SENTIMENT ANALYSIS OF SOCIAL MEDIA DATA ON COVID-19

Adwita Arora[1], Krish Chopra[1], Divya Chaudhary[2], Ian Gorton[2] and Bijendra Kumar[1]

[1]Netaji Subhas University of Technology, India
[2]Northeastern University, USA

ABSTRACT

The COVID-19 pandemic has forced people to resort to social media to express their thoughts and opinions, which could be analysed further. In this paper, we aim to analyse the impact of the COVID-19 pandemic on social media users by Sentiment analysis of data collected from popular social media platforms, Twitter and Reddit. The textual data is preprocessed and is made fit for proper sentiment analysis using two unsupervised methods, VADER and TextBlob. Special care is taken to translate tweets or comments not in the English language to ensure their proper classification. We perform a comprehensive analysis of the emotions of the users specific to the COVID pandemic along with a time-based analysis of the trends, and a comparison of the performance of both the tools used. Geographical distribution of the sentiments is also done to see how they vary across regional boundaries.

KEYWORDS

Sentiment analysis, social media analysis, natural language processing, covid-19

1. INTRODUCTION

Social Media is everywhere today. It has taken over traditional forms of media and communication and has become the central and focal point of personal, political and cultural discourse. It connects people from all around the world, a fact deemed especially true with the COVID pandemic confining people to their homes leaving them with social media as one of the only forms of communication.

Research shows that there has been a 50-70% increase in Internet usage from the onset of the COVID-19 pandemic and 50% of that time was spent by users engaging in social media (Mark Beech, 2020) **[1]**. This highlights the importance of analysing a user's social media activity to infer how the pandemic has affected the mentality around the world since people suffering from mental health ailments often resort to discussing their feelings on social media platforms like Twitter, Reddit, Facebook etc.

Sentiment Analysis is the field of study about the analysis of people's opinions, emotions and attitudes towards certain topics or entities present in a given text entry (Barreto et al. 2021) **[2]**. The main problem with sentiment analysis on social media text is its informal nature, with most of the data extracted rife with misspellings, poor grammar, improper use of words, Internet slang, emojis, and emoticons.

David C. Wyld et al. (Eds): NLPML, AIAP, SIGL, CRIS, SEC, COSIT, DMA -2023
pp. 13-28, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.130802

Sentiment analysis has applications in multiple fields - assessing consumer behaviour through analysis of product reviews, as well as analysis of the stock market and politics (Yousef et al. 2014) **[3]**.

Social media is another great example of sentiment analysis or opinion mining since in the age of the internet people use it to discuss their opinions and beliefs freely on a variety of topics on different platforms.

Researchers typically use one of two approaches for sentiment analysis - machine learning methods that employ supervised learning on labelled datasets and lexicon-based methods which rely on a predefined dictionary of words and related sentiments (Ribeiro, Filipe N., et al. 2016) **[4]**.

For this paper, we have chosen to perform sentiment analysis using lexicon-based approaches. Specifically, we would be using **VADER (Valence Aware Dictionary for sEntiment Reasoning)**, for its exceptional performance concerning social media analysis (Hutto, C., & Gilbert, E.,2014) **[5]** and **TextBlob**. VADER is a rule-based lexicon especially fine-tuned for micro-blogging contexts, which makes it a golden choice for sentiment analysis on social media data. It has been shown to perform exceptionally well on social media data, which is why it is such an attractive choice (Hutto, C., & Gilbert, E.,2014) **[5]**.

TextBlob is a popular Python module for performing a variety of Natural Language Processing (NLP) tasks, for example, Part-of-speech tagging, Sentiment Analysis, Spelling Correction, Tokenization etc. The TextBlob sentiment analyzer calculates average polarity and subjectivity over each word in the text by using a dictionary consisting of common adjectives and their scores. These scores are taken from SentiWordNet, a lexical resource for sentiment analysis (Baccianella, S., Esuli, A., & Sebastiani, F.,2010) **[6]**.

This paper aims to analyse social media posts, specifically Twitter 'tweets' and Reddit comments, and perform Sentiment Analysis to find out sentiments expressed by users relating to COVID-19. With the users of social media increasing from around the world, special care has been taken to consider language since people feel most comfortable expressing themselves in their native tongue.

Our dataset consists of close to one hundred thousand tweets and comments that have been collected using the Twitter and Reddit APIs. Our analysis has classified these tweets into positive, negative and neutral, with TextBlob also providing subjectivity scores.

2. Related Work

Social media sites have been a great area of interest for Sentiment Analysis and Opinion Mining. Below we present some papers that have carried out a similar analysis to ours.

(Kharde & Sonawane, 2016) **[7]** provides a survey of various techniques used in the sentiment analysis of Twitter data. (Melton et al., 2021) **[8]** focused on public sentiment on Reddit by using sentiment analysis and topic modelling on data from 13 Reddit communities focused on the COVID vaccine. Our analysis not only includes a broader community employing data from both Reddit and Twitter but also focuses on topics beyond the vaccine. Identification of keywords that indicate the different sentiments regarding the pandemic and a time-based analysis of sentiments is central to our study.

