Sabaragamuwa University Journal
Volume 20 Number 1 December 2022
ISSN 1391-3166 eISSN 2386-2041
http://dx.doi.org/10.4038/suslj.v20i1.7822
© 2022 Sabaragamuwa University of Sri Lanka

Analysing Public Perception and Sentiment on the COVID-19 Booster Dose through Social Media

Thamodya Madhumani*[1], Abishethvarman Vadivel[1], Banujan Kuhaneswaran[1], Prasanth Senthan[2] and Samantha Kumara[1]
[1]Faculty of Computing, Sabaragamuwa University of Sri Lanka, Belihul Oya, Sri Lanka
[2]Faculty of Applied Sciences, Sabaragamuwa University of Sri Lanka, Belihul Oya, Sri Lanka
*thamodya01@gmail.com

Abstract

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has emerged as a severe public health crisis, reshaping societies worldwide since its inception in December 2019. Amidst these challenges, vaccination emerged as a pivotal strategy to curb transmission and restore normalcy. The journey to develop and distribute vaccines encountered obstacles but ultimately led to significant progress. Vaccine hesitancy persists, impacting the uptake of booster shots. Therefore, this research analyses the acceptance of booster shots by examining the tweets that reflect public perceptions. Over 12,000 tweets provided insights into public perceptions, highlighting the need for nuanced approaches to address vaccine hesitancy. Leveraging topic modelling algorithms such as Latent Dirichlet Allocation (LDA), the study identified pertinent topics and keywords, shedding light on variations in booster shot acceptance across different income categories. It further analysed the public acceptance of booster shots using machine learning techniques like Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM). The ANN algorithm outperformed the LSTM, which is known for itś semantic understanding. Persistent vaccine hesitancy remains a formidable challenge in the battle against COVID-19. Middle-income countries admit to the booster shot compared to high-income countries.

Keywords: Booster shots, COVID-19 pandemic, Social media analysis, Vaccine hesitancy

1. Introduction

Titled in the annals of history as a major public health crisis, the COVID-19 pandemic, caused by the SARS-CoV-2 virus (Hasöksüz, Kilic, & Saraç, 2020), has gripped the world with its relentless and unforgiving onslaught since it was first reported on December 31, 2019. The World Health Organization (WHO) recognised the magnitude of the emerging crisis, declaring COVID-19 a global health emergency on January 30, 2020, and later elevating it to a worldwide pandemic on March 11, 2020. These landmark announcements marked the beginning of an era defined by significant shifts in socioeconomic landscapes and personal lifestyles around the globe. SARS-CoV-2 (Lamers & Haagmans, 2022) did not discriminate in its wave of destruction. Every sector felt the virus's impact, from international conglomerates to local businesses. The virus (Wu, Wu, Liu, & Yang, 2020) shuttered economic engines, even threatening the stability of developed nations. The ensuing lockdowns brought numerous hardships, straining healthcare systems as they grappled with increasing caseloads. The calamitous aftermath of the pandemic-induced global shutdown resulted in dire economic repercussions and significantly affected individual lives, irrespective of their social or professional status.

Schools and colleges transitioned to online platforms, forcing students to adapt to virtual learning. A substantial portion of the global workforce transitioned to remote work, provided their occupations permitted. Stringent lockdown measures sequestered people within their homes for extended periods, leading to shortages of essential commodities and disruptions in public transportation services. The underprivileged, particularly the homeless, bore the brunt of the lockdown measures, further compounding their daily struggles. To counteract the relentless spread of COVID-19, WHO initiated public health measures in collaboration with nations and regions in the Western Pacific. Vaccination emerged as a critical strategy to protect populations, reduce mortality, and minimise societal disruptions. Many countries in the Western Pacific Region started rolling out their COVID-19 vaccination programs once they received the necessary regulatory permissions.

Finding a viable vaccine was fraught with challenges, yet science prevailed. On December 11, 2020, Pfizer's vaccine received(Patel, Kaki, Potluri, Kahar, & Khanna, 2022) emergency use

approval from the Food and Drug Administration (FDA), marking the beginning of a new chapter in the fight against the virus. The development and distribution of the vaccine were expedited, requiring extraordinary international cooperation. However, the success of this endeavour hinged on the population's willingness to vaccinate. Nature Medicine Trusted Source conducted surveys in October 2020 across 19 countries, revealing a notable vaccine hesitancy among the population. Among those surveyed, 71.5% indicated they would consider getting the COVID-19 vaccine, and only 48.1% would do so if their employer recommended it. By October 2021, healthcare workers worldwide had administered almost 7 billion doses of the COVID-19 vaccine. Despite these significant strides, vaccine hesitancy persisted.

