Texila International Journal of Public Health
ISSN: 2520-3134
DOI: 10.21522/TIJPH.2013.12.03.Art034

Public Health Surveillance for COVID-19 Using Twitter Sentiment Analysis

Kashim I. A
Public Health, Texila American University, Central University of Nicaragua

Abstract

By analyzing sentiments on Twitter/X during the COVID-19 pandemic, this study adds to the existing body of literature on new techniques to analyze social media's 'big data'. In this study, a sentiment classification model was developed to accurately predict public sentiments on Twitter/X. Twitter's Apify crawler was used to scrape a sample of English language tweets from the social media platform using hashtags:#covid-19vaccine,#coronavaccine,#vaccinesaveslives, #getvaccinated, #coronavirusvaccine. The study was from December 2020 to November 2021. Data preprocessing techniques were first applied followed by a hybrid approach for sentiment analysis (both lexicon based natural language processing and machine learning). Approval for the study was received from Texila American University, ethical concerns were limited as no personalized data or human subjects were used. Data processing and analysis using PYTHON involved the use of Natural language processing techniques (VADER) to classify the sentiment and predict accuracy of the model developed. Trend analysis showed that as tweets increased with each month, tweets became negative with fear and anxiety being commonly expressed emotions. Over the study period, there was a statistically significant difference in sentiment polarity (positive p=0.000; negative p=0.02). VADER analysis predicted that this trend (increasing negative polarity with time) is likely to continue in future epidemics as correlates with existing machine learning models (Random Forest and Support Vector Machine) both validated this trend. It is evident that sentiment analysis techniques can be leveraged for the purpose of public health disease surveillance and to enable the identification of trends, forecasting public perception and behavior.

Keywords: *Covid-19, Machine Learning, Pandemic, Sentiment Analysis.*

Introduction

In recent years, social media platforms have become integral in shaping public opinions and reflecting societal trends. The COVID-19 pandemic has brought to the forefront the importance of understanding the discourse surrounding this global crisis. Twitter/X, with its vast user base and real-time nature makes it a valuable source for analysis of public sentiment related to the pandemic. By analysing 'big data', researchers can get valuable insights into public perceptions, concerns, and behaviours. Twitter/X has played a crucial role as a source of health information, allowing for the exchange of experiences and ideas that are related to prevention and treatment.

Since individuals have turned to social media platforms to voice their sentiments and experiences, the dynamic interplay of emotions has indicated broader societal anxieties and coping mechanisms in response to the pandemic. In particular, social media has served dual roles as both a source of connection and a vector for misinformation, complicating the public's understanding of the crisis [1]. This underscores the vital need for effective public health surveillance and communication to guide their emotions. Studies have suggested that with a deeper understanding of science and reminders to check source accuracy, people can

Received: 19.07.2024 ***Accepted: 12.09.2024*** ***Published on: 30.09.2024***
Corresponding Author: irkashim@gmail.com

differentiate between true and false information, which improves their decisions related to information sharing on social media [2-7].

There is a growing need for methodologies that can be leveraged to extract insights from the massive volume of real-time internet data. Researchers can harness the power of ML to analyze vast amounts of Twitter/X data efficiently, and rapidly gain valuable insights into the complex dynamics of public opinion during significant events like the COVID-19 pandemic. This ability to rapidly generate insights is a strength of Natural Language Processing (NLP) and machine learning (ML) techniques. Exploring the intersection of social media sentiment and public health surveillance can inform health policy decisions and strategies in real time. Moreover, social media as a complementary surveillance tool can be useful for tracking public sentiment across affected countries and demonstrating coping mechanisms across geographic regions during crises [8, 9]. A study that utilized this approach to analyse the emotional indicators extracted from COVID-19 Geo-Tagged tweets demonstrated significant mental health implications across diverse regions, suggesting that policymakers should consider public emotional distress as a factor in their strategies [10].

Ultimately, integrating innovative surveillance methods can enhance the capacity to manage public health challenges effectively, and enable public health officials to better understand community concerns and tailor interventions particularly during public health crises, such as the COVID-19 pandemic. The immediacy and interactivity of social media allow for real-time monitoring of public opinions and behaviors related to health risk factors. For instance, recent analysis has demonstrated how Twitter sentiment can be indicative of shifts in cancer risk factors, particularly during the COVID-19 pandemic, with findings revealing positive sentiments related to physical activity and nutrition, despite rising anxiety levels linked to smoking and alcohol use [11].

