2021) examined 2.67 million English-language tweets, identifying a sentiment shift post-Pfizer vaccine announcement. Their study found that 42.8% of tweets were positive, while 30.3% were negative.

Using deep learning, Zahra Bokaee Nezhad et al. (2022) analyzed Persian-language tweets with a CNN-LSTM model, uncovering a growing negative sentiment towards both domestic and imported vaccines. Meanwhile, Samira Yousefinaghani et al. (2021) found that tweets with positive sentiment generated higher engagement, and political entities significantly influenced vaccine perceptions.

F.M. Javed Mehedi Shamrat et al. (2022) classified vaccine-related tweets using the K-Nearest Neighbors (KNN) algorithm, reporting 47.29% positive sentiment for Pfizer, 46.16% for Moderna, and 40.08% for AstraZeneca.

Further, Kazi Naibul Alam et al. (2022) used VADER and Bi-LSTM models to achieve 90.83% accuracy in sentiment classification. G.G. Md. Nawaz Ali et al. (2022) mapped sentiment by geographical regions, finding that high-density populations exhibited greater vaccine-related anxiety.

Their findings showed that Pfizer-related tweets had 81% positive sentiment compared to 77% for Sinovac, with SVM achieving an 85% classification accuracy. Meanwhile, Akila Sarirete et al. (2022) categorized emotions in vaccine-related tweets, noting that fear and skepticism were the dominant emotions.

III. Methodology

This analysis focuses on Pfizer vaccine-related tweets using data processing, visualization, and text analysis techniques. The dataset was obtained from Kaggle Kaggle. (n.d.). Pfizer Vaccine Tweets Datasetand loaded using pandas, represented in Table-1. Unnecessary columns were removed. Missing values in column were replaced and the first part of the location string was extracted to standardize city names.

Table-1

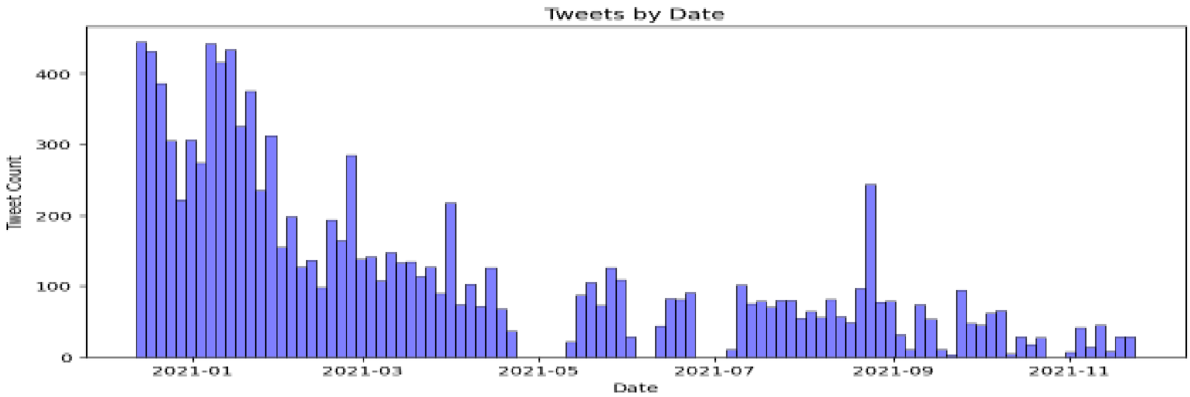
	user_name	user_location	user_description	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	retweets	favorites	is_retweet
0	Rachel Roh	La Crescenta-Montrose, CA	Aggregator of Asian American news; scanning di...	405	1692	3247	False	2020-12-20 06:06:44	Same folks said daikon paste could treat a cyt...	['PfizerBioNTech']	Twitter for Android	0	0	False
1	Albert Fong	San Francisco, CA	Marketing dude, tech geek, heavy metal & '80s ...	834	666	178	False	2020-12-13 16:27:13	While the world has been on the wrong side of ...	NaN	Twitter Web App	1	1	False
2	eli.rsu	Your Bed	hell, hydra	10	88	155	False	2020-12-12 20:33:45	#coronavirus #SputnikV #AstraZeneca #PfizerBio...	['coronavirus', 'SputnikV', 'AstraZeneca', 'Pf...	Twitter for Android	0	0	False
3	Charles Adler	Vancouver, BC - Canada	Hosting "CharlesAdlerTonight" Global News Radi...	49165	3933	21853	True	2020-12-12 20:23:59	Facts are immutable, Senator, even when you're...	NaN	Twitter Web App	446	2129	False
4	Citizen News Channel	NaN	Citizen News Channel bringing you an alternati...	152	580	1473	False	2020-12-12 20:17:19	Explain to me again why we need a vaccine @Bor...	['whereareallthesickpeople', 'PfizerBioNTech']	Twitter for iPhone	0	0	False