Unequal access and distribution of vaccines across high-, middle-, and low-income countries, however, threatened to derail efforts to control the spread of the virus. By September 2021, 70% of people in high-income countries had been vaccinated against COVID-19, while only just over 2% in low-income countries had received the vaccine. Middle-income countries had vaccinated about 50% of their population, but Figure 1 was just 30% in low-income nations. This inequality fuellled the rampant spread of the virus, leading to increased morbidity and mortality rates. As the world grappled with this disparity, a new challenge emerged: the Omicron variant (B.1.1.529) (Fan et al., 2022). As identified by the WHO on November 26, 2021, this variant necessitated booster shots due to its potential for increased transmissibility or virulence. While initially, the booster shots were primarily given to the elderly and individuals with underlying health conditions, the advent of Omicron led to a broader application of booster shots to the general population aged 16 and above, provided they had received their second dose at least three months earlier.

Despite the pressing need for booster shots in light of the emerging variants, public hesitancy increased. This phenomenon was primarily due to the lack of understanding and mistrust regarding the rapid development of these booster shots. The health department recognised the need for public acceptance and urged people to get the booster vaccine.

The COVID-19 vaccination campaign ignited substantial debate globally. With the emergence of social media as a critical communication medium, these debates found a prolific platform for expression. Consequently, significant research(Ong, Pauzi, & Gan, 2022) has been conducted into vaccine hesitancy, with many studies utilising social media data, particularly from Twitter. However, studies focused explicitly on booster shot acceptance are scarce.

Previous research(Lai et al., 2021) on booster acceptance used methods ranging from cross-sectional surveys among healthcare workers in the United States to online self-administered questionnaires gauging the psychosocial drivers of booster shot hesitancy. However, none of these studies considered social media data in their analysis of booster shot acceptance. Given social media's vast reach and immediate feedback potential, this study aims to fill this gap by analysing data from popular platforms like Twitter and YouTube.

The impact of COVID-19 transcends geographical boundaries and demographic distinctions. The necessity to get vaccinated, albeit reluctantly for some, has become a harsh reality in this pandemic era. However, the introduction of the COVID-19 booster shot as a third dose was met with significant reluctance. This study focuses on the public perception of the third booster dose(Delcea, Cotfas, Crăciun, & Molănescu, 2022) of the COVID-19 vaccine. Furthermore, it compares vaccine hesitancy across high-, middle-, and low-income countries using a dataset of 11,359 tweets. These tweets were categorised as positive, negative, neutral, or informative, providing a comprehensive perspective on global booster shot acceptance. This approach could inform governments in low-income nations about vaccine acceptance, allowing them to take appropriate steps in refining their vaccination strategies.

The study highlights the importance of communication and its impact on public health measures in this digital age. Analysing social media discourse on booster shot acceptance in the context of countries' income status can provide insights that can help shape effective health communication strategies, promote vaccination, and eventually end this global crisis. This approach enables a deeper understanding of public perceptions toward vaccines and the degree of their acceptance.

© 2022 Sabaragamuwa University of Sri Lanka

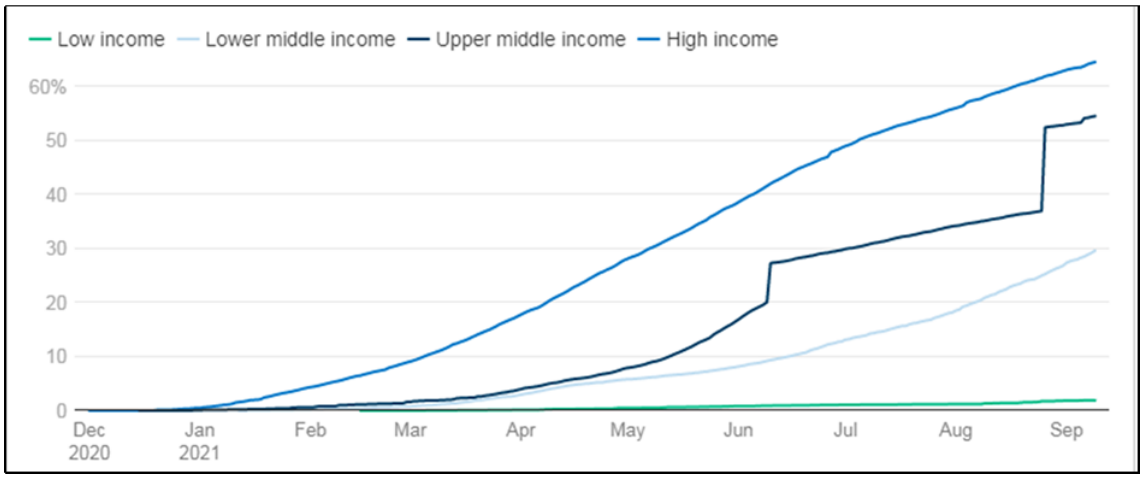

Figure 11: Share of the population that has received the vaccine by income Sourced from Tracking Global COVID-19 Vaccine Equity: An Update

2. Literature Review

As the COVID-19 outbreak and death toll increased, the COVID-19 vaccination program gained popularity globally. Therefore, there has been considerable debate about vaccination since the inception of the vaccine industry. Social media has emerged as an essential aspect of human life in the current digital era. Consequently, the vaccination produces a vast amount of original social media content. As a result, people who support and oppose vaccinations share their views on social media. This has led to numerous studies on COVID-19 vaccination reluctance. Some are based on Twitter data (Luo, Chen, Cui, & Liao, 2021), while others are based on Facebook, YouTube, and WhatsApp communications comments. Parents' hesitation to vaccinate their children has been evaluated on social media (Sumengen, Ozcevik, Kursun, & Ocakci, 2021). Additionally, several social media sites examined vaccine reluctance in light of geolocation(Nurdeni, Budi, & Santoso, 2021) and cultural context(Luo et al., 2021). However, the authors found very few boosters shot acceptance-related research.

