International Journal of Computing and Digital Systems
Vol. XXX, No. XXX, 2023

Use of AI Applications in order to Learn the Sentiment Polarity of Public Perceptions: A Case Study of the COVID-19 Vaccinations in the UAE

Abdulrahman Radaideh
Engineering Management,
University of Sharjah, UAE
u18105563@sharjah.ac.ae
ORCID#: 0000-0002-6184-9177

Fikri Dweiri
Industrial Engineering and Engineering Management,
University of Sharjah, UAE
fdweiri@sharjah.ac.ae
ORCID#: 0000-0001-5360-2142

Abstract— **Artificial Intelligence (AI) has revolutionized predictive, forecasting, and classification capabilities, finding applications across diverse domains. This paper investigates AI's role in analyzing sentiment polarity within the public's perceptions of post-COVID-19 vaccination providers in the United Arab Emirates (UAE). AI's prevalence in medical research, particularly for low-computation tasks, underscores its significance. Amidst abundant medical resources, concerns about the safety and efficacy of various COVID-19 vaccines persist. Social media platforms serve as dynamic spaces where individuals share vaccination experiences, shaping public perceptions. Recognizing AI's pivotal role in influencing business perception, this study exploits AI to extract insights into current vaccine perceptions. This methodology employs data mining to analyze textual data, classifying and clustering social media posts into distinct groups. Emotional labeling discerns sentiments associated with each vaccine. The dataset includes tweets on Pfizer-BioNTech, Sinopharm, Sputnik V, and Oxford-AstraZeneca—chosen for their availability in the UAE, where around 70% of the population received vaccinations by June 2021. This analysis aims to understand public sentiment, identifying preferences and concerns. Findings offer valuable insights into how these vaccines are perceived in the UAE, contributing to the broader discourse on AI, public perception, and healthcare. The study informs decision-makers and health authorities in refining communication strategies and addressing public concerns.**

Keywords— COVID-19; Sentiment analysis; public perception; SMOTE; LSTM, RNN, Pfizer-BioNTech; Sinopharm; Sputnik V; Oxford-AstraZe

I. Introduction

COVID-19 has been affecting many countries ever since the first case was recorded back in December of 2019. As of June of 2021, about 223 countries have been affected by it with a total of 181 million cases being registered and about 4 million deaths [25]. Due to the severity of it, it had become a necessity to get people across the world vaccinated so that we can come out of this pandemic safe and sound [1]. While precautionary measures can help stop the spread, vaccines are a crucial weapon against COVID-19 as they can help build antibodies against it and future variants of it as well. Due to the increased need for the creation of the vaccine, it has become a race between major biotechnology firms to develop and mass-produce a vaccine for the COVID-19. With the distribution of vaccines being started, it is very important to examine the acceptance of COVID-19 vaccinations not just with its efficacy rate which tells us about how effective the vaccine is in protecting the people from COVID-19 but also the public perception and recommendation over it on social media.

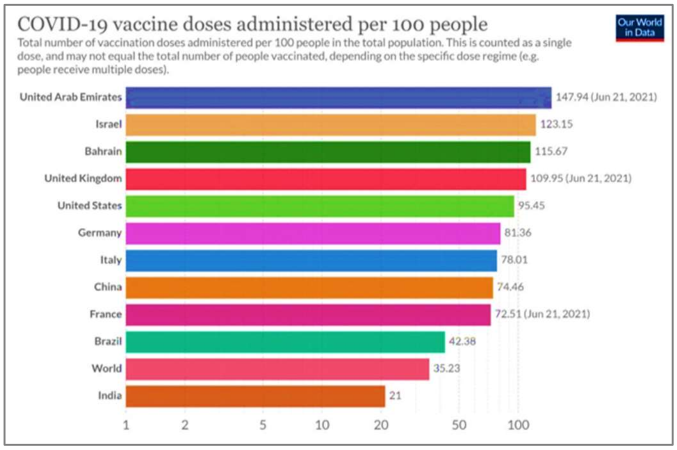
With sites like Facebook, Twitter and YouTube, social media has become a huge part of our life thanks to the ease in communication and also for how easy it has made to spread information and also misinformation. Since the rise of COVID-19, we have seen an increase in conspiracy theories related to its which consists of it being part of an elaborate plan by the Chinese government to it being a war against religions among others [2, 3]. This has led to ones made for the vaccines as well with the most common theory being spread around is that production and development of the vaccines are being funded by Bill Gates so that he can put microchips inside the human body through the vaccine to control their emotions and behavior [2].

These can present a problem as people would blindly believe in them rather than doing their own research and differentiate fact from fiction. As such, we need to devise methods that will allow us to analyze and interpret the posts that users make and also to determine what a given perception of a given vaccine is. This paper deals with an extensive analysis of the data that we have collected within the small duration the vaccination process in order to analyze the public perception of a given vaccine.

The research begins with the collection of tweets from various uses that will allow us to get an idea of what the users think about the vaccines themselves. The data is then cleansed and filtered out so that we can get only those tweets that are relevant to our research. The goal of this is to get a clear and unbiased overview of how the vaccinations are being perceived and as such, we need to be sure that the data we take helps in providing an unbiased opinion on the matter. This will be limited to the country of UAE since around 51% of the population is vaccinated as of May 2021, which makes them the second country to achieve this feat, the first being Israel and as of June 2021, about 150 doses have been administered per 100 people, which is found to be higher than Israel's and Bahrain's [25]. This can be seen in Figure 1 as well.

