HSAM: Hybrid Sentiment Analysis Model for COVID-19 Contact Tracing Applications

Raghubir Singh [1] and Neeraj Kumar [2]

[1]University of Bath
[2]Affiliation not available

October 30, 2023

Abstract

To understand the public's perception of COVID-19 tracing applications, previous studies were primarily based on exploratory research, surveys or machine learning methods, which are semantically weak and time-consuming. To increase the reliability of this analytical methodology, hybrid-based Twitter sentiment analysis can be applied. In this paper, we propose a hybrid model for sentiment analysis by using Valence Aware Dictionary for Sentiment Reasoning (VADER) + Support Vector Machine (SVM). We demonstrate from the numerical analysis that a VADER and SVM-based hybrid model provides the best performance with 82.3% accuracy, 0.84 precision, 0.83 recall and 0.82 F1-score. The use of hybrid-based methods is shown to be effective in analysing the public's perception towards COVID-19 contact tracing applications using tweets collected from the UK, USA and India. Positive responses clearly outweighed negatives responses towards contact tracing, but this was contradicted by the low uptake of apps in all three nations. Our analysis, however, showed that neutral responses were 52% of the collected tweets; these tweets did not express positive or negative opinions, and subsequent tweets from the same users could not be verified, thus limiting the number of analyzed tweets available.

HSAM: Hybrid Sentiment Analysis Model for COVID-19 Contact Tracing Applications

Raghubir Singh[†] and Neeraj Kumar [‡]
[†]Department of Computer Science, University of Bath, United Kingdom.
rs3022@bath.ac.uk
[‡]Thapar University, Patiala, Pb, India
neeraj.kumar@thapar.edu

***Abstract*—To understand the public's perception of COVID-19 tracing applications, previous studies were primarily based on exploratory research, surveys or machine learning methods, which are semantically weak and time-consuming. To increase the reliability of this analytical methodology, hybrid-based Twitter sentiment analysis can be applied. In this paper, we propose a hybrid model for sentiment analysis by using Valence Aware Dictionary for Sentiment Reasoning (VADER) + Support Vector Machine (SVM). We demonstrate from the numerical analysis that a VADER and SVM-based hybrid model provides the best performance with 82.3% accuracy, 0.84 precision, 0.83 recall and 0.82 F1-score. The use of hybrid-based methods is shown to be effective in analysing the public's perception towards COVID-19 contact tracing applications using tweets collected from the UK, USA and India. Positive responses clearly outweighed negatives responses towards contact tracing, but this was contradicted by the low uptake of apps in all three nations. Our analysis, however, showed that neutral responses were 52% of the collected tweets; these tweets did not express positive or negative opinions, and subsequent tweets from the same users could not be verified, thus limiting the number of analyzed tweets available.**

Keywords*—*COVID-19, Global Health, Sentiment Analysis, Lexicon and hybrid-based models

NOMENCLATURE

F1-Score: A measurement balance of precision and recall.

FN: The number of positive instances wrongly classified as negative.

FP: The number of negative instances wrongly classified as positive.

TN: The number of negative instances accurately classified as negative

TP: The number of positive instances accurately classified as positive.

I. INTRODUCTION

THE Coronavirus Disease 2019 (COVID-19) has caused at least 6.7 million deaths as of January 2023. It has been declared a pandemic and a threat to global health [1]. To track and reduce the spread of COVID-19, many countries have developed local digital contact tracing applications [2]. The primary purpose of these applications is to inform users when they have been close to infected patients, thus encouraging close contacts to go into self-isolation to stop the spread of COVID-19 [3].

Research has shown that a pandemic could have been stopped if at least 60% of the population used a contact tracing application [4]. Despite the importance of contact tracing, the uptake rate worldwide is low. For example, as of end-December 2020, the uptake rate in the United Kingdom (UK), the United States (US) and India is 28% [5], 14% [6] and 13.5% [7] respectively. The uptake rate in each country is calculated by the number of application downloads mathematically divided by the population in said country. This is a cause for concern because the effectiveness of the application is tied to the uptake rate [5], [8]. Although most countries have ceased COVID-19 restrictions, understanding the public's sentiments towards contact tracing applications is still important to prepare the world for the next disease outbreak of a highly transmissible communicable disease.

A. Motivation

Several research studies have analysed opinions towards COVID-19 contact tracing applications. Several opinions point to privacy and security issues, application effectiveness [9] and social influence [10]. However, most of the research focused on exploratory research [11], [12], survey analysis [9] and manual analysis of users' reviews [13]. These methods are labour-intensive and require collecting primary data, which is time-consuming. To address these limitations; artificial intelligence methods can be applied to analyse the public's perception in a shorter time and with much less human labour.

Online users are generating more opinionated textual data with the ease of accessibility to social media [14]. In recent years, sentiment analysis in microblogs has become an emerging field as it yielded valuable user insights [15]. Twitter is a significant source for sentiment data collection as it is one of

Corresponding author: Raghubir Singh, Department of Computer Science, University of Bath, United Kingdom. Email: rs3022@bath.ac.uk

the most prominent social media platforms, and it allows users to express their opinions freely [16], [17]. Sentiment analysis is a study of opinions and feelings towards an entity. An entity can be anything, such as an organisation, an individual or a product. Sentiment analysis is a valid tool for studying the public's perceptions as it has produced high performance [18]. The outcome of sentiment analysis is the polarity classification of an entity, and it has been applied across many domains, including product reviews [19] healthcare [20] and fake news detection [21].

Figure 1 shows a timeline of serious disease-related mortality from the Third Plague Outbreak in nineteenth-century China to the Covid-19 pandemic. Unlike the Spanish Flu epidemic, COVID-19 occurred in the modern digital age where mass communication is both rapid and very widespread. Unlike HV/AIDs, COVID-19 deaths were mostly in just three years, 2020-2022. The rapidly increasing death toll, therefore, was a major challenge to how contemporary societies can (and did) react to a global pandemic of a contagious disease.

A recent study surveyed 43 research articles and summarised the approaches to Twitter sentiment analysis. Sentiment analysis is more challenging than the traditional classification method, especially when analysing Twitter data, as each tweet has a limit of 280 characters. The approaches include machine learning methods, lexicon-based methods and hybrid-based methods. The study concluded that most of the surveyed research applied machine learning algorithms [22]. In other studies, machine learning methods are found to be semantically weak [23] and time-consuming due to the manual labelling of training datasets for each application domain [24]. These limitations are mitigated by combinations with lexicon-based methods to create hybrid-based methods [25].

