Textual Analysis of Vaccine Hesitancy Trends During The COVID-19 Pandemic

Maryem Ismail, Qasim Niaz, Imran Ali, Shakeeb Ahmed Khan, Hina Mehboob

Department of Computer Science, University of Southern Punjab, Multan, Pakistan

Abstract: Sentiment Analysis (SA) is a very prominent area of research in Natural Language Processing (NLP). It involves the extraction and identification of subjective information. In this study, vaccine hesitancy trends were discussed during the COVID-19 pandemic. People expressed their opinions about COVID-19 hesitancy on social media platforms such as Facebook and Twitter. This study includes a detailed review of the methods used by different researchers to identify and extract the sentiments of people about hesitancy of the COVID-19 vaccine. Textual datasets extracted from social media platforms, which contain tweets related to vaccine hesitancy trends were also discussed. This study concludes all the methods, techniques, and models that were adopted by researchers using textual datasets taken from social media. Finally, the conclusion of this study provides detailed insights and understanding of people sentiments who are hesitant about COVID-19 vaccine.

Keywords: Sentiment Analysis, NLP, Machine Learning, Text Mining, COVID-19.

Email: maryemismail@isp.edu.pk

1. Introduction

Sentiment refers to the expression of opinions, emotions, and attitudes towards a particular topic or subject. It can be positive, negative, or neutral Catelli, et al [1] Analysis is the process of breaking down a complex system or concept into smaller parts to better understand it Sarkar, et al [2] In the context of SA, it involves examining text data to identify and extract subjective information. Dake, et al [3] SA is a technique that involves the use of natural language processing, text mining, and Machine Learning (ML) to identify and extract subjective information from textual data. Naithani, et al [4]

Media refers to the various channels through which information is communicated to a large audience. This includes television, radio, newspapers, magazines, and the internet. Tojaliyev, et al [5] Social media refers to online platforms that allow users to create and share content with others. Examples include Facebook, Twitter, Instagram, and LinkedIn. Social media allows people to interact with each other by sharing content such as text, images, and videos Tarigan, et al [6] Users can also engage with each other by commenting on posts and sending direct messages. People show their emotions on social media by expressing their opinions and feelings through text, images, and videos. Benrouba, et al [7] This can include positive emotions such as

happiness and excitement, as well as negative emotions such as anger and sadness. Manokara, et al [8]

Twitter is a social media platform that allows users to post short messages called tweets. Tweets are limited to 280 characters and can include text, images, and videos. People show their emotions on Twitter by expressing their opinions and feelings through tweets. This can include positive emotions such as happiness and excitement, as well as negative emotions such as anger and sadness. Abdusattorova, et al [9] We can extract positive and negative comments from Twitter using SA techniques. These techniques involve analyzing text data to identify subjective information such as opinions and emotions. Saranya, et al [10]

COVID-19 is an infectious disease caused by the SARS-CoV-2 virus. It was first identified in Wuhan, China in December 2019 and has since become a global pandemic. People show their emotions about COVID-19 by expressing their opinions and feelings about the disease through various channels such as social media. Devkota, et al [11] The COVID-19 pandemic has had a significant impact on our world. It has led to widespread illness, death, economic disruption, and changes in social behavior. We can extract positive and negative comments from Twitter regarding COVID-19 using SA techniques. These techniques involve analyzing text data to identify subjective information such as opinions and emotions. Braig, et al [12]

Vaccines are made by using a weakened or inactivated form of a virus or bacteria to stimulate an immune response in the body. This response helps the body recognize and fight off the actual virus or bacteria if it is encountered in the future. We can extract COVID-19 vaccinated people who hesitate by text mining technique using SA techniques. Li T. a., et al [13] These techniques involve analyzing text data to identify subjective information such as opinions and emotions related to vaccination. People were afraid of the COVID-19 vaccine due to concerns about its safety and efficacy. There were also concerns about potential side effects of the vaccine. Bendau, et al [14]

Natural language programming (NLP) is a field of computer science that focuses on developing algorithms that allow computers to understand human language. Text mining technique involves analyzing large amounts of unstructured textual data to identify patterns, trends, and insights that would be difficult or impossible for humans to identify manually. Dhawas, et al [15]

We can detect emotions and sentiments from Twitter using text mining techniques such as SA. These techniques involve analyzing text data to identify subjective information such as opinions

and emotions related to a particular topic or subject. William, et al [16] Some popular text mining technique models include Latent Dirichlet Allocation (LDA), Support Vector Machines (SVM), Naive Bayes Classifier (NBC), Decision Trees (DT), Random Forests (RF), Gradient Boosting Machines (GBM), Long Short-Term Memory Networks (LSTM), Convolutional Neural Networks (CNN), etc. A hybrid model of text mining technique combines two or more models to improve performance or accuracy. For example, combining Latent Dirichlet Allocation (LDA) with Support Vector Machines (SVM) can help in achieving better results.