Fig. 1. Vaccination Doses (Administered per 100 people)

A. Research Aim



This research aims to identify and assess the public perception of vaccinations by the people. It is done to determine the facts from misinformation or fake news. It is used to assess the public perception of the vaccination based on how it performed, which is used to understand the effectiveness of the vaccine other than seeing the efficacy rate of the vaccine.

B. Research Objectives

The objectives of this paper consist of:

- To assess the public perception of the vaccines in the UAE using artificial intelligence algorithms.
- To analyze and interpret the data that we collect from social media.
- To eliminate misinformation and bias from the reality by giving a clear assessment of how the vaccines performed.

C. Research Questions

- Where And How to Collect the Data?
- What Has Been the Topic of Discussion Amongst People?
- What Is the General Perception of Vaccines Amongst People?
- What Are the Most Commonly Used Words That Are Associated with A Given Vaccine?

II. MOTIVATION AND JUSTIFICATION OF THE PROPOSED WORK

The motivation behind this work has been to portray how false information could lead to presenting misleading information to the general audience. It is especially the case with social media influences and news media sites that are powerful and capable to manipulate a large audience. With billions of users concurrently active in today's age, any information spread across the internet can be accepted as a fact at its face value without conducting any research. It is especially the case in the present scenario where people consider the conspiracy theories as facts rather than forming their own opinions. This can justify and provide a key reason as to why research like this should exist as it can help in getting an idea on what people think about the vaccine rather than hearing rumors or false information.

To increase the rate of vaccination, there needs to be a way for us to spread real information amongst people. However, there's a sizable portion that does not take the severity of the vaccine seriously. Instead, they would share false information and conspiracy theories. There has been an increase in anti-vaxxers over the years and has increased when the COVID-19 pandemic started. Many types of research have suggested that exposure to negative sentiment and an increase in misinformation and lies about the vaccine can lead to people being hesitant towards taking the vaccine, which could decrease the vaccination uptake amongst people [4,5,6].

In their research, Romer and Jamieson [4] noted that the belief in conspiracy theories usually relate to socially significant events. In this case, the people incline towards the novel coronavirus to unknown and sinister actions associated with it. As previously mentioned, conspiracy theories can undermine any kind of progress that practitioners make towards increasing the likelihood of vaccination due to how difficult it can be to rebut against it [7]. Any belief in one of them can lead to having an association with another that is similar to the one they believe in because of the implication it provides and how it relates to the original conspiracy theory. Some theories spread mass fear towards the side effects of vaccines such as the previously mentioned microchip theory and maybe those vaccinations are causing people to die more than without taking them can attribute to people not vaccinating themselves.

III. LITERATURE REVIEWS

Sharing on social media has been on the rise in recent years, with an estimate of millions of posts is being shared each day. According to Ovadia [8], they are about 6000 tweets per second posted on Twitter which equals about 500 million tweets per day and 200 billion tweets per year. It provides Artificial intelligence researchers with textual data for research on various purposes such as NLP or to data analysts to evaluate patterns that are associated with the. Several kinds of research have been done on sentimental analysis on political space, entertainment space, financial field, medicine, and many more areas using data from social media. In this section, we wish to show how sentimental analysis has been applied on different topics and the results obtained from such applications.

Ravikumar [9], in his research, aimed to filter the data and apply NLP techniques and calculate sentiment polarity detected in the user tweets during the 2014 Brazilian World Cup. To achieve this, he used a machine-learning algorithm for the normalization of data and used natural language processing techniques like word tokenization, stemming and lemmatization, Part of speech (POS) Tagger, Named Entity Recognition (NER) and parser to extract emotions from each tweet. The resultant assigned polarity obtained was then further analyzed using Naïve Bayes and Support Vector Machine (SVM) to compare both results for Naïve Bayes and SVM and prove the best approach for sentiment analysis on social media. By applying NLP techniques, the researcher was able to gain the sentimental polarity of the user's tweets and thus was able to assign emotional values.

Shihab and Yang [10] aimed to perform sentimental analysis using ordinal regression and machine learning algorithms. The tweets were first preprocessed then features extracted to help with the analysis. Multinomial logistic regression (SoftMax), Support Vector Regression (SVR), Decision Trees (DTs) and Random Forests (RF) for sentiment analysis classification in the proposed framework. Results experiment showed that the proposed model can ordinal regression in Twitter using machine learning methods with accuracy.

Severyn and Moschitti [11] proposed a system for sentiment analysis of Twitter. The work contributed to the new model for initializing the parameter weights of the convolutional neural network, which was crucial for training an accurate model while avoiding the need for new features. An unsupervised neural language model is used to train word embedding and is trained by an unsupervised group of tweets. The conventional neural network is used to refine the embedding on the supervised corpus.

The components used in the proposed work are activations, sentence matrix pooling, SoftMax and convolutional layers. To train, Stochastic Gradient Descent (SGD) and non-convex function optimization algorithms were used to calculate the gradients backpropagation algorithm. Dropout was used to enhance the regularization of the neural networks. The deep learning model is applied on two tasks: message-level task and phrase-level task from Semeval-2015 to predict polarity and achieve high outcomes.