B. Our Contributions

This paper presents a hybrid sentiment analysis model (HSAM) for COVID-19 tracing applications. The main contributions of this paper are threefold:

- Investigation of the suitability of hybrid-based methods in studying perception towards contact tracing applications.
- Different approaches to determining whether the public associates contact tracing applications positively or negatively and to understand the factors contributing to these sentiments.
- Developing detailed statistical evaluation metrics to compare hybrid-based and non-hybrid-based methods

C. Structure of the Paper

The remainder of the paper is organised as follows. Section II discusses current work and different approaches in sentiment analysis. Section III discusses the choice of research methods and how the research questions can be studied. Section IV presents our analytical methodology. Section VI presents our analysis of the tweet data. Section VI presents a detailed analysis of sentiment word association in our research outputs. Lastly, Section VII concludes our results and provides a brief outlook for future work in sentiment analysis.

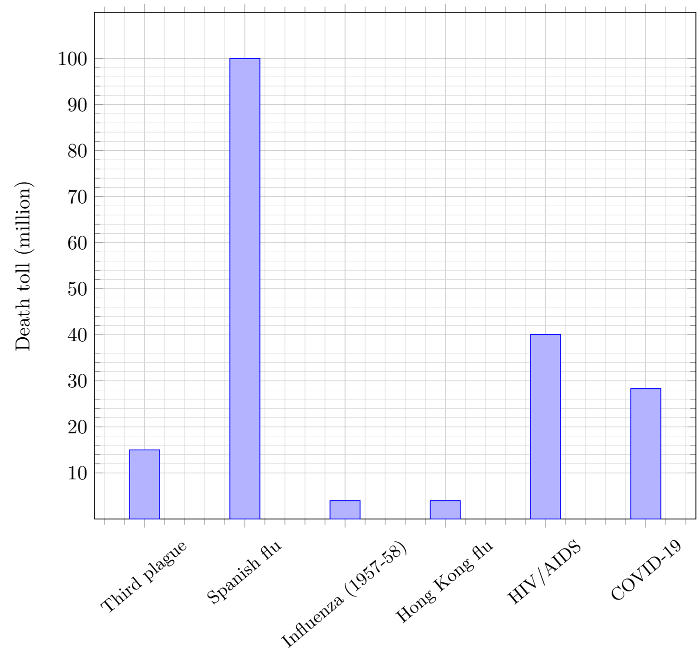

Figure 1. Examples of pandemics include the Spanish flu in 1918-1919 and the ongoing COVID-19 pandemic [26]

II. Related Work

This section provides an overview of current sentiment analysis works and evaluation results using machine learning algorithm approaches, lexicon-based approaches and hybrid-based approaches. As there is limited research related to sentiment analysis in COVID-19 contact tracing applications, this section discusses necessary sentiment analysis research that is both related and unrelated to contact tracing applications. Table I shows the related work from previous research

Sentiment analysis studies opinions and feelings towards an entity, which can be an organisation, an individual or a product. Sentiments expressed can be classified into three polarities – positive, neutral or negative. Figure 2 shows the different types of machine learning and lexicon-based approaches.

A. Machine learning algorithms approach

Machine learning in sentiment analysis uses supervised machine learning techniques. It uses training and testing datasets for prediction. The training set is transformed into vectors, and the class labels are annotated. Previous research was frequently based on supervised machine learning approaches, especially Support Vector Machine (SVM), Naïve Bayes [27], and enhanced algorithms developed by the researchers.

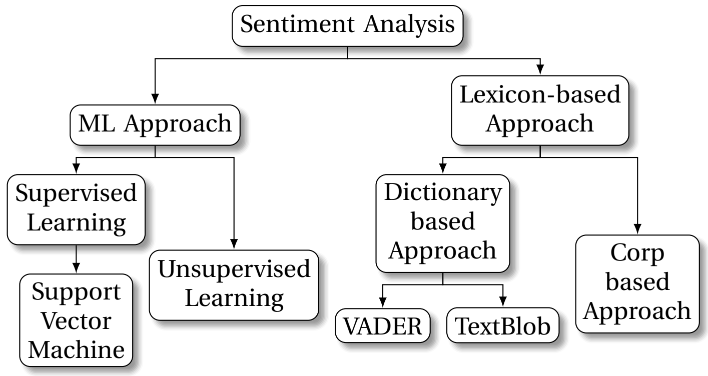

Figure 2. Sentiment Analysis Classification Techniques

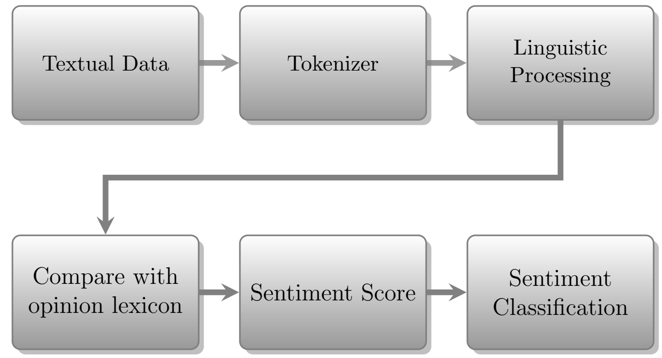

Figure 3. Lexicon-based sentiment analysis

One of the most cited machine learning algorithm papers was written by Pang et al [28]. This experimental study examined the effectiveness of feature engineering (unigrams, bigrams, part-of-speech, word position, term frequency, and term presence) on machine learning techniques. The research also studied human versus machine learning classifications and attempted to classify sentiments of movie reviews using SVM, Naïve Bayes and Maximum Entropy. First, humans were asked to classify the reviews and this achieved 58% and 64% accuracy with two proposed word lists. The same dataset was then pre-processed and used with machine learning algorithms when the accuracy reached 82.9% (SVM), 78.7% (Naïve Bayes) and 77.7% (Maximum Entropy).

[29] studied ensemble machine-learning algorithms. The focus of this research was on classifying emotions in tweets. Tweets were pre-processed and converted into vectors through term frequency–inverse document frequency (TF-IDF), which was based on frequency ratio. This paper built on previous works by taking emoticons into consideration and replacing them with representative words. For example, smiley face ":)" was replaced with the word "positive". SVM and Adaboosted Decision Tree (ADT) were applied separately, and the outputs were trained with a Decision Tree hybrid-based model. The accuracy of SVM, ADT and a hybrid approach was 82%, 67% and 86%, respectively. The result is comparable to an earlier 82.9% accuracy for SVM [29]. Another important conclusion was that performance can be improved by stacking machine learning algorithms instead of just relying on one algorithm.

B. Lexicon-based approach

Lexicon-based approaches make use of a pre-built list of terms to calculate polarity, and no training data is required. The dictionaries are primarily created in English and a translation function is required if the dataset is in another language. When the same models are applied across different domains, lexicon-based approaches tend to perform better than machine-learning methods [23]. Figure 3 shows the framework in lexicon-based sentiment analysis.