Various types of tools can be used in a hybrid model of text mining technique, depending on the specific requirements and objectives of the research. Some commonly used tools include programming languages such as Python or R, libraries like sci-kit-learn and NLTK, and frameworks like TensorFlow and PyTorch. To detect sentiments of COVID-19 vaccinated people who were hesitant to use text mining technique, we can apply SA algorithms to analyze their social media posts or tweets By examining the language and sentiment expressed in these texts, we can gain insights into their emotions and attitudes towards COVID-19 vaccination. Praveen, et al [17]

The research problems in this study include identifying and extracting relevant tweets related to COVID-19 vaccine hesitancy, developing an accurate SA model for classifying positive and negative sentiments, and addressing challenges such as noisy data, sarcasm, and language variations Li Q. a., et al [18]

The research objectives of this study are to analyze sentiments expressed on social media platforms like Twitter regarding COVID-19 vaccine hesitancy, develop a text mining model for SA, identify factors contributing to vaccine hesitancy, and provide insights for public health organizations to address concerns and improve vaccination rates. (Predictors of vaccine hesitancy: implications for COVID-19 public health messaging, et al [19]

The future scope of this research topic includes exploring advanced ML techniques for SA, incorporating real-time data from social media platforms, investigating the impact of vaccination campaigns on sentiment trends, and developing personalized interventions to address vaccine hesitancy. Khan, et al [20]

SA is a technique that involves the use of natural language processing, text mining, and ML to identify and extract subjective information from textual data. In this research paper, we aim to perform SA on tweets related to COVID-19 vaccine hesitancy using text mining techniques.

Qorib, et al [21] We will explore how people interact with each other on social media, how they express their emotions, and how we can extract positive and negative comments from Twitter. We will also discuss the impact of COVID-19 on our world and how it has changed our lives. Najm, et al [22] Furthermore, we will examine how vaccines are made and how they work to remove the virus. We will then focus on how we can extract COVID-19 vaccinated people who hesitate by text mining technique. We will discuss why people were afraid of the COVID-19 vaccine and what are the research problems and objectives. Finally, we will provide a summary of the text mining techniques used in this research paper and discuss the future scope of this research topic. Panja, et al [23]

2. Literature Review

The Covid-19 pandemic has presented numerous challenges for policymakers. In this study, a new research framework is suggested, which involves analyzing over 400,000 research abstracts from the last 2.5 years. To handle this large volume of abstracts, an innovative text-mining technique called coherent topic clustering is employed. It can be said that scholars tend to produce research at a faster pace amidst these challenging circumstances Lam, et al [24] This study summarized the latest trends in the design of COVID-19 chatbots from 5 different perspectives. This information can assist developers in selecting a chatbot archetype that is forward-looking and capable of meeting the increasing demands for improved pandemic response in the future Lim, et al [25] The COVID-19 pandemic, believed to have originated in Wuhan, China in 2019, prompted global measures like handwashing and masking. In early 2021, COVID-19 vaccination efforts began, sparking debates. Vishwakarma, et al [26] This study, relying on social media data, evaluates three sentiment computation methods (Azure ML, VADER, Text-Blob) and five learning algorithms (Random Forest, Logistic Regression, Decision Tree, Linear SVC, Naïve Bayes) with different vectorization methods. Vocabulary normalization employed stemming, lemmatization, and their combination. The study indicates a gradual decrease in COVID-19 vaccine hesitancy over time. Combining stemming and lemmatization enhances model performance. The Text-Blob sentiment score with TF-IDF vectorization and Linear-SVC classification yields the best performance. Combining Count-Vectorizer and TF-IDF decreases model accuracy. Vishwakarma, et al [26]

Sheikhattari, et al [27] examines COVID-19 vaccine hesitancy through SA of tweets using text mining techniques. It analyzes 280,000 tweets from Ohio and Michigan during the early stages

of the vaccination rollout (January 2021 to March 2021). Natural Language Processing, particularly LDA, identifies 11 topics and 8 sub-topics, revealing temporal and spatial trends in public sentiment, with concentrations in urban areas with higher COVID-19 case rates. Insights gained can inform public health strategies and local authorities' responses to future pandemics. Sheikhattari, et al [26]

The COVID-19 pandemic has profoundly shaped grocery shopping behaviors and necessitated government policies and supermarket regulations. This study, using text-mining techniques and time series analysis of UK supermarket tweets during the pandemic's first lockdown, investigates customer sentiment changes and their relationship with supermarkets' crisis responses. Li, et al [28] suggests that effective social media crisis communication can aid retail managers in understanding customer perceptions and highlights the importance of leveraging social media data to enhance crisis strategies and service quality assessment in the grocery sector. Li, et al [28]

Domgaard, et al [29] emphasizes the rich potential of social media, especially Twitter, as a valuable information source for public health organizations and decision-makers in understanding vaccine hesitancy. It highlights the efficient data collection process and preprocessing steps undertaken. The analysis reveals statistically significant relationships among various emotions, underscoring the complex interplay of sentiments on social media. Notably, the study employs neural network architectures to achieve impressive accuracy in classifying sentiments and emotions related to COVID-19, with the BERT model demonstrating the highest accuracy of 96.71%. Domgaard, et al [29]