(Chen, Zihan & Sokolova, Marina, 2021) **[9]** focused on the r/Depression subreddit for sentiment analysis of COVID-related posts which gives the study almost a singular focus as they look at data extracted from a community especially focused on depressive content. Our work meanwhile looks at r/Coronavirus and r/Health as well as Twitter data with no such community focus. This helps us get data that targets different social media demographics. This research focuses on using topic modelling to extract the central topics of each text and assign them a sentiment based on that, while our study assigns the sentiments straightaway using the VADER and TextBlob tools. (Manguri, Kamaran & Ramadhan, Rebaz & Mohammed Amin, Pshko,2020) **[10]** focuses on Twitter data for sentiment analysis of COVID-related tweets but only uses #COVID-19 and #coronavirus hashtags while our approach uses many more hashtags and keywords, which are given in the analysis section. The research above also has tweets from a limited period of 7 days, while our research focuses on a dataset of Reddit comments ranging from 2020 to 2022, allowing us to understand the trends of sentiments during the pandemic. (Shofiya & Abidi, 2021) **[11]** performed sentiment analysis using a hybrid approach on tweets containing keywords pertaining to social distancing and specific to Canada. The research also focuses on developing a supervised model by labelling the dataset using the SentiStrength tool. At the same time, we are more concerned about the trends in the sentiment of the data we collected.

(Marec & Likic, 2022) **[12]** They used a lexicon-based approach to analyse the sentiment associated with COVID vaccines. This research focuses on sentiments regarding vaccine efficacy as well as vaccine refusal among the surveyed people. Our approach is more focused on the overall sentiment of people on the COVID-19 pandemic over its various facets such as the lockdown, new variants and quarantine. (Barkur, Vibha, Kamath, 2020) **[13]** focuses on a localised dataset consisting of Tweets scraped specifically for Indian users. The research here was an early foray into the sentiment analysis of users at the onset of COVID-19 as well as an exclusive focus on tweets targeted towards the lockdown. Our research focuses on sentiment regarding COVID-19 as a whole, including new variants, lockdowns and quarantine. (Vijay et al., 2020) **[26]** focuses on analysing sentiments of COVID-19 in a state-wise and month-wise manner and studying the variations across geography and time. Our research also does something similar, but it also involves a larger period (from 2019-2022) and focuses on countries instead of individual states. The larger period gives us a better understanding of how the sentiments of various people varied following the changes in the effects of the pandemic across the world. A global analysis instead of a country-focused one gives us a wider view of how the pandemic affected different regions of the world.

3. ANALYSIS

The first phase of our analysis is the collection of data. For this purpose, we are using Twitter and Reddit APIs. Post collection, we translate the posts and tweets into English. The data thus collected is extremely noisy and requires preprocessing. The preprocessed and translated data is finally fed to our sentiment analysis algorithms, TextBlob and VADER. Each step of our analysis is described in detail below:

3.1. Collecting the Data using Social Media Apis

The data was scrapped using the Twitter Python API and PushShift.io API for Reddit comments. This data was selected based on the presence of certain keywords and hashtags:

Twitter Hashtags: #coronavirus, #covid19, #lockdown, #quarantine, #omicron, #deltavariant and #pandemic

Reddit Keywords: Isolation, Coronavirus, Quarantine, Omicron, Delta variant, Covid, Lockdown.

Around 50,000 English and 10,000 multilingual tweets were collected using the Twitter API and the Tweepy module. Around 42,000 Reddit comments were collected in English from the subreddit r/Coronavirus and r/Health. Reddit comments collected varied from 2020-01-21 to 2022-05-19 with Reddit comments earlier than 2020 being removed from the analysis. The PushShift.io API library was used for collecting Reddit Data

3.2. Cleaning and Preprocessing the Data

The data collected from the web is riddled with noise. The following techniques were employed for the cleaning and preprocessing of the collected data:

1. **Removal of hashtags and URLs:** Hashtags and URLs were removed from the corpus by the usage of regular expressions.
2. **Conversion of text to lowercase:** All text was converted from uppercase to lowercase for the purpose of sentiment analysis.
3. **Removal of duplicate rows:** Duplicate rows were removed from the dataset to remove any redundancy in our analysis
4. **Replacement of emojis and emoticons:** Emojis and emoticons were replaced with their corresponding words to aid us in sentiment analysis

```
EMOTICONS = {
    u":-\)":"Happy face or smiley",
    u":\)":"Happy face or smiley",
    u":-\]":"Happy face or smiley",
    u":\]":"Happy face or smiley",
    u":-3":"Happy face smiley",
    u":3":"Happy face smiley",
    u":->":"Happy face smiley",
    u":>":"Happy face smiley",
    u"8-\)":"Happy face smiley",
```

Figure 1: Some of the emoticons

5. **Removal of stop words:** Stopwords such as 'the',' that',' and' and many more were removed by referring to the NLTK stopwords corpus as they don't provide any valuable information for sentiment analysis (Sarica & Lua, 2021) **[14]**.
6. **Tokenization of text:** The text is tokenized by splitting the examples in the text corpus into individual words (word level tokenization) to prepare for the lemmatization of the text (Webster, Jonathan & Kit, Chunyu, 1992) **[15]**.
7. **POS tagging:** POS stands for part of speech tagging (whether a particular token represents a noun, adjective, verb or adverb). This process is also vital to ensure that the lemmatization of text is performed correctly.
8. **Lemmatization of text:** Here lemmatization refers to reducing different forms of a word to a root word. Example: The word running, runs and ran are all converted to run. This helps our sentiment analysis procedure as we deal with common root words only (Balakrishnan, Vimala, Ethel Lloyd-Yemoh, 2014) **[16]**.

For the purposes of tokenization, POS tagging and lemmatization, the Natural Language Toolkit (NLTK) library in Python was used.