Few studies on booster acceptability have been conducted globally. One of these studies included a cross-sectional survey to ascertain COVID-19 vaccine reluctance among US healthcare professionals. An online English language survey instrument was created using the REDCap electronic data-capturing software from the University of New Mexico[6]. An online self-administered questionnaire (SAQ) was utilised in another study to gather data on participant demographics, COVID-19 infection, vaccination anamneses, willingness to receive COVID-19 vaccine BD, and the psychosocial factors that influence vaccine booster hesitancy(Klugar, Riad, Mohanan, & Pokorná, 2021). Another study that evaluated vaccination literacy, vaccine confidence, trust, and overall sentiments regarding vaccines used a web-based 48-item psychometrically valid survey(Yadete et al., 2021). Several researchers collected data on booster uptake using conducting polls. However, social media data on booster shot acceptance was not considered. As a result, the authors decided to use the well-known social media platforms Twitter and YouTube to gather data. Many of the studies mentioned above are specifically dependent on geolocations. Some are exclusively accessible to medical professionals. But in this case, the authors consider the nation's economic predicament. Therefore, the authors intend to use social media analysis to assess the acceptance of COVID-19 booster shots and compare the levels of acceptance between high-income countries and low and middle-income countries (Ghazy et al., 2022; Lupu & Tiganasu, 2023; Solís Arce et al., 2021).

The arrival of two types of vaccinations in Indonesia, Sinovac and Pfizer, elicited differing reactions from Indonesian society. This research focuses on a sentiment analysis of the two types of vaccines based on Twitter data. Data from October to November 2020 were crawled and analysed to determine citizen opinion. Sinovac and Pfizer datasets were separated from the rest of the data. Both datasets were manually categorized as

© 2022 Sabaragamuwa University of Sri Lanka

positive, negative, and neutral categories (Nurdeni et al., 2021).

This project aims to evaluate COVID-19 vaccine-related tweets and produce a report based on the findings. The propagation of the virus and the daily increase in cases are the driving forces behind this research. On a dataset of 980,557 tweets scraped from Twitter based on a specific set of 14 keywords namely "Covaxin," "BNT162b2," "SputnikV," and "mRNA-1273," the method used for topic modelling is a new statistical model in Natural Language Processing (NLP) that belongs to the machine learning toolbox and artificial intelligence toolbox, Latent Dirichlet Allocation (LDA) (Praveen, Ittamalla, & Deepak, 2021) was used to determine popular themes, along with VADER and TextBlob.(Rahul, Jindal, Singh, & Meel, 2021)

Researchers used social media posts as data for the research. They scraped the social media posts from Indian residents discussing the COVID-19 vaccine using Python. Study 1 used sentiment analysis to examine how Indian residents' perceptions of the COVID-19 vaccine evolved during the COVID-19 pandemic. In Study 2, researchers used topic modelling to better understand the general public's key concerns regarding the COVID-19 vaccination. As a result, with just over 35% of people showing positive sentiment towards vaccinations, the Indian government should eliminate vaccine anxiety before executing the mass immunisation program(Praveen et al., 2021).

To compare public opinion on the COVID-19 vaccination, this study uses a cultural sensitivity approach and analyzes two popular social media platforms in the United States (Twitter) and China (Weibo). The semantic analysis results show that social media users in both countries shared common themes such as local vaccination programs, priority groups, COVID-19 variant problems, and the worldwide pandemic. Variation in cultural traits between the two countries accounted for these differences. The researchers concentrated on perceiving public health challenges in cross-cultural contexts and how social media can support health informatics research and the exploration of public attitudes during times of public health crisis(Luo et al., 2021).

Researchers aim to understand the general public's feelings and concerns about COVID-19 vaccinations and how these feelings and concerns change over time. They collected tweets about COVID-19 vaccines using Twitter Application Programming Interface (API) from December 2020 to April 2021, resulting in over 857,000 tweets after data cleaning. This dataset was subsequently subjected to sentiment-based topic modelling. They estimated the topic distribution to depict the topic evolution and utilised the coherence score to find the ideal topic number for topic modelling. The primary topics of positive tweets were (a1) recognition of the necessity of getting vaccinated (59.47 per cent), (a1) being grateful for the expectation of receiving the vaccine (9.92 per cent), and (a3) feeling pleased about vaccine administration (7.1 per cent).

