Sentiment Analysis on USA vs. New Zealand on Health and Safety Mandates During Early Stages of COVID -19 Pandemic

Joshua DALES[a], Farhaan MIRZA[a], and Amr ADEL[a & b]
[a] *Auckland University of Technology*, *Auckland New Zealand*
[b] *Yoobee Colleges, School of Information Technology, Auckland, New Zealand*

Abstract. The Coronavirus pandemic has surprised the world and social media was extremely used to express frustrations and development of the cases found. Social media tools, such as Twitter, show a comparable impact with the number of tweets related to COVID -19 indicating remarkable development in a limited ability to focus time. The purpose of this paper is to investigate the impact of Coronavirus on the United States of America (USA) and New Zealand (NZ), and how that is reflected in a sentiment analysis through the examination of American and New Zealand tweets. We have gathered tweets from a March 2020 – August 2020 and used sentiment extraction on the tweets. The major finding of this sentiment extraction is the fact that the overall average sentiment over the 5-month period stayed in a negative range in the USA and NZ. This paper aims to analyze these trends, identify patterns, and determine whether these trends were caused by the COVID -19 pandemic or outside sources. One trend that was analyzed was the spike of COVID -19 results in relation to the number of protests occurring in the USA.

Keywords. Sentiment Analysis, COVID -19, Coronavirus, Data Request, Twitter.

1. Introduction

COVID -19 is a vital area to investigate in this current time. This will be analyzed through a Sentiment Analysis of 10,000 tweets during the period of March to July. Sentiment analysis is the process of organizing and grouping large amounts of text together based on certain facts contained in the text. The text is categorized into groups based on feelings such as positive, neutral, and negative. By grouping the text into several categories, it can provide a greater understanding of the material. The topic area for this research is «Sentiment Analysis of COVID -19 and how the understanding of it changed general sentiment towards the virus». The motivation of this research focuses on the sentiment analysis on twitter data because it gives us the option to see how sentiments towards COVID -19 grew or reduced over a period. This research has the potential to aid future researchers in understanding how normal people will react to a pandemic. The findings of this research are useful for assessing a specific research gap. The research gap is a social comparison between America and New Zealand. Governmental branches in America have made several attempts to suggest a health and safety mandate that were created to reduce the spread of COVID -19. Section 3 focuses on the research methodologies and how the data was obtained and validated.

Section 4 goes into the research findings and identifies themes, patterns, and areas for analyzing. Section 5 is a discussion part focusing on possible reasons for why the USA has such high numbers of COVID -19 cases. Furthermore, it analyses why New Zealand does not have such high cases. Finally, Section 5 goes into a comparison and identifies what New Zealand did that the USA might not have been able to do. Section 6 focuses on the Conclusion of this research.

2. Literature Review

Previous studies have used sentiment analysis & social media to discuss how public has reacted to public health issues. For example, one existing solution that has attempted to use data mining techniques to understand the sentiment of collective outspoken citizens is the research done by Darliansyah et al in their report of Long-Term trends in public sentiment in Indian demonetization policy. It is unlikely that the results found in analyzing data contributing to COVID -19 will be as clear cut as Demonetization trends, but through this we can begin to understand how large audiences from a variety of countries can react to drastic change. That report analyzed the sentiments of 41,724 tweets by Indian citizens through a period of 9 months; what they found was that only 33% of the data was positive with another 32% being negative and only 34% of the data being neutral to the changes.

Another study reported on similar areas of study and conducting sentiment analysis to understand greater audiences. A report done by Barkur, et al. (2020) [10]. Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India, shows evidence that in India, that many citizens responded positively to the government lockdowns, roughly 25000 tweets show a positive lean-to going into a lockdown. What this report shows is that depending on the country-of-origin reactions to a countrywide lockdown will differ greatly. The primary negative sentiment towards the imposed lockdown was one of fear, a sentiment that was shared by 10,000 other citizens.

A third report conducted by Matosevic and Bevanda (2020) [9] that studied a Sentiment analysis of tweets about COVID -19 disease during the pandemic. What this report shows is that in countries that handled COVID -19 in safe way their citizens tweet emotion rating was "Trust" whereas in countries that were struggling with COVID -19 their populations tweets tended to range on the "fear" side of the emotion spectrum. This report studied outgoing tweets from USA, UK, Spain, Italy, Sweden and Germany, this report made a distinction between politicians and the public and determined that the public and politicians shared similar emotions when tweeting about COVID -19.