NLP and ML approaches have enabled researchers, program planners and public health experts to uncover nuances and sentiments that would not be immediately apparent through manual analysis, thus enabling data-driven decision making and recognizing how public perspective changed during the ongoing pandemic [12]. This has been instrumental for developing focused interventions, appropriate policies and improved communication strategies to promote public health. However, using Twitter data for research, particularly in the context of public health surveillance, raises significant ethical considerations that demand careful scrutiny. Researchers must navigate issues of privacy and consent, as tweets are public by nature yet encompass personal expressions of sentiment and experience. The need to balance the utility of this data for understanding public health responses during emergencies, as exemplified by the COVID-19 pandemic, against the potential for harm or misinterpretation is paramount. Ethical frameworks must be established to ensure that data usage respects individuals rights and accurately reflects the context of the messages. By addressing these ethical dimensions, researchers can promote responsible practices that uphold the integrity of public health discourse.

ML algorithms can accurately determine the sentiment behind emoticons and decipher the meaning of abbreviations in tweets. A range of computational methods are used in Quantitative Text Analysis (QTA) [13] to convert textual data or natural language into structured format for sentiment analyses. High-quality preprocessing of data mitigates bias and improves the accuracy of sentiment analysis.

This study contributes to the existing body of literature by offering a comprehensive statistical and ML approach for COVID-19 sentiment analysis on Twitter/X. The aims are

1) to provide valuable insights into public perceptions, behaviors and concerns related to the pandemic and 2) recognize how these perceptions and behaviors evolved over the study period.

Materials and Methods

In accordance with Twitter/X public policy, registered users provide consent for collection of some data for aggregate analytics and research purposes. Apify's Twitter Scraper was used to collect tweets, without the need for a Twitter account. The relevant hashtags and the time frame for the study were specified. The following hashtags were used: #covid-19vaccine, #coronavaccine, #vaccinesaveslives, #pfizercovidvaccine, #astrazenecacovidvaccine, #modernacovidvaccine, #sinopharmcovidvaccine, #getvaccinated, #coronavirusvaccine. They were organized by month within a time frame of December 2020 to November 2021. The extracted tweets were accessed in a CSV format.

VADER Sentiment Analysis. VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media and works well on texts from other domains. See Fig 1 for the conceptual framework utilized in this study.

Using Python libraries for natural language programming, pre-processing techniques were applied to improve the quality of the data, reduce noise and prepare it for the machine learning models. Preprocessing steps included filtering, tokenization, stemming, and stop-word removal. All links, special characters, and usernames were removed from the tweets. English tweets with the above hashtags were loaded into the DataFrame in Python. Each tweet was assigned a sentiment label (positive, negative, and neutral). Tweets with positive sentiment (value greater than zero) indicated an optimistic attitude towards COVID-19 vaccination program, while tweets with negative sentiment (value not greater than zero) represented a predominantly pessimistic attitude towards the vaccination program. Tweets with neutral sentiment (value equal to zero) indicated a lack of strong emotion or opinion. One way ANOVA was used to compare sentiments across the study period.

Steps in using VADER sentiment analysis are outlined below:

Algorithm 1: Sentiment Analysis of Twitter Data (SATD)
Input: Twitter Data (TD) using *APIFY (for Twitter)*
Output: Sentiments of TD

1. Take COVID -19 TD from 1 June 2021 to 31 December 2021
2. Preprocess and clean TD using *Pandas*
3. Classify sentiment using *VADER* sentiment function
4. Sentiment _Value = *VADER* (sentence).sentiment [0].

If Sentiment Value: = 0
Return 0 //"Neutral"
If Sentiment Value > 0:
Return 1 //"Positive"
else:
Return 2 //"Negative"

5. Compute Sentiment Value
6. Use *VADER* to measure accuracy of sentiment expression in the future

Steps in Supervised Machine Learning using both Sector Vector Machines (SVM) and Random Forest (RF).

1. Split the data into training sets (80%) and test sets (20%) using Most Persistent Feature Selection (MPFS) method

2. Develop baseline classifier model using 2 training and testing models - Random Forest (RF) and Support Vector Machines (SVM)
3. Conduct regression analysis in the calibration phase of the trained datasets.
4. Compare the performance of the baseline classifier models using five statistical metrics in the verification phase on selected data sets.

Machine learning (ML) methods for sentiment analysis rely on the feature selection of labeled data sets to perform the classification task. Using Supervised Learning Algorithms (SLA) a classifier model was developed to train features (words or phrases in the text) on 80% of the data, which is then used to classify the unseen test data (20%). We used Most Persistent Feature Selection (MPFS) method [14]. and developed a baseline classifier model using Random Forest (RF) and Support Vector Machines (SVM).

The popular machine learning algorithm RF is renowned for its adaptability and capacity to manage problems involving both classification and regression [15]. SVR is a data-driven learning machine for solving any classification, regression analysis, prediction, and pattern recognition problems [16]. Using two models for combining in the study enables the convergences of information from several sources on the surface features of the data.