Several months after the immunisations became publicly available, researchers revisited HCW attitudes toward COVID-19 vaccines. Only 7.9% of this research participants hesitated to take the vaccine's first or second dosage. Higher vaccine hesitancy was associated with younger age (18–40 years) and lower educational attainment (GED or less), whereas self-identified Asian racial identification was associated with increased acceptability of COVID-19 immunisation. Nearly two-thirds of respondents feared that vaccination would be ineffective against new strains and that booster doses would be required. However, only 14.3 per cent of vaccine-hesitant respondents were willing to receive a hypothetical booster dose. Vaccine apprehension, on the other hand, was shown to have demographic predictors that were similar to those previously described; some US HCWs' reluctance to receive booster doses could reflect a broader hesitancy toward other types of vaccination (Pal et al., 2021).

This research focused on examining the potential drivers of COVID-19 vaccination booster hesitancy (VBH) across Czech healthcare personnel. Between November 3 and 11, 2021, a national cross-sectional survey was conducted using an online self-administered questionnaire (SAQ) to investigate the participants' demographic characteristics, desire to receive COVID-19 vaccine booster dose (BD), and the psychosocial drivers of VBH. Because the study sample was designed to represent the target demographic, it demonstrated high BD acceptability (71.3%) among Czech HCWs, with 12.2% remaining hesitant and 16.6% opposing the currently available BD.

© 2022 Sabaragamuwa University of Sri Lanka

These findings align with those of other recent studies from Central Europe. Participants who were medical professionals, males, and older were more likely to accept BD than those who were not. Medical professionals, men, and older individuals were more willing to accept BD than allied health professionals, women, and younger individuals. When engaging with HCWs and other population groups, the BDs' perceived safety and ethical concerns of vaccine justice should be adequately addressed. The recommendation to postpone the COVID-19 vaccination mandate in favour of emphasising these altruistic considerations in public health messages is based on the altruistic grounds for BD acceptance, which include family protection, patient safety, and community health protection (Klugar et al., 2021).

This study will assess COVID-19 vaccine booster reluctance and its associated determinants in a nationally representative sample. Vaccination literacy, confidence, faith, and overall views regarding vaccines were measured using a web-based 48-item psychometrically validated survey. Chi-square (with a post hoc contingency table analysis) and independent-sample t-/Welch tests were used to assess the data. Nearly 62 per cent of the 2138 participants planned to receive booster doses, while the rest were undecided about getting the COVID-19 vaccination. In comparison to their non-hesitant peers, the vaccine-booster-hesitant group was more likely to be unvaccinated and less likely to intend to have their children immunised. Compared to non-hesitant individuals, a considerably more significant proportion of booster-hesitant persons showed little to no faith in the COVID-19 vaccination information provided by public health/government authorities (Yadete et al., 2021).

3.Methodology

The primary aim of this study is to discern public sentiment toward the COVID-19 booster shot using Twitter data. The research employed three distinct machine learning algorithms to detect hesitancy around the booster shot, rather than relying on a singular algorithm to achieve this. The data for this investigation were globally sourced from Twitter. While the research follows a linear process, the implementation of the study is multifaceted and complex.

3.1 Research design

The objective is to scrutinise Twitter content through machine learning techniques, focusing on the sentiment expressed—positive, negative, neutral, or informative. The methodology encompassed four stages: Initially, data was collected using specific keywords. A preprocessing stage followed this to prepare the dataset. Next, the analysis algorithms categorised the preprocessed data based on sentiment.

Lastly, the accuracy of each algorithm applied in this study was meticulously evaluated. The overall methodology is shown in Figure 2.

3.2. Data collection

Twitter remains a leading platform for the dissemination of daily updates and information. This research explicitly targets tweets that advocate for the COVID-19 booster shot. Given Twitter's vast user base, with approximately 206 million individuals disseminating nearly half a billion tweets daily, it provides a substantial data source for analysis. From December 2021 to April 2022, over 2,000 tweets related to the COVID-19 booster shot were harvested using selected keywords prevalent in discussions.

However, pinpointing tweets of interest within a specific topical domain can be formidable. Tweets pertinent to booster shots were extracted using a targeted keyword search strategy to refine the data. The Twitter API, accessible via a developer account, proved invaluable in this data extraction process. It allows for the collection of tweets containing designated keywords or hashtags. For this study, fifteen keywords were identified as critical, including #Covaxin, #BoosterShot, #covidbooster, #FullyVaccinated, #booster, #coronavirus, #COVID19, #Vaccine, #COVISHIELD, #boosted, #triplevaxxed, and #ModernaBooster. The collected tweets were then downloaded into an Excel spreadsheet using Postman and transformed into a CSV file for analysis, utilising the Python library Pandas for data manipulation.

Data collection includes information about the location, tweets, and hashtags users utilised in relation to booster shots globally between December 2021 and April 2022. The tweets from the different locations are displayed in Table 1.