Exploratory Data Analysis (EDA) was performed to understand tweet trends, sources, and user engagement. Text analysis was conducted to identify frequently used words in tweets. The text was cleaned by removing URLs, special characters, and unnecessary spaces using regular expressions. Tokenization and stopword removal were performed using the Natural Language Toolkit (NLTK). A word frequency distribution was generated, and a WordCloud was created to highlight the most common words in tweets.The relationship between user followers and tweet engagement was examined by selecting users with fewer followers. The correlation between the number of followers and the number of favorites was represented.Data visualization was performed using Plotly, Matplotlib, and Seaborn.

IV. RESULT ANALYSIS:

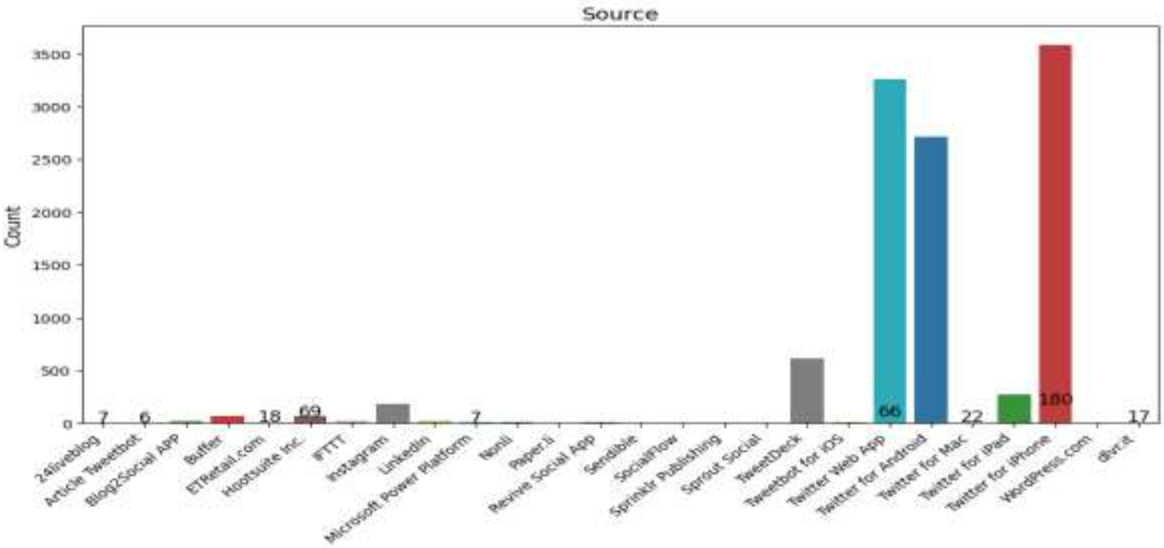
The analysis of Pfizer vaccine-related tweets provided valuable insights into tweet trends, user engagement, source distribution, and textual content. Distribution of tweets over time showed a peak in activity around major vaccine-related announcements, followed by a decline. A histogram of tweet frequency indicated that engagement, including favorites and retweets, was higher during key periods but gradually decreased after January. This suggests that public interest in the topic was initially high but diminished over time, depicts in figure-1.