The 21st-century rise of social media platforms has transformed communication dynamics, enabling widespread expression of emotions and opinions. With the surge in raw data generated across cultures, the need for meaningful information extraction has become paramount Lemor, et al [30] The COVID-19 pandemic exacerbated the spread of misinformation on social media, necessitating research into mitigating this issue. Lemor, et al [30] This study presents a sentiment-based model for detecting genuine COVID-19 news in Arabic text from Twitter in Gulf countries, utilizing ML and SMOTE for imbalanced datasets. Findings indicate negative sentiments prevalent during the pandemic, offering valuable insights for government responses to control the virus's spread. Lemor, [30]

The SA of COVID-19 vaccine discussions on Twitter, a topic often overshadowed in pandemic-related social media research. It employs the VADER model to categorize emotions and gauge sentiment in tweets from December 15th, 2020, to December 31st, 2021. Utilizing Latent Dirichlet Allocation (LDA), the study uncovers varied sentiments concerning. Chinese and international vaccines. It also identifies factors influencing sentiment, such as daily case numbers and key communication network issues, while highlighting ten major public concerns and shedding light on vaccine trust. Catalan-Matamoros, et al [31]

Asaad, [32] Represents the COVID-19 pandemic, with its severe health and economic impacts, has spurred extensive discussions on social media platforms, fostering both information dissemination and misinformation. This study introduces TClustVID, an intelligent clustering-based classification and topic extraction model for analyzing COVID-19-related tweets. Employing datasets from the IEEE Data-port repository, the model outperforms traditional methods, offering improved SA and topic extraction. TClustVID efficiently identifies prevailing public opinions and attitudes regarding COVID-19 and infection prevention strategies, providing valuable insights amid the pandemic's discourse. Asaad, et al [32]

While extensive research has focused on the COVID-19 pandemic's impact, this study addresses the relatively understudied area of public sentiment. Analyzing COVID-19 tweets from Canadian and U.S. cities using RoBERTa, Vader, and NRC methods, it assesses sentiment intensity and temporal variations. Jain, et al [33] Comparing sentiments related to anti-epidemic measures, including "masks," "vaccine," and "lockdown," causal relationships are explored, revealing fluctuating public sentiments and associations with specific measures, offering valuable insights into the evolving public response to the pandemic. Jain, et al [33]

The public attitudes toward COVID-19 vaccination boosters in Malaysia, employing SA of 788 relevant tweets. Utilizing latent Dirichlet allocation (LDA), three vital topics are identified: (1) type of booster, (2) booster effects, and (3) vaccination program operation. The study identifies key determinants, such as "Pfizer" and "mix," strongly influencing sentiments, showcasing the significance of public discourse on booster vaccinations in shaping perceptions. This research offers a valuable framework for healthcare topic retrieval from social media, facilitating informed decision-making. Alghamdi, et al [34]

TextBlob and VADER analyzers on historical tweets to gauge emotional responses to the COVID-19 pandemic in Nigeria. Collecting over a million tweets using the hashtag 'COVID-19,'

SA reveals varying degrees of positivity, neutrality, and negativity in public sentiment. The study emphasizes the potential for leveraging social media data to aid decision-making, combat misinformation, and gain insights into public opinion and challenges related to COVID-19, providing valuable inputs for organizations and governments worldwide. Gesser-Edelsburg, et al [35]

The COVID-19 pandemic underscored the vital role of efficient vaccine supply chain management. This study, utilizing social media analytics, examines the global impact of COVID-19 vaccine supply chain challenges. Data from Twitter and the Global Database of Events, Language, and Tone (GDELT) reveal significant disruptions, particularly in the healthcare systems of developing countries. The findings emphasize the urgency of enhanced coordination in vaccine supply chain management to ensure equitable access and highlight the utility of social media analytics for informed decision-making by policymakers and industry leaders. Belhadi, et al [36]

In the era of Sina Weibo, understanding public sentiment and opinion on this social platform has become critical. This study introduces the BERT-FGM-BiGRU model for emotion analysis in Chinese text data from Sina Weibo during COVID-19. Utilizing BERT for word vectors, BiGRU for feature extraction, and FGM for data perturbation, the model outperforms traditional methods, providing valuable insights for government opinion monitoring. The analysis highlights the spatial relationship between pandemic severity and public sentiment, revealing a transition from negative to positive emotions as the pandemic evolved. Fang, et al [37]

SA of Twitter data from India to assess perceptions of COVID-19 immunization. From January 2021 to March 2023, SA reveals predominantly positive sentiments, with support for vaccination prevailing. Nonetheless, some negative sentiments concerning vaccine hesitancy, side effects, and mistrust emerge. Demographic analysis unveils variations in sentiment across gender, age, and location, emphasizing the importance of tailored communication strategies to address hesitancy and promote vaccination within specific demographics, providing valuable insights into India's COVID-19 vaccination landscape. Limaye, et al [38]

The psychological impact of the COVID-19 pandemic by analyzing Twitter data. Using a convolutional neural network optimized through an arithmetic optimization algorithm, the approach achieves a high accuracy rate of 95.098% in classifying tweets as positive, negative, or

neutral, outperforming existing methods. This demonstrates the potential of SA on social media platforms like Twitter to assess public sentiments and psychological effects during the pandemic.