© 2022 Sabaragamuwa University of Sri Lanka

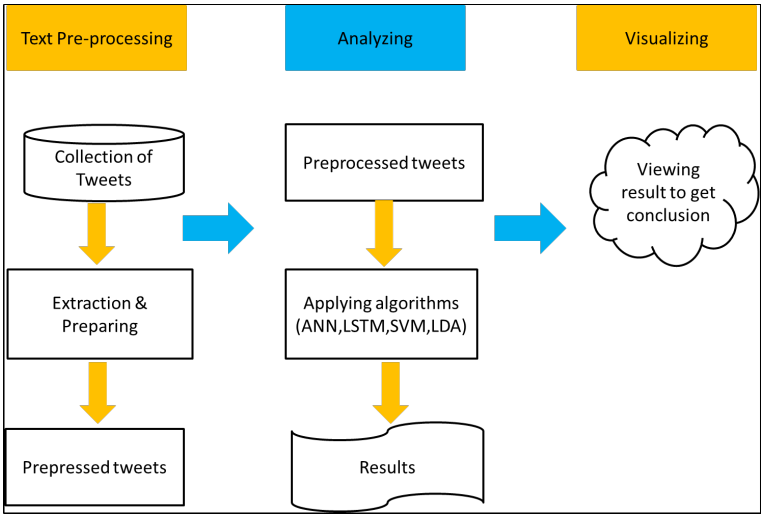

Figure 12: Research methodology

3.3 Software and hardware requirements and programming language

i. Intel ® Core ™ i5-8250U CPU

ii. RAM - 8GB

iii. Google Collaborator

iv. Python programming language

3.4. Data preparing

Twitter is a social media platform that allows users to share information. To make the tweets more interesting, these tweets frequently include attachments like documents, links, photographs, multimedia, emoticons, and other emotional cues. To prevent any distractions that can bias the results, it's crucial to have a clean dataset for the study that solely contains the text of the tweets. Preprocessing the tweets and eliminate non-textual parts can be accomplished using the Natural Language Toolkit's (NLTK) neat text module to accomplish this. It's vital to note that this phase is essential to obtain reliable results from the analysis. It is represented in the provided image.

3.4.1 Removing unwanted content

In data preparation, removing undesired content refers to cleaning up the data before it is analysed. This entails eliminating irrelevant tweets, hashtags, mentions, URLs, and other non-textual components that can bias the analysis's findings. This stage is crucial because it ensures that the data collection is concentrated on the subject under investigation and that accurate data support the analysis. These were accomplished using the neattext package from the Natural Language Toolkit (NLTK), which can be used to preprocess tweets and remove undesirable material.

3.4.2 Removing punctuation

Eliminating punctuation from the content of the tweets that can skew the analysisfindings is referred to as removing punctuation. This stage is crucial because it ensures that the data collection is concentrated on the subject under investigation and that accurate data support the analysis. Commas, exclamation points, question marks, and other punctuation can modify the meaning of the text or make it challenging to analyse. Therefore, all of the symbols denoted as punctuation (‘”, ”@”, ”#”, ”%”, ” ”, ”&”, ”*”, ”-”, ” -”, ”+”, ”=”, ”,”, ”.”, ”/”, ”?”, ”¿”, ”—”,”:”,”;”) have been removed in here.

3.4.3 Removing stop words

Stop words, such as ”a,” ”an,” ”the,” ”is,” ”are,” and similar terms, are removed from the text to make the analysis more meaningful. To reduce noise and boost performance, ”stop words” are often eliminated from the data set before processing. This phase is crucial since it lowers the datas dimensionality, facilitating quicker processing and more accurate analysis. Removing stop words will help ensure that the analysis is grounded in relevant facts and that the data set is more narrowly focused on the topic under study. Therefore, all the stop words were eliminated during the data preparation phase.

© 2022 Sabaragamuwa University of Sri Lanka

Table 14: A few records of the original data set

Location	Text
Michigan State University, MI, USA	The COVID-19 vaccine has been made available around the world. Now, COVID-19 booster shots have begun rolling out.â€¦ https://t.co/URtNusJtM3
Washington, USA	We decided to do the other #MRNA brand for our booster FWIW. Why not? #health #prevention #vaccine #Pfizer #Modernaâ€¦ https://t.co/DVtZ2NJAhp
Ansari Nagar, New Delhi	After a lot of hard work, our team sees success. Published in @TheLancetInfDis @TheLancet #COVAXIN #vaccineâ€¦ https://t.co/PwJT5h98AF
England, United Kingdom	Currently riding the covid booster side effect wave #Covid_19 #Moderna https://t.co/xrNEozLoEI
Cambridge, England	I have just had a booster shot. In the end it was Moderna as that's what was in stock today. Relieved to have had tâ€¦ https://t.co/5kADBzWWVL
Indore, India	#CDC updated it's guidelines on #COVID19 #booster doses compared to last version which recommended booster for highâ€¦ https://t.co/iUcn6b4io9
Rhode Island, USA	Received my #ModernaBooster today @Walgreens. I didn't feel the didnâ€™t at all! My arm is a bit sore but thatâ€™s it. Pleasâ€¦ https://t.co/C9RUHLNDGV
Sri Lanka	"@rjrebert @MoH_SriLanka Your correct. Many countries have proposed a 6 month interval for the booster. The answer iâ€¦ https://t.co/rFuhhQCjQp"

3.4.4 Tokenising

The technique of tokenising involves dividing the text of tweets into tokens, which might be words, phrases, or even individual letters. This phase is crucial for several reasons, including the ability to examine the meaning of each word, phrase, or character separately, perform more accurate analysis, and reduce the dataś dimensionality, making it more straightforward to process and analyse. The data will be more valuable and informative for subsequent analysis after being tokenised.