Through understanding the report by Davison (2017) [11] we can understand how the public's opinion shifts. Through this understanding, we can identify and analyze how data opinions shift over time. This is due to how more evidence or facts become relevant. Based on this report, we can theorize that in the beginning of the COVID -19 pandemic the public was not affected, nor did they think it was a problem. But as COVID -19 started to spread globally and more information came out public opinion would shift, and they would become more interested in stopping the spread. It can be hypothesized that such sentiments would be positive in the face of policies that would be implemented to slow the spread of COVID -19. Additionally, negative sentiment would be based on those of the public who either do not believe COVID -19 is a problem or by others who are not happy with the policies and what change they would bring.

Through the reviewed literature this paper was able to recognize and understand how sentiment analysis can aid in understanding how a countries sentiment towards a specific topic can shift as new information is introduced. Evidence provided by Davison (2017) [11] reports that opinions shift over time as more knowledge and facts are provided. This trend is also observed in Barkur, et al. (2020) [10] where a sentiment analysis was conducted in regards to the opinions of the citizens of India where more information was provided the sentiment shown was seen in a positive light. Furthermore, this trend is seen in the report by Matosevic and Bevanda (2020) [9] who did a sentiment analysis on USA, UK, Spain, Italy, Sweden, and Germany. This report shows that where the pandemic was better handled there were more tweets that gave sentiment results of "Trust" and "Anticipation" whereas Spain a country that is not handling the pandemic well has a lot of tweets the express "fear" more prevalently than any other emotion. Finally, what can be noted from these scholars is that the less the public knows about an event can cause misinformation to spread.

This information is helpful in the way that it will be able to reflect on the differences between America's response to COVID -19 and compare it to New Zealand's response to COVID -19 during the 5-month period of March-July of 2020. Additionally, this research will assist in understanding the benefits of following a strict health and safety mandate when facing a global pandemic versus not following any health and safety suggestions.

This research aims to investigate a research gap in sentiment analysis of countries afflicted with COVID -19. By comparing the difference of following government mandates versus not following government mandates. This gap is relevant due to the overwhelming numbers of deaths and cases that have appeared in America. This paper aims to understand how the American people react to government enforced mandates and analyze any outstanding circumstances that occurred throughout the 5-month period. Following that, this paper will analyze why sentiments toward COVID -19 raised or fell during this period.

3. Methodology

In this study, twitter data was captured using a python script called "GetOldTweets3" created by Mottl (2019) [8] which allows for the capture of historical data. After the tweets were captured, that data was run through the tool "Rapid Miner" and a sentiment analysis was applied to the extracted tweets. Due to budget constraints only 10,000 tweets over a period of 5-months were analyzed. The tweets are categorized as very positive, positive, neutral, negative, and very negative tweets.

3.1. Data Acquisition

The areas that this study will research are identifying trends in data mined from twitter based on hot topic key words. By analyzing data pulled from twitter based on these keywords, we can notice trends and identify certain statistics. From these statistics this study, we will investigate what caused certain sentiments to form in the following months. The months that were selected were March 2020 to July 2020, was done based on the number of tweets that could be analyzed and categorized by the Rapid Miner. Each month was restricted to only pull 2000 tweets giving us a total of 10,000 tweets over a 5-month period.

The data for this investigation was gathered through a python script that is known as "GetOldTweets3", due to the way that Twitter gathering historical twitter data is incredibly difficult. "GetOldTweets3" worked around these problems and allowed for large twitter data extraction. The code has some limitations, but it allows for the large amount of twitter data to be extracted. The process for twitter data extraction is demonstrated in Figure 1.

Figure 1. Data Request Process

To fully analyze the data, the raw twitter data had to go through a "cleaning" process. The main point of this process was to get all the data into one excel file with the correct date attached to the correct twitter data. When all the data was in one file, the sentiment extraction process could begin. The sentiment analysis program analyzed the data and outputted what sentiment level each tweet carried.

3.2. Sentiment Analysis

Following the cleaning process, RapidMiner was used to analyze the data. The data was analyzed through a specific add-on known as the "Operators Toolbox" which splits the incoming data into positive and negative results. From these results we can identify several key factors such as, '0' being mainly neutral remarks, remarks that were ranked between 1 to 10 and -1 to - 10 being positive and negative.
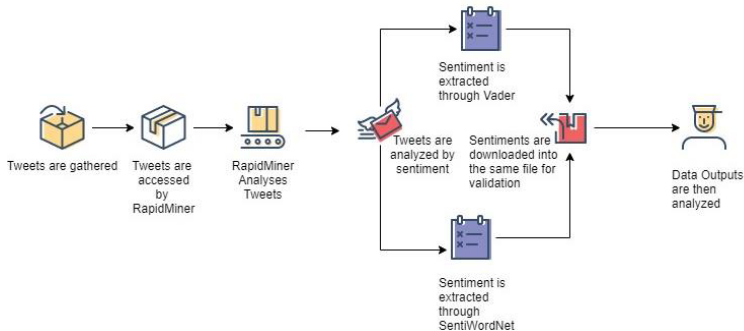