With the global prominence of COVID-19 on Twitter, this study employs content analysis and Latent Dirichlet Allocation to extract and analyze topics from over five million tweets across 32 countries. The analysis reveals that trends in negative topics on Twitter are significantly correlated with the number of confirmed COVID-19 cases in most countries, shedding light on the alignment between public sentiment and the pandemic's progression.

Amid the global COVID-19 pandemic, this study examines public sentiment on vaccine-related topics, specifically considering the influence of gender. Analyzing tweets on various vaccine brands, the research utilizes Deep Learning (DL) models, with LSTM achieving the highest accuracy at 85.7%. This systematic analysis provides insights into how gender impacts public perception of COVID-19 vaccines, shedding light on the nuances of vaccine-related discussions on social media platforms like Twitter.

The challenge of SA in the context of "E-learning" and COVID-19-related misinformation on Twitter. The proposed approach employs denoising AutoEncoder, an attentional mechanism, and ELM-AE with LSTM for SA classification, achieving an F1-score of 0.945, outperforming existing approaches. By effectively tackling noise and integrating features, this method contributes to a clearer understanding of sentiments and misinformation in the education sector during the COVID-19 pandemic.

The impact of COVID-19 on financial stock markets by proposing a SA-based stock market prediction system. By analyzing COVID-19 news and articles, the system aims to predict stock movements during the pandemic, specifically for stocks like TSLA, AMZ, and GOOG. Feature selection and spam tweet reduction techniques are applied to enhance prediction quality, making this system a hybrid approach combining text mining and social media data mining for improved predictions.

Explores public sentiment regarding COVID-19 vaccines on Twitter and Reddit, analyzing data from January 1, 2020, to March 1, 2022. Using a fine-tuned DistilRoBERTa model, sentiments of approximately 9.5 million tweets and 70 thousand Reddit comments were assessed. Twitter exhibited more negative sentiment, while Reddit had a more positive sentiment. Despite these differences, both platforms displayed similar sentiment trends during key vaccine-related developments in the pandemic. The study highlights the importance of using these platforms to

enhance vaccine confidence and counter misinformation, emphasizing the need for tailored educational strategies to address public sentiments and tackle misinformation. Do Nascimento, et al [39]

A Bidirectional Long Short-Term Memory (Bi-LSTM) NLP method for sentiment classification on COVID-19-related discussions from Twitter and Reddit. Bi-LSTM's bidirectional context generation showed improved results over traditional LSTM models. The study emphasizes the potential of NLP techniques for understanding public sentiment, mitigating negative messages, combating misinformation, and informing health decision-making during the pandemic. Wang, et al [40]

The impact of the COVID-19 pandemic and the role of Twitter in disseminating information. It analyzes two sets of tweets, one from early 2019 to March 2020 and another from December 2019 to May 2020. While the latter period shows more positive and neutral tweets about COVID-19, the former highlights a prevalence of negative sentiments. The study employs DL classifiers and a Gaussian membership function-based fuzzy rule base to identify sentiments in tweets, achieving an accuracy of up to 81%. The research underscores the importance of effective communication during a pandemic.

Investigates COVID-19 content sharing on Twitter. Analyzing over 57,000 tweets mentioning COVID-19, it employs Natural Language Processing techniques, including topic modeling, named entity-relationship, emotion and SA, and linguistic feature extraction. Findings reveal that tweets with named entities, negative emotions, references to mental health, optimism, and longer length are more likely to be retweeted. Conversely, tweets with numerous hashtags and user mentions tend to have lower retweet rates. This analysis sheds light on factors influencing COVID-19 content dissemination on social media.

The rapid development of COVID-19 vaccines, misinformation on social media has fueled vaccine hesitancy. This study analyzed COVID-19 vaccine discussions across 13 Reddit communities from December 2020 to May 2021. Results showed predominantly positive sentiment with a focus on vaccine side effects rather than conspiracy theories. Although sentiment remained stable, indications of vaccine hesitancy emerged. These findings emphasize the importance of targeted messaging and interventions to enhance vaccine confidence and counteract misinformation.

The COVID-19 pandemic had a significant impact on people's emotions, especially evident on social networks like Twitter. This study focused on assessing the sentiment of the Mexican population during a severe wave of the pandemic. A semi-supervised approach, combining lexical-based labeling and Transformer-based models, was employed. Multiple models, including Spanish language-specific Transformers, were trained and compared with traditional classifiers like Support Vector Machines and Naive Bayes. The exclusive Spanish Transformer model demonstrated superior performance in measuring COVID-19 sentiment among the Mexican Twitter community.