The ensemble classifiers were trained on bigram as well as trigram features in the classifier model and fitted to train the data in SVM, where a nonlinear kernel was then used to capture the nonlinear structure of the data. The accuracy of regression analysis in both SVM and RF, was compared with the baseline classifier models using five statistical metrics: Pearson correlation coefficient (PCC), Willmott Index (WI), mean absolute error (MAE), mean absolute percentage error (MAPE) and Positive Bias (PBIAS) [see Equation 1 showing the formula used].

Results

Out of 194,378 tweets in English, a total of 144,911 (74.5%) were studied using selected hash tags related to COVID -19 vaccination. Overall, the most frequently expressed sentiment toward vaccination was neutral followed by negative and then positive (mean values: neutral 6126.8; negative 3282.2; positive 2666.9) [Table 1]. The results of one-way ANOVA using VADER demonstrated a significant difference between the mean positive sentiment expressed by Twitter users over time (df=1, F-statistic of 29.406, p=0.000) and significant difference in mean negative sentiment (df=1, F-statistic of 7.599, p=0.02). However, Twitter users predominantly expressed neutral sentiments regarding COVID-19 vaccination, with no significant differences compared to positive and negative sentiments (df=1, F-statistic of 0.304, p-value = 0.593) [Table 2]. VADER analysis forecasted a lower probability of expressing positive sentiment (ranging from 15% to 17%) and a higher probability of expressing negative sentiment (ranging from 30% to 35%). Interestingly, the probability of expressing neutral sentiment was highest among all categories (ranging from 48% to 50%).

We used SVM and RF for both classification and regression. Regression analysis showed that every month the number of tweets with negative sentiment score increased (p=0.02) and positive sentiment score decreased (p=0.00). There was an inverse relationship between number of tweets with neutral sentiment score that was not significantly different from other sentiments (p=0.593).

The performance analysis of these ML models was generated based on five indices: [Table 3].

Pearson's Correlation Coefficient (PCC), which measures the linear relationship between predicted and actual sentiment values. PCC values close to one show strong correlations.

Willmot Index (WI), which assesses the agreement between predicted and observed

sentiment values. WI values close to one suggest accurate prediction.

Mean Absolute Percentage Error (MAPE), which calculates the average percentage difference between predicted and actual values. MAPE values close to zero signify better accuracy.

Mean Absolute Error (MAE), which measures the average absolute difference between predicted and actual values. MAE values close to zero indicate high accuracy.

Percentage Bias (PBIAS), which assesses the overall bias in predictions with positive values indicating overestimation and negative values indicating underestimation. A value close to zero signifies balanced predictions.

Calibration Phase

Support Vector Regression (SVR) formulates prediction as a convex optimization problem, ensuring that it finds the global minimum during training. This avoids convergence issues and produces more stable and reliable results.

Neutral sentiment (NRS)

SVR-NRS:

1. PCC: 0.972 (Very high correlation).
2. WI: 0.800 (High agreement).
3. MAPE: 0.208 (Very low average percentage error).
4. MAE: 0.002 (Minimal absolute errors).
5. PBIAS: -0.041 (Slight underestimation).

RF-NRS:

1. PCC: 0.956 (Very high correlation).
2. WI: 0.734 (Moderate agreement).
3. MAPE: 0.359 (Low average percentage error).
4. MAE: 0.002 (Minimal absolute errors).
5. PBIAS: -0.002 (Balanced predictions).

Both RF-NRS and SVR-NRS had very high PCC values (0.956 and 0.972 respectively) indicating strong correlation. RF-NRS had a WI value (0.734) and SVR-NRS (0.800) indicating a close match between predictions and actual value. SVR-NRS had the lowest MAPE (0.208). Results were classified into very high, high, moderate and small depending on the different metrices. For PBIAS classification was underestimation, balance, and overestimation of the bias.

Negative sentiment (NS)

SVR-NS:

1. PCC: 0.909 (Very high correlation).
2. WI: 0.786 (High agreement).
3. MAPE: 0.926 (Moderate average percentage error).
4. MAE: 0.003 (Small absolute errors).
5. PBIAS: -0.189 (Underestimation).

RF-NS:

1. PCC: 0.862 (High correlation).
2. WI: 0.571 (Moderate agreement).
3. MAPE: 1.663 (Decent average percentage error).
4. MAE: 0.004 (Small absolute errors).
5. PBIAS: -0.255 (Underestimation).

Both models indicated a slight underestimation of PBIAS (SVR-NS, -0.189; RF-NS, -0.255).

Positive sentiment (PS):

SVR-PS:

1. PCC: 0.859 (High correlation).
2. WI: 0.785 (High agreement).
3. MAPE: 0.373 (Low average percentage error).
4. MAE: 0.002 (Minimal absolute errors).
5. PBIAS: 0.052 (Balanced predictions).