[30] reported a lexicon-based comparison study on VADER, SentiStrength, SentiWordNet, AFINN-111 and Liu-Hu Lexicon. Two publicly available Twitter datasets of a different domain were used for comparison: Stanford Twitter Dataset Test Set (tweets on people, companies and products) and Sandars Twitter Dataset Test Set (tweets on technology companies). The research aimed to apply the same models to datasets with different domains. Analysis of the Stanford dataset showed the best model was VADER with 72% accuracy, followed by SentiStrength with 67% accuracy, AFINN-111 and Liu-Hu Lexicon with 65% accuracy and SentiWordNet with 53% accuracy [30]. VADER performed best in positive classification, whereas AFINN-111 performed best in negative classification. For Sandars dataset, VADER achieved 65% accuracy and AFINN-111 achieved 62% accuracy. Liu-Hu performed the worst at 58% accuracy [30]. This shows that VADER performs well in both datasets of different domains.

[31] was a lexicon-based comparative analysis using Stanford Core NLP Sentiment Analyser, Textblob and VADER. To study domain-specific and general-purpose topics, the research used a healthcare dataset and general-purpose datasets. It found that all three methods performed poorly on the domain-specific dataset: Stanford NLP Sentiment Analyser achieved 31.2% F1- score, VADER achieved 48.7% F1-score and TextBlob achieved 21.6% F1-score. For the general-purpose dataset, the F1-score Values for Stanford NLP Sentiment Analyser, VADER and TextBlob were 53.3%, 71.6% and 57%, respectively.

The study [32] aimed to analyse public sentiments on COVID-19 contact tracing applications in the Republic of Ireland. It collected Twitter data from 1st January 2020 to 31st December 2020, and filtered the tweets using the keywords "covid19", "covid19ireland" and "covidprivacy". The study collected 1,420 tweets related to the COVID-19 contact tracing application and compared the performance of TextBlob and Senti-Foclóir on a test dataset of 216 tweets.

C. Hybrid-based approaches

Hybrid-based approaches are built from a combination of machine learning and lexicon-based approaches. The lexicon-based approach is first applied to textual data to classify sentiment polarity, and the classified data is used as the training dataset for machine learning algorithms. The purpose of hybrid-based is to mitigate the limitations and tap into the strengths of both methods, as discussed earlier.

Only one research work [33] has focused on hybrid-based approaches to analyse COVID-19 contact tracing applications. The research extracted relevant datasets from Twitter and Facebook geotagged in the UK from the period 1st March 2020 to 31st October 2020. It proposed three models – 1) VADER, 2) TextBlob and 3) an average-weighting ensemble of VADER and TextBlob combined with deep learning Bidirectional Encoder Representations from Transformers (BERT). The results showed lexicon-based models were able to classify positive sentiments with a better accuracy whereas BERT had higher accuracy for negative and neutral sentiments. Accuracy for VADER was 62.5%, TextBlob was 59.7% and BERT was 61.0%. The ensembled hybrid approach had a significant accuracy increase at 71.6%. The classifier found 76% positive and 12% negative tweets [33].

This study will add to existing research by studying and comparing the suitability of using hybrid-based sentiment analysis on COVID-19 contact tracing applications. The research will cover a broader range of sentiments by studying more extended periods and more keywords.

III. Methodology

This research takes on a positivism and quantitative methodology, using theoretical and experimental methods. Theory helps to analyse and evaluate existing sentiment analysis approaches, and experimental analysis is conducted to answer the research questions. The programming language Python 3 in Jupyter Notebook [1] is used.

To study and compare the suitability of hybrid-based approaches, the study will first compare the performance of hybrid-based sentiment analysis methods against non-hybrid-based sentiment analysis methods. The results of the models will lend reliability in understanding the sentiments. Twitter data is collected and analysed using lexicon-based approaches VADER and TextBlob, and a hybrid of both with the SVM algorithm. Figure 4 outlines the proposed framework, which will be explained in the following sections.

A. Data Collection

A Twitter Search Application Programming Interface (API) was used to collect Twitter data. Relevant tweets were collected between the period 1st March 2020 – 31st December 2021 to explore the public's perception from the start of the pandemic to towards end of the pandemic (when everyone was used to living with COVID-19). Tweets in this time period covered a wide range of sentiments. Tweets with geotags from the UK, the US and India were collected as these countries provided a sufficient number of Tweets in English: 5,505 tweets were collected, excluding duplicates and retweets [34]. After removing

[1] Jupyter: https://jupyter.org/

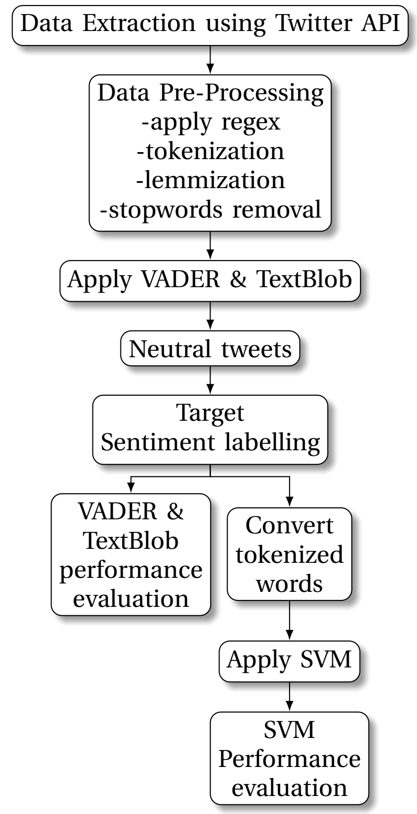

Figure 4. Proposed research methodology framework

the neutral tweets, there were 2,645 tweets, as shown in Table II. Tweets were randomised, and 200 tweets from each country dataset were selected for manual labelling for training and testing datasets, forming a total of 600 randomised tweets. All positive and negative labels were mapped to 1 (positive) and -1 (negative). An analysis was done on a country level instead of a global level, as comparing sentiments across different continents was important; this also prevented any country with a significantly high portion of particular sentiments from skewing the overall results.

Keywords were used with Python logical operation to ensure only relevant tweets were collected. Specific keywords were associated with local contact tracing applications. To name a few, some of the popular applications created are NHS COVID-19 from the UK, COVIDWISE from the US and Aarogya Setu from India. Important keywords used for local contact tracing include: "covidapp", "NHS" and "app", "covid" use the full list along to "notification".

To build the model, four classifiers were applied: Lexicon-based: VADER [2], Lexicon-based: TextBlob2 [3], Hybrid-based: VADER and SVM3 [4], and Hybrid-based: TextBlob and SVM.