After these steps were performed, the text was joined again to get a modified text example for sentiment analysis.

3.3. Translation of Non-English Text

We identify and translate the cleaned text using the Google Translate API. We even identify code-mixed Hindi-English and Spanish-English using CodeSwitch, an NLP tool which can be used for Language identification, POS (part of speech) tagging, named entity recognition and sentiment analysis of code-mixed data **[17]**.

The text is then translated using the Google Translate API for python.

3.4. Location Extraction

We use the spaCy module available in python to perform an entity extraction from the location feature, which is then converted to an exact location using the Geopy module for python.

3.5. Sentiment Analysis

We then feed the cleaned data to VADER, a rule-based, unsupervised Sentiment Analyzer and TextBlob for sentiment classification into positive, negative and neutral sentiments based on the user's emotions. The VADER lexicon returns a polarity score for the data analysed which is used to classify the sentiment of the text. The TextBlob module returns both polarity and subjectivity scores, which help us analyse the sentiment intensity and how subjective the text is. These results are displayed and conclusions are drawn from them in the following sections.

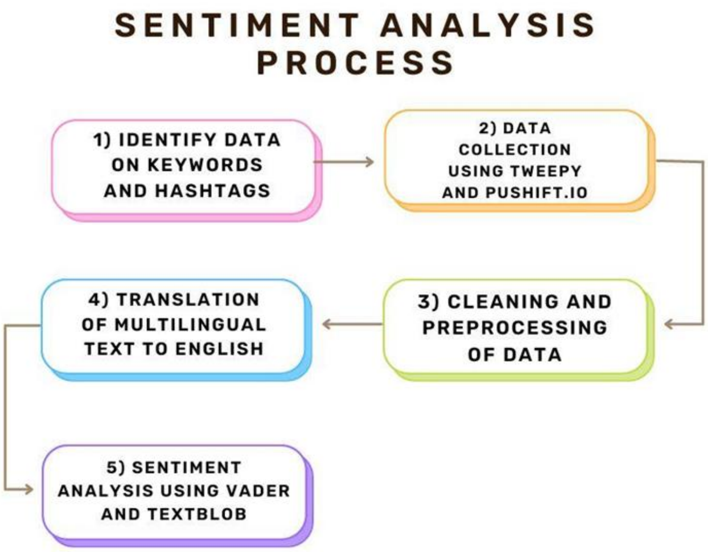

Figure 2: A flow chart of our sentiment analysis process

4. RESULTS

Each tweet, post and comment now has two sentiment scores associated with it, one from VADER and the other from TextBlob. The VADER lexicon assigns each tweet with a compound score - a number between -1 and 1 with -1 indicating extreme negative sentiment and 1 indicating extreme positive sentiment. The examples above 0 were given a Positive label while those scoring below 0 were given a Negative label. The rest of the examples were labelled as neutral.

4.1. Word Clouds

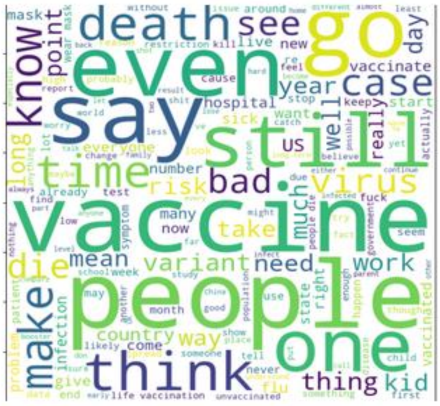

Figure 3: Negative Word Cloud for Reddit

Figure 4: Negative Word Cloud for Twitter

Figure 5: Positive Word Cloud for Reddit

Figure 6: Positive Word Cloud for Twitter

Word clouds are prepared for both negative and positive sentiments for Twitter and Reddit. In the positive word cloud we can see words like ‘good’, ‘great’, ‘wear mask’,’ positivity’, and ‘success’ among others appear a lot in the positively labelled corpus. The Reddit word cloud seems to have more neutral words while the Twitter word cloud has a good amount of positive words.

One interesting word appears in the positive word cloud - positivity. When looking at this word from a normal standpoint it generally has a positive sentiment. However, in terms of the pandemic, Test Positivity Rate (TPR) refers to the number of people who have tested positive for COVID amongst 100 tests (Vaman RS, Valamparampil MJ, Augustine AE, 2020) **[18]**. So contextually, the word “positivity” does not necessarily have a positive sentiment. This highlights the importance of context in performing sentiment analysis.

A similar discussion applies to the word “positive” which in the COVID context of “testing positive” does not have a positive sentiment associated with it.

Words in the Reddit dataset seem to share a lot with the words in the positive word cloud and negative word cloud. This may be because the most frequently occurring words (disregarding the stopwords) may be neutral, while the actual words that influence the polarity of the comment may be present only in small amounts in the dataset.

The negative WordCloud shows that ‘death’, ‘risk’, ‘minister’ and ‘worry’ are used commonly in tweets about the COVID-19 pandemic. It is visible from the plot above that words like ‘death’, ‘worry’ and ‘crisis’ have a higher frequency of occurrence within text classified with a negative sentiment. A similar result was observed using TextBlob analysis as well. Some other words like ‘seriously’, ‘attention’ and ‘disease’ appeared frequently in negative tweets along with ‘suffer’, ‘sadly’ and ‘prevention’ also present in significant frequencies. Surprisingly, words like ‘death’, ‘risk’ and ‘worry’ appeared in tweets labelled positive by TextBlob, revealing a possible inaccuracy. This inaccuracy is further seen in the subsequent analysis done as shown in the next few sections.