3.4.5 Stemming

Stemming is the process of condensing words to their stem, fundamental form, or root. This step is crucial because it makes the data set more consistent and straightforward to analyse by combining several word variations (such as "running," "runner," and "ran") as one. Additionally, stemming reduces the dimensionality of the data, enabling more accurate analysis. Stemming will help to lower the dataś dimensionality and increase the data set's usefulness and informativeness for additional analysis.

3.4.6 Split the data set into train and test

It is possible to assess the model's performance on previously unexplored data and ensure it generalises well to new examples by dividing the data into training and test sets. The usual split is between 70 and 80 per cent of the data for training and between 20 and 30 per cent for testing. Therefore, in this study, the prediction model was constructed utilising the selected machine learning techniques after completing the data preprocessing that came before is finished. The data was split into a training and testing data set before building the machine learning model. Here, the authors divided the total into 30% for model testing and 70% for model training.

3.5 Implementation of machine learning algorithm

The supervised machine learning algorithms used in the current study are the Support Vector Machine (SVM) learning method, the Artificial Neural Network (ANN) algorithm, and the Long Short Term Memory (LSTM) algorithm

© 2022 Sabaragamuwa University of Sri Lanka

(supervised machine learning algorithm). This study focused on LDA as a topic modelling algorithm for comparison.

4 Results and Discussions

4.1 Performance Metrics

The effectiveness of an analysis model using Twitter data about booster acceptance can be assessed using several performance indicators. The most popular metrics are Accuracy, Precision, Recall, F1-score, Root Mean Squared Error, etc.

4.2 Module results

Table 2 provides a comparative analysis of the three algorithms utilised in our study: Artificial Neural Networks (ANN), Long Short-Term Memory Networks (LSTM), and Support Vector Machines (SVM). It showcases the accuracy, precision, and recall metrics for each method.

Accuracy in our context denotes the algorithm's ability to correctly classify tweets into positive, negative, neutral, or informative categories. A higher accuracy percentage reflects better performance in accurate tweet classification. In this study, ANN leads with an accuracy of 93.693%, followed by LSTM at 85.764% and SVM at 80.647%.

Precision measures the algorithm's success in correctly categorising genuinely commendable tweets. Higher precision indicates fewer false positives. ANN outperforms with a precision score of 0.925, followed by LSTM at 0.722 and SVM at 0.533.

Recall is the metric that assesses the algorithm's proficiency in identifying all relevant tweets. A higher recall value signifies fewer false negatives. In this study, all three algorithms—ANN, LSTM, and SVM—achieved recall scores above 0.70, with ANN achieving the highest at 0.813.

ANN is the superior algorithm across accuracy, precision, and recall metrics. While LSTM shows commendable accuracy, it trails in precision and recall. While the least effective of the three, SVM demonstrates the importance of choosing the right algorithm based on specific performance criteria, as shown in Figure 3.

This studyś findings indicate a comparatively low concern over vaccine hesitancy in high-income countries, as reflected in the reduced volume of Twitter discussions about the COVID-19 booster shot. In contrast, middle-income countries demonstrate a heightened interest in the acceptability of the booster shot, as evidenced by the significant increase in related tweets.

The research then focused on a comparative analysis between high-income and middle-income countries. To facilitate this comparison, the authors employed the Latent Dirichlet Allocation (LDA) algorithm for topic modelling. The LDA algorithmś results summary delineates the differing responses and discussions surrounding booster shots in high-income versus middle and low-income countries, as listed in Table 3.

4.3 Discussion

This research has analysed COVID-19 booster shot-related tweets within an economic context by employing topic modelling. The study demonstrates that Twitter, a widely utilised platform, is vital for gauging public sentiment toward health promotion strategies aimed to increasing vaccination rates while reducing hesitancy and opposition. Understanding public emotions and viewpoints on vaccines enables public health authorities to bolster positive discourse in affirmative posts and challenge the aggressive rhetoric in negative posts that often perpetuate misinformation.

Furthermore, public health bodies could strategically leverage Twitter and other social media platforms to amplify supportive messaging, reduce adversarial narratives, and decisively restrict accounts, such as bots that disseminate anti-vaccination content, thereby fostering a more conducive environment for vaccine uptake. The findings suggest that discourse on vaccine research, public opinion, and the effectiveness and implementation of vaccines resonates strongly with the public's demand for information during the COVID-19 pandemic, thus gaining traction on Twitter. Such discourses can provide critical, timely insights that help alleviating uncertainty during health crises. Vaccine initiatives could benefit from placing a greater emphasis on these topics to enhance the spread of information through social media channels.