Figure 2. Proposed Architecture of Sentiment Extraction and Validation

3.3. Data Validation

Data validation is the process of ensuring that the data that is being used for analytics is accurate as mentioned by SafeSoftware (2020). Furthermore, data validation process as shown in Figure 2 is a critical move in the data workflow as it ensures that the data that is being used is accurate. The data went through this process firstly through a sentiment analysis tool known as Operator's Toolbox which has access to three sentiment analysis models. The data went through the data analysis model Vader first. The Vader model aligned text based on their content and gave a sentiment score based on the content contained in the text. To validate Vader's results this report used a second sentiment analysis tool from the Operator's Toolbox known as SentiWordNet which gave out similar results with some deviation. By validating the twitter sentiment data, this investigation can be assured that all data used based on the tweets have been correctly analyzed by the software to provide accurate analysis based on the results.

4. Findings

The data that was captured through the months of March to July 2020 in Figure 3 show us that the sentiment towards coronavirus depends on the sentiment analysis type and tends to be "low" nearing the neutral in March before descending in July to being mainly negative. The data sets when spread out over a 5-month period tell us that a major turning point can be viewed between June and July, this point is constant in both Vader and SentiWordNet analysis. This means that an event occurred during these two months. Another event can be observed occurring more visibly in the Vader analysis due to the decrease in sentiments between April and May. The observed events in the data findings can be related to two major events that occurred in the USA during this time. The events that occurred were anti-mask and anti-lockdown protests during April-May and Black Lives Matter protests during June-July, it is during these events that we can see fluctuations in the data.
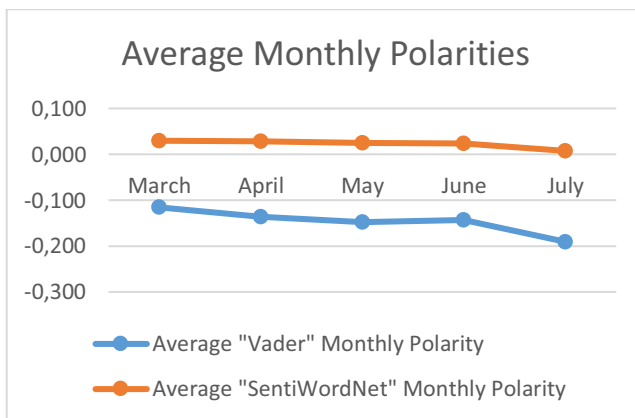

Figure 3. Average Monthly Polarity

4.1. Twitter Data

As seen in the table below, tweets have been analyzed based on their contents and categorized by their sentiment value. Based on the data that this research has gathered; several themes can be identified from Figure 3 such as that during the month of March, we can note that there is a steady decrease in the average polarity results. From this steady decrease we can identify that there must have been several events that occurred between May and July that caused a steady decrease of the overall polarity results. Furthermore, from the data results we can identify that a significant event occurred in March to April that caused the negative polarity results before another event occurred in May to cause a sharp decrease in results.

Table 1. *A table of tweets detailing different types of polarity*

Tweet	SentiWordNet Polarity	Vader Polarity	Average Polarity	Polarity Type
So hypocritical, when you KNOW damned well that #5G is the culprit, not #coronavirus itself. 5G directly affects human cells with radio waves, so that underlying conditions are escalated. This is a KILLING TECHNOLOGY! It needs to be stopped!	-3.9	-5.9	-4.9	Very Negative
I am DONE being nice about corona virus . If you think I wasn't very nice before, just wait. WEAR A FUCKING MASK YOU FUCKING MORONS. DEAD SERIOUS. YOU ARE KILLING PEOPLE WHILE BEING A BITCH. ASHMA ISNT AN EXCUSE NOT TO WEAR ONE NEITHER IS GLASSES. I WOULD MORE THAN HAPPY TO FIND	-11.3	-10.4	-10.9	Very Negative
Mother and brother of NYC hoops legend Sebastian Telfair, reportedly both have died from COVID -19. #COVID 19#CoronaVirus #COVID _19	-0.1	-0.7	-0.4	Negative
Uhh do you not read mystery books? Did you not watch the promise neverland? Literally ANYTHING. You can have a theory on. For example the whole 5G coronavirus theory. None of these idiots went out here to a 5G tower and tested it for radiation but that theory still exists.	.3	-.9	-.3	Negative

5. Discussion

5.1. America Discussion

The major aims that this paper have identified are, how did the American population react to Health and Safety mandates as well as lockdowns initiated by the government? Did any outstanding events occur during the 5-month period that could cause a spike in the sentiment analysis? What caused sentiments about COVID -19 to rise or fall during this period?

