Twitter Sentiment Analysis of Covid-19 Vaccine Brands

TAL STERNBERG, Dartmouth College, USA
PATRICK NICCOLAI, Dartmouth College, USA
ANDREA ROBANG, Dartmouth College, USA
XIAOYU WEI, Dartmouth College, USA

CCS Concepts: • **Applied computing** → **Consumer health**.

Additional Key Words and Phrases: COVID, vaccines, sentiment analysis, Twitter

ACM Reference Format:
Tal Sternberg, Patrick Niccolai, Andrea Robang, and Xiaoyu Wei. 2022. Twitter Sentiment Analysis of Covid-19 Vaccine Brands. 1, 1 (November 2022), 17 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

From early 2020, The COVID-19 pandemic has endangered world public health. The endpoint of the COVID-19 pandemic is either herd immunity or the widespread availability of an effective vaccine[1]. Throughout the pandemic, many pharmaceutical companies and research institutions in various countries have been producing vaccines and to date, there are many different brands of vaccines available in various countries around the world. The global progress of the COVID-19 vaccine has also generated much discussion on social media platforms about multiple factors, such as vaccine protection and efficacy. In this article, we would like to analyze the relevant information on social media to understand the people's sentiment towards various vaccine brands.

Social media is a key data source for analyzing sentiment of various populations. The Covid-19 vaccines have been polarizing and highly debated on social media, namely Twitter. Our goal is to perform sentiment analysis using BERTsent [10] in order to compare with prior work using the AFINN lexicon [15], and to see how sentiments for these vaccines change over time. We aim to how vaccine sentiment changes over time for different brands, and to explore the driving causes of vaccine sentiment.

The continually updated Coronavirus (COVID-19) Tweets Dataset [4] provides us an unprecedented opportunity to analyze COVID-19 vaccine brands trends throughout the world. A deeper understanding of sentiment trends by brand can help governments, health organizations, and distributors figure out how to change marketing and get more people vaccinated. Understanding sentiment trends is the first step in an aim to maximize sentiment for each vaccine.

(1) Project Objectives:
   This project aims to leverage BERTsent [10] and BERTweet [14] to measure sentiments for various vaccine brands. BERT outperforms lexicon based sentiment analysis **??**, and thus our

Authors' addresses: Tal Sternberg, tal.g.sternberg.23@dartmouth.edu, Dartmouth College, Hanover, New Hampshire, USA; Patrick Niccolai, patrick.m.niccolai.gr@dartmouth.edu, Dartmouth College, Hanover, New Hampshire, USA; Andrea Robang, andrea.s.robang.24@dartmouth.edu, Dartmouth College, Hanover, New Hampshire, USA; Xiaoyu Wei, xiaoyu.wei.gr@dartmouth.edu, Dartmouth College, Hanover, New Hampshire, USA.

XXXX-XXXX/2022/11-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

aim is to get more accurate sentiment trends so that they can be harnessed to study how certain vaccine brands must change their marketing.
To accomplish these goals, our approach is to compare AstraZeneca/Pfizer/Moderna sentiments over time to a baseline, the results from Marcec et al. **??**, and then make novel exploration of Sinopharm and Sputnik vaccine sentiment over time.

(2) Baseline and Contribution:
In a paper by Marcec et al. [13] from 2021, a sentiment analysis of AstraZeneca, Pfizer and Moderna COVID- 19 vaccines was conducted. This paper will be the baseline with which we compare our research. Additionally, we will do the same analysis on Sputnik and Sinopharm vaccines which were excluded from Marcec's study.
Our main contribution will be to attempt to reproduce Marcec et al.'s results about vaccine sentiment trends over time using the more powerful BERTsent model, and then to get novel results for the Sputnik V and Sinopharm vaccines. We will then do topic modeling to attempt to explain our sentiment results.

2 RELATED WORK

2.1 Baseline and Using Sentiment Analysis for Social Media Data

Our paper is based off a paper entitled Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID- 19 vaccines by Marcec et al. which was published in July of 2021 [13]. They found that the sentiments for Pfizer and Moderna were stable over time, while the sentiments for AstraZeneca had a significant negative drop over the period of time between December 2020 and March 2021. They used AFINN lexicon to classify sentiment and compared the probability the sentiment was positive across tweets to compare. See Section 2.2 for more information on lexicon based approaches to sentiment analysis.

Another paper on COVID-19 related sentiment analysis comes from Schmitzberger and his team [18] that used social media data across multiple platforms including Twitter, Reddit, and Facebook to categorize COVID-19 related posts as neutral, negative, and positive then further extract themes from those three buckets to create strategies to increase COVID-19 vaccine acceptance.

A study done by Piedrahita-Valdés et al. looked at vaccine hesitancy on social media from June 2011 to April 2019 [16]. This study is pre-covid vaccine and explores vaccine hesitancy towards other vaccines such as flu. They used an option mining analysis to discover that peaks of positive tweets occurred every April and decreased every weekend. Negative tweets followed the opposite pattern.

2.2 AFINN Lexicon

The lexicon approach to sentiment analysis works by first creating a lexicon, which is simply a list of words and their sentiment scores. In the case of the AFINN lexicon, Nielson created a list of 2,477 words that were manually given a sentiment score ranging between -5 (negative sentiment) and +5 (positive sentiment) [15]. This lexicon was created specifically for Twitter, so it has many slang words and abbreviations commonly used on Twitter.

To use this lexicon for sentiment analysis on a Tweet, one can simply take a sum of the sentiment scores from the lexicon for every word in the Tweet which is also in the lexicon and divide by the number of words. This is the approach that Marcec et al. took. Lexicon based approaches to natural language processing have been shown to be less effective than machine learning based approaches [8], so we hope to improve upon this paper by applying a machine learning sentiment analysis model to the same domain.

2.3 BERT and BERTweet

Transformers are an architecture of NLP models which can be used, among other things, to translate text from one language to another. The Transformer does this in two steps. First, an encoder takes the natural language and embeds it into a feature representation which retains the meaning of the original text. The next step is for a decoder to take this embedding and re-create natural text.