[2] VaderSentiment: https://github.com/cjhutto/vaderSentiment
[3] Hybrid-based: VADER and SVM3: https://textblob.readthedocs.io/en/dev/
[4] SVM: https://scikit-learn.org/stable/modules/svm.html

Table I. Summary of performance results from previous research works from literature

Work	Domain	Algorithms	Accuracy	F1-Score
[28]	Movie Reviews	SVM Naïve Bayes Maximum Entropy	82.90% 78.70% 77.70%	-
[29]	Wide range of tweets	SVM ADT Ensemble Decision Tree	82% 67% 84%	82% 67% 84%
[30]	Standford dataset: Tweets about people products and companies	**Standford Twitter Dataset:**		
		VADER	72%	-
		SentiStrength	67%	-
		SentiWordNet	53%	-
		AFINN-111	65%	-
		Liu-Hu Lexicon	65%	-
	Standford dataset: Tweets about techs Companies	VADER	65%	-
		SentiStrength	58%	-
		SentiWordNet	59%	-
		AFINN-111	62%	-
		Liu-Hu Lexicon	58%	-
[31]	Social media data on general topics and healthcare	**General-purpose reviews:**		
		Standfor Core NLP	-	53.30%
		VADER	-	71.60%
		TextBlob	-	57%
		Healthcare reviews		
		Stanford Core NLP	-	31.20%
		VADER	-	48.70%
		TextBlob	-	21.60%
[32]	Tweets on COVID-19 contact tracing application	TextBlob Senti-Focóir	64.81% 65.74%	-
[33]	Twitter and Facebook data on covid-19 contact tracing applications	VADER TextBlob Bert Ensemble hybrid approach	62.50% 59.70% 61.10% 71.60%	-

Country	Tweets Collected	Tweets left (after removing neutral sentiments)
UK	1921	989
USA	2064	1076
India	1520	520
	5505	2645

Table II. Twitter data sets for analysis

B. Data pre-processing

Data pre-processing methods were iterated many times by studying individual tweets and the polarity results. The process differs slightly between VADER and TextBlob as they are built with a different pre-defined dictionary.

Regular expression4, commonly known as Regex, is first applied to clean the text. Links, numbers, additional whitespaces, and Twitter-specific syntax such as “” and its usernames are removed. All special characters are removed for TextBlob [34]. For VADER, special characters such as “!” and “?” that can affect the compound score are retained. For example, “awesome” and “awesome!!” have a compound score of 0.62 and 0.69, respectively. The more positive the sentiment is, the higher the compound score is. In terms of hashtags, only the “#” symbol is removed, and the accompanying word, which typically contains important sentiment, is retained [35]. Sometimes hashtags may consist of a combination of words stringed together, and sentiment analysers can ignore them and classify them as 100 neutral. All casewords for TextBlob are changed to lowercase as casewords do not affect the polarity score [36]. Tokenization is then applied to prepare the corpus for subsequent preprocessing.

IV. Numerical Results

This section compares and discusses an analysis of proposed models, presents findings of the public’s perception of COVID-19 contact tracing applications, and provides a brief discussion of the limitations

A. Evaluation Metrics

To evaluate the reliability of the proposed models, evaluation metrics used were accuracy, precision, recall and F1-score. These metrics were calculated based on true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) [24], [37]–[39]

Accuracy measures the percentage of correctly predicted classes relative to the entire dataset. The equation is:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

Precision measures the percentage of true positives relative to the total number of predicted positives. A score of 1.0 indicates every positive prediction was correct. The equation is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

A recall measures the percentage of true positives relative to the total number of actual positives. A score of 1.0 indicates every negative prediction is correct. The equation is:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

F1-score is a measurement balance of precision and recall. The higher the score, the better performance of the model is. For classification problems where both precision and recall are important, F1-score should be maximised. Since this research study aims to analyse all sentiments, F1-score is more indicative of the model performance than precision or recall. The formula is:

$$\text{F1-score} - \text{score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{4}$$

B. Performance Evaluation

Accuracy for VADER across all countries was improved by 1 – 2% when case words were not changed to lowercase. Performance for precision, recall and F1-score improved by 0.01 – 0.04. This contrasts with past VADER-based research, which changed all words to lowercase. VADER can calculate the intensity of words through case words. For example, “the app is brilliant” was classified as 55.9% positive, whereas “the app was BRILLIANT” was classified as 60.2% positive. On the other hand, case words do not matter for TextBlob.

When stemming was used, accuracy for VADER and TextBlob across all countries dropped by 15 – 20%. Stemming is a method that removes derivational suffixes and inflections. For example, ‘risky’ was stemmed from ‘risk’, and ‘worrying’ was stemmed to ‘worry’. Stemming causes spelling mistakes and wrong interpretations. This was supported by [40], a lemmatization and stemming comparison research that showed lemmatization yielded better performance than stemming. The final models in this study used lemmatization, and accuracy increased by 0.3 – 0.5%.

Comparison between feature vectorization Count-Vectorizer and TF-IDF was evaluated. The count vectorizer was based on the occurrences of words, whereas TF-IDF assigned more weights to unique words [41]. Table III-IV show the performance comparison. The results revealed TF-IDF yielded higher performance by 1.5 – 4.5% in five experiments. The only exception was the TextBlob and SVM models for India, where Count-Vectorizer yielded higher performance by 2.6%. As the length of a tweet contains only up to 280 characters, many users shortened their tweets as much as possible, thus reducing word occurrences. This provides a challenge to see any significant differences. This was supported by [39], a machine learning comparative analysis of bag-of-words and TF-IDF on Twitter data, which showed only a 1% difference in accuracy.

C. Models Evaluation

Table V-VI show the performance of the four classifiers. Regarding lexicon-based approaches, VADER and TextBlob had reasonable accuracy in classifying Twitter data. VADER achieved 69.8% accuracy in the UK tweets, 79.1% in US tweets and 73.9% in India tweets. TextBlob achieved 66% accuracy in the UK tweets, 74.9% in the US tweets and 70.7% in India tweets. The result was comparable to other Twitter lexicon-based research, which had 72% accuracy using VADER [30] and 64.81% accuracy using TextBlob [32]. VADER and TextBlob generally performed well as they not only depend on a lexical dictionary, but they were also sensitive to degree modifiers. The more positive words would receive a higher positive score, and the more negative words would receive a more negative score. VADER and TextBlob take into account negation words and emoticons. For example, both analysers classify the smiley face ":)" as positive.

Looking at the evaluation metrics, VADER had a higher precision value, which means VADER can predict the "negative" class more accurately than TextBlob. VADER results showed the "negative" class had precision scores of 0.92 for the UK tweets, 0.83 for the US tweets and 0.82 for India tweets. TextBlob results show lower precision scores at 0.90 for UK tweets, 0.70 for US tweets and 0.77 for India tweets. VADER analyses conjunctions and word order better than TextBlob. Using conjunctions such as "but" indicates a shift in sentiment, and VADER places a higher emphasis on the second part of the sentence [42]. This was consistent with previous research [43]. For example, "The app is great, but it is bad for data privacy" was classified as negative by VADER and positive by TextBlob.

In all four models, a low recall in the "negative" class label was found in the results for all three countries. There was an imbalanced class due to fewer negative sentiments compared to positive sentiments. It also showed there is a high number of false negatives of the "negative" class. In other words, many tweets were wrongly classified as "positive" when they should be "negative". This primarily arises from tweets with sarcasm that use positive words to express negativity. For example, "I'm interested in how the app works. Could be a scammer's delight" and "this app needs a Bluetooth function. Good luck teaching that to my mother-in-law". However, the high precision in the 'negative' class showed when the tweets were classified as negative; they were mostly accurate. In other cases, the challenge in sentiment analysis across when one word can express different meanings [44]. For example, the word "like" can express positivity; however, it can also be used as a filler word. Consider the sentence, "she told me to download the app. I was, like, erm... no..." This is wrongly classified as positive due to the word "like".