Ale et al. [12] proposed implementing sentimental analysis using deep learning techniques. They argued that sentiment analysis faced challenges because of a lack of sufficient data in the field of NLP. To solve this, they argued that deep learning and sentiment analysis can techniques can be merged due to their automatic learning capabilities.

They used SVM and Naïve Bayes models for twitter specific lexicon to compete with existing studies. The artificial neural network (ANN) model used for classification and outperformed the existing models. The proposed model while used in sentimental analysis proved to be more accurate and efficient. Deep learning models proved to be better than SVM and normal neural networks as they have more hidden layers as compared to the normal neural network.

In their research, McGregor [13] showed how much social media plays a key role in understanding public perception along with how social media has been used by mainstream media to manipulate the audience to present the narrative that they want people to know. It is pretty evident from the 2016 US Elections, which was the core part of their analysis where the news media and outlets used public opinions to find new ways to manipulate the odds in the favor of former President of the United States, Donald J. Trump.

IV. METHODOLOGY

This section will highlight the methods and the materials that will be used to ensure that the research we have done is according to the scope and the criterion we have set before conducting the research. As such, our data will be limited to assessing the vaccines that are in the UAE, i.e. Pfizer-BioNTech, Sinopharm, SputnikV, and Oxford-AstraZeneca and the data sources will be limited to collecting data from Twitter since the reviews from it can be easily processed by the model.

Inclusion/Exclusion criteria need to be set as well to ensure that we can maintain and work on the data as ethically as we can. It will consist of making sure that any kind of data that does not fit with our criteria is omitted from it. It includes not analyzing tweets from news media outlets or any social media influencer because there can be some form of business in there which can badly impact the analysis. In addition, any retweets from those news media sites or people sharing those links will be omitted as well as that is just them spreading the news they saw on those sites. Any irrelevant tweets such as false information about the efficacy of the vaccine will be omitted as well because that can just be conspiracy theorists trying to reduce the likelihood of people getting vaccinated or to spread false rumors about it as well. We will try to include those posts that are made where the people go through their experiences after getting the vaccination as that will help in getting an insight on which one they got. Along with that, we will get an insight into any side effects they received from getting the vaccination as how it impacted their lives as well. Lastly, the data collected will be from 1st of February, 2021 when the vaccination process started to late April, 2021 as it will give us a healthy amount of tweets that will help in conducting this analysis. Since we need to collect 16,000 tweets, collecting from a possible millions of tweets should be easy, even with the API limit or with the limitation of SNScrape.

A. Tools that will be used

- Python as our main programming language
- Twitter API or SNScrape to collect and store the tweets.
- Pandas, Numpy, NLTK and Scikit Learn (to clean, process and standardize the data)
- Matplotlib and Seaborn (to visualize and analyze the data and the model itself)
- Tensorflow (to build the model)

B. Data Collection and Preparation

The data collection will be done by collecting 16000 reviews from Twitter. The data for this will be as a text. About 4000 tweets will be allotted to each of the given vaccination to give every vaccine a fair chance before conducting the research. The tweets will be narrowed down at random so that we can eliminate any form of bias from our side and we report the facts found in the research.

These will be collected through the mix of scraping the data from the web and through using the Twitter API. The reason a mixed approach is used is that Twitter's API is restrictive in terms of how much data we can scrape out of it so using other methods is a necessity to collect data for our use case [14]. In the case of an audio or a video post, we will scrape or transcribe the text in it which will then be cleared up to be used as a way to analyze what the user is saying in that video or audio [15].

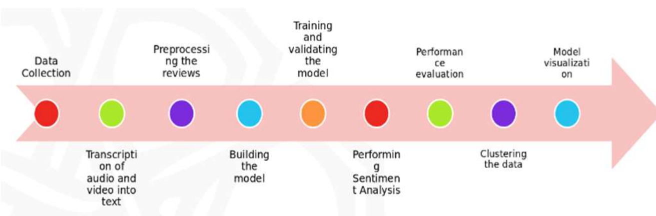


Fig. 2. Data Processing and Analysis

The data collected is followed the exclusion/inclusion criteria we have set before collecting the data. For this, we will scrape off any unnecessary links or texts so that the dataset does not affect the result in any way. Substitutions are placed for common spelling mistakes so that the classifier does not confuse. This process is represented in graphical form in Figure 2 and 3 as well. in this, it gives a general overview on how we will perform the crucial steps that will help in collecting the data, cleaning and preprocessing the data and also preparing it to be clustered, segmented and be used to train the model as well.