BERT stands for Bidirectional Encoder Representations from Transformers. Prior to BERT, most Transformers analyzed a sequence of text from left-to-right or right-to-left, i.e. in one direction. BERT analyzes text bidirectionally, which allows it to have a stronger understanding of the context of a piece of natural language. In order to train a model to see bidirectionally, BERT employs Masked Language Modeling (MLM). MLM involves masking random words in a sentence and training a model to predict those words. In this way, BERT learns not only to predict a word to the right or to the left of a sequence, but anywhere in the sequence, making BERT bidirectional and more powerful than other NLP techniques [6].

The original BERT model was trained on the BookCorpus dataset [25]. This is a dataset containing the text of over 11,000 books. Pretraining on this dataset is useful for many tasks, but ineffective for analyzing Twitter data. This is due to the vastly different characteristics of Tweets compared to other texts, including informal grammar and short length. To address this, Nguyen et al. introduced BERTweet [14]. BERTweet uses the same architecture as BERT, with a RoBERTa model pre-training procedure [12], but is trained on a dataset of Tweets. Specifically, the training data for BERTweet included 850 million Tweets, along with an addition 23 million Tweets relating to Covid-19.

Because BERT and BERTweet are types of Transformer, they each contain an encoder which transforms natural language into a feature representation. This encoder is what BERTsent is built off of.

2.4 BERTsent

The method we use is BERTsent, derived from BERTweet, as derived in Lamsal et al.[10]. This models employs transfer learning by taking the encoder from the pretrained BERTweet model and doing fine tuning to create a classifier. The additional training data for fine tuning is a datset of Tweets with labeled sentiments from [20]. This dataset was created for the SemEval-2017 Sentiment Analysis in Twitter task, created by [17]. This task challenged people to create NLP models for Twitter sentiment analysis, and for this challenge SemEval provided training data which consisted of Tweet texts and their labeled seniment scores. The sentiment scores were manually created, and were in the range {Highly negative, Negative, Neutral, Positive, Highly positive}. The Tweets in this dataset were all Tweets relating to trending topics during 2017, when the challenge took place. Note that these Tweets are not related to Covid or vaccines.

BERTsent outputs a tuple with three values: the probability of negative sentiment, the probability of neutral sentiment, and the probability of positive sentiment.

3 METHOD

3.1 Data Collection

Our data comes from the IEEE Open Access Covid-19 Tweets Dataset [4]. The dataset contains Tweet IDs of Covid-related tweets ranging from December 2019 to September 2021. Using the tweepy library [3] and Twitter's Developer API we searched Twitter for the Tweets. For each Tweet, we downloaded it if it was specifically related to vaccines, which we determined by searching the Tweet text for vaccine keywords, such as "vaccine" or "moderna". The location and date distributions of our Tweets can be seen in Figure 1 and Figure 2. For context, most of the tweets used in our

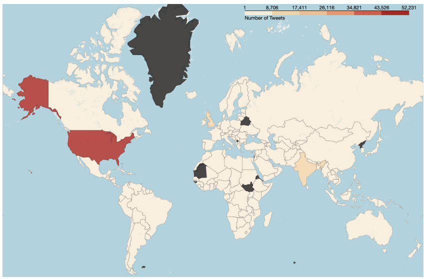


Fig. 1. Distribution of Tweets by Country

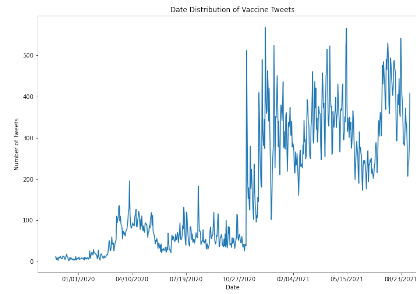


Fig. 2. Distribution of Tweets by Date

methodology are made by users from the United States with a spike in posts between December 2020 and August 2021.

We downloaded the Tweet text, along with the date and time the Tweet was published and the location where it was published. So, for every vaccine-related Tweet, we add to our dataset the tuple: (t, d, l, i) where t is the Tweet text, d is the date and time of the Tweet, l is the location of the Tweet, and i is the Tweet ID. At this stage our full dataset is:

$$\{(t_1, d_1, l_1, i_1), ..., (t_n, d_n, l_n, i_n)\}$$

where $n = 112,220$, the number of Tweets we downloaded.

After downloading vaccine-related tweets, the final processing step was to remove non-English Tweets. Because BERTweet and BERTsent were trained only on English data, we decided to exclude non-English Tweets from our analysis. To do this, we used Google's langdetect [21]. This library provides a Naive Bayes classifier pre-trained on texts from 49 different languages which has been shown to have precision over 99% in predicting the language of a given text [11].

3.2 Data Description

A breakdown of our dataset by language can be seen in Table 1. We omit any languages with less than 1000 Tweets from the table. For our analysis we only used English Tweets, with left us with a dataset of over 98,000 Tweets. This was 87.6% of our full dataset.

For our analysis we divided the Tweets by manufacturer. This was done by searching a Tweet text for either the name of the manufacturer or the WHO official name [23] of that manufacturer's vaccine. For example, a Tweet was considered to be about the Moderna vaccine if the keywords "moderna" or "spikevax" were present in the Tweet's text. The distribution of the amount of Tweets for each of the six manufacturers we studied is seen in Table 2. This table includes the Johnson & Johnson vaccine, but we did not include this in our analysis because of the small sample size for that particular vaccine.

The majority of the Tweets (over 90%) did not mention a specific brand. This means that the Tweet contained the word "vaccine", "vaccination", or "vax", but did not mention any manufacturer name or specific vaccine name.