4.2. Subjectivity v/s Polarity

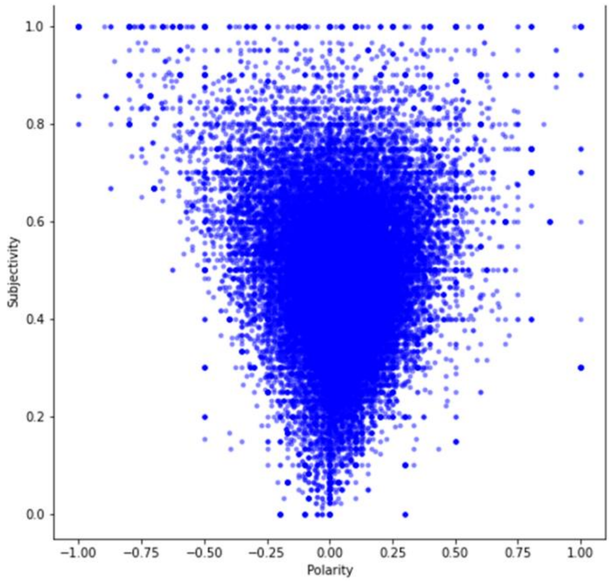

Figure 7: Scatter Plot for Reddit

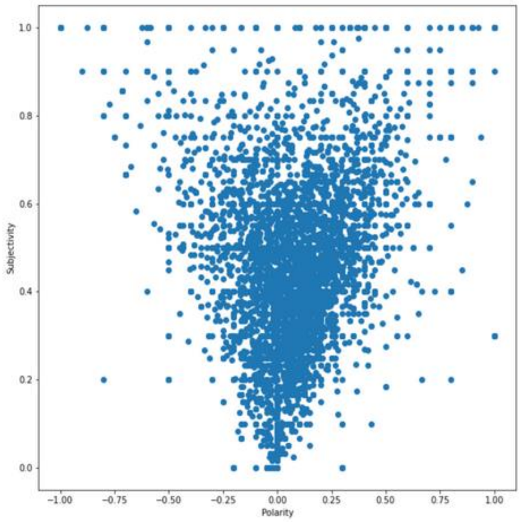

Figure 8: Scatter Plot for Twitter

The scatter plot for polarity vs subjectivity given above shows that most comments lie in a close range of about -0.4 to 0.4 for polarity. Highly polar comments (that are either highly negative or highly positive) are also more subjective as they are more expressive of a user's opinion, while comments more neutral in nature tend to be greatly concentrated in the lower regions of subjectivity since factual comments which are more objective are neutral in a tone most of the time. Highly polar comments are as expected rare in the dataset and as our analysis on the Time based variation of polarity demonstrates later, the mean tendency of polarity is close to +0.1 and -0.1.

The scatter plots across both datasets appear to behave the same, with no significant differences between the shapes the distributions take.

4.3. Time-based Analysis of Polarity

4.3.1. Time Series for TextBlob

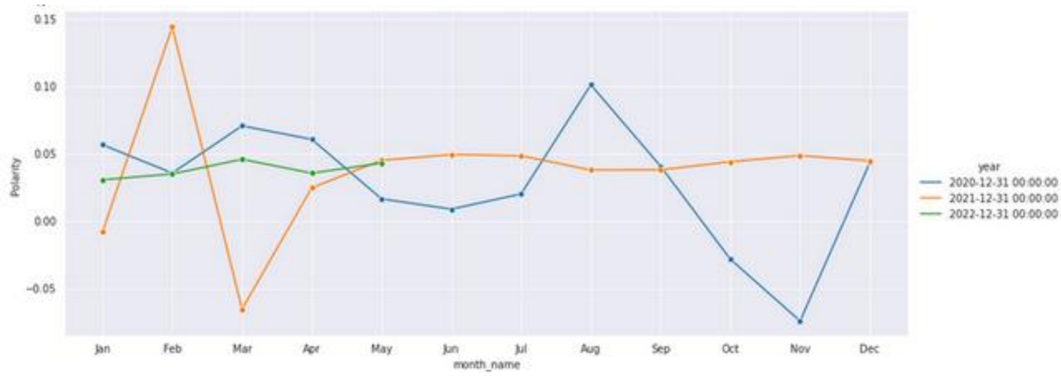

Figure 9: Time Series for TextBlob

Given above is a time-based plot for polarity scores extracted from TextBlob for 12 months, with 3 such plots for the 3 years we have under consideration.
The dataset begins on 21-01-2020 and ends on 2022-05-19.

Some interesting points to note:

1. February 2020 being highly positive seems like an inaccuracy. This period was the onset of the pandemic in many countries and it seems weird that this period would see an average sentiment of positivity (WHO Director-General's remarks at the media briefing on 2019-nCoV on 11 February 2020) **[19]**.

2. August sees a climb in positive sentiment as the relaxation of lockdown begins in many countries around the world and the fear of the virus's destructive capability was being superseded by a growing feeling of it being nothing more than a simple seasonal flu.

3. November 2020 sees a huge dip in sentiment, though considering that the average sentiment score is still -0.05, the actual sentiment isn't highly negative. The reason for this sudden dip seems to be just a random fluctuation.

4. The start of 2021 has a sharp increase in positive sentiment regarding COVID. This may have happened because of the relaxation of lockdown in most countries by that point (Rathod S et al., 2021) **[20]**.