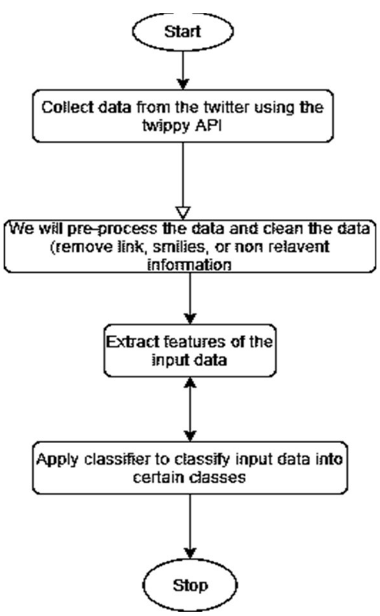


Fig. 3. Twitter Data Extraction and Sentiment Analysis

V. RESULTS

Once the data is processed and scraped, the data is first classified through Sentiment Analysis in order to label it positive, negative or neutral. This will help in ensuring that we are able to Based off these labels, this is then passed through another classifier, this time it is SMOTE in order to cluster the texts based on the features that are found inside the corpus of text. Doing this will help us in understanding the public perception of the users on the vaccinations. Once the model is built, we will then use it to segment the data based off the labels that we have set in order to get an idea of how many are there that are considered positive, negative or neutral in nature. The model will be trained using the data we have collected and also a mix of data that we may collect from sites such as Kaggle should they have the data we need. In addition, this will be used to classify any new data that we will pass in order to get an idea on how well it will be able to determine whether the corpus sent is positive, negative or neutral.

A. Analysis and building of the model

Once the data is processed and scraped, first the data is classified through Sentiment Analysis in order to label it positive, negative or neutral. Based off these labels, this is then passed through another classifier, this time it is SMOTE in order to classify it based on the features that are found inside the corpus of text. Lastly, the data will be clustered using SMOTE in order to cluster the tweets based on the features that it has and the sentiment that it provides. To assess the performance of it, we will separate about 30% of the data that we collect in order to use the unseen data to assess how well it can classify a given corpus of text.

The higher the validation accuracy and how close it is to the accuracy of the model, the more we will be sure that the model is trained well and that it can accurately label the text we throw at it. It can be used to examine large chunks of text which can help in contributing the perception of a given vaccine amongst the people of UAE.

B. Empirical Data Analysis

Once the data is cleaned and filtered out, we can analyze and process the tweets according to the questions that we aim to answer. It will ensure that we are able to develop a good understanding on what the people have to say about a given vaccine that they have received along with an idea on whether they believe it is good or not. There are duplicate tweets that are present in it such as some have mentioned both Sinopharm and Pfizer-BioNTech in their tweet while encouraging their followers to get vaccinated. These can be considered for this research as they are talking more about getting vaccinated rather than giving their opinion on the given vaccine. Similarly, tweets that only mention the fact that they are vaccinated but have used one of the tags in their tweets is also considered as we can safely assume that they used that given vaccine. In addition, we decided to remove the usernames of the individuals so that their anonymity remains and so that we can only focus on working with the tweet itself. Some of the examples of the data that we had to work with consist of the following:

TABLE 1: SAMPLE DATA FOR ASTRAZENECA VACCINE

Date	Tweet body	Search Query
2021-03-31 12:35:33+00:00	Hope Consortium: African ministers say doubts about AstraZeneca led to vaccine hesitancy	AstraZeneca
2021-03-31 23:27:46+00:00	JnJ now guys it's just a mixup next batch is in next week be ready johnsonandjohnson AstraZeneca	JnJ, AstraZeneca
2021-03-31 23:36:21+00:00	Just a warning to those doubtful of this story; the astra zeneca jab has made many people sick with an array of troubling symptoms, myself included. I've not yet had the 2nd, but some friends have; several vomited, all had crushing headaches, swollen faces, great lethargy"	AstraZeneca
2021-03-31 23:35:49+00:00	Somewhere, the AstraZeneca folks are feeling cheered-up that someone else effed up for once.	AstraZeneca

From these, we can see that the perception for AstraZeneca was perceived bad because of the present blood clotting issue that is found in it. This has caused a negative image towards it after it was revealed that patients can experience this and in extreme cases, die due to it. This has sort of decreased over time when Johnson and Johnson vaccine faced similar reports, which is being referred to in the last tweet in table 1. Other sample data that can be shown is in the following:

TABLE 2: SAMPLE DATA FOR PFIZER-BIONTECH VACCINE

Date	Tweet body	Search Query
2021-03-31 22:53:59+00:00	1 shot down Feeling very blessed vaccinated PfizerBioNTech	Pfizer
2021-03-31 22:43:08+00:00	I'm finally completely vaccinated Even though I'm grateful for science every day today is particularly special! PfizerBioNTech	Pfizer
2021-03-31 18:24:42+00:00,	Got my first COVID vaccine dose this morning. GetItDone PfizerBioNTech	Pfizer
2021-03-31 18:13:45+00:00	Bring on Twins baseball I'm fully vaccinated as of today Twins targetfield PfizerBioNTech	Pfizer

TABLE 3: SAMPLE DATA FOR SPUTNIKV VACCINE

Date	Tweet body	Search Query
2021-03-31 21:33:46+00:00	As for me I prefer SputnikV, however I have payment for one dose only hence I have to mix	SputnikV
2021-03-31 21:03:48+00:00	It saddens alot because in that crowd I can assure you 50 voted for and look at what they're going Through A lady fainted today as she was crossing the bridge to	SputnikV
2021-03-31 21:02:56+00:00	SputnikV is deliberately not being registered in the EU because of pressure from USA and UK Otherwise the vaccine have no issues.	SputnikV, vaccine