However, it is essential to note that account verification does not necessarily equate to content veracity. Consequently, misinformation spread by verified accounts may challenge vaccine promotion efforts. The researchers have proffered interpre-

© 2022 Sabaragamuwa University of Sri Lanka

Table 14: Result of the three algorithms used by the study

Algorithm	Accuracy	Precision	Recall
ANN	93.693	0.925	0.813
LSTM	85.764	0.722	0.714
SVM	80.647	0.533	0.610

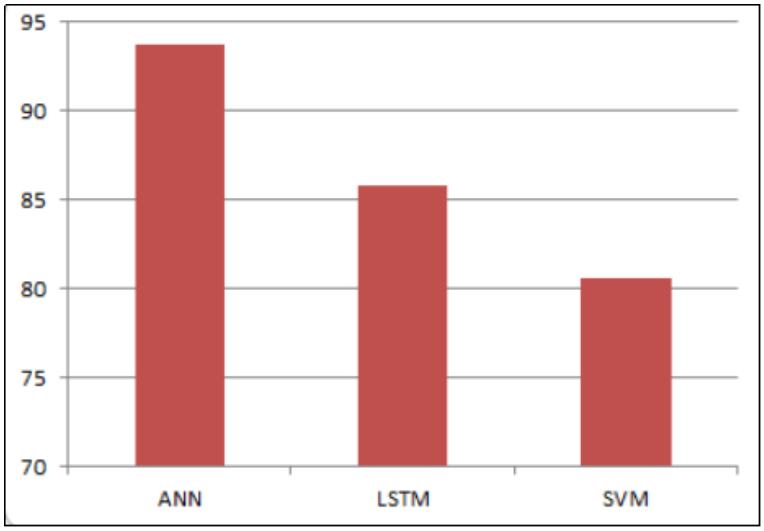

Figure 13: Accuracy of the three algorithms

tations pertinent to high and middle-income countries based on the topics generated.

Admittedly, several challenges were encountered in applying this research method. Vaccine campaigns should consider engaging and incentivising various verified accounts, both institutional and personal, to counter misinformation and disseminate accurate information. Nonetheless, the current study is not without its limitations. Although English is a prevalent language on social media, it does not capture the entire spectrum of vaccine-related discourse. The data set used to categorise anti-vaccine, pro-vaccine, and vaccine-hesitant tweets may not encompass the full breadth of vaccine-related expressions on Twitter. The data query was confined to specific terms related to COVID-19 and vaccines, potentially omitting relevant content. Moreover, while efforts were made to categorise user geographic data, the self-reported locations in user profiles may not always accurately indicate the actual locations from which tweets originate.

One proposed solution is implementing content warnings for posts containing misinformation, allowing users to conceal or block such material. Nevertheless, these measures alone are insufficient to ensure widespread acceptance of the COVID-19 vaccine. There is a pressing need for comprehensive strategies that not only counteract misinformation but also proactively foster informed decision-making among the public.

5 Conclusion

This study has utilised three principal unsupervised machine learning methods: Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks to analyse public acceptance of COVID-19 booster vaccinations. Through these methods, topics generated via topic modelling were further examined, with the predictive models demonstrating varied levels of efficacy. The SVM method yielded the lowest accuracy at 80.64% in determining acceptance of booster shots, whereas the ANN algorithm achieved the highest accuracy at 93.69%.

Latent Dirichlet Allocation (LDA), a topic modelling algorithm, was employed to identify salient topics. The prevalence of keywords within these topics correlated with the acceptance levels of booster shots across high-income and middle-income countries. The analysis revealed that hesitancy towards booster shots is less prevalent

© 2022 Sabaragamuwa University of Sri Lanka

Table 14: Summary of the topics

Topics with High-income countries
Topic -1 - COVID-19 Vaccine Updates: Approval of Moderna, COVAXINE, Sinovac as third dose.
Topic -2 - COVID-19 3rd dose vaccination: Efficacy of COVAXIN, Bharat Biotech, Sputnik V, and COVISHIELD.
Topic -3 - Getting COVID-19 vaccines as third dose: Moderna, COVAXIN, Pfizer, Sinovac, Sputnik V, and Doses Needed as 3rd dose.
Topic -4 - Getting the COVID-19 3rd dose: Moderna, COVAXIN, Pfizer, AstraZeneca, and Latest Updates on Dosages.
Topics with Low- and middle-income countries
Topic -1 - COVID-19 3rd dose vaccination: Approval of COVAXIN, Sputnik V, Moderna, Sinopharm, Pfizer, COVISHIELD, and Bharat Biotech
Topic -2 - COVID-19 3rd dose vaccination: Efficacy of COVAXIN, Bharat Biotech, Sputnik V, and COVISHIELD.
Topic -3 - Getting the COVID-19 3rd dose/ booster: Availability of COVAXIN, COVISHIELD, Sputnik V, and Other Vaccines.
Topic -4 - COVID-19 booster shot vaccination: COVAXIN, COVISHIELD, Sputnik V and the Role of the Government, Bharat Biotech and Hospitals.

in high-income countries than in their middle-income counterparts. Moreover, the frequency of tweets expressing vaccine hesitancy is lower in high-income countries.