3.3 Applying BERTsent

Using the transformers library [2], we downloaded the pretrained BERTsent model. We also download the BERTweet tokenizer from the same source, which breaks the full Tweet text into

Language	Number of Tweets	Percent of full Dataset
English	98344	0.876
French	3627	0.032
Italian	2538	0.022
Spanish	1916	0.017
Indonesian	1034	0.009
Other	4761	0.042
Total	112220	1

Table 1. Language Distribution of Dataset.
Only Languages with over 1000 Tweets are shown.

Manufacturer	Number of Tweets	Percent of full Dataset
None	101789	0.907
Pfizer	5140	0.045
Moderna	1975	0.017
AstraZeneca	1832	0.016
Sputnik V	701	0.006
SinoPharm	676	0.006
Johnson & Johnson	107	0.001
Total	112220	1

Table 2. Manufacturer Distribution of Dataset.

tokens and removes things like URLs or Twitter user mentions so that the Tweet can be analyzed by the sentiment analysis model. The tokenizer also handles emojis by replacing the emoji with text. For example the crying face emoji is replaced with ":crying_face:".

We use the tokenizer on the Tweet text and then pass the tokenized text to BERTsent. The model returns the sentiment score tuple s which is (p_-, p_0, p_+) where p_- is the probability the Tweet sentiment is negative, p_0 is the probability the Tweet sentiment is neutral, and p_+ is the probability the Tweet sentiment is positive. We add this to the Tweet's data, and our full dataset becomes:

$$\{(t_1, d_1, l_1, i_1, s_1), ..., (t_n, d_n, l_n, i_n, s_n)\}$$

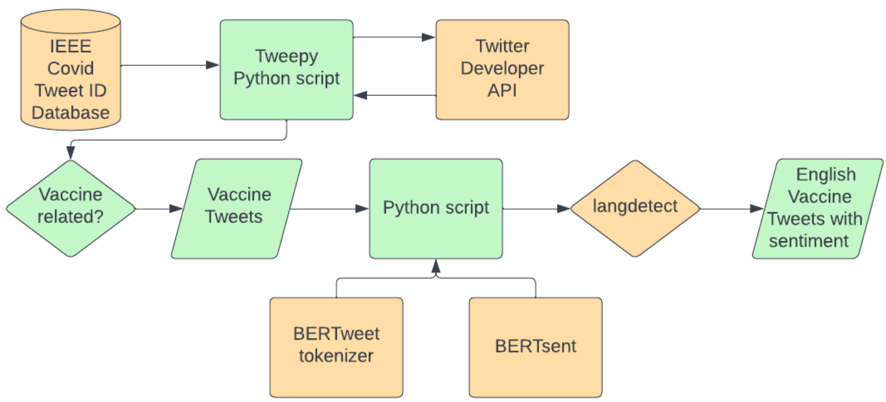
The full data processing pipeline can be seen in Figure 3. Also see Section 6.1 in the Appendix for some example Tweets. Note that the order of the BERTsent output is probability of negative sentiment, then probability of neutral sentiment, then probability of positive sentiment.

3.4 Outline for Analysis

3.4.1 Comparing to AstraZeneca/Pfizer/Moderna Sentiments Over Time to Baseline.

(1) For our pre-processing, we select only the Pfizer, Moderna, and AstraZeneca data. We only select data that is binned between 1 December 2020 to 31 March 2021, to compare to the baseline of Marcec et al.'s study.
(2) For classification of sentiment we will be using the daily average sentiment, similar to Marcec et al. and comparing those values over time across different vaccine brands. We define the sentiment value of a Tweet as $p_+ - p_-$, where p_+ is the probability the sentiment is positive and p_- is the probability the sentiment is negative.

Fig. 3. Data Processing Pipeline



(3) Because our sentiment scores differ from the values in Marcec et al. due to the difference in how the scores are generated, we use statistical tests to examine the trends in these scores, which are comparable to the results of Marcec et al.

3.4.2 Novel Exploration of Sinopharm and Sputnik Vaccine Sentiment Over Time.

(1) Here, we select only the Sputnik and Sinopharm data.
(2) Again, we will use BERTSent from BERTweet to analyze the sentiment of tweets.
(3) With the sentiment generated from the model and tweets binned by month, we will explore the trends of these vaccines over time.

3.5 Further Statistical Tests

3.5.1 Kruskal-Wallis Test. To further analyze our data we use the Kruskal-Wallis statistical test [9]. This is a test that is used to determine if multiple groups of samples are drawn from the same distribution. The test works by calculating the H score, which is defined as

$$H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j} \right] - 3(n+1)$$

where n is the total number of samples, c is the number of groups, T_j is the sum of the jth sample for each group (considering samples in order), and n_j is the size of the jth sample. We chose the Kruskal-Wallis test because it does not assume samples are drawn from a normal distribution, and it does not require the groups to be the same size, making this test appropriate for our dataset.

After calculating H, we compare it to the chi-square critical value, where the degrees of freedom is the number of groups minus 1. If H is greater than the chi-square value, we can conclude that the distributions are significantly different.

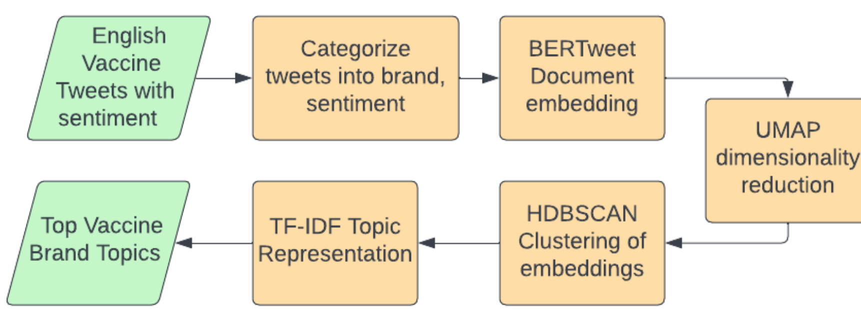
We used scipy's stats.kruskal function [19], which returns the H score, along with the Kruskal-Wallis p-value. This value, which we'll call p_k, is the chi-square p-value for which the H score equals the chi-square critical value. Using this interpretation, the p_k value gives us the probability of getting our results if the null hypothesis is true. In our case, the null hypothesis is that the groups of samples are from the same distribution, so a small p_k means we reject this hypothesis, and that the samples are from significantly different distributions. We consider a p-value of 0.05 to be statistically significant, so for any Kruskal-Wallis test with a p_k value < 0.05, we consider those distributions significantly different.