The emergence of variants has profoundly impacted daily life worldwide. Vaccine development offers hope, but discussions on social media about vaccine effectiveness and safety lack systematic analysis. This study delves into Twitter discussions on COVID-19 vaccines, examining hot topics and emotional sentiments across countries and vaccine brands. Results indicate trust in vaccine effectiveness, alongside concerns about side effects and shortages. Leveraging Natural Language Processing, this analysis contributes valuable insights for policymakers navigating the complex landscape of social media discourse during the pandemic.

Combining SA and multivariate statistics, to assess the evolving social mood during Spain's COVID-19 vaccination journey. Analyzing a substantial dataset of Spanish tweets over nearly two years, the research uncovers emotional trends, opinion clusters, and public sentiments, providing a comprehensive understanding of the vaccination process's social dynamics. This innovative approach enhances our perception of social reality, aiding in more effective problem resolution within the context of Spain's COVID-19 vaccination efforts.

Investigates citizens' emotional responses and concerns during the early days of the COVID-19 pandemic in New York State using a dataset of 65,313 tweets. The analysis reveals a predominance of negative sentiment along with a strong focus on issues related to trust, fear, anticipation, and social solidarity. Citizens expressed concerns about infection status and death. Timely and transparent communication regarding disease management is crucial for ensuring psychological stability among the population during health crises like COVID-19.

The COVID-19 pandemic, declared by the World Health Organization, has rapidly spread globally, with limited academic studies on its impact on mental health and economies. Understanding these effects is crucial for researchers to guide efforts in mitigating the pandemic's consequences. Twitter, with its vast user base and daily tweets on diverse topics, is a

valuable source for SA research. This study leverages Twitter data to analyze sentiment regarding various COVID-19 vaccinations. Using Twitter's API and Tweepy, the research retrieves and processes tweets, generating rules and variables for SA, represented through donut graphs.

The COVID-19 pandemic, declared by the World Health Organization in March 2020, prompted widespread social distancing measures to curb its spread. The effectiveness of such measures relies on public behavior and adherence. Twitter analysis has emerged as a powerful tool for understanding public sentiment and reactions during crises. With its growing user base, Twitter offers real-time insights into how individuals perceive and respond to institutional and media communications. Hashtags, in particular, facilitate the aggregation of individual perspectives into a collective voice. Understanding and guiding public perception is crucial for effective crisis management and communication strategies.

Social media platforms like Twitter have become a rich source for SA due to their high emotional content. This study focuses on SA of 90,000 tweets using both DL (LSTM) and ML (SVM) techniques. LSTM outperformed SVM with an accuracy of 87% compared to 86%. When considering binary classes of positive and negative, both LSTM and SVM achieved 90% accuracy, highlighting the effectiveness of these methods in SA on Twitter data.

Twitter's role in crisis communication during the COVID-19 pandemic in the tourism industry. It analyzes 123,868 tweets using word association and SA. The findings highlight Twitter's effectiveness in crisis communication, mitigating negative perceptions and aiding communication strategies for companies and organizations dealing with tourism crises. This research contributes to understanding the value of Twitter in managing the impact of crises on the tourism sector.

Explores the impact of COVID-19-related news on financial markets, particularly during the initial stages of the pandemic when information was abundant and complex. It analyzes over 200,000 online articles from three news platforms and employs ML techniques to extract sentiment. The results reveal a significant positive relationship between sentiment scores and the S&P 500 market, underscoring the influence of news sentiment on market expectations. Additionally, the study highlights variations in sentiment components and news categories from different platforms in their relation to market returns.

Addresses gaps in COVID-19 vaccine hesitancy research by examining a focused social media conversation. Using data from September 1, 2020, through December 31, 2020, it identifies trust-related topics, declining vaccine acceptance, and a parallel pandemic of distrust. Positive sentiment is associated with healthcare professionals and government organizations. Pfizer elicits both positive and negative emotions in vaccine-hesitant discussions, with negative sentiment dominating once vaccines are available. The study suggests targeted communication strategies to accelerate vaccine acceptance and mitigate hesitancy, focusing on personal anecdotes of safety, effectiveness, and family recommendations.

A tweet polarity classifier with four categories: positive, neutral, negative, and no opinion. It employs a genetic algorithm for feature extraction based on entropy and semantic context, customized to each word studied in the reviews. The algorithm outperforms previous literature works, achieving an M F1 of 0.640 and an accuracy of 0.689, primarily due to its flexible feature extraction approach.

Utilizes Twitter data to analyze travel mode choices during the pandemic in New York City. Natural language processing (NLP) models are employed to classify tweets related to different travel modes and perform SA. Positive attitudes toward buses, bikes, and private vehicles align with a shift away from subways. Negative tweets express concerns about mask usage on public transit. Regression analysis indicates that users in the service industry were more influenced by MTA subway service during the pandemic.

A SA system using three Naïve Bayes classification models to detect emotions in tweets. It achieves high accuracy, with the Multinomial Naïve Bayes model outperforming others, reaching 91.6% accuracy for the first dataset and 87.6% for the second. The research aims to expand with larger data and other SA methods to predict users' sentiments about various topics, including COVID-19, to improve model accuracy.