When the lexicon-based methods were combined with SVM, the results revealed the best model was VADER + SVM analysing UK tweets with 82.3% accuracy and weighted 0.84 precision, 0.83 recall and 0.82 F1-score. This was followed by TextBlob + SVM analysing UK tweets with 81.8% accuracy and weighted 0.82 precision, 0.82 recall and 0.81 F1-score. The performance results were reasonable and comparable with past hybrid-based sentiment analysis, which has 79.78% accuracy [45]. The hybrid-based models also tended to outperform its lexicon-based counterparts. When hybrid models were used, there is 12.5% – 15.8% increase in accuracy for UK tweets, 1% – 4.3% increase in accuracy for US tweets, and 7.1% – 8.6% increase in accuracy for India tweets.

Overall, the results revealed hybrid models performed better than lexicon-based counterparts as they benefit from the advantages of the combined methods. Hybrid-based derives high accuracy from machine learning methods and benefits from the stability of lexicon-based dictionaries. The stability comes from how polarity scores were calculated based on a pre-defined dictionary. Words in the pre-defined dictionary could be added and changed to reflect current trends.

D. Sentiment polarity

Fig 5 shows the distribution of positive and negative tweets analysed. There is a similar distribution of positive and negative tweets between the countries. There was a similar distribution of positive and negative comments between the three countries. Overall, the percentage of positive comments (67.4%) was lower than the UK-only data sources (76% positive) in [33].This may have reflected the longer sampling time use din this study than in [33]

V. Sentiment Word Association

Tables VII-XII show top-ranked unigrams, bigrams and trigrams with positive and negative sentiments. Interesting word associations that add value to the research were discussed in this section.

A. Privacy Concern

"Privacy" was the top-ranked word in positive and negative sentiments in all three countries. Privacy issues have been raised worldwide and privacy is considered one of the critical factors that dissuade people from using the application, and this is consistent with previous research [46]. In the UK, "privacy"

	VADER + SVM (Count-Vectorizer)				VADER + SVM (TF-IDF)			
Country	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-score	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-score
UK	78.3%	0.78	0.78	0.78	82.8%	0.82	0.81	0.80
USA	76.9%	0.76	0.77	0.76	80.1%	0.83	0.80	0.76
India	81.0%	0.80	0.81	0.79	81.0%	0.83	0.81	0.77

Table III. Performance evaluation of VADER + SVM using TF-IDF and Count-Vectorizer

	TextBlob+ SVM (Count-Vectorizer)				TextBlob + SVM (TF-IDF)			
Country	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-score	Accuracy	Weighted Precision	Weighted Recall	Weighted F1-score
UK	80.3%	0.80	0.80	0.79	81.8%	0.82	0.81	0.81
USA	76.9%	0.79	0.77	0.74	79.2%	0.79	0.79	0.76
India	79.3%	0.77	0.79	0.77	76.7%	0.77	0.77	0.73

Table IV. Performance evaluation of TextBlob + SVM using TF-IDF and Count-Vectorizer

Country	**Accuracy**	**Precision**			**Recall**			**F1-score**		
		Negative	Positive	Weighted	Negative	Positive	Weighted	Negative	Positive	Weighted
UK	69.8%	0.92	0.57	0.79	0.55	0.93	0.70	0.69	0.70	0.70
USA	79.1%	0.83	0.78	0.80	0.51	0.94	0.79	0.63	0.85	0.78
India	73.9%	0.82	0.70	0.76	0.55	0.90	0.74	0.66	0.79	0.73

Table V. Performance evaluation of VADER

Country	**Accuracy**	**Precision**			**Recall**			**F1-score**		
		Negative	Positive	Weighted	Negative	Positive	Weighted	Negative	Positive	Weighted
UK	66.0%	0.90	0.54	0.76	0.50	0.92	0.66	0.64	0.68	0.66
USA	74.9%	0.70	0.77	0.74	0.51	0.88	0.75	0.59	0.82	0.74
India	70.7%	0.77	0.68	0.72	0.51	0.87	0.71	0.61	0.76	0.70

Table VI. Performance evaluation of TextBlob

accounted for 22% of the positive sentiments and 20% of the negative sentiments. In the US, “privacy” accounted for 38% of the positive sentiments and 42% of the negative sentiments. In India, “privacy” accounted for 18% of the positive sentiments and 20% of the negative sentiments.

Among the negative sentiments, there was distrust towards the government and the fear of them mishandling data. Unigram “government” accounted for 14% of UK and 17% of India tweets. “State” accounts for 13%” of US tweets. In the UK, Twitter users did not trust the 30 government-affiliated NHSX

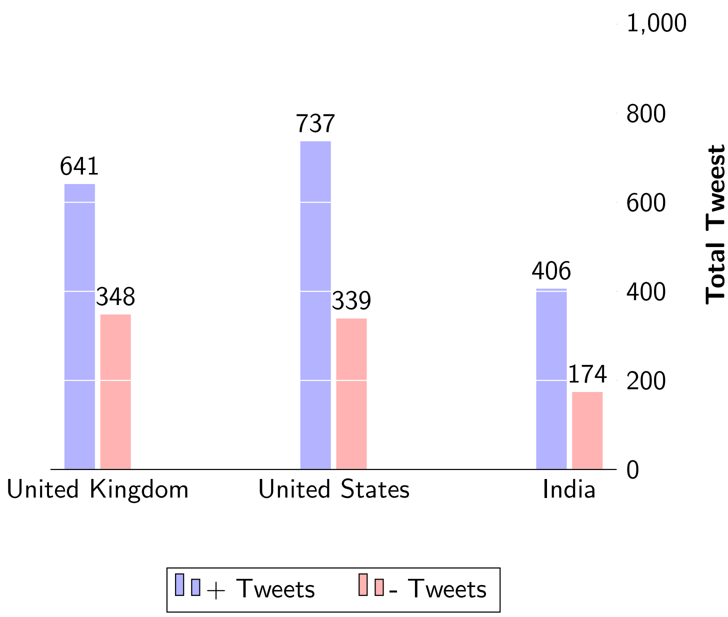

Figure 5. Distribution of positive and negative sentiments

applications due to its centralised system. They preferred the decentralised system developed by Apple and Google.

On the other hand, Twitter users in the US did not even trust Apple and Google applications. "Google" and "Apple" accounted for 4% and 3% of negative sentiments. An American survey by [47] revealed a decreasing level of trust towards technology companies over the last decade due to several consumer data protection violations. Among the three countries, US Twitter users expressed the most concern about privacy. Many ngrams were associated with privacy, such as "privacy concern", "privacy law", "privacy security", "data trust", and "HIPAA". HIPAA is a Health Insurance Portability and Accountability Act law in the US that protects patients' health data.