The major areas of debate and negative polarity results are tweets about mask wearing and lockdowns. During the months of April to May, there were many reports of Anti-Mask and Anti-Lockdown protests that occurred. According to Aratani (2020) [4] many of the governmental leaders have been providing mixed messages about face masks. Democrats have been more vocal about wearing them and many democratic governors made it mandatory to wear face masks. In contrast Republican leaders have been more hesitant to mandate masks.

Additionally, Aratani (2020) [4] notes that a lot of protests face masks because they call them "Muzzles" or see them as their freedom being eroded. One more vocal leader raised anger at the face mask being a way of "Throwing God's wonderful breathing system out the door" Aratani (2020) [4] reports. Even though both the CDC (2020) [6] and The World Health Organization (2020) [5] have evidence of masks and lockdowns slowing and removing COVID -19 many Anti-Maskers and Lock-downers dislike or even rebuke the thought of wearing a mask or going into lockdowns. According to

McKelvey (2020) [7] who interviewed a health professor at Morgan State University quotes " We're seeing politics and science literally crashing"

It is through this reason that the rise of COVID -19 cases can be observed to be rising in Figure 4. It is through this that we can acknowledge the reason for sentiment value of April and May being at -0.1 due to the rise of COVID cases. Further evidence of this is observed in table 1, where a sample of tweets can be seen about masks and lockdowns. It is through this table that we can observe how anti-mask/lockdown sentiments could have affected the rise of new COVID -19 cases in America.

The data gathered from the sentiment analysis provides a timeline of rises and falls in how people are feeling which they express through tweeting. From these results we can note that major events are related to the low sentiment analysis scores recorded. The results that were provided have determined that several events that occurred during the 5-month period of March to July appear to be related to several infection rate increases. The events that have been identified are the Black Lives Matter, Police Brutality, Anti-Mask and Anti-Lockdown Protests that sparked between the months of March and July of 2020. This evidence is seen in Figure 4.
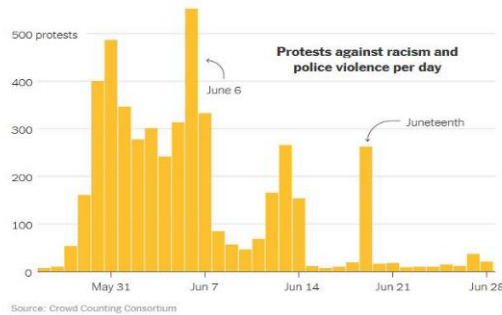

Figure 4. Number of Protests May 31st - June 28th[1]

Another event that occurred that led to mass infection is due to the protests involving Anti-Mask and Anti-Lockdowns. One of the more vocal leaders raised anger at the face mask's being a way of "Throwing God's wonderful breathing system out the door" Aratani (2020) [4] reports. From the sentiment analysis we can extrapolate from the results is that due to the mixed messages given by the various politicians. According to Aratani (2020) [4] many of the governmental leaders have been providing mixed messages about face masks. Democrats have been more vocal about wearing them and many democratic governors made it mandatory to wear face masks. In contrast Republican leaders have been more hesitant to mandate masks.

5.2. New Zealand's Response to COVID -19

In Comparison to how the USA has handled the COVID -19 crisis New Zealand initiated a nationwide lockdown. This lockdown allowed the people of New Zealand to reduce the infection rate to zero in a quick succession with minimal repercussions to setting a nationwide lockdown (Kunzmann, 2020). Comparably in America it would be harder to

---

[1] L. Buchanan, Q. Bui, and J. Patel, The New York Times, Crowd Counting Consortium.

initiate a nationwide lockdown according to Watson (2020) [3] "is that because there is no federal lockdown authority or a quarantine authority there is no standard or precedent that they can impose". Furthermore, according to Watson's (2020) [3] interviews, President Trump can only shutdown incoming travel at best Mr. Trump can recommend that people stay home. Additionally, according to the Federal Emergency Management Agency (FEMA) "States and Cities are responsible for announcing any type of restrictions or safety measures" What this means is that states are cities are given the ability to be pass restrictions and other measures on their own time, this means that if the state officials do not agree with the recommendations, they can choose whether or not to initiate such changes.

6. Conclusion

This paper has discussed how the sentiment of Americans changed between the months of