RF-PS:

1. PCC: 0.469 (Moderate correlation).
2. WI: 0.455 (Fair agreement).
3. MAPE: 0.738 (High average percentage error).
4. MAE: 0.004 (Small absolute errors).
5. PBIAS: 0.477 (Overestimation).

RF-PS showed very poor results with WI=0.455 and PBAIS=0.477 while SVR-PS attained better results with WI=0.785 and PBAIS= 0.052. Similarly, SVR-PS and RF-PS showed reduced MAE values (0.002 and 0.004 respectively).

Verification Phase

SVR-PS, SVR-NS and SVR-NRS as well as RF-NS and RF-NRS showed high PCC values indicating a high level of accuracy in predicting sentiment. In this instance, RF-PS did not show strong correlation. SVR-PS and SVR-NRS have high WI values, suggesting that their predictions match well with the actual sentiment values. Low WI values were noted for RF-PS and RF-NS which show that there is low agreement with the actual sentiment values.

Lower MAPE and MAE values are desirable, as they indicate smaller prediction errors. SVR-PS, SVR-NS and SVR-NRS achieved low MAPE and MAE values, indicating their superior performance in this aspect. RF-NRS had lower MAPE and MAE values than RF-PS and RF-NS.

A PBIAS close to zero indicates balanced predictions. SVR-PS has a slightly negative PBIAS, indicating a slight underestimation. Hence, SVR-PS and SVR-NRS stand out as the most accurate and reliable approaches in this study as a result.

Discussion

Our study contributes to the ongoing discourse on the use of sentiment analysis for opinion mining from social media platforms and deploying new innovative approaches to respond to public health emergencies. It proffers a role for artificial intelligence in the utilization of innovative ML and NLP as tools for public health surveillance by offering a snapshot of public sentiment in the period, uncovering the trends and thus enabling a more nuanced understanding of public perception to the COVID 19 pandemic and attitude to vaccination over the study period. As methodologies advance on surface learning of the characteristics and patterns of features.

The trend analysis was based on the use of NLP techniques for sentiment analysis and underscores the dynamics that unfolded within the realm of social media during that challenging time. Overall, neutrality was prevalent with no significant divergence from other non-neutral sentiment expressions. As tweet volume increased, there was an increase in negative sentiment, coupled with a decrease in positive and neutral sentiments over time. The ML algorithms predicted higher likelihood of negative sentiment compared to positive sentiment over time, with a significant portion of the population adopting a neutral stance, possibly indicating a state of uncertainty or a reserved approach to tweeting about their real opinions of vaccination. In the distinctive approach deployed in this study, machine learning predictive tools were used to decipher and make predictions about positive sentiment (PS), negative sentiment (NS), and neutral sentiment (NRS). Overall, our results show that SVM was a more accurate approach in all sentiment predictions with minimized prediction errors. The integration of advanced data preprocessing techniques enhanced the accuracy of sentiment classification, underscoring the feasibility of employing social media analytics in public health surveillance. VADER analysis predicted that negative sentiment potentially related to vaccine hesitancy, is likely to continue in future outbreaks.

In the future, public health management will be increasingly driven by big data and human behavior forecasting. Understanding public perceptions of the COVID-19 pandemic and vaccination can provide real-time insights for informed decision-making and the development of targeted responses and communication strategies [17]. Extensive research has been conducted on public opinions and emotions toward COVID-19 vaccines, employing various methodologies over time to identify key themes. For instance, a study from India found that transparent and informative messaging can increase the proportion of positive comments on Twitter [18]. Other studies have highlighted the need to build trust in vaccines by countering misinformation campaigns, with a rapid review showing that

such campaigns are negatively associated with vaccine uptake [19-21]. An analysis of persuasion techniques has revealed that celebrities are used to post anti-vaccine messages on Twitter, often focusing on conspiracy theories and spreading fear and concerns about vaccine safety and choice [22]. Trust has emerged as a prevalent positive emotion regarding COVID-19 vaccines [23]. Sentiment analysis across age and gender demographics has indicated that individuals over 40 generally hold more positive views, emphasizing the need for appropriate messaging aimed at younger age groups, particularly males [24]. This suggests the importance of adolescent-friendly messaging on popular social media platforms. User behaviors toward vaccination are noted to be influenced by community and individual factors, varying with sociocultural contexts, social circumstances, and personal experiences [25]. These insights have emphasized the importance of responsive public health messaging to address sentiment shifts, dispel misinformation, and enhance the overall confidence in vaccination initiatives as vaccination programs evolve [26].