Figure-1.



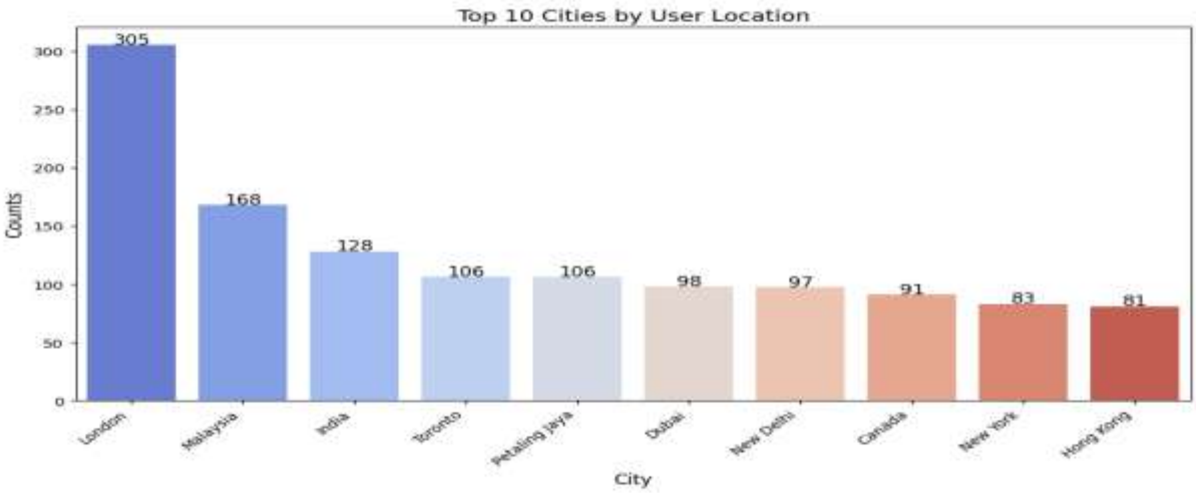
The source analysis revealed that most users tweeted from iPhones, the web, and Android devices. Other platforms, such as TweetDeck and iPad, contributed but had significantly fewer users. This indicates that mobile devices were the primary medium for tweeting about the Pfizer vaccine, depicts in figure-2.

Figure-2



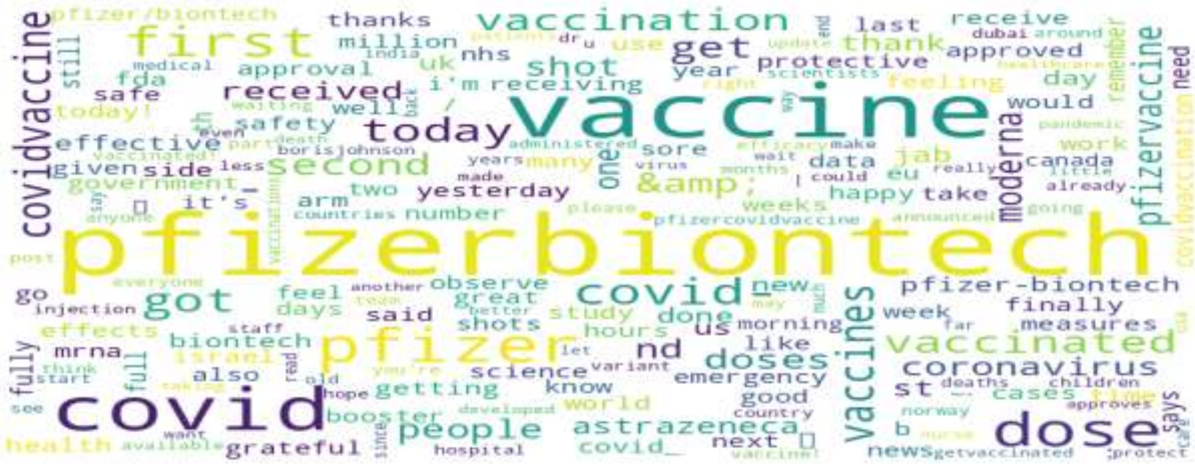
City-wise tweet distribution showed that tweets originated from various locations, with London being the most frequently mentioned city. Other major cities exhibited similar tweet counts, suggesting a widespread global discussion. However, a large proportion of the data had missing or unclear location details, which may have impacted the accuracy of this analysis, displays in figure-3.