An extensive analysis of over 12,000 tweets has provided a broad perspective on public sentiment regarding the COVID-19 booster shot. The findings suggest that a more detailed understanding could be attained by analysing currently popular forms of communication, such as memes, which offer insights into public attitudes toward booster vaccinations. The study uncovered public perceptions and the thematic essence conveyed through the LDA, presenting an opportunity for future research to expand upon the methodology by incorporating various filtering categories within the analysis. Furthermore, integrating sentiment analysis could yield additional depth in assessing public sentiment.

References

Delcea, C., Cotfas, L.-A., Crăciun, L., & Molănescu, A. G. (2022). New wave of COVID-19 vaccine opinions in the month the 3rd booster dose arrived. *Vaccines, 10*(6), 881.

Fan, Y., Li, X., Zhang, L., Wan, S., Zhang, L., & Zhou, F. (2022). SARS-CoV-2 Omicron variant: recent progress and future perspectives. *Signal transduction and targeted therapy, 7*(1), 1-11.

Ghazy, R. M., Abdou, M. S., Awaidy, S., Sallam, M., Elbarazi, I., Youssef, N., . . . Adam, M. F. (2022). Acceptance of COVID-19 vaccine booster doses using the health belief model: a cross-sectional study in low-middle-and high-income countries of the East Mediterranean region. *International journal of environmental research and public health, 19*(19), 12136.

Hasöksüz, M., Kilic, S., & Saraç, F. (2020). Coronaviruses and sars-cov-2. *Turkish journal of medical sciences, 50*(9), 549-556.

Klugar, M., Riad, A., Mohanan, L., & Pokorná, A. (2021). COVID-19 vaccine booster hesitancy (VBH) of healthcare workers in Czechia: National cross-sectional study. *Vaccines, 9*(12), 1437.

Lai, X., Zhu, H., Wang, J., Huang, Y., Jing, R., Lyu, Y., . . . Fang, H. (2021). Public perceptions and acceptance of COVID-19 booster vaccination in China: a cross-sectional study. *Vaccines, 9*(12), 1461.

Lamers, M. M., & Haagmans, B. L. (2022). SARS-CoV-2 pathogenesis. *Nature reviews microbiology, 20*(5), 270-284.

Luo, C., Chen, A., Cui, B., & Liao, W. (2021). Exploring public perceptions of the COVID-19 vaccine online from a cultural perspective:

© 2022 Sabaragamuwa University of Sri Lanka

Semantic network analysis of two social media platforms in the United States and China. *Telematics and Informatics, 65*, 101712.

Lupu, D., & Tiganasu, R. (2023). COVID-19 vaccination and governance in the case of low, middle and high-income countries. *BMC Public Health, 23*(1), 1073.

Nurdeni, D. A., Budi, I., & Santoso, A. B. (2021). *Sentiment analysis on Covid19 vaccines in Indonesia: from the perspective of Sinovac and Pfizer.* Paper presented at the 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT).

Ong, S.-Q., Pauzi, M. B. M., & Gan, K. H. (2022). *Text mining and determinants of sentiments towards the COVID-19 vaccine booster of Twitter users in Malaysia.* Paper presented at the Healthcare.

Pal, S., Shekhar, R., Kottewar, S., Upadhyay, S., Singh, M., Pathak, D., . . . Sheikh, A. B. (2021). COVID-19 vaccine hesitancy and attitude toward booster doses among US healthcare workers. *Vaccines, 9*(11), 1358.

Patel, R., Kaki, M., Potluri, V. S., Kahar, P., & Khanna, D. (2022). A comprehensive review of SARS-CoV-2 vaccines: Pfizer, moderna & Johnson & Johnson. *Human vaccines & immunotherapeutics, 18*(1), 2002083.

Praveen, S., Ittamalla, R., & Deepak, G. (2021). Analysing the attitude of Indian citizens towards COVID-19 vaccine–A text analytics study. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 15*(2), 595-599.

Rahul, K., Jindal, B. R., Singh, K., & Meel, P. (2021). *Analysing public sentiments regarding the COVID-19 vaccine on Twitter.* Paper presented at the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS).

Solís Arce, J. S., Warren, S. S., Meriggi, N. F., Scacco, A., McMurry, N., Voors, M., . . . Adeojo, O. (2021). COVID-19 vaccine acceptance and hesitancy in low-and middle-income countries. *Nature medicine, 27*(8), 1385-1394.

Sumengen, A. A., Ozcevik, D., Kursun, H. Y., & Ocakci, A. F. (2021). Vaccine-hesitant parents' reasons for choosing alternative protection methods in Turkey. *Journal of Research in Nursing, 26*(6), 540-553.

Wu, D., Wu, T., Liu, Q., & Yang, Z. (2020). The SARS-CoV-2 outbreak: what we know. *International journal of infectious diseases, 94*, 44-48.

Yadete, T., Batra, K., Netski, D. M., Antonio, S., Patros, M. J., & Bester, J. C. (2021). Assessing the acceptability of COVID-19 vaccine booster dose among adult Americans: a cross-sectional study. *Vaccines, 9*(12), 1424.

© 2022 Sabaragamuwa University of Sri Lanka