3.5.2 Post Hoc Games-Howell Test. The Kruskal-Wallis test tells us if multiple groups of samples are drawn from different distributions, but it does not give any information about which of the groups are different or how they differ from one another. To address this we, do post hoc testing using the Games-Howell test. The is test is similar to Kruskal-Wallis, but the Games-Howell test performs a pairwise analysis on the groups of samples. This way, we can learn exactly which pairs of distributions are statistically significantly different. Like the Kruskal-Wallis test, the Games-Howell test does not assume that the groups have samples have the same variance or are the same size, making the test suitable for our dataset.

We use the Pingouin [22] library's implementation of the Games-Howell test. This implementation returns p-value for each pair in the pairwise test, which we'll call p_g. Similar to p_k, p_g gives the probability of getting the results given the null hypothesis, which is that the two distributions are the same. So, a small p_g means that we can conclude that the two distributions are different.

3.6 Topic Modeling

Fig. 4. Topic Modeling Pipeline



To get a better sense of why each vaccine received the average sentiment score they did for those used in our study and the baseline (Moderna, Pfizer, AstraZeneca), we implemented topic modeling on tweets grouped by vaccine brand and sentiment (negative or positive). Topic modeling is a machine learning technique to cluster words by meaning or "topic". This was used to gain a better sense of the language being used in the tweets relative to each vaccine brand and to give some insight into reasons for trends over time. The full topic modeling pipeline can be seen in Figure 4.

We used BERTopic and grouped out tweet data by vaccine type and overall sentiment as characterized by BERTsent (positive or negative). We used the following as stopwords: "covid", "moderna", "pfizer", "astra", "astrazeneca", "sputnik","sinovac", "sinopharm", "vaccine", "covid19", "coronavirus", "vaccine", "vaccines". These stopwords in addition to the standard english stopwords of Python library scikit-learn's CountVectorizer were removed from the texts since their inclusion to our topics would not contribute anything of significance. Stopwords are removed in the topic modeling process because they may be articles like "i", "the", "they", etc. Stopwords related to COVID-19 were also removed because after previous methods of tweet filtering, we are sure that all of our tweets are related to COVID-19 and different vaccine brands, so their inclusion in the topic modeling process would be redundant.

We then trained the BERTopic model [7] following the steps of the pipeline, by first embedding documents using the pre-trained transformer-based language model BERTweet to represent our tweets as embeddings or as high-dimensional vectors wherein semantically similar words are near each other in their vector representation. The dimensions of these vectors are then reduced using UMAP (Uniform Manifold Approximation and Projection) to avoid the effect of high-dimensionality wherein as dimensions increase, data points become more equidistant rendering our distance-based method of clustering unproductive. The low-dimensional embeddings are then clustered using HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), such that each cluster represents a different topic. HDBSCAN is a soft clustering technique that reduces the impact of outliers by increasing the distance between outliers, thus lowering the probability that they impact our top dense clusters. Lastly, the model finds the words to represent each of these clusters or topics by calculating each word's c-TF-IDF (class-based term frequency–inverse document frequency) which is calculated by taking the frequency of the word multiplied with the log of 1 plus the average number of words per topic divided by the frequency of the word across all clusters. This scoring method increases the likelihood that distinct topics with little overlap is generated. The topic modeling output includes the top five words for at most five topics for each vaccine brand and a c-tf-id score was assigned to each word in a given topic [7].

4 RESULTS

4.1 Sentiment Over Time Compared to Baseline for Moderna, AstraZeneca, and Pfizer

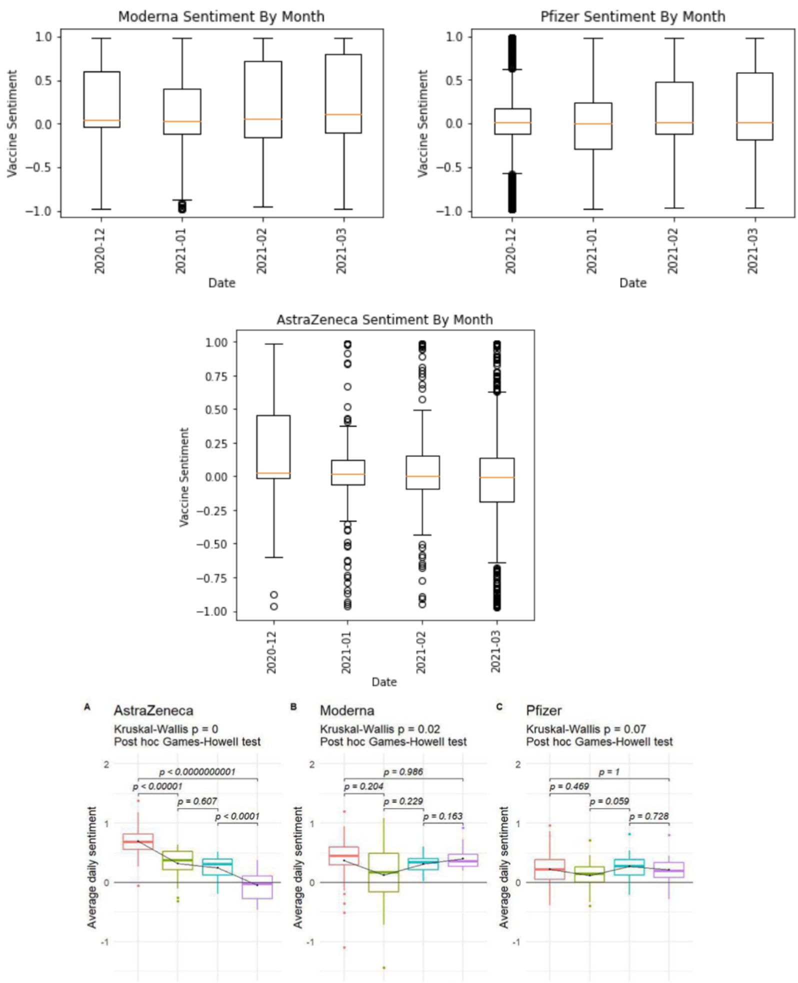
The first aspect of our methodology was to compare our sentiment trends over time output with that of Marcec et al. who performed a similar sentiment analysis study, though using a different sentiment analysis method, for the brands they studied which are AstraZeneca, Pfizer, and Moderna. This comparison between our study and theirs will give us confidence that our results are accurate, and will let us validate the conclusion of Marcec's paper as well.