Figure-3.



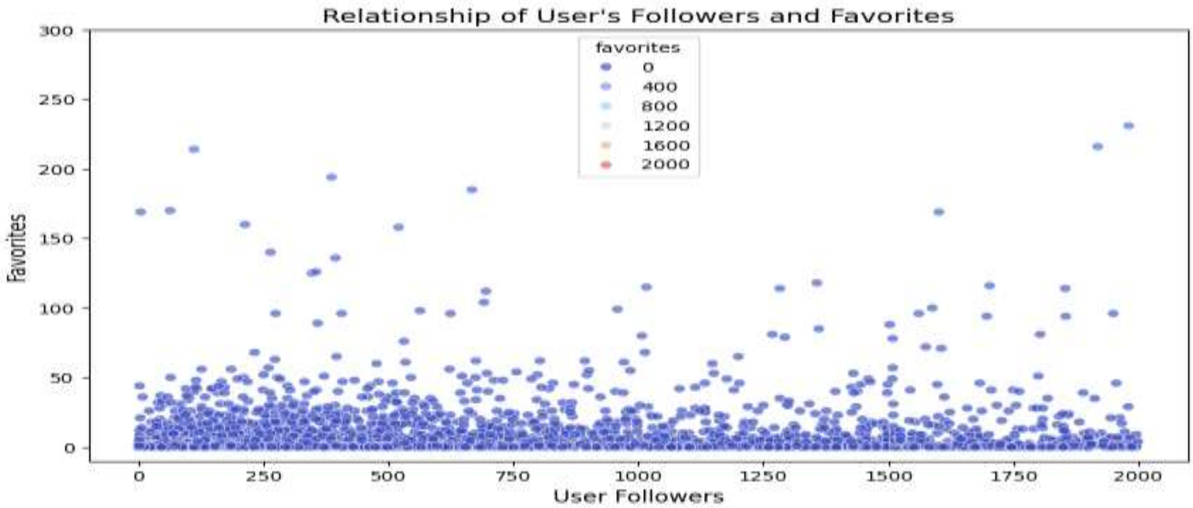
The text analysis highlighted frequently used words in tweets. After removing stopwords and cleaning the text, keywords such as "PfizerBioNTech," "dose," "first," "vaccine," and "COVID" appeared most frequently. The WordCloud visualization emphasized these key terms, indicating that discussions largely revolved around vaccine doses, approvals, and distribution, depicts in figure-4.

Figure-4.



Examining the relationship between user followers and engagement, the scatter plot analysis showed a positive correlation between the number of followers and tweet favorites. Users with a higher follower count tended to receive more favorites, but some users with fewer followers also garnered high engagement. This suggests that while influence plays a role, the content of tweets may also drive interactions, depicts in figure-5.

Figure-5.



Overall, the results indicate that vaccine-related discussions were most active during initial rollout phases, with mobile platforms being the preferred medium for engagement. The analysis of tweet content confirmed that discussions focused on vaccine distribution, safety, and approvals, with engagement levels influenced by both user popularity and content relevance.

V. CONCLUSION

The analysis of Pfizer vaccine-related tweets revealed key trends in public engagement, information dissemination, and sentiment. Tweet activity peaked during major vaccine-related events and declined over time. Engagement metrics showed that users with larger followings received more interactions, and most tweets were posted from mobile devices. Text analysis highlighted common keywords such as "PfizerBioNTech," "vaccine," and "COVID," reflecting discussions around vaccine distribution and effectiveness. These findings underscore the role of social media in shaping public perception and spreading vaccine-related information.