Social media can also serve as a platform to disseminate accurate information about the pandemic, as well as preventive measures, diagnostic and treatment protocols, to the public, healthcare professionals and scientists [27, 28]. Crowdsourcing scientific knowledge has been noted as an effective way to examine news [29]. The World Health Organization has shared infographics that address myths about the pandemic [30]. Several studies have recommended that health agencies, health information experts, scientists and journalists should take ownership in the fight against misinformation [31-33]. Other studies have reported that tagging misinformation as going against public health guidance increases transparency [34, 35]. During the pandemic, several Health ministries around the world used online communication, especially social media platforms like Twitter and Facebook, to provide information, communicate warnings to the public, and influence behavior during the COVID-19 pandemic. Turkey's Ministry of Health's (MoH) adopted a highly successful social media communication strategy involving content analysis of the content shared via its official Twitter, Facebook, and Instagram accounts [36].

A combination of behavior change strategies and ML algorithms can therefore improve prediction of human behaviors and solve public health response to global crisis including but not limited to pandemics, climate change, and poverty. Policy makers can utilize ML and deep learning techniques to promote health and well-being as well as achieve the SDG goals. By harnessing the power of ML for explaining societal trends, public perceptions, beliefs and concerns toward crisis situations, an evidence base can be developed to guide the creation of appropriate policies. For instance, by detecting trends in sentiment over time, decision-makers can adapt their messaging to address evolving concerns and effectively engage with the public. The rule – based sentiment analysis techniques (VADER) employed in this study provided valuable insights into the general sentiment surrounding COVID-19 on Twitter [37], offering a quantitative understanding of the emotions and opinions expressed by Twitter users and aiding in gauging the overall perception of the pandemic within the online community. Moreover, sentiment analysis in the context of COVID-19 pandemic can also help to identify misinformation and disinformation, enabling authorities to combat the spread of false narratives and enhance public trust. Machine learning can be automated using deep learning of features to classify other datasets and thus open new automated paradigms for public health surveillance.

Limitations: This study has some important limitations, hence further analysis and validation are essential to fully understand the

practical implications of these results. First, variability of how individual users express their opinions makes text classification in machine learning models very challenging. Second, Twitter users may not be representative of the entire population since only a small percentage may be active on this platform, and they can be different from users on other social media platforms in other countries. This means that generalizing the online data is a risk. Finally, the granularity of data is limited by what can be captured by Apify's Twitter Scraper and the duration of the study.

Equations

(1)

Name	Formula	Range
PCC	$\text{PCC}=\frac{\sum(S_0-S_{om})(S_p-S_{pm})}{\sqrt{S_p-S_{pm})^2-(S_0-S_{0m})^2}}$	$(-\infty < \text{NSE} < 1)$
WI	$WI = 1-\frac{\sum\lvert S_{(p)}-S_{(o)}\rvert}{\sum(\lvert S_{(p)}-S'_{(o)}\rvert+\lvert S_{(o)}-S'_{(o)}\rvert)}$	$(-\infty < \text{R} < 1)$
MAE	$\frac{MAE=\sum_{i=1}^{N}\lvert S_{(p)}-S_{(o)}\rvert}{N}$	$(0 < \text{MAE} < \infty)$
MAPE	$MAPE=\frac{100}{N}\sum_{I=1}^{N}\left\lvert\frac{S_{(O)}-S_{(P)}}{S_{(o)}}\right\rvert$	$(0 < \text{MAPE} < 100)$
PBAIS	$PBIAS=\frac{\sum_{i=1}^{N}(S_{(o)}-S_{(p)})}{\sum_{i=1}^{N}S_{(p)}}$	$(-\infty < \text{PBAIS} < \infty)$

Where S_0, $S_{(o)}$ are the observed value, S_{om}, $S'_{(o)}$, as the observed mean value, $S_p, S_{(p)}$, is the predicted value, S_{pm}, as the predicted mean value.

Conclusion

NLP and ML techniques rapidly provide invaluable insights from real-time social media data to better understand public attitude towards COVID-19 as part of gauging perceptions and behaviors and can contribute immensely to public health surveillance. This can be immensely useful for developing targeted interventions, appropriate policies and communication strategies to promote public health. From the sentiment analysis, it is evident that a wide range of emotions has been expressed on Twitter/X with negative perceptions becoming prevalent and this trend is likely to continue in the future. In this analysis, SVM outperformed RF in determining the accuracy of sentiment polarity. NLP and ML approaches can be helpful to stakeholders involved in creating data-driven, responsive health communication programs to mitigate the impact of future pandemic.