While our sentiment values differ significantly from the sentiment values calculated in Marcec et al, this is due to the fact that the methodology we used differs from the tool Marcec used and thus the measurements of sentiment are calculated in across different ranges. For each tweet we analyze, BERTsent produces a set of values that represent the probability of the tweet having a positive, negative, or neutral sentiment. For our study, we used those values to get a sentiment score by subtracting the negative sentiment probability from the positive sentiment probability, resulting in a range from -1 to +1. The lexicon based analysis by Marcec handles this differently by assigning values to each word in a tweet and calculating the average sentiment by adding and dividing the sentiments by word and normalizes to a value between -5 and +5. In summary, these values differing is of little consequence because they do not represent comparable values so we will only focus on trends over time.

In Figure 5, we explore the trends of these results. While the values for sentiment may differ, the trends roughly align with those of the Marcec study. Moderna and Pfizer have an average sentiment that are more consistently positive compared to AstraZeneca, the sentiment of which declines over time.

In Figure 5, the box plot for Moderna, Pfizer, and AstraZeneca all have a median average sentiment that remains generally consistent over the four months. Moderna and Pfizer, however, have an upper quartile, specifically the 75th percentile, that seems to be increasing over time especially in the February and March of 2021. This means more of the tweets over time have a positive sentiment analysis with their distribution increasing as well. As for AstraZeneca, the opposite behavior is observed. As seen in its boxplot, the 25th percentile decreases over time, suggesting that more negative tweets were included in the analysis and that a downward trend may be present.

Fig. 5. Comparison of Distributions obtained using BERTsent to those using AFINN Lexicon[13]



4.2 Statistical Test Results

4.2.1 Main Takeaways. The results of our Kruskal-Wallis and Games-Howell tests can be seen in Tables 5, 6, 7, and 8.

First, for each brand we perform the Kruskal-Wallis test where the groups of samples are grouped by month, from December 2020 - March 2021, as seen in Table 5. We find that across the four months, Pfizer and AstraZeneca's monthly distribution of sentiment did change significantly, while Modera, Sputnik V, and Sinopharm did not. Note that Moderna has more samples in our dataset than AstraZeneca, so these results are not due to a small sample size, at least for Moderna.

The Games-Howell test results for the same data divisions are shown in Table 7. This more detailed analysis tells us for which pairs of months were the distributions significantly different. As expected based on the Kruskal-Wallis test, for Moderna there are no pairs of months with significantly different distributions. Interestingly, for Pfizer the only pair of months with signinficantly different distributions is December and March, the first and last month we analyzed. This suggests that the sentiment for Pfizer changed slowly over a long period. For AstraZeneca, most pairs of months have significantly different distributions, suggesting that the sentiment for AstraZeneca is the most volatile of the three brands.

We then perform the tests where the data is grouped by brand within a month. We exclude Sputnik V and Sinopharm from these tests in order to compare with the results with Marcec et al[13]. Table 6 shows the results of the Kruskal-Wallis test on this data division. We see that for December, February, and March, the three brands had significantly different distributions. For February, this was not the case.

The Games-Howell test results for the same data divisions are shown in Table 8. We find that December and March have the most pairwise significant differences, while January and February have none at all. We also find that AstraZeneca and Pfizer were the most frequently different, being significantly different in both December and March.

4.2.2 Comparing to Prior Results. The same statistical tests we used were also performed by Marcec et al. [13], so we compare our results to the results of that paper. The comparison yields interesting results. When the data is divided by month, the results of our Games-Howell tests are very similar to the results in Marcec et al., but our Kruskal-Wallis tests give differing results. When the data is divided by brand, the opposite is true: our Kruskal-Wallis tests give similar results to Marcec et al., but the Games-Howell tests do not.

4.3 Novel Results for Sinopharm and Sputnik

In our analysis of the average sentiment by month calculated by BERTsent, Table 6 displays the values for Sinopharm and Sputnik, two vaccines that were excluded from the previous study. These are more negative than our results for the previous 3 vaccines explored in the baseline study. There is an overall negative sentiment for these vaccines.

In Figure 6 we can see that the Sputnik and Sinopharm vaccine sentiment trends are consistent over the 4 month period. They mostly have a negative sentiment value with a smaller range between its quartiles.

This is further supported by Figure 7a, which displays that aside from a high initial positive sentiment for Sputnik, its sentiment in the given time frame is predominantly negative. The same figure illustrates that Sinopharm seems to have a slightly better sentiment score compared to Sputnik.