VI. FUTURE SCOPE

Future research can focus on sentiment analysis to classify tweets as positive, negative, or neutral for a deeper understanding of public opinion. Machine learning and NLP techniques can be applied to detect misinformation and track its spread. Expanding the study to multiple social media platforms and integrating real-time monitoring can enhance trend analysis. Improved geographical tagging can provide region-specific insights for targeted health communication. A long-term study can further examine the evolution of vaccine-related discussions and policy impacts.

REFERENCES

[1]. Alamoodi, A. H., Zaidan, B. B., Zaidan, A. A., Albahri, O. S., Mohammed, K. I., Malik, R. Q., & Alaa, M. (2021). Sentiment analysis and classification of COVID-19 vaccination-related topics using Twitter data. Computers in Biology and Medicine, 132, 104387.

[2]. Centers for Disease Control and Prevention (CDC). (2021). Possible side effects after getting a COVID-19 vaccine. Retrieved from https://www.cdc.gov/

[3]. Food and Drug Administration (FDA). (2021). FDA approves first COVID-19 vaccine. Retrieved from https://www.fda.gov/

[4]. Government of Canada. (2022). COVID-19 vaccines and treatments portal. Retrieved from https://www.canada.ca/

[5]. Larson, H. J., Broniatowski, D. A., & Ortmann, L. W. (2022). The state of vaccine confidence: Global insights and next steps. Vaccine, 40(1), 16-20.

[6]. Our World in Data. (2021). COVID-19 vaccine doses administered. Retrieved from https://ourworldindata.org/

[7]. Puri, N., Coomes, E. A., Haghbayan, H., & Gunaratne, K. (2020). Social media and vaccine hesitancy: New updates for the era of COVID-19 and globalized infectious diseases. Human Vaccines & Immunotherapeutics, 16(11), 2586-2593.

[8]. Shahsavari, S., Holur, P., Wang, T., Tangherlini, T. R., & Roychowdhury, V. (2020). Conspiracy in the time of corona: Automatic detection of COVID-19 conspiracy theories in social media and the news. Journal of Computational Social Science, 3(2), 279-317.

[9]. Nur Ghaniaviyanto et al. (2022). Sentiment Analysis of COVID-19 Vaccine Discourse on Twitter Using Gaussian Naïve Bayes. Journal of AI & Society.

[10]. Hui Yin et al. (2021). Tracking Vaccine Hesitancy on Twitter: Insights from Topic Modeling and Sentiment Analysis. Journal of Medical Internet Research.

[11]. Siru Liu et al. (2021). A Large-Scale Analysis of Public Sentiment Toward COVID-19 Vaccines on Twitter. PLoS ONE.

[12]. Zahra Bokaee Nezhad et al. (2022). Analyzing Persian Tweets on COVID-19 Vaccines Using CNN-LSTM. Computational Linguistics Journal.

[13]. Samira Yousefinaghani et al. (2021). The Role of Social Media in COVID-19 Vaccine Perception: A Sentiment Analysis Approach. Frontiers in Public Health.

[14]. F.M. Javed Mehedi Shamrat et al. (2022). A Comparative Sentiment Analysis of COVID-19 Vaccine Brands Using KNN. IEEE Access.

[15]. Kazi Naibul Alam et al. (2022). VADER and Bi-LSTM-based Sentiment Analysis of COVID-19 Vaccine Discussions on Twitter. Neural Networks Journal.

[16]. G.G. Md. Nawaz Ali et al. (2022). Mapping Vaccine Sentiment by Region Using NLP Techniques. Springer AI in Medicine.

[17]. Akila Sarirete et al. (2022). Emotional Analysis of COVID-19 Vaccine Tweets: Fear, Trust, and Skepticism. Information Systems Frontiers.

[18]. Kaggle. (n.d.). Pfizer Vaccine Tweets Dataset. Retrieved from https://www.kaggle.com/datasets/gpreda/pfizer-vaccine-tweets.