Twitter data analyzes vaccine popularity during the ongoing COVID-19 pandemic. Various ML and DL models were employed for SA, with the deep neural network model achieving the highest accuracy at 97.87%. This research provides valuable insights for vaccine producers and decision-making regarding their products amid widespread discussions on social media platforms.

The consistency of public health communications during the COVID-19 pandemic. SA of 131 public health briefings in Alberta, Canada, reveals consistent messaging regardless of case count

fluctuations. Specific words are associated with positive and negative sentiments, highlighting message patterns during different phases of the pandemic. This research underscores the utility of SA in objectively assessing health communications during crises.

A novel NLP infoveillance workflow based on BERT to understand public concerns and sentiments during health emergencies like COVID-19. It achieves high classification accuracy for content topics and SA. The research highlights significant variations in public awareness and sentiment towards COVID-19 topics over time and space, providing valuable insights for public health communication and intervention strategies.

Leveraged Twitter data to understand the impact of COVID-19 in Indonesia. SA, employing TF-IDF and Lexical methods, revealed public opinions and emotions related to the virus's spread and societal effects. The TFBS method showed the highest accuracy (0.85), outperforming other methods. Precision, recall, and F-score measurements supported its effectiveness. This research sheds light on public sentiment during the pandemic, aiding in informed decision-making and response strategies.

Vaccine development has taken center stage. To gauge public sentiment and behavior towards vaccines, this study utilizes social media data, specifically tweets. TextBlob and BERT models were employed to analyze sentiment, with BERT outperforming, achieving high precision, recall, and F-scores for both positive and negative sentiment classification, shedding light on public perception of vaccine campaigns.

In light of the COVID-19 pandemic, analyzing public sentiment on social media platforms, especially Twitter is crucial for effective crisis management. This study gathered real-time tweets related to coronavirus from Indian and European users and employed various feature extraction methods. The ensemble classifier, consisting of a Gated Recurrent Unit (GRU) and Capsule Neural Network (CapsNet), achieved high prediction accuracy, indicating its effectiveness in classifying sentiments such as anger, sadness, joy, and fear, with an accuracy of 97.28% for Indian users and 95.20% for European users.

DL methods analyze public sentiment and topics related to the COVID-19 pandemic in the United Kingdom through Twitter data. Over 500,000 tweets from 48 UK cities over two years were analyzed. The findings show changing sentiments and emotions over time, with approximately 60% of tweets being positive, and topics evolving with the pandemic's progression. This research provides valuable insights into how public emotions and topics shift

geographically and temporally during a health crisis, which can inform policy responses in specific regions.

Focuses on SA of public opinions regarding social distancing during the COVID-19 pandemic using Twitter data. The study employs preprocessing techniques and a Deep Belief Neural Network (DBN) with pseudo-labeling. By introducing the pseudo-boost mechanism, the approach improves classification performance and keyword extraction accuracy. The results indicate that using DBN in conjunction with bigram N-gram models outperforms other methods with a 90.3 percent accuracy. This approach has practical implications for decision-makers and healthcare professionals in tailoring pandemic responses to public sentiment in different locations.

The significant societal impacts and a strong demand for social responses. People increasingly turn to social media to share health information and opinions, seeking interaction and news about the pandemic Nabity-Grover, Cheung, and Thatcher; Riaz, Wang, and Guo. However, misinformation also spreads on social media, raising concerns about information accuracy.

During the coronavirus pandemic, the logistics sector faced economic challenges. This study employed text-mining techniques to analyze service consultant sites within logistics. Frequent terms included "supply," "chain," and "COVID-19." The analysis identified key terms like "kpis," "inventory," and "trust," indicating strategies for cost reduction and inventory management. Notably, "supply" had positive associations, while "pandemic" was linked negatively. The study found that logistics consultants emphasized trust and offered insights into pandemic-driven cost reduction and inventory management.

Vaccine development has been critical. Vaccine hesitancy, fueled by misinformation on platforms like Twitter, poses a challenge. This study, using over 8000 tweets from January 2021, analyzed sentiment towards COVID-19 vaccines. The results showed a relatively high rate (30%) of negative sentiment, with fear dominating. Trust was prevalent on the positive side. Logistic Regression and Naïve Bayes achieved around 90% accuracy in sentiment classification. This highlights the importance of addressing vaccine hesitancy through accurate information dissemination.

Examined how pre-vaccination social communication, encompassing personal acquaintances, news reports, and social media, influenced post-vaccination side-effect experiences. It also explored the role of pre-vaccination expectations in mediating these relationships. The findings

showed that exposure to pre-vaccination social media posts and perceptions of severity from personal acquaintances predicted both pre-vaccination side-effect expectations and post-vaccination experiences. Additionally, pre-vaccination expectations acted as a mediator between social communication and vaccine side effects. This study suggests that modifying side-effect expectations through these channels could impact post-vaccination outcomes, not only for COVID-19 vaccines but also for other widely discussed vaccines and medications.

3. Discussion

SA, as employed in this research, leverages natural language processing, text mining, and ML to discern subjective information within textual data. Our study is specifically directed towards conducting SA on tweets related to COVID-19 vaccine hesitancy through text mining techniques.