Table 1. Description of Twitter dataset (TD)

Period	**Total tweets**	**#English tweets**	**%English tweets**	**Sentiment**		
				Positive	**Negative**	**Neutral**
Dec.,2020	12,776	9,592	75.08%	3213	2,099	4,280
Jan., 2021	15,658	12,084	77.17%	3485	2,570	6,029
Feb., 2021	14,913	10,635	71.31%	2,833	1,852	5,950
Mar., 2021	19,510	14,942	76.59%	4,514	3,222	7,206
Apr., 2021	21,134	16,238	76.83%	3,627	3,865	8,746
May, 2021	15,496	10,637	68.64%	2,780	2,186	5,671
Jun., 2021	14,855	9,952	66.99%	2,156	2,354	5,442
Jul., 2021	16,501	12,235	74.15%	1,811	4,308	6,116
Aug., 2021	17,768	13,676	76.97%	1,902	4,732	7,042
Sep., 2021	16,486	12,287	74.53%	1,922	4,400	5,965
Oct., 2021	16,547	12,384	74.84%	2,139	4,225	6,020
Nov., 2021	12,734	10,249	80.49%	1,621	3,573	5,055
TOTAL	194,378	144,911	74.55%	32,003	39,386	73,522

Table 2. ANOVA using VADER

Sentiment Expression		Sum of Squares	df	Mean Square	F statistic	Sig.
Positive Sentiment	Between Groups	6602316.750	1	6602316.750	29.406	.000
	Within Groups	2245264.167	10	224526.417		
	Total	8847580.917	11			
Negative Sentiment	Between Groups	5067400.333	1	5067400.333	7.599	.020
	Within Groups	6668851.333	10	666885.133		
	Total	11736251.667	11			
Neutral Sentiment	Between Groups	418880.333	1	418880.333	.304	.593
	Within Groups	13758247.333	10	1375824.733		
	Total	14177127.667	11			

Table 3. Comparison of Support Vector Machine and Random Forest models

	Calibration phase					**Verification Phase**				
	PCC	WI	MAPE	MAE	PBIAS	PCC	WI	MAPE	MAE	PBIAS
RF-PS*	0.469	0.455	0.738	0.004	0.477	0.367	0.318	1.399	0.006	1.365
SVR-PS	**0.859**	**0.785**	**0.373**	**0.002**	**0.052**	**0.969**	**0.825**	**0.275**	**0.001**	**-0.021**
RF-NS**	0.862	0.571	1.663	0.004	-0.255	0.980	0.310	0.909	0.007	0.452
SVR-NS	**0.909**	**0.786**	**0.926**	**0.003**	**-0.189**	**0.938**	**0.589**	**0.267**	**0.002**	**0.112**
RF-NRS***	0.956	0.734	0.359	0.002	-0.002	0.947	0.713	0.297	0.002	0.146
SVR-NRS	**0.972**	**0.800**	**0.208**	**0.002**	**-0.041**	**0.958**	**0.822**	**0.255**	**0.001**	**0.020**

*PS=positive sentiment; **NS=negative statement; ***NRS=neutral sentiment

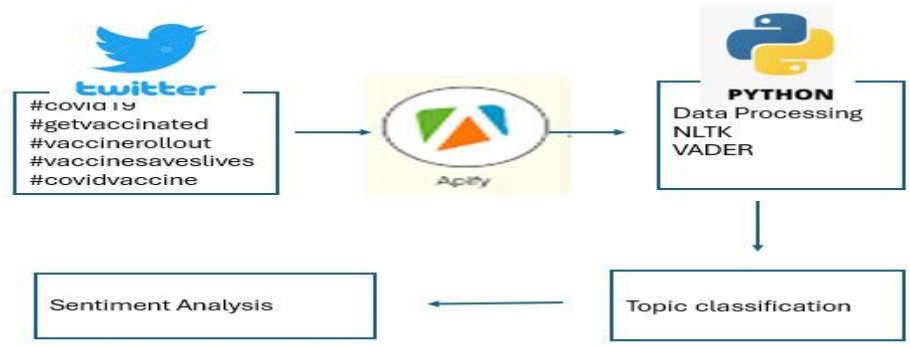

Figure 1. Framework for Sentiment Analysis

Conflict of Interest

The author declares that there is no conflict of interest.

Acknowledgements

The author wishes to acknowledge Shamsudeen and Hamid Hamza Maccido, for their dedicated efforts in aspects related to data collection, statistical and machine learning analysis. Many thanks to my supervisor, Prof M. Oche, my colleague, Asma Qureshi for her profound efforts to provide timely revisions to ensure that the manuscript is finalized.

References

[1]. Cho, H., Li, P., Ngien A., Tan, M. G., Chen, A., Nekmat, E., 2023, The Bright And Dark Sides of Social Media Use During COVID-19 Lockdown: Contrasting Social Media Effects Through Social Liability vs. Social Support. *Comput Human Behav*. 146:107795. Doi: 10.1016/j.chb.2023.107795. PMID: 37124630; PMCID: PMC10123536.