TABLE 4: SAMPLE DATA FOR SINOPHARM VACCINE

Date	Tweet body	Search Query
2021-03-15 19:14:02+00:00	If you had Sinopharm vaccine you can test positive even after getting the 2nd dose but it will be very mild, it gives you full immunity after completing 14 days after 2nd dose. You can get severe illness from 1st dose too so still be careful	Pfizer
2021-03-31 22:43:08+00:00	They combined two trials for phase 3 without standardized dosing schemes. But there is a lot more data now and that doesn't point to an issue. And in such an old type of vaccine you wouldn't expect issues I took Sinopharm as part of a phase 3 trial on the same logic btw	Sinopharm
251,2021-03-15 17:27:04+00:00	I wasn't convinced but went cos of my mum. Sinopharm made more sense to me but ended up getting Pfizer and had quite an ugly reaction - what essentially felt like a panic attack with 6 days of recovery post 2nd jab. Unknown times but yes I think it is best to get vaccinated ASAP.	Sinopharm

From looking at these tables, we can easily deduce that Pfizer has a generally better public perception compared to others due to either its high efficacy rate or due to the fact that it was readily available in countries such as US, UK and in the UAE. This has led to it being more used and talked about compared to other vaccines which have been unavailable to those areas either due to political reasons or due to the fact that these are found in small stocks. The former is true SputnikV and Sinopharm due to the fact the restrained relationships with Russia and China and the latter is true for the case of AstraZeneca.

In addition to this, we have had also analyzed the data itself so that we are able to provide a clear overview on the frequency of words that are used all across the dataset, which will give an idea on what the tweets consisted of and what were the general talking points in those. For this, we will eliminate any of the common words such as "vaccine", "vaccination" or the name of the vaccine as we only need to examine what other words were commonly used alongside it. In addition, any mention of countries will be omitted as well as that would not tell us anything significant of the tweets that are being made on the vaccine. For this, the top three most used words (MUW) that were found in the tweets for each of the vaccine are according to the below:

TABLE 5: WORD FREQUENCY DISTRIBUTION OF THE TWEETS IN THE DATASET

Vaccine	1st MUW	Frequency	2nd MUW	Frequency	3rd MUW	Frequency
Pfizer-BioNTech	dose	1234	first	1107	Moderna	738
AstraZeneca	people	466	blood	409	Jab	298
Sinopharm	pharmaceutical	3815	firm	3703	Arrives	1125
SputnikV	doses	340	approved	207	People	199

- People getting their first dose of Pfizer vaccine or people comparing or suggesting others to go with Moderna vaccine.
- People being concerned over the blood clotting issue that is found in AstraZeneca but this is from the more recent tweets.
- People discuss about the availability of the vaccine in pharmaceutical firms for Sinopharm vaccine.
- Lastly, people talking about the doses for Sputnik to be approved for use despite the political tension that is present between Russia and countries like the USA, UK and the UAE.

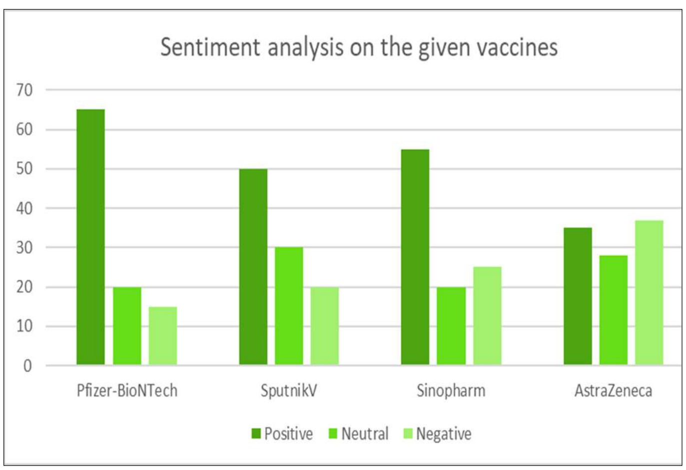


Fig.4. Graph Depicting the Perception of the Given Vaccines

The tweets which the sentiment analysis done can be summed up and clustered as seen in Fig 5.

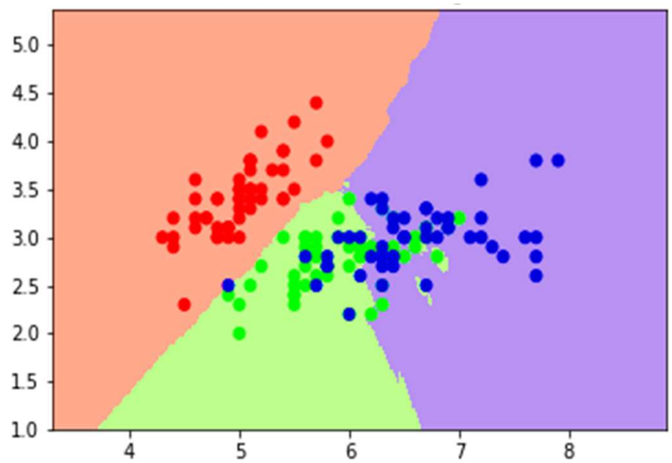


Fig.5. Clustering Result

C. Assessment of the algorithms used

As previously mentioned before, we will use a sentiment classifier to do this along with perform a clustering algorithm, i.e., SMOTE in order to classify these tweets as positive, negative or neutral. As such, we will need to develop the sentiment classifier that we can compare with the SMOTE algorithm.