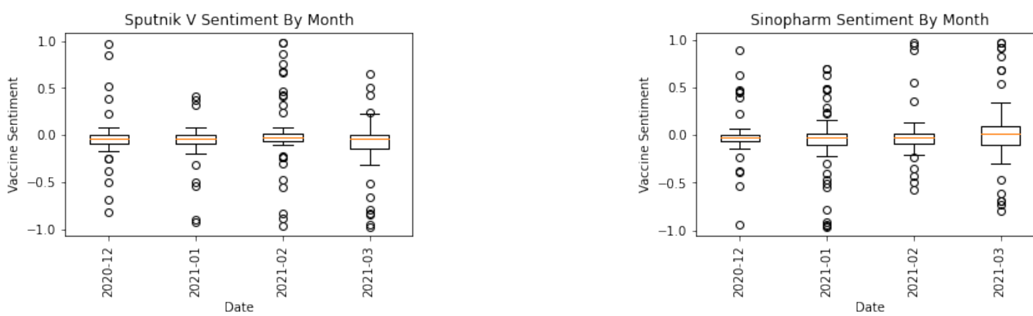
Figure 7b is similar to Figure 7a and shows the general trend of monthly average sentiment except for the brands Pfizer, Moderna, and AstraZeneca. Most of the average sentiment of these three vaccines are positive, with notable negative values on around May 2020 for AstraZeneca and October 2020 and July 2021 for all three vaccines.

4.4 Topic Modeling

Figures 8, 9, and 10 show the results of this topic modelling. Moderna shows more generalized negative terms related to painful side-effects or the growing pandemic, whereas Pfizer's negative words are associated with Trump, the government, the UK, and FDA approval status. AstraZeneca's negative words are related to the UK, scandals in Brazil, and the risk of blood clots as a fatal side

Date	BERTsent Sputnik	BERTsent Sinopharm
December 2020	-0.045195	-0.062139
January 2021	-0.102311	-0.051280
February 2021	0.001502	0.047934
March 2021	-0.105742	0.057342

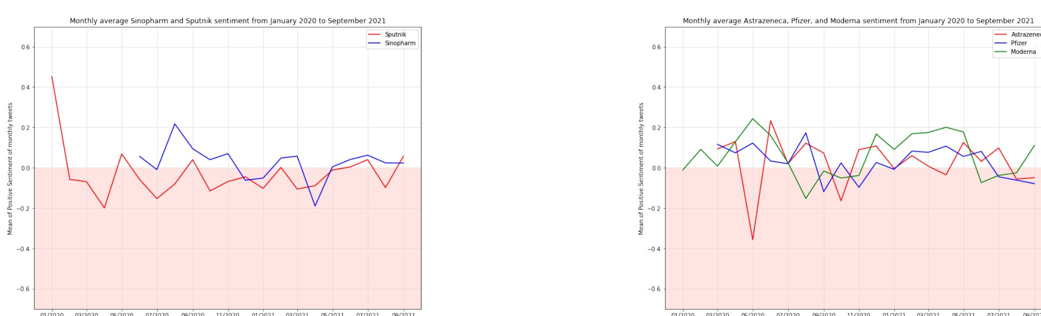
Table 3. BERTsent sentiments for Sputnik and Sinopharm.



(a) Sputnik Trends Over Time

(b) Sinopharm Trends Over Time

Fig. 6. Novel Results for Sputnik V and Sinopharm



(a) Monthly average sentiment of Sinopharm and Sputnik Over Entire Data Range

(b) Monthly average sentiment of Moderna, Pfizer, and AstraZeneca Over Entire Date Range

Fig. 7. Monthly Average Sentiment for Each Brand

effect. No further analysis of these results were conducted for this study, but the results themselves offer more contextual information for the possible drivers of positive and negative sentiment for the different vaccine brands. As for the positive sentiments, for all three brands these were mostly associated with words possibly representing happiness at receiving the vaccine, approval and efficacy rates in different countries and for different ages, as well as news of vaccine of distribution.

Fig. 8. Topic Modelling for Moderna

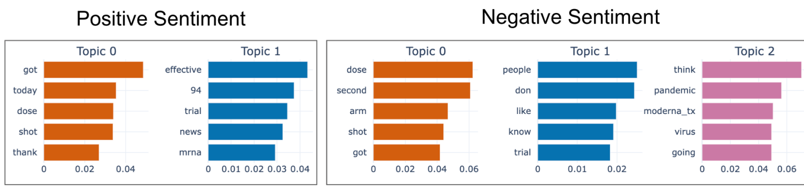


Fig. 9. Topic Modelling for Pfizer

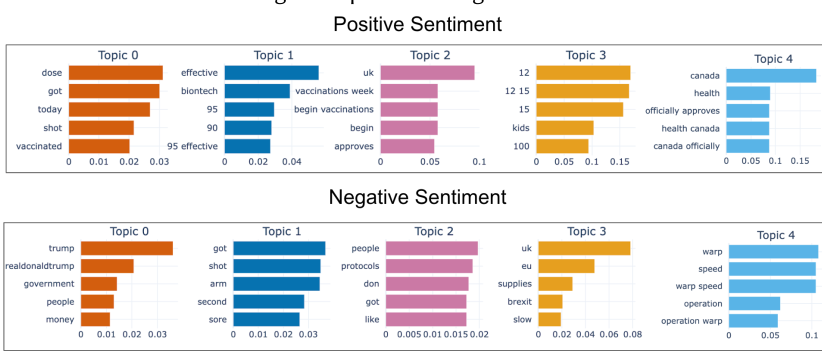
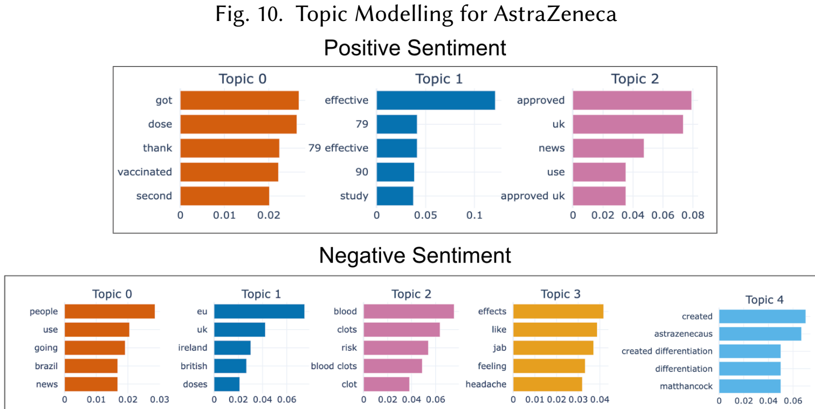