[2]. Sahni, H. Sharma, H. 2020, Role of Social Media During the COVID-19 Pandemic: Beneficial, Destructive, or Reconstructive. *International Journal of Academic Medicine* 6(2):p 70-75, Doi: 10.4103/IJAM.IJAM_50_20

[3]. Abbas, J., Wang, D., Su, Z., and Ziapour, A., 2021, The Role of Social Media in the Advent of COVID-19 Pandemic: Crisis Management, *Mental Health Challenges and Implications*. Risk Management and Healthcare Policy. 14, 1917–1932. Available online: https://doi.org/10.2147/rmhp.s284313

[4]. Jaffery, T. N., Shan, H., Gillani, R., Hassan, U., Sehar, B., 2022, COVID 19 Vaccination Related Misconceptions and Myths. *J Islamabad Med Dental Coll*. 11(2):120-126.

[5]. Boon-Itt, S., Skunkan, Y., 2020, Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. *JMIR Public Health Surveill*, 6(4):e21978. Doi:10.2196/21978

[6]. Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. 2020, Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, 31(7), 770-780. https://doi.org/10.1177/0956797620939054

[7]. Capraro, V., & Celadin, T. 2023, "I Think This News Is Accurate": Endorsing Accuracy Decreases the Sharing of Fake News and Increases the Sharing of Real News. *Personality and Social Psychology Bulletin*, 49(12), 1635-1645. https://doi.org/10.1177/01461672221117691

[8]. Mohammad, A., Kausar, A. S, Nasar, M., 2021, Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak, *International Journal of Advanced Computer Science and Applications*, 12(2).

[9]. Tsai MH, Wang Y., 2021, Analyzing Twitter Data to Evaluate People's Attitudes towards Public Health Policies and Events in the Era of COVID-19.

Int J Environ Res Public Health.18(12):6272. Doi: 10.3390/ijerph18126272. PMID: 34200576

[10]. Wang, S., Dhakal, S., and Upadhyay, B., 2024, Sentiment Analysis and Emotion Detection of COVID-19 Geo-Tagged Twitter Data," in 2024 9th International Conference on Big Data Analytics (ICBDA), Tokyo, Japan, pp. 180-185. Doi: 10.1109/ICBDA61153.2024.10607368

[11]. Christodoulakis, N., Abdelkader, W., Lokker, C., Cotterchio, M., Griffith, L. E., Vanderloo, L. M., Anderson, L. N., 2023, Public Health Surveillance of Behavioral Cancer Risk Factors During the COVID-19 Pandemic: Sentiment and Emotion Analysis of Twitter Data. *JMIR Form Res.* 2023 Nov 2;7:e46874. Doi: 10.2196/46874. PMID: 37917123; PMCID: PMC10624214.

[12]. Aldosery, A., Carruthers, R., Kay, K., Cave, C., Reynolds, P., Kostkova, P., Enhancing Public Health Response: A Framework For Topics And Sentiment Analysis of COVID-19 in the UK using Twitter and the Embedded Topic Model. *Front Public Health*. 2024 Feb 21;12:1105383. Doi: 10.3389/fpubh.2024.1105383. PMID: 38450124

[13]. Nielbo, K. L., Karsdorp, F., Wevers, M., Lassche, A., Baglini, R. B., Kestemont, M., & Tahmasebi, N., 2024, Quantitative Text Analysis. *Nature Reviews Methods Primers*, 4(1), 1-16. https://doi.org/10.1038/s43586-024-00302-w

[14]. Sangam, Savita & Shinde, Subhash, 2019, A Novel Feature Selection Method Based on Genetic Algorithm for Opinion Mining of Social Media Reviews: Third International Conference, ICICCT 2018, New Delhi, India, Revised Selected Papers. 10.1007/978-981-13-5992-7_15.

[15]. Farooq, F., Amin, M. N., Khan, K., Sadiq, M. R., Javed, M. F., Aslam, F., & Alyousef, R., 2020, A Comparative Study of Random Forest And Genetic Engineering Programming For The Prediction of Compressive Strength Of High Strength Concrete (HSC). *Applied Sciences* (Switzerland), 10(20), 1–18. https://doi.org/10.3390/app10207330

[16]. Vapnik, V. N. 1998, Statistical Learning Theory Wiley-InterScience*, New York*, ISBN: 978-0-471-03003-4.

[17]. Ghani, R., 2021, Integrating Sentiment Analysis and Machine Learning to Gauge Public Perceptions of COVID-19. *Journal of Medical Internet Research*, 23(8), e31220.