5. This is followed by a sharp decrease in March and April. This might be due to the onset of the 2nd wave of COVID in many countries, which hit most countries harder than the first wave (Choudhary OP et al., 2021)**[21]**.

6. The period after that has been very stable with the sentiment showing little variability and remaining firmly between 0.05 and 0. At this point, the COVID-19 pandemic was firmly in the realm of being considered the new normal now and discussion about it ceased to be as prevalent as before.

More importantly, the mean polarity for each month remains firmly in a region of 0.15 to -0.05, meaning that the Reddit dataset collected is more likely to be neutral or positive than negative. Even when it is positive, there are no periods of it being overwhelmingly positive, as well as no periods where sentiments have been significantly negative.

4.3.2. Time Series For VADER

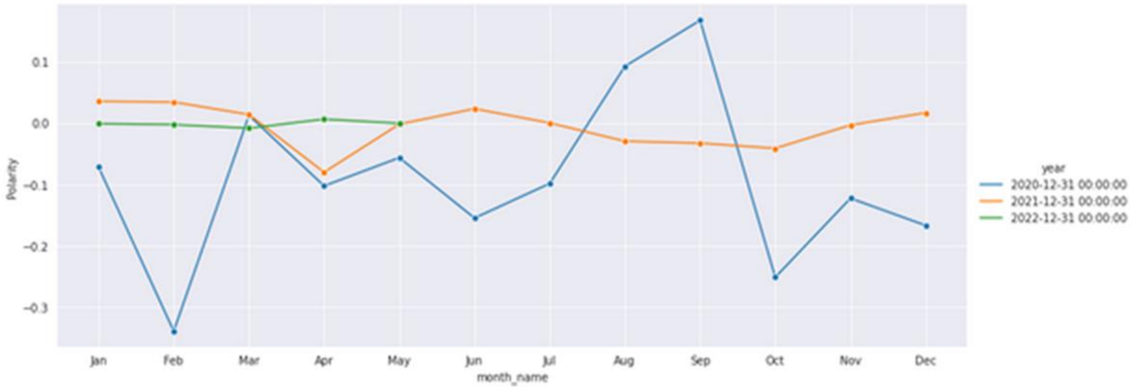

Figure 10: Time Series for VADER

VADER begins with it having a very up-down 2020, before settling into an almost constant polarity from 2021 onwards. It also ranges between 0.1 and -0.3, which shows that VADER has a higher variation in sentiment scores than TextBlob. Another point to note is how VADER and TextBlob seem to have opposite tendencies, TextBlob is more likely to classify comments as positive while VADER is more likely to treat those comments as negative. The reason for this variation is something worth investigating.

Some notable things:

1. Feb 2020 is highly negative sentiment-wise, which was not the case for TextBlob. This makes sense though as this was the period where the COVID pandemic was truly setting in for the first time in most countries and it was beginning to be acknowledged as a global crisis. This further helps the claim that TextBlob is prone to classify text as positive.

2. August and September climbed up in positive sentiment due to the relaxation of lockdown norms in many countries during this period. This is something that both on TextBlob and VADER seem to agree upon.

3. The sentiments at the end of 2020 ended on a negative note while they picked themselves up right up in a more positive direction at the onset of 2021. The new year brought up new hope that the worst of the pandemic was behind us now.

4. Another dip is seen in sentiments during April of 2021. Again this coincides with the second wave of COVID-19, which as seen earlier is something that TextBlob also agrees upon with VADER.

5. After the dip in April 2021, most comments hover at the neutral region for the rest of 2021 and carry into 2022.

We see that VADER tends to classify comments more as negative, while TextBlob varies mostly between positive and neutral. The Feb 2020 rise in sentiment for TextBlob feels inconsistent with the actual sentiment that was prevalent during that period, while VADER seems to give a more accurate negative dip. Except for this anomaly, all major fluctuations in the two sentiment analyzer's results seem to match each other, even if their magnitudes may be different.

4.4. Distribution of Tweets and Comments Across Different Sentiments

4.4.1. Sentiment Distribution for Twitter

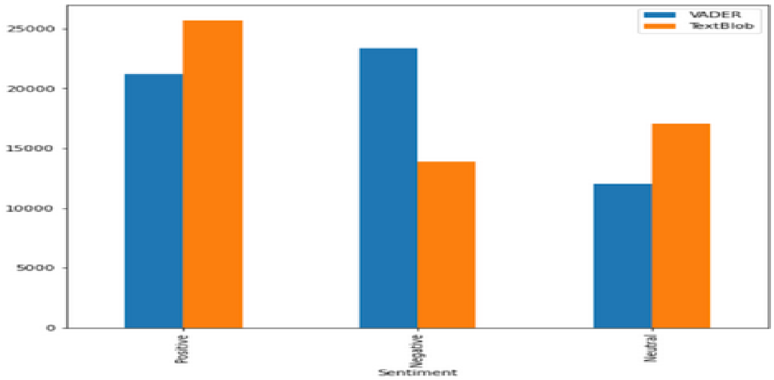

Figure 11: Sentiment Distribution for Twitter

VADER predicts tweets to be more negative than positive or neutral regarding COVID, while TextBlob regards them in a more positive light. This lines up with the inaccuracy we saw above of words like 'death,' worry' and 'risk' appearing in tweets labelled positive, which implies that TextBlob is more likely to classify tweets as positive. Researchers in the past have also found inaccuracies with TextBlob (Aryal, Ranjan Raj and Bhattarai, Ankit, 2021) **[27]** (Melton, Chad A., et al., 2021) **[28]**,