Fig. 10. Topic Modelling for AstraZeneca

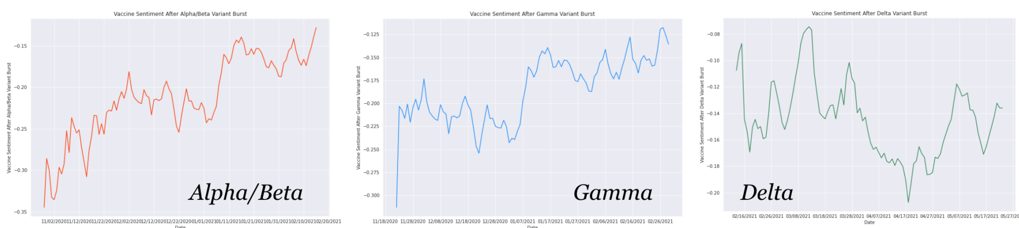


4.5 Vaccine Sentiment Analysis for Covid19 New Variant Burst

Since the outbreak of the epidemic, there are many variants of COVID-19 and we would like to explore the sentiment towards the vaccine at the time when different variants were prevalent to see if there is correlation among the vaccine brands. Data on the outbreak cycles of the different variants were referred from the official website of the World Health Organization.[24] Since our dataset only contains tweets up to March 2021, when the last variant is Delta, We analyze the sentiment during the period of the three variants of Alpha/Beta, Gamma and Delta. Among them, the Alpha and Beta outbreaks were almost at the same time, so they were studied together as the same period.

First, we applied BERTsent to analyze the sentiment regarding all vaccines (regardless of brand) during the three time periods. We used the outbreak time as the midpoint and analyzed the data for the 50 days before and after. Overall, sentiment was on the rise during both the Alpha/Beta and Gamma periods and there were some shocks in the emotional response during the Delta period. The Figure 11 shown the trends in the 3 period. Then we performed a comparative analysis of the

Fig. 11. All brands Vaccine Sentiment Analysis after Different Variants Bursts



different sentiments of the three vaccines(Moderna, Pfizer and AstraZeneca) within each of the three periods. Since the Russia's Sputnik V and the China's Sinopharm had a relative sparse dataset on Twitter to compare with, we split them up and put them in different graphs for comparison. (Figure 13 shows Sputnik and Sinopharm).

We can clearly see that the three vaccines have different sentiment expressions during the outbreak of the three different variants. As in Figure 12, the midpoint is approximately the date of the variant outbreak. In the Alpha/Beta burst time, sentiment about all three vaccines rose slightly. Sentiment of the three vaccines tends to be smooth, stable and similar in the Gamma period. During the Delta period, the sentiment regarding AstraZeneca was significantly lower than the other two vaccines.

Fig. 12. AstraZeneca, Pfizer and Moderna Sentiment Analysis after Different Variants Bursts

Fig. 13. Sputnik and Sinophram Sentiment Analysis after Different Variants Bursts

The changes in sentiment about Sputnik and Sinopharm tend to be almost similar in all three periods as it shows in Figure 13. This may be related to the fact that both vaccines are from non-English speaking countries and BERT can only perform English language analysis.

In other respects, when performing this step of the BERT analysis, it was difficult to analyze their general trends in a straightforward manner due to the high level of variance in the data. So we used the Exponentially Weighted Moving Average(EWMA) calculation in our codes which can show the volatility of the data in a more accurate way.

5 DISCUSSION

While the overall average sentiment was roughly the same by median month by month, AstraZeneca sentiment decreased over time while Pfizer and Moderna stayed consistent, agreeing with the baseline trends [13]. In our novel exploration of Sputnik and Sinopharm sentiment trends using BERTsent, we found that these vaccine sentiments trended slightly lower than the 2 explored in the previous study, and both stay consistent over the 4 month period. The trends in average sentiment were also more compressed for these 2 vaccines – leaving smaller interquartile ranges than those for Moderna, AstraZeneca, and Pfizer

In future exploration we would hope to build our own model based on BERTsent but perhaps trained on some hand labeled tweets. This was beyond the scope of our project timeline. Additionally, we would aim to compare current sentiments (2022) of AstraZeneca, Pfizer, and Moderna to current sentiments of Sinopharm and Sputnik vaccines.

5.1 Limitations

One limitation of our study is the size of our dataset. While the initial dataset is large with over 110,000 Tweets, as seen in Table 2 over 90% of the dataset was not used in our analysis because those Tweets were not related to a specific brand of vaccine. With more time, more Tweets could be downloaded, likely leading to more significant analysis. Similarly, our study focused only on Tweets from December 2020 - March 2021 in order to compare our results to Marcec et al. [13].

A future extension of our work could use our data processing pipeline and statistical analyses to examine trends in vaccine sentiment over a broaded data range.

Another limitation is that BERTsent analyzes the sentiment of a whole Tweet, not the sentiment towards vaccines within that Tweet. For example, the text "I hate people who are anti vaccine" is given a 0.969 probability of negative sentiment from BERTsent [5], but the sentiment toward the vaccine is positive. To address this, a dataset of Tweets would have to be labeled such that the label corresponds to the sentiment toward the vaccine (i.e. the above example would be labeled as positive sentiment), and then fine tuning of BERTweet would have to be done on that dataset. Such a model would give more accurate results for vaccine sentiment, but was outside the scope of this project.