For this, we will use TensorFlow to build the algorithm to classify and determine the sentiments of the tweets that we have collected (Large-Scale Machine Learning in the Earth Sciences, 2018). TensorFlow is an open-source library that is made by Google that is used to build models from scratch or to use the likes of Keras as a front end to simplify the building process.

For this, we will train the model to perform sentiment analysis using the existing dataset that is used to train such models. For this, we will use the dataset that we collected from the internet to classify it as positive, negative or neutral. The model that we created looked like the following in Table 6.

TABLE 6: MODEL SUMMARY

Layer	Output Shape	Params
Embedding	(None, 1000, 64)	640000
Dropout	(None, 1000, 64)	0
Dropout_1	(None, 1000, 64)	0
LSTM	(None, 100)	66000
Dense	(None, 1)	101

From these, we can deduce that a total of 706,101 trainable parameters are included in it, which will help in making sure that the model is pretty accurate. The model works by first embedding the given text or string to it being a noun, pronoun or among others. This is then trained and passed on to a LSTM network that will read through it sequence by sequence so that it can detect the emotion of the text or what is the sentiment of it in this case. The diagram for the model will look something as what is shown in the Figures 6 and 7 shown below:

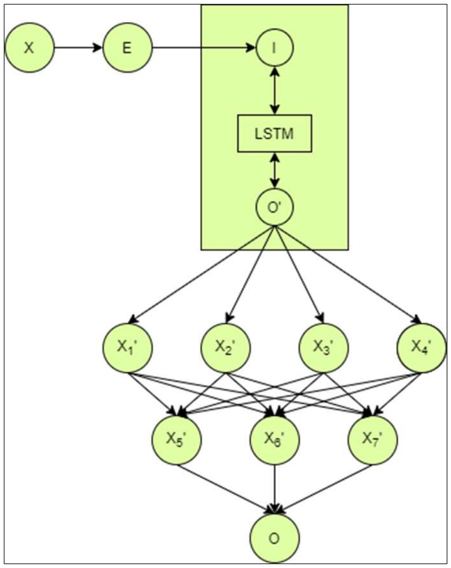


Fig.6. Basic diagram of the model

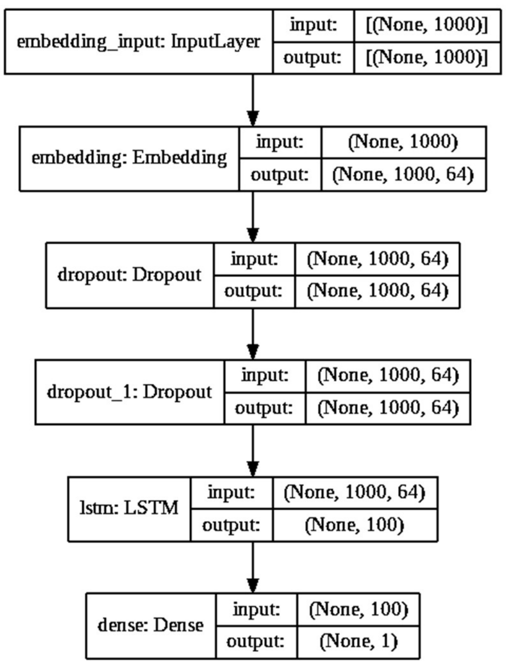


Fig 7. Graph Plot of the Model with the Variables and Input Parameters

In this diagram shown in Figure 6, the following labels are used the following parts of the model

- X – this is used to refer to the input data that is being transferred to the model. this
- X_n' – this refers to the hidden layers that are received from the model and into the dense layer. Each of these nodes consist of an activation function that will help in adjusting the parameter of the function so that we are able to get the right result.
- I – this refers to the input layer of the LSTM network that we have used. In this, the input received is from the embedding layer that will consist of the features that are found inside of the LSTM layer.
- O' – the output layer of the LSTM layer from which we collect the output that we have obtained after processing it in the hidden layers of the LSTM model. These differ from something along the likes of RNN with the use of a forget gate inside the hidden layer that is used to ensure that the network does have a vanishing gradient problem.

In this, LSTM consists the input layer, output layers and the hidden layers which consists of the forget gate. To prevent overfitting, we have added a dropout rate to it. A dropout acts as a regularizer that helps in making sure that the models are properly trained and that the risk of overfitting is reduced. This works by stopping a set number of neurons at random in the network from training so that the activated ones can train better (Srivastava et al., 2015). In our case, we have trained it to work under a dropout rate set at 40%, which means that 40% of the subsequent layers will be deactivated at random at each epoch so that the rest of the 60% will be trained and further improved. This will help in better regularizing the LSTM network so that it can better analyze the sentiment of the model. In our case, we have added two dropouts as it can potentially help in better training the model as we will be first deactivating a set percent of the neuron and then proceed to deactivate a portion of the remaining ones. This model will be passed through 50 epochs which means that the data will be passed through the model for about 50 times which will help it get familiar with the patterns found in a positive, neutral or a negative corpus. The diagram in Figure 7 gives us a general overview on what the model will look like along with the number of parameters that it will have. This consists of the input parameters of the layers and the output parameters of the layer. This information can help in understanding how many neurons will work in a network.