[18]. Sharma, S. S, Kaur, D., Chawla, T. K., Kapoor, V., 2021, Information Sharing through Twitter by Public Health care Institution during COVID-19 Pandemic: A Case Study of AIIMS, Raipur. *Indian J Community Health* [Internet]. 33(1):189-92. https://iapsmupuk.org/journal/index.php/IJCH/article/view/2035

[19]. Niu, Q., Liu, J., Kato, M., Shinohara, Y., Matsumura, N., Aoyama, T., Nagai-Tanima, M., 2022, Public Opinion and Sentiment Before and at the Beginning of COVID-19 Vaccinations in Japan: Twitter Analysis. *JMIR Infodemiology,* 2(1):e32335. Doi: 10.2196/32335

[20]. Kisa, S., Kisa, A., 2024, A Comprehensive Analysis of COVID-19 Misinformation, Public Health Impacts, and Communication Strategies: Scoping Review, *J Med Internet Res*; 26:e56931 Doi: 10.2196/56931

[21]. Skafle, I, Nordahl-Hansen, A, Quintana, D. S, Wynn, R., Gabarron, E., 2022, Misinformation About COVID-19 Vaccines on Social Media: Rapid Review. *J Med Internet Res*. 4;24(8):e37367. Doi: 10.2196/37367. PMID: 35816685

[22]. Scannell, B. J., Drum, C., & Hine, C., 2021, Persuasion Techniques Used In Anti-Vaccine Twitter posts during the COVID-19 Pandemic. *Health Communication*, 36(11), 1368-1376.

[23]. Lyu. J. C, Han, E. L, Luli, G. K, 2021, COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *J Med Internet Res*., Jun 29;23(6):e24435. Doi: 10.2196/24435. PMID: 34115608

[24]. Cheng, T., Han, B., Liu, Y., 2023, Exploring Public Sentiment And Vaccination Uptake of COVID-19 Vaccines In England: A Spatiotemporal And Sociodemographic Analysis of Twitter data. *Front Public Health*. 17;11:1193750. Doi: 10.3389/fpubh.2023.1193750. PMID: 37663835

[25]. Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R., & Bettinger, J. A., 2013, Vaccine Hesitancy: An overview. *Human Vaccines & Immunotherapeutics*, 9(8), 1763–1773. https://doi.org/10.4161/hv.24657

[26]. Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., Shah, Z., 2020, Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *J Med Internet Res*, 22(4):e19016. Doi: 10.2196/19016

[27]. González-Padilla, D. A., Tortolero-Blanco, L., 2020, Social Media Influence in the COVID-19 pandemic. *Int Braz J Urol.,* 46:120-4. 10.1590/S1677-5538.IBJU.2020.S121

[28]. Cuello-Garcia, C., Pérez-Gaxiola, G., van Amelsvoort, L., 2020, Social Media can have an Impact On How We Manage And Investigate the COVID-19 Pandemic. *J Clin Epidemiol*. 127:198-201. 10.1016/j.jclinepi.2020.06.028

[29]. Pennycook, G, Rand, D. G. 2019, Fighting Misinformation On Social Media using Crowdsourced Judgments Of News Source Quality. *Proc Natl Acad Sci* U S A, 116(7):2521-2526. Doi: 10.1073/pnas.1806781116. PMID: 30692252

[30]. World Health Organization: coronavirus disease (COVID-19) advice for the public., 2022, Accessed: August 18, 2024: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters.

[31]. Islam, M. S, Sarkar, T., Khan, S. H., et al. 2020, COVID-19-related Infodemic And Its Impact on Public Health: A Global Social Media Analysis. *Am J Trop Med Hyg*. 103:1621-9. 10.4269/ajtmh.20-0812

[32]. Romer, D., Jamieson, K. H., 2020, Conspiracy Theories As Barriers To Controlling The Spread of COVID-19 in the U.S. *Soc Sci Med.* 263:113356. 10.1016/j.socscimed.2020.113356

[33]. Naeem, S. B., Bhatti, R., Khan, A., 2021, An Exploration Of How Fake News Is Taking Over Social Media And Putting Public Health At Risk. *Health Info Libr J.*, 38:143-9. 10.1111/hir.12320

[34]. Joseph, A M., Fernandez, V., Kritzman, S., et al. 2022, COVID-19 Misinformation on Social Media: A Scoping Review. *Cureus* 14(4): e24601. Doi:10.7759/cureus.24601

[35]. Baker, S. A., Wade, M., & Walsh, M. J. 2020, The Challenges of Responding To Misinformation during A Pandemic: Content Moderation and The Limitations of The Concept of Harm. *Media International Australia*, 177(1), 103-107. https://doi.org/10.1177/1329878X20951301

[36]. Kılıç, N, Dikmen, E, Akşak, E, 2023, Managing Pandemic Communication Online: Turkish Ministry of Health's Digital Communication Strategies During COVID-19. *International Journal of Communication.* Vol. 17.

[37]. Hutto, C. J. & Gilbert, E. E., 2014, VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.