REFERENCES

[1] Shahcheraghi et al. *An overview of vaccine development for COVID-19.* 2020.
[2] Thomas Wolf et al. *HuggingFace's Transformers: State-of-the-art Natural Language Processing.* 2019.
[3] *An easy-to-use Python library for accessing the Twitter API.* https://www.tweepy.org/.
[4] *Coronavirus (COVID-19) Tweets Dataset.* https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset.
[5] *Example of BERTsent incorrect classificaiton.* https://huggingface.co/rabindralamsal/BERTsent?text=I+hate+people+who+are+anti+vaccine.
[6] Santiago González-Carvajal and Eduardo C. Garrido-Merchán. *Comparing BERT against traditional machine learning text classification.* 2021.
[7] Maarten Grootendorst. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure". In: *arXiv preprint arXiv:2203.05794* (2022).
[8] Olga Kolchyna et al. *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination.* 2015.
[9] William H. Kruskal and W. Allen Wallis. "Use of Ranks in One-Criterion Variance Analysis". In: *Journal of the American Statistical Association* 47.260 (1952), pp. 583–621. DOI: 10.1080/01621459.1952.10483441.
[10] Rabindra Lamsal, Aaron Harwood, and Maria Rodriguez Read. *Twitter conversations predict the daily confirmed COVID-19 cases.* 2022.
[11] *langdetect Presesntation.* https://www.slideshare.net/shuyo/language-detection-library-for-java.
[12] Yinhan Liu. *RoBERTa: A Robustly Optimized BERT Pretraining Approach.* 2019.
[13] Robert Marcec and Robert Likic. *Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID- 19 vaccines.* 2021.
[14] Dat Quoc Nguyen. *BERTweet: A pre-trained language model for English Tweets.* 2020.
[15] Finn Årup Nielsen. *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs.* 2011.
[16] Hilary Piedrahita-Valdés. *Vaccine Hesitancy on Social Media: Sentiment Analysis from June 2011 to April 2019.* 2019.
[17] Sara Rosenthal, Noura Farra, and Preslav Nakov. "SemEval-2017 Task 4: Sentiment Analysis in Twitter". In: *Proceedings of the 11th International Workshop on Semantic Evaluation.* SemEval '17. Vancouver, Canada: Association for Computational Linguistics, Aug. 2017.
[18] Florian F Schmitzberger. *Identifying Strategies to Boost COVID-19 Vaccine Acceptance in the United States.* 2022.
[19] *scipy.stats.kruskal.* https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.kruskal.html.

[20] *SemEval Sentiment Analysis on Twitter.* https://alt.qcri.org/semeval2017/task4/.
[21] Nakatani Shuyo. *Language Detection Library for Java.* 2010. URL: http://code.google.com/p/language-detection/.
[22] Raphael Vallat. *Pingouin: statistics in Python.* 2018. DOI: 10.21105/joss.01026.
[23] *WHO Official Vaccine Names.* https://covid19.trackvaccines.org/agency/who/.
[24] *WHO-tracking-SARS-CoV-2-variants.* https://www.who.int/activities/tracking-SARS-CoV-2-variants.
[25] Yukun Zhu et al. "Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books". In: *2015 IEEE International Conference on Computer Vision (ICCV).* 2015, pp. 19–27. DOI: 10.1109/ICCV.2015.11.

6 APPENDIX

6.1 Example Tweets with Sentiment

Tweet ID	Tweet Text	BERTsent output
1359146653803044872	Morning vaccine distribution Mayors briefing with General (Ret'd) Rick Hillier, Chair of the Ontario COVID-19 Vaccine Distribution Task Force. #COVID19Vaccine #COVID19Ontario #onpoli	[[0.00400772 0.9647413 0.03125099]]
1352728656607666184	'I feel great': Bill Gates shares photo of himself getting first dose of COVID-19 vaccine https://t.co/Uvng62DkzH	[[0.00136127 0.14636794 0.8522708]]
1422730837271920640	@[redacted] I just think it's hypocritical to be pro choice and force vaccines on people.	[[0.942726 0.05450472 0.00276926]]

Table 4. Examples of Tweets with sentiment scores from BERTsent

6.2 Statistical Test Results

Manufacturer	Our p_k-value	Marcec et al. [13] p_k-value
Moderna	0.3989	0.02
Pfizer	0.0047	0.07
AstraZeneca	0.0025	0
Sputnik V	0.0685	N/A
Sinopharm	0.2470	N/A

Table 5. Kruskal-Wallis Tests for each brand for the months Dec 2020-Mar 2021

Month	Our p_k-value	Marcec et al. [13] p_k-value
December 2020	0.00002	0
January 2021	0.047	0.01
February 2021	0.282	0.45
March 2021	0.0001	0

Table 6. Kruskal-Wallis tests for each month in Dec 2020-Mar 2021 [13]
We only include Moderna, Pfizer, and AstraZeneca in order to compare our results with results from Marcec et al.

Brand	Month 1	Month 2	Our p_g-value	Marcec et al. p_g-value
AstraZeneca	Dec	Jan	0.033	0.00001
	Jan	Feb	0.891	0.607
	Feb	Mar	0.071	0.0001
	Dec	Mar	0.002	0
Moderna	Dec	Jan	0.200	0.204
	Jan	Feb	0.771	0.229
	Feb	Mar	0.996	0.163
	Dec	Mar	0.977	0.986
Pfizer	Dec	Jan	0.630	0.469
	Jan	Feb	0.063	0.059
	Feb	Mar	0.985	0.728
	Dec	Mar	0.039	1

Table 7. Comparison of Games-Howell tests across months with Marcec et al. [13]

Month	Brand 1	Brand 2	Our p_g-value	Marcec et al. p_g-value
December 2020	AstraZeneca	Moderna	1	0.003
	AstraZeneca	Pfizer	0.014	0
	Moderna	Pfizer	0	0.287
January 2021	AstraZeneca	Moderna	0.832	0.166
	AstraZeneca	Pfizer	0.918	0.007
	Moderna	Pfizer	0.308	1
February 2021	AstraZeneca	Moderna	0.771	0.071
	AstraZeneca	Pfizer	0.989	0.025
	Moderna	Pfizer	0.902	0.595
March 2021	AstraZeneca	Moderna	0.002	0
	AstraZeneca	Pfizer	0.011	0.0001
	Moderna	Pfizer	0.698	0.001

Table 8. Comparison of Games-Howell tests across brands with Marcec et al. [13]