When training the model out, we found out that the model has a pretty high accuracy of about 98%, which shows that it is pretty familiar with the data itself. Meanwhile the validation accuracy was found to be around 88%. There could be a chance that the model has been over-fitted where the data is too familiar with the data and any unfamiliar text fed to it will be met with an uncertainty. The graphs of these can be seen in Figures 8 and 9 which shows the accuracy of the model as well as the loss that it calculated as well.

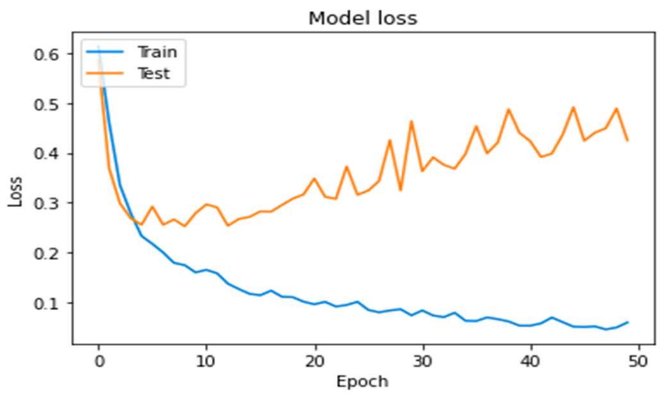


Fig. 8. Graph Depicting the Training and Test Accuracy of the Model

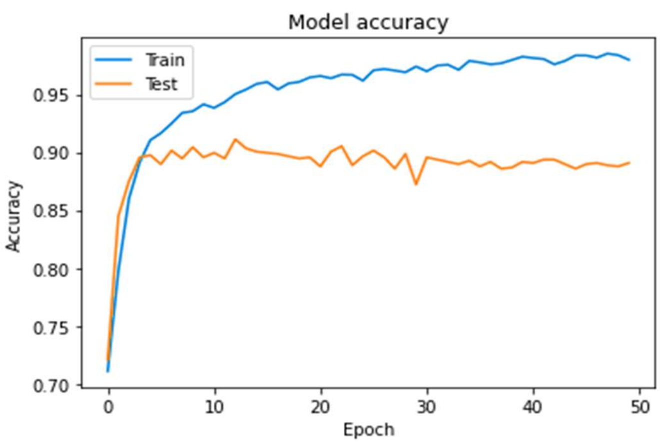


Fig. 9. Graph Depicting the Training and Test Loss of the Model

From the loss that we calculated; we can examine that the model performed as well as we intended it to be. The high amount of loss on test indicates that there are some accuracy problems in the model such as it can generate more false positives and negatives but overall, we can see that the model will be able to accurately detect the sentiment of the text that is being passed at it.

This will be compared to the SMOTE algorithm that we will use to cluster these tweets based on the sentiment that it provides. To simplify our process, we will use the module of SMOTE from the library imbalanced-learn, which is like SciKit-Learn but it provides already made algorithms for oversampled data, which also includes SMOTE. Comparing the two together, we can give the result in the following table:

Table 7: Comparison with Our Algorithm And SMOTE

Algorithm	**Test Accuracy**	**Validation Accuracy**
Our Model	86%	88%
SMOTE	82%	84%

As we can see in Table 7, our model performed generally better because SMOTE is usually used when data is highly imbalanced, unlike our dataset where the labeling is somewhat balanced in general. This can help in making sure that we are able to classify an imbalanced data. However, SMOTE can be used to train the model far better as it can better cluster the data compared to classifying with the help of Sentiment Analysis.

VI. CONCLUSION

In this work, we have analyzed public sentiment in the UAE about the COVID-19 vaccine using Twitter data. The findings from Twitter data analytics were verified and validated by the actual vaccination data from the USA We have analyzed the progression of public sentiment from early February to late April of 2021. We have found positive sentiment to be dominant

RESEARCH FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGEMENT

This research was made in part with the help of my peers and supervisors who helped in making this paper and conducting the research. I will also like to thank my friends who kept me motivated me towards finishing this paper.

REFERENCES

[1] M. A. e. al, "Energy Choices in Alaska: Mining people's Perception and Attitudes from Geotagged Tweets,," *Renewable and Sustainable Energy Reviews, vol. 124, p. 109781, May 2020,* 2020.

[2] A. a. N. Elmitwally, "A Comprehensive Study for Arabic Sentiment Analysis (Challenges and Applications),," *EgyptianInformaticsJournal,vol.21,no.1,pp.7–12,Mar.2020,doi: https://doi.org/10.1016/j.eij.2019.06.001,* 2020.

[3] G. A. e. al., "AI in the media and creative industries," *New European Media , vol. 01,* 2019.

[4] I. C.-P. J. F. S.-R. a. C. A. I. O. Araque, "Enhancing Deep Learning Sentiment Analysis with Ensemble Techniques in Social Applications," *Expert Systems with Applications, vol. 77,,* 2017.

[5] M. E. B. a. A. Kabiri, "HOMPer: A New Hybrid System for Opinion Mining in the Persian Language," *Journal of Information Science, vol. 46, no. 1, pp. 101–117, ,* 2019.

[6] B. S. Y. X. a. C. H. E. Cambria, "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intelligent Systems, vol. 28, no. 2,* 2018.