4.4.2. Sentiment Distribution for Reddit

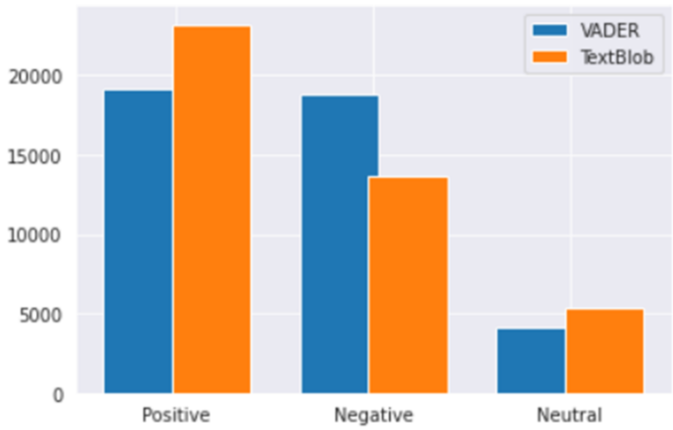

Figure 12: Sentiment Distribution for Reddit

In the Reddit dataset, VADER doesn't have a large difference between positive and negative comments, but TextBlob again gives more positive results than negative, lining up with what we saw with the Twitter dataset and the performance of TextBlob in it. It is worth investigating

4.5. Location Analysis

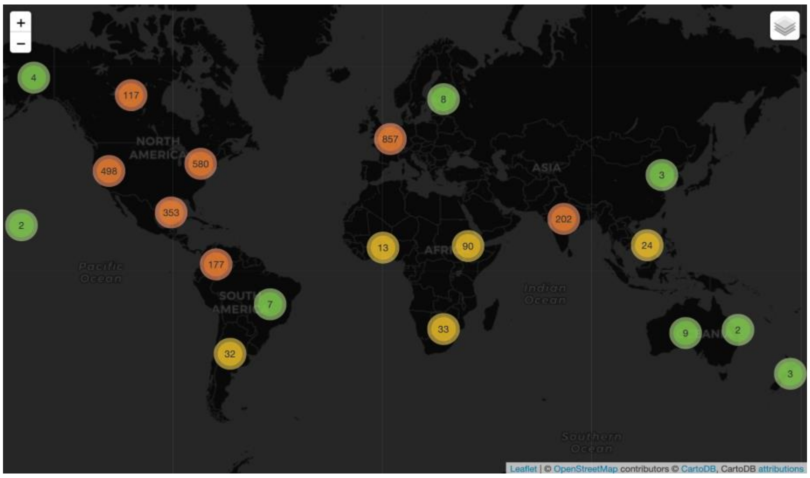

Figure 13: Locations of Tweets Collected

We used NER (Named Entity Recognition) to extract locations from the features collected. To ensure balance among regions, we considered only the 10000 multilingual tweets while extracting locations. While these locations are spread across the globe, eight major hot spots are identified - United States, Canada, United Kingdom, France, India, Germany, Mexico and Venezuela.

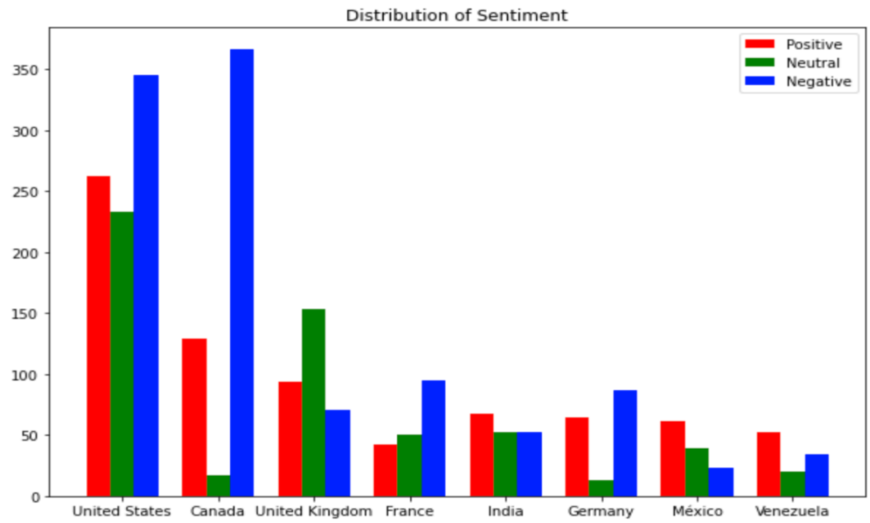

Figure 14: Distribution of sentiment for different countries

The sentiment distribution in the United States, Canada, France and Germany show a higher number of negative sentiments as compared to positive sentiments, while the United Kingdom, India, Mexico and Venezuela show higher positive sentiments.

We see a higher number of negative sentiments in developed countries probably due to the growing concern over the different variants of COVID - particularly the Delta variant and more recently the Omicron variant and the immunity offered by vaccines currently being administered over these variants (Salsabil Islam, Towhidul Islam, Md. Rabiul Islam, 2022) **[22]**. This is also supplemented by the general spread of misinformation regarding the vaccines and the pandemic as a whole with several forums online being dedicated to spreading misleading articles and videos (Ahmed, Wasim, et al., 2020) **[29],** (Kouzy, Ramez, et al., 2020) **[30]** and (Roozenbeek, Jon, et al., 2020) **[31]**. Studies have also shown the presence of hesitancy and scepticism towards vaccines, the concern about side effects and the spread of misinformation regarding the pandemic which could add to the negative sentiments (Lanyi et al., 2022) **[23]** and (Imhoff & Lamberty, 2020) **[24]**.