In India, there was a hacking incident in May 2020, and this caused further distrust in the application's security. Unigrams "hacker" and "fake" account for 3% of the negative sentiments. A previous study on India's perception towards contact tracing applications revealed that negative user sentiment increased 11% after the cyber security incident [48].

Among the positive sentiments, some users mentioned there was no breach of privacy, while others think privacy was a trade-off for health protection. Some users tweeted sentiments like "Sceptical about government's action on privacy. But we need all tools to fight this virus" and "Bunch of my friends were torn between privacy and community". In the UK, "Scotland" accounts for 8% of the positive unigrams and "protect Scotland" ranks second in positive bigrams sentiments. Upon studying the tweets, it was found Twitter users believe the Scottish application was not a threat to privacy. For example, users tweeted "stores zero personal info" and were "solid and privacy-centred". This was supported by an exploratory analysis of COVID-19 contact tracing applications, revealing that the Protect Scotland application was user-friendly and effective and rated the highest among other European contact tracing applications [12].

B. Effectiveness of application

All Twitter users in the three countries mentioned the application helped to stop the spread of COVID-19, and users encouraged others to download the application. In the UK, the word "download" ranked fourth in unigrams and accounted for 11% of the positive tweets and "help" accounts for 9%. In the US, "help" ranked second in unigrams and accounts for 11% of the positive tweets and "protect" accounts for 5%. Users also mentioned the application would only be "effective" when more people use it. In India, Twitter users expressed the application as a "solution" to the pandemic. As the application was needed to register vaccine appointments in India, users also encouraged others to download the application for this purpose.

C. Reliability of application

In the UK and India, the application was found unreliable by some users. They mentioned getting a false notification when they were not in close contact, while others did not receive a notification when someone close to them was infected with COVID-19. "Failed" accounts for 6% of the negative sentiments in the UK. In India, "confusion" accounts for 3% negative sentiments.

D. Economic disruption

In India, users were unhappy with the mandatory usage of contact tracing as it caused economic disruption to the poor. "Mandatory" and "poor" comprised 8% of the negative sentiments. As the Indian government made contact tracing applications mandatory for all workers and people in the COVID-19 "containment" zone [49], this caused some users to think the lower socioeconomic groups would suffer financially if they were requested to self-isolate.

E. Limitations

A limiting factor of this research is the lack of context-based analysis [50]. Contextual content on Twitter consists of the preceding tweet history that a user is replying to, an external link, or an image of an internet meme. Consider the following conversation. User1: "Total cost of NHS contact-tracing app set to top £35 million. This is unacceptably high!" User2: "I agree. These figures are phenomenal!"

Without looking into the context, User2's tweet is classified as positive. However, the actual sentiment should be negative as the user agrees and replies to a negative tweet. Improvements can be made to future research by retrieving contextual data using Twitter IDs or a crawler to retrieve the title header of the external link.

Bigrams & **Trigrams**	**Frequency**	n (%)	**Unigrams**	**Frequency**	n (%)
apple, google	33	5%	privacy	141	22%
protect Scotland	23	4%	data	73	11%
England, wales	11	2%	work	68%	11%
test trace	11	2%	download	68	11%
privacy concern	10	2%	new	58	9%
help stop spread	9	1%	help	55	9%
technical spec	9	1%	Scotland	53	8%
download today	8	1%	good	52	8%
privacy notice	7	1%	google	52	8%
self-isolate	7	1%	government	47	7%

Table VII. Top ranked unigrams, bigrams and trigrams positive sentiments in the UK

Bigrams & **Trigrams**	**Frequency**	n (%)	**Unigrams**	**Frequency**	n (%)
apple, google	15	4%	privacy	71	20%
world-beating	9	3%	government	48	14%
care home	9	3%	nhsx	36	10%
privacy concern	7	2%	data	29	8%
cyber security	7	2%	new	25	7%
test trace	7	2%	test	23	7%
went wrong	7	2%	health	22	6%
data privacy	7	2%	work	22	6%

Table VIII. Top ranked unigrams, bigrams and trigrams negative sentiments in the UK

Bigrams & **Trigrams**	**Frequency**	n (%)	**Unigrams**	**Frequency**	n (%)
exposure notification	52	7%	privacy	282	38%
Apple, google	25	3%	help	80	11%
public health	12	2%	data	63	9%
save life	11	1%	health	51	7%
social distancing	9	1%	protect	36	5%
slow spread	9	1%	technology	40	5%
privacy law	9	1%	system	26	4%
united states	9	1%	information	25	3%
respect privacy	8	1%	apple	41	6%

Table IX. Top ranked unigrams, bigrams and trigrams positive sentiments in the USA

Bigrams & **Trigrams**	**Frequency**	n (%)	**Unigrams**	**Frequency**	n (%)
privacy concern	36	11%	privacy	141	42%
privacy law	8	2%	state	44	13%
apple google	7	2%	data	34	10%
go wrong	6	2%	health	19	6%
privacy security	5	1%	game	16	5%
public health	5	1%	mygovcuomo	14	4%
white house	5	1%	google	15	4%
data trust	5	1%	apple	11	3%
difficult time	4	1%	HIPAA	10	3%

Table X. Top ranked unigrams, bigrams and trigrams negative sentiments in the USA

Bigrams & **Trigrams**	**Frequency**	n (%)	**Unigrams**	**Frequency**	n (%)
download aarogya	17	4%	vaccine	75	18%
stay safe	16	4%	government	71	17%
vaccine centre	16	4%	safe	61	15%
vaccine centre ready	11	3%	download	55	14%
android ios	10	2%	privacy	46	11%
social distancing	9	2%	pmoindia	30	7%
show solution	8	2%	updated	21	5%
built realtime	8	2%	security	17	4%
use aarogya setu	7	2%	solution	10	2%

Table XI. Top ranked unigrams, bigrams and trigrams negative sentiments in India

Bigrams & **Trigrams**	**Frequency**	n (%)	**Unigrams**	**Frequency**	n (%)
faced poor	8	5%	privacy	35	20%
security act nfsa	8	5%	vaccine	33	19%
disruption caused	8	5%	government	30	17%
national good security	8	5%	security	19	11%
setu mandatory	4	2%	pmoindia	14	8%
spy could use	4	2%	fake	6	3%
govt fails	4	2%	confusion	5	3%

Table XII. Top ranked unigrams, bigrams and trigrams negative sentiments in India

Data was collected using the Twitter search query "place_country" parameter, which limited this study to Twitter users who geotagged on the country level. More data can be collected from users who geotagged on city-level, such as London or New York. However, when attempting to use the parameter "place" for this case, the codes threw several errors. Due to time limitations, a decision was made to move on to the next stage. Future work may benefit from having a larger sample size. As data was collected from Twitter, the research analyses perceptions from people who have access to the internet. People without access to the internet or smartphones may have also contributed to the low application uptake rate.