The significance of this research lies in its ability to elucidate how individuals engage in discourse on social media platforms, their emotional expressions, and the extraction of positive and negative sentiments from Twitter data. By analyzing such sentiments, we can gain insights into public perceptions and concerns regarding COVID-19 vaccination, thereby contributing to a better understanding of societal attitudes and aiding in targeted interventions.

The global impact of the COVID-19 pandemic is undeniable, altering the course of daily life across the world. In this context, understanding vaccine hesitancy becomes crucial, as it directly impacts vaccination efforts, public health outcomes, and societal recovery. By exploring the impact of COVID-19 on our world, the paper sets the stage for comprehending the multifaceted reasons behind vaccine hesitancy.

The subsequent examination of vaccine development and mechanisms of action provides essential background knowledge. It helps in demystifying the science behind vaccines, which is critical for addressing misconceptions and concerns held by vaccine-hesitant individuals.

Our central focus then shifts to the application of text mining techniques to identify COVID-19 vaccinated individuals who express hesitancy. This aspect of the research delves into the nuances of how hesitancy is expressed in online conversations and what factors contribute to this skepticism. It represents a pivotal step in harnessing the power of text mining to extract meaningful insights from vast datasets.

Addressing the 'why' behind vaccine hesitancy becomes imperative. By investigating the underlying reasons, we aim to uncover the root causes and motivations behind individuals'

reluctance to get vaccinated against COVID-19. These insights can inform public health campaigns and policy decisions.

Twitter serves as a rich source of user-generated data, making SA crucial for corporate and political decision-making. The COVID-19 pandemic has intensified the need to track public sentiment on this issue. This research employs DL and class label analysis to tackle SA challenges with limited data, particularly focusing on multiclass sentiment classification. It concludes that standard neural networks outperform bidirectional ones, achieving a validation accuracy of 92.5% and an F1 measure of 0.92, demonstrating the model's effectiveness in this context.

The COVID-19 pandemic has had profound emotional impacts, especially on patients. This study delves into the emotional experiences of COVID-19 patients and healthy controls using interviews and text mining. It identifies differences in emotional expression, revealing that COVID-19 patients had less complex and shorter narratives, with a specific focus on fear, disgust, and happiness. This analysis may aid in predicting mental health disorder symptoms in COVID-19 patients based on their emotional discourse.

4. Conclusion

In conclusion, this research paper has ventured into the realm of SA using text mining techniques, with a specific focus on COVID-19 vaccine hesitancy. Through an interdisciplinary approach that combines natural language processing and ML, we have aimed to uncover the sentiments and attitudes expressed in Twitter data.

Our investigation has touched upon the broader context of the COVID-19 pandemic and its profound impact on societies worldwide. By shedding light on the development and functionality of vaccines, we have laid the foundation for a comprehensive understanding of vaccine hesitancy.

The core contribution of this research lies in the application of text mining to identify and analyze hesitancy among those already vaccinated. We have explored the reasons behind this hesitancy, delving into the complex interplay of factors that influence public opinion.

In summary, this research not only enriches our understanding of SA but also offers valuable insights into the dynamics of vaccine hesitancy in the context of a global health crisis. We hope that these findings will guide future research endeavors and contribute to more effective public health strategies in addressing vaccine hesitancy. The future scope of this research includes

continued exploration of evolving sentiment trends and the refinement of text mining techniques for even deeper insights.

References

[1]. Catelli, R. a. Lexicon-based sentiment analysis to detect opinions and attitudes towards COVID-19 vaccines on Twitter in Italy. Computers in Biology and Medicine, 158, 106876, (2023).

[2]. Sarkar, P. a. Applied mathematical modeling in evolutionary biochemistry. Scandinavian Journal of Information Systems, 68-75, (2023).

[3]. Dake, D. K. Using sentiment analysis to evaluate qualitative students' responses. Education and Information Technologies, 4629-4647, (2023).

[4]. Naithani, K. A. Realization of natural language processing and machine learning approaches for text-based sentiment analysis. Expert Systems, e13114, (2023).

[5]. Tojaliyev, A. Media Communication Terms and Their Specific Features in Modern Linguistics. Current approaches and new research in modern sciences, 31-40, (2023).

[6]. Tarigan, I. M. Understanding Social Media: Benefits of Social Media for Individuals. Jurnal Pendidikan Tambusai, 7(1), 2317-2322, (2023).

[7]. Bendau, A. a. COVID-19 vaccine hesitancy and related fears and anxiety. International immunopharmacology, (2021).

[8]. Manokara, K. A. display rules differ between positive emotions: Not all that feels good looks good. Emotion, 23, 243, (2023).

[9]. Abdusattorova, B. Revealing Emotion of Joy and Sadness. Intellectual Education Technological Solutions and Innovative Digital Tools, 165-169, (2023).

[10]. Saranya, S. a. A Machine Learning-Based Technique with Intelligent WordNet Lemmatize for Twitter Sentiment Analysis. Intelligent Automation & Soft Computing, 36, (2023).