[7] Y. L. F. Z. X. S. P. a. K. K. E. Cambria, "SenticNet 6: Ensemble Application of Symbolic and Subsymbolic AI for Sentiment Analysis," *Proceedings of the 29th ACM International Conference on Information & Knowledge Management, vol. 28,,* 2020.

[8] N. C. a. P. Wang, "Advanced Combined LSTM-CNN Model for Twitter Sentiment Analysis," *in Proceedings of 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), IEEE,* 2018.

[9] M. T. A. A. S. M. A. M. E. B. a. X. Z. U. A. Chauhan, "A Comprehensive Analysis of Adverb Types for Mining User Sentiments on Amazon Product Reviews," *Wide Web, vol. 23, no. 3, pp. 1811–1829,* 2020.

[10] S. H. I. A. A. A. A. N. O. D. I. A. S. M. Y. I. S. A. e. a. Muhammad, "Afrisenti: A twitter sentiment analysis benchmark for african languages," *arXiv preprint arXiv:2302.08956,* 2023.

[11] S. S. S. J. C. a. M. S. Park, "Mind games: A temporal sentiment analysis of the political messages of the Internet Research Agency on Facebook and Twitter," *New Media & Society 25, no. 3 ,* 2023.

[12] R. S. P. C. C. C. P. a. M. E. R. S. P. C. C. C. P. a. M. E. Catelli, "Lexicon-based sentiment analysis to detect opinions and attitude towards COVID-19 vaccines on Twitter in Italy.," *Computers in Biology and Medicine 158,* 2023.

[13] N. A. B. S. V. J. B. a. R. B. Braig, "Machine Learning Techniques for Sentiment Analysis of COVID-19-Related Twitter Data.," *IEEE Access 11,* 2023.

[14] O. M. C.-F. W. C.-F. a. K. I. Carvache-Franco, "Topic and sentiment analysis of crisis communications about the COVID-19 pandemic in Twitter's tourism hashtags," *Tourism and Hospitality Research 23, no. 1 ,* 2023.

[15] S. R. J. a. H. Ramasangu, "Classification of Cognitive States using Task-Specific Connectivity Features," *Eng. Technol. Appl. Sci. Res., vol. 13, no. 3, pp. 10675–10679,* 2023.

[16] S. M. K. M. U. F. a. .. W. M. Anwer, "Attack Detection in IoT using Machine Learning," *Eng. Technol. Appl. Sci. Res., vol. 11, no. 3, pp. 7273–7278,* 2021.

[17] V. P. V. a. V. K. Yadav, "Long short term memory (LSTM) model for sentiment analysis in social data for e-commerce products reviews in Hindi languages," *International Journal of Information Technology 15, no. 2 ,* 2023.

[18] O. A. A.-R. J. L. H. S. S. B.-C. J. Z.-P. J. A. Y. a. M. C.-C. Iparraguirre-Villanueva, "The public health contribution of sentiment analysis of Monkeypox tweets to detect polarities using the CNN-LSTM model," *Vaccines 11, no. 2,* 2023.

[19] A. M. E. P. a. A. S. Xu, "Sentiment Analysis On Twitter Posts About The Russia and Ukraine War With Long Short-Term Memory," *Sinkron: jurnal dan penelitian teknik informatika 8, no. 2 ,* 2023.

[20] M. W. A. A. W. A. S. K. M. U. C. I. a. T. R. G. Bibi, "A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis.," *Pattern Recognition Letters 158,* 2022.

[21] N. S. J. A. P. S. B. K. S. A. L. A. K. Y. a. J. T. Leelawat, "Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning," *Heliyon 8,* 2022.

[22] M. S. a. R. R. Neethu, "Sentiment analysis in twitter using machine learning techniques.," *n 2013 fourth international conference on computing, communications and networking technologies (ICCCNT),* 2013.

[23] A. a. R. O. Alslaity, "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions," *Behaviour & Information Technology ,* 2022.

[24] P. Q. T. a. H. B. N. M. D. Nguyen, "An Application of Analytic Network Process (ANP) to Assess Critical Risks of Bridge Projects in the Mekong Delta Region"," *Eng. Technol. Appl. Sci. Res., vol. 13, no. 3, pp. 10622–10629,* 2023.

[25] S. J. A. M. B. R. S. a. R. T. Mann, "Twitter sentiment analysis using enhanced BERT.," *In Intelligent Systems and Applications: Select Proceedings of ICISA 2022,,* 2022.

[26] G. a. A. S. Kaur, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *Journal of Big Data 10, no. 1,* 2023.

[27] S. K. Shrivastav, "Exploring the application of analytics in supply chain during COVID-19 pandemic: a review and future research agenda.," *Journal of Global Operations and Strategic Sourcing 16, no. 2,* 2023.

[28] H. S. S. S. M. T. M. a. P. M. Madhu, "Detecting offensive speech in conversational code-mixed dialogue on social media: A contextual dataset and benchmark experiments.," *Expert Systems with Applications 215,* 2023.

[29] F. Sufi, "Algorithms in low-code-no-code for research applications: a practical review.," *Algorithms 16, no. 2,* 2022.

[30] S. C. M. F. G. A. a. E. N. Fanni, "Natural language processing.," *In Introduction to Artificial Intelligence, pp. 87-99. Cham: Springer International Publishing,* 2023.