The positive collection of tweets in developing countries like India could be made in response to the arrival of vaccines combating the Coronavirus as well as the ease of lockdowns, reduction of COVID-related restrictions and the falling number of positive cases over the past few months from when the Twitter data has been collected. India especially shows a higher positive sentiment which can be regarded as the successful inoculation drive across the country as well as the reopening of public places and a general return to the state of normalcy.

The overwhelmingly negative sentiment in Canada could be in response to the government's strict COVID mandates, including full vaccination for truck drivers, which led to a protest convoy and blockades (Mark Scott, 2022)**[32]** and (Catharine Tunney, 2022)**[33]**.

5. CONCLUSIONS

The problem of figuring out the emotion a user is trying to express on Social media is a vital one. The variety of emotions accessed on various platforms vary in accordance with real-life events and COVID-19 was probably the biggest pandemic in recent memory. Thus our task was to figure out how the pandemic affected the sentiments of people around the world. Tweets in multiple languages were seen in the dataset, one of the most prominent being Spanish and Hindi, and hence special care was taken to ensure they were translated and treated carefully while performing sentiment analysis

VADER sentiment analysis for both the Twitter and Reddit datasets shows that tweets had a good ratio of positive and negative tweets and comments, with negative tweets having a slight edge in the Twitter dataset while being almost equal to the count of positive sentiment comments in the Reddit dataset. Neutral sentiment tweets and comments comprise a lower fraction.

TextBlob analysis on the other hand shows an overwhelming number of positive sentiment tweets and comments with a slight increase in the number of neutral tweets as well in both datasets. We detected a possible inaccuracy in the classification done by TextBlob as WordCloud analysis revealed that words like 'death', 'risk' and 'worry' appeared in tweets labelled positive by TextBlob. This was further shown in the time-centric analysis of the Reddit dataset, as February 2020 had a mean positive sentiment using TextBlob while a negative sentiment using VADER. The authors feel that it is more likely that the period had an overall negative sentiment due to the fact that this period saw the start of the pandemic in many parts of the world. The inaccuracies with using TextBlob for sentiment analysis of social media data have also been observed by researchers in the past and may be further looked into in future works.

A time-based analysis of the dataset was done to present the time-based variation of the sentiments for each month in 16 months starting from January 2020 to May 2022. The results show that there were large fluctuations in sentiments in 2020. 2021 and the start of 2022 have had a stable sentiment score.

The polarity v/s subjectivity plot shows that most comments and tweets are rarely very highly positive or negative, and when they are they tend to be highly subjective in nature. Both the

Twitter and Reddit dataset show the same kind of trend while comparing subjectivity and polarity.

6. FUTURE WORK

The importance of sentiment analysis of social media sometimes necessitates that the methods we use for them should give accurate results (Neri et al., 2012) **[25]**. The volume of social media data means that checking for inaccuracies by reviewing social media data one by one is almost impossible to do.

Along with this, datasets collected from the Internet inherently come without labels and hence our paper deals exclusively with unsupervised methods.

In the future, we would like to approach this problem using supervised learning models. The reasons for this are twofold: With a labelled dataset we can see how accurate our predictions for the sentiments are and hence also see which kind of text makes sentiment analyzers scratch their heads. Secondly, lexicon-based analyzers are general-purpose tools in nature. Even with VADER being a tool geared explicitly towards accurate sentiment analysis of social media data, a machine learning approach can be trained on a specific dataset to make accurate predictions on that kind of data. We would also like to perform analysis keeping in mind the context of the situation since a direct mapping of sentiments to data cannot be possible without contextual sensitivity.

REFERENCES

[1] Beech, M. (2020, March 26). *Covid-19 pushes up internet use 70% and streaming more than 12%, first figures reveal*. Forbes. Retrieved January 15, 2023, from https://www.forbes.com/sites/markbeech/2020/03/25/covid-19-pushes-up-internet-use-70-streaming-more-than-12-first-figures-reveal/?sh=30dac3933104

[2] Barreto, S., Moura, R., Carvalho, J., Paes, A., & Plastino, A. (2021). Sentiment analysis in tweets: an assessment study from classical to modern text representation models. *arXiv*. https://doi.org/10.48550/arXiv.2105.14373

[3] Hassan Yousef, Ahmed & Medhat, Walaa & Mohamed, Hoda. (2014). Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Engineering Journal. 5. 10.1016/j.asej.2014.04.011.

[4] Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*(1), 1-29. https://doi.org/10.1140/epjds/s13688-016-0085-1

[5] Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[6] SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining (Baccianella et al., LREC 2010)

[7] Kharde, V. A., & Sonawane, P. S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *arXiv*. https://doi.org/10.5120/ijca2016908625

[8] Melton, C. A., Olusanya, O. A., & Ammar, N. (2021). Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. *arXiv*. https://doi.org/10.1016/j.jiph.2021.08.010

[9] Chen, Zihan & Sokolova, Marina. (2021). Sentiment Analysis of the COVID-related r/Depression Posts.

[10] Manguri, Kamaran & Ramadhan, Rebaz & Mohammed Amin, Pshko. (2020). Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks. Kurdistan Journal of Applied Research. 54-65. 10.24017/covid.8.

[11] Shofiya C, Abidi S. Sentiment Analysis on COVID-19-Related Social Distancing in Canada Using Twitter Data. Int J Environ Res Public Health. 2021 Jun 3;18(11):5993. doi: 10.3390/ijerph18115993. PMID: 34204907; PMCID: PMC8199732.

[12] Marcec R, Likic R, Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines Postgraduate Medical Journal 2022; 98:544-550

[13] Barkur G, Vibha, Kamath GB. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. Asian J Psychiatr. 2020 Jun;51:102089. doi: 10.1016/j.ajp.2020.102089. Epub 2020 Apr 12. PMID: 32305035; PMCID: PMC7152888.

[14] Sarica S, Luo J (2021) Stopwords in technical language processing. PLOS ONE 16(8): e0254937. https://doi.org/10.1371/journal.pone.0254937

[15] Webster, Jonathan & Kit, Chunyu. (1992). Tokenization as the initial phase in NLP. 1106-1110. 10.3115/992424.992434.

[16] Balakrishnan, Vimala and Ethel Lloyd-Yemoh. “Stemming and lemmatization: A comparison of retrieval performances.” (2014).

[17] https://codeswitch.readthedocs.io/en/latest/#sentiment-analysis

[18] Vaman RS, Valamparampil MJ, Augustine AE. Using Test Positivity Rate (TPR) as an Indicator for Strategic Action in COVID-19: A Situational Analysis in Kerala, India. Indian Journal of Clinical Medicine. 2020;10(1-2):31-35. doi:10.1177/26339447211054234

[19] World Health Organization. (n.d.). *Who director-general's remarks at the media briefing on 2019-ncov on 11 February 2020*. World Health Organization. Retrieved January 15, 2023, from https://www.who.int/director-general/speeches/detail/who-director-general-s-remarks-at-the-media-briefing-on-2019-ncov-on-11-february-2020

[20] Rathod S, Pallikadavath S, Graves E, Rahman MM, Brooks A, Soomro MG, Rathod P, Phiri P. Impact of lockdown relaxation and implementation of the face-covering policy on mental health: A United Kingdom COVID-19 study. World J Psychiatry. 2021 Dec 19;11(12):1346-1365. doi: 10.5498/wjp.v11.i12.1346. PMID: 35070782; PMCID: PMC8717029.

[21] Choudhary OP, Priyanka, Singh I, Rodriguez-Morales AJ. Second wave of COVID-19 in India: Dissection of the causes and lessons learnt. Travel Med Infect Dis. 2021 Sep-Oct;43:102126. doi: 10.1016/j.tmaid.2021.102126. Epub 2021 Jun 16. PMID: 34144178; PMCID: PMC8214078.

[22] Islam S, Islam T, Islam MR. New Coronavirus Variants are Creating More Challenges to the Global Healthcare System: A Brief Report on the Current Knowledge. Clinical Pathology. January 2022. doi:10.1177/2632010X221075584

[23] Lanyi K, Green R, Craig D, Marshall C. COVID-19 Vaccine Hesitancy: Analysing Twitter to Identify Barriers to Vaccination in a Low Uptake Region of the UK. Front Digit Health. 2022 Jan 24;3:804855. doi: 10.3389/fdgth.2021.804855. PMID: 35141699; PMCID: PMC8818664.

[24] Imhoff R, Lamberty P. A Bioweapon or a Hoax? The Link Between Distinct Conspiracy Beliefs About the Coronavirus Disease (COVID-19) Outbreak and Pandemic Behavior. Social Psychological and Personality Science. 2020;11(8):1110-1118. doi:10.1177/1948550620934692

[25] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros and T. By, "Sentiment Analysis on Social Media," 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012, pp. 919-926, doi: 10.1109/ASONAM.2012.164.

[26] T. Vijay, A. Chawla, B. Dhanka and P. Karmakar, "Sentiment Analysis on COVID-19 Twitter Data," 2020 5th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2020, pp. 1-7, doi: 10.1109/ICRAIE51050.2020.9358301.

[27] Aryal, Ranjan Raj, and Ankit Bhattarai. "Sentiment Analysis on Covid-19 Vaccination Tweets using Naïve Bayes and LSTM." Advances in Engineering and Technology: An International Journal 1.1 (2021): 57-70.

[28] Melton, Chad A., et al. "Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence." Journal of Infection and Public Health 14.10 (2021): 1505-1512.

[29] Ahmed, Wasim, et al. "COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data." Journal of medical internet research 22.5 (2020): e19458.

[30] Kouzy, Ramez, et al. "Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on Twitter." Cureus 12.3 (2020).

[31] Roozenbeek, Jon, et al. "Susceptibility to misinformation about COVID-19 around the world." Royal Society open science 7.10 (2020): 201199.

[32] Scott, M. (n.d.). *Ottawa Truckers' convoy galvanizes far-right worldwide*. POLITICO. Retrieved January 15, 2023, from https://www.politico.com/news/2022/02/06/ottawa-truckers-convoy-galvanizes-far-right-worldwide-00006080

[33] Tunney, C. (2022, February 15). *Federal Government invokes emergencies act for first time ever in response to protests, blockades | CBC News*. CBCnews. Retrieved January 15, 2023, from https://www.cbc.ca/news/politics/trudeau-premiers-cabinet-1.6350734