[11]. Devkota, B. a. Social Media Reactions to COVID-19 and Its Countermeasures. Rethinking Cities for Resilience and Growth in the Post-COVID-19 World, (2023).

[12]. Braig, N. a. Machine learning techniques for sentiment analysis of COVID-19-related Twitter data. IEEE Access, 11, 14, (2023).

[13]. Li, T. A. Using data mining technology to analyze the spatiotemporal public opinion of COVID-19 vaccine on social media. The Electronic Library, (2022).

[14]. Bendau, A. a. COVID-19 vaccine hesitancy and related fears and anxiety. International immunopharmacology, (2021).

[15]. Dhawas, P. a. (n.d.). Big Data Analysis Techniques: Data Preprocessing Techniques, Data Mining Techniques, Machine Learning Algorithm, Visualization. Big Data Analytics Techniques for Market Intelligence, 183-208.

[16]. William, P. a. Natural Language processing implementation for sentiment analysis on tweets. Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022, (2023).

[17]. Praveen, S. a. Analyzing the attitude of Indian citizens towards COVID-19 vaccine--A text analytics study. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 595-599, (2021).
[18]. Li, Q. a. Detecting Negative Sentiment on Sarcastic Tweets for Sentiment Analysis. International Conference on Artificial Neural Networks, (2023).
[19]. Predictors of vaccine hesitancy: implications for COVID-19 public health messaging. International journal of environmental research and public health, 8054, (2021).
[20]. Khan, S. a. Longitudinal analysis of behavioral factors and techniques used to identify vaccine hesitancy among Twitter users: Scoping review. Human Vaccines & Immunotherapeutics, 2278377, (2023).
[21]. Qorib, M. a. Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset. Expert Systems with Applications, 118715, (2023).
[22]. Najm, A. a. COVID--How will it continue to change our lives? Joint Bone Spine, 105572, (2023).
[23]. Panja, S. Information Retrieval Systems in Healthcare: Understanding Medical Data Through Text Analysis. Transformative Approaches to Patient Literacy and Healthcare Innovation, (2024).
[24]. Lam, B. S. Do Scholars Respond Faster Than Google Trends in Discussing COVID-19 Issues? An Approach to Textual Big Data. Health Data Science, 04, 01116, (2024).
[25]. Lim, W. A. General Characteristics and Design Taxonomy of Chatbots for COVID-19: Systematic Review. Journal of Medical Internet Research, e43112, (2024).
[26]. Vishwakarma, A. a. COVID-19 vaccination perception and outcome: society sentiment analysis on Twitter data in India}. Social Network Analysis and Mining, 13, 84, (2023).
[27]. Sheikhattari, P. a. Lessons learned to improve COVID-19 response in communities with greatest socio-economic vulnerabilities. BMC Public Health, 23, 659, (2023).
[28]. Li, X. a. Exploring customer concerns on service quality under the COVID-19 crisis: A social media analytics study from the retail industry. Journal of Retailing and Consumer Services, 103157, (2023).
[29]. Domgaard, S. K. Misinformation and Vaccines: Using Media Literacy to Improve Health Decision-Making, (2023).
[30]. Lemor, A. a. Exploring the role of uncertainty, emotions, and scientific discourse during the COVID-19 pandemic. Policy and Society, puae010, (2024).
[31]. Catalan-Matamoros, D. a.-S. Crisis Communication during COVID-19: English, French, Portuguese, and Spanish Discourse of AstraZeneca Vaccine and Omicron Variant on Social Media. Vaccines, 11, (2023).
[32]. Asaad, C. a. When Infodemic Meets Epidemic: a Systematic Literature Review. arXiv preprint arXiv: 2210.04612, (2022).
[33]. Jain, G. a. Public sentiments toward the COVID-19 pandemic: Insights from the academic literature review and Twitter analytics. International Journal of Stress Management, (2023).
[34]. Alghamdi, A. M. Social Media Analysis to Enhance Sustainable Knowledge Management: A Concise Literature Review. Sustainability, 9957, (2023).

[35]. Gesser-Edelsburg, A. How to make health and risk communication on social media more "social" during COVID-19. Risk Management and Healthcare Policy, 3523-3540. (2021).
[36]. Belhadi, A. a. Building supply-chain resilience: an artificial intelligence-based technique and decision-making framework. International Journal of Production Research, 60, 4487-4507, (2022).
[37]. Fang, F. a. Network structure and community evolution online: behavioral and emotional changes in response to COVID-19. Frontiers in public health, 9, 813234, (2022).
[38]. Limaye, R. J. What message appeal and messenger are most persuasive for COVID-19 vaccine uptake: Results from a 5-country survey in India, Indonesia, Kenya, Nigeria, and Ukraine. PloS one, e0274966, (2022).
[39]. Do Nascimento, I. J.-M.-O. Infodemics and health misinformation: a systematic review of reviews. Bulletin of the World Health Organization, 544, (2022).
[40]. Wang, Y. a. A case study of using natural language processing to extract consumer insights from tweets in American cities for public health crises. BMC Public Health, 935, (2023).