VI. Conclusions and Future Work

In this paper, sentiment analysis was applied to the COVID-19 contact tracing application Twitter textual data and results were compared to study the suitability of using hybrid-based methods. The results showed that hybrid-based methods perform better than their lexicon-based counterparts, with a mean increase of 8% accuracy in the studied data. VADER + SVM outperformed TextBlob + SVM. The best model is VADER + SVM with 82.3% accuracy, 0.84 precision, 0.83 recall and 0.82 F1-score. Fine-tuning data pre-processing steps is required to optimise model performance. In the analysis of lexicon-based methods, VADER tends to outperform TextBlob, with a mean increase of 3.7% accuracy. The best model is VADER with 79.1% accuracy, 0.80 precision, 0.79 recall and 0.78 F1-score. Higher performance in VADER results in higher performance in VADER + SVM model. Previous analyses on COVID-19 contact tracing applications were largely based on exploratory and survey methods which is time-consuming and labour-intensive. The results showed that hybrid-based sentiment analysis is an efficient and reliable method for future work in this area.

The analysis showed there were more positive than negative sentiments. The positive sentiments for UK, US and India users were 64.8%, 68.5% and 70%, respectively. Several sentiments were found. Privacy is the top-ranked word that was associated with positive and negative tweets. An interesting finding is that some users are willing to trade off privacy for health protection.

On the other hand, there is a high portion of negative sentiments showing some users do not trust their government in data handling. The analysis further reveals there is perceived effectiveness of the application in stopping the disease spread. Other users encountered software issues and perceived the application to be unreliable when it sent inaccurate exposure notifications. In India, some users are concerned about disruption to their livelihoods if they are sent into self-isolation from contact tracing.

Having a general understanding towards these applications will help the health sector better prepare for the next disease outbreak. To improve application uptake, government bodies need to earn people's trust and mitigate privacy issues. Users generally have positive sentiments about the Protect Scotland application. Further studies can examine the software and hardware differences between Protect Scotland and other applications to encourage usage. Software issues should also be solved to prevent false positive exposure notifications.

However, the high proportion of positive sentiments is inconsistent with the low application uptake rate. As discussed in previous sections, several other research also revealed their studies contained more positive sentiments [10]–[12]. Therefore, the reason for the low uptake rate still remains unclear to a certain extent. As mentioned in Section III, the initial analysis showed that 52% of the tweets were neutral. Neutral tweets contained mostly queries and product information and were removed for subsequent analysis. The reason for the low uptake rate may come from users' nonchalant neutral attitude, and this acts as a barrier to adopting the application.

Future work can expand on collecting more data from several social media sources and study whether or not users on other platforms generally feel neutral as well. Results discussed show sentiment similarities and differences in UK, US and India. This demonstrates cultural and geographical differences between countries. This provides insights for government bodies to strategize their policy locally to increase user uptake.

References

[1] W. H. Organisation, "COVID-19," March. 15, 2019. [Online]. Available: https://covid19.who.int/

[2] S. Munzert, P. Selb, A. Gohdes, L. F. Stoetzer, and W. Lowe, "Tracking and promoting the usage of a COVID-19 contact tracing app," *Nature Human Behaviour*, vol. 5, no. 2, pp. 247–255, 2021.

[3] J. Abeler, M. Bäcker, U. Buermeyer, H. Zillessen *et al.*, "COVID-19 contact tracing and data protection can go together," *JMIR mHealth and uHealth*, vol. 8, no. 4, p. e19359, 2020.

[4] M. Abueg, R. Hinch, N. Wu, L. Liu, W. Probert, A. Wu, P. Eastham, Y. Shafi, M. Rosencrantz, M. Dikovsky *et al.*, "Modeling the combined effect of digital exposure notification and non-pharmaceutical interventions on the COVID-19 epidemic in washington state," *MedRxiv*, 2020.

[5] C. Wymant, L. Ferretti, D. Tsallis, M. Charalambides, L. Abeler-Dörner, D. Bonsall, R. Hinch, M. Kendall, L. Milsom, M. Ayres *et al.*, "The epidemiological impact of the nhs COVID-19 app," *Nature*, vol. 594, no. 7863, pp. 408–412, 2021.

[6] M. Consult, "National tracking poll," *The Morning Consult COVID-19 Vaccine Dashboard*, pp. 1–38, 2023. [Online]. Available: http://bit.ly/3RYX3PN

[7] K. Vikram, "Not Many takers for Aarogya Setu App, Delhi and Chandigarh see Highest Users," vol. 25, 2021. [Online]. Available: https://www.newindianexpress.com/nation/2021/mar/21/not-many-takers-for-aarogya-setu-appdelhi-and-chandigarh-see-highest-users-2 html

[8] M. Walrave, C. Waeterloos, K. Ponnet *et al.*, "Adoption of a contact tracing app for containing COVID-19: A health belief model approach," *JMIR public health and surveillance*, vol. 6, no. 3, p. e20572, 2020.

[9] M. Bano, D. Zowghi, and C. Arora, "Requirements, politics, or individualism: What drives the success of COVID-19 contact-tracing apps?" *Ieee Software*, vol. 38, no. 1, pp. 7–12, 2020.

[10] J. Jamieson, D. A. Epstein, Y. Chen, and N. Yamashita, "Unpacking intention and behavior: Explaining contact tracing app adoption and hesitancy in the united states," in *CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–14.

[11] G. Samuel, S. Roberts, A. Fiske, F. Lucivero, S. McLennan, A. Phillips, S. Hayes, and S. Johnson, "COVID-19 contact tracing apps: UK public perceptions," *Critical Public Health*, vol. 32, no. 1, pp. 31–43, 2022.

[12] V. Garousi, D. Cutting, and M. Felderer, "Mining user reviews of COVID contact-tracing apps: An exploratory analysis of nine european apps," *Journal of Systems and Software*, vol. 184, p. 111136, 2022.

[13] K. Rekanar, J. Buckley, S. Buckley, M. Abbas, S. Beechum, M. Chochlov, B. Fitzgerald, L. Glynn, K. Johnson, J. Laffey *et al.*, "Sentiment analysis of user feedback on the hse contact tracing app," 2020.

[14] C. Diamantini, A. Mircoli, D. Potena, and E. Storti, "Social information discovery enhanced by sentiment analysis techniques," *Future Generation Computer Systems*, vol. 95, pp. 816–828, 2019.

[15] R. Singh, R. Singh, and A. Bhatia, "Sentiment analysis using machine learning technique to predict outbreaks and epidemics," *Int. J. Adv. Sci. Res*, vol. 3, no. 2, pp. 19–24, 2018.

[16] K. Ali, H. Dong, A. Bouguettaya, A. Erradi, and R. Hadjidj, "Sentiment analysis as a service: a social media based sentiment analysis framework," in *2017 IEEE international conference on web services (ICWS)*. IEEE, 2017, pp. 660–667.

[17] H. Bhavsar and R. Manglani, "Sentiment analysis of twitter data using python," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 3, pp. 510–511, 2019.

[18] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams engineering journal*, vol. 5, no. 4, pp. 1093–1113, 2014.

[19] M. Z. Asghar, F. M. Kundi, S. Ahmad, A. Khan, and F. Khan, "T-saf: Twitter sentiment analysis framework using a hybrid classification scheme," *Expert Systems*, vol. 35, no. 1, p. e12233, 2018.

[20] A. Alabrah, H. M. Alawadh, O. D. Okon, T. Meraj, and H. T. Rauf, "Gulf countries' citizens' acceptance of COVID-19 vaccines—a machine learning approach," *Mathematics*, vol. 10, no. 3, p. 467, 2022.

[21] J. C. Reis, A. Correia, F. Murai, A. Veloso, and F. Benevenuto, "Supervised learning for fake news detection," *IEEE Intelligent Systems*, vol. 34, no. 2, pp. 76–81, 2019.

[22] A. Giachanou and F. Crestani, "Like it or not: A survey of twitter sentiment analysis methods," *ACM Computing Surveys (CSUR)*, vol. 49, no. 2, pp. 1–41, 2016.

[23] I. Gupta and N. Joshi, "Enhanced twitter sentiment analysis using hybrid approach and by accounting local contextual semantic," *Journal of intelligent systems*, vol. 29, no. 1, pp. 1611–1625, 2020.

[24] Y. Kirelli and S. Arslankaya, "Sentiment analysis of shared tweets on global warming on twitter with data mining methods: a case study on turkish language," *Computational Intelligence and Neuroscience*, vol. 2020, 2020.

[25] O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A hybrid approach to the sentiment analysis problem at the sentence level," *Knowledge-Based Systems*, vol. 108, pp. 110–124, 2016.

[26] J. Piret and G. Boivin, "Pandemics throughout history," *Frontiers in microbiology*, vol. 11, p. 631736, 2021.

[27] P. Kaviani and S. Dhotre, "Short survey on naive bayes algorithm," *International Journal of Advance Research in Computer Science and Management*, vol. 04, 11 2017.

[28] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*. Association for Computational Linguistics, Jul. 2002, pp. 79–86. [Online]. Available: https://aclanthology.org/W02-1011

[29] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, "Sentiment analysis of tweets using machine learning approach," in *2018 Eleventh international conference on contemporary computing (IC3)*. IEEE, 2018, pp. 1–3.

[30] M. Al-Shabi, "Evaluating the performance of the most important lexicons used to sentiment analysis and opinions mining," *IJCSNS*, vol. 20, no. 1, p. 1, 2020.

[31] L. He and K. Zheng, "How do general-purpose sentiment analyzers perform when applied to health-related online social media data?" *Studies in health technology and informatics*, vol. 264, p. 1208, 2019.

[32] P. Lohar, G. Xie, M. Bendechache, R. Brennan, E. Celeste, R. Trestian, and I. Tal, "Irish attitudes toward COVID tracker app & privacy: sentiment analysis on twitter and survey data," in *The 16th International Conference on Availability, Reliability and Security*, 2021, pp. 1–8.

[33] K. Cresswell, A. Tahir, Z. Sheikh, Z. Hussain, A. D. Hernández, E. Harrison, R. Williams, A. Sheikh, A. Hussain *et al.*, "Understanding public perceptions of COVID-19 contact tracing apps: Artificial intelligence–enabled social media analysis," *Journal of medical Internet research*, vol. 23, no. 5, p. e26618, 2021.

[34] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," *HP Laboratories, Technical Report HPL-2011*, vol. 89, pp. 1–8, 2011.

[35] S. Elbagir and J. Yang, "Twitter sentiment analysis using natural language toolkit and vader sentiment," in *Proceedings of the international multiconference of engineers and computer scientists*, vol. 122, 2019, p. 16.

[36] S. Taj, B. B. Shaikh, and A. F. Meghji, "Sentiment analysis of news articles: a lexicon based approach," in *2019 2nd international conference on computing, mathematics and engineering technologies (iCoMET)*. IEEE, 2019, pp. 1–5.

[37] M. Alazab, A. Awajan, A. Mesleh, A. Abraham, V. Jatana, and S. Alhyari, "COVID-19 prediction and detection using deep learning," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 12, no. June, pp. 168–181, 2020.

[38] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Computer Science*, vol. 127, pp. 511–520, 2018.

[39] M. Mujahid, E. Lee, F. Rustam, P. B. Washington, S. Ullah, A. A. Reshi, and I. Ashraf, "Sentiment analysis and topic modeling on tweets about online education during COVID-19," *Applied Sciences*, vol. 11, no. 18, p. 8438, 2021.

[40] V. Balakrishnan and L.-Y. Ethel, "Stemming and lemmatization: A comparison of retrieval performances," *Lecture Notes on Software Engineering*, vol. 2, pp. 262–267, 01 2014.

[41] S. Kaur, P. Kumar, and P. Kumaraguru, "Automating fake news detection system using multi-level voting model," *Soft Computing*, vol. 24, no. 12, pp. 9049–9069, 2020.

[42] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the international AAAI conference on web and social media*, vol. 8, no. 1, 2014, pp. 216–225.

[43] M. A. Mudassir, Y. Mor, R. Munot, and R. Shankarmani, "Sentiment analysis of COVID-19 vaccine perception using nlp," in *2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2021, pp. 516–521.

[44] Y.-J. Tai and H.-Y. Kao, "Automatic domain-specific sentiment lexicon generation with label propagation," in *Proceedings of International Conference on Information Integration and Web-based Applications & Services*, 2013, pp. 53–62.

[45] V. Chaithra, "Hybrid approach: naive bayes and sentiment vader for analyzing sentiment of mobile unboxing video comments," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 4452–4459, 2019.

[46] E. Y. Chan and N. U. Saqib, "Privacy concerns can explain unwillingness to download and use contact tracing apps when COVID-19 concerns are high," *Computers in Human Behavior*, vol. 119, p. 106718, 2021.

[47] E. Hargittai, E. M. Redmiles, J. Vitak, and M. Zimmer, "Americans' willingness to adopt a COVID-19 tracking app," *First Monday*, vol. 25, no. 11, p. online, 2020.

[48] I. Garg, D. Kiran, and I. Sharma, "Sentimental analysis of "aarogya setu"," in *2020 international conference on smart innovations in design,*

environment, management, planning and computing (ICSIDEMPC). IEEE, 2020, pp. 263–267.

[49] A. Clarance. (2020) Aarogya Setu: Why India's Covid-19 contact tracing app is controversial. [Online]. Available: https://www.bbc.co.uk/news/world-asia-india-52659520

[50] A. Vanzo, D. Croce, and R. Basili, "A context-based model for sentiment analysis in twitter," in *Proceedings of coling 2014, the 25th international conference on computational linguistics: Technical papers*, 2014, pp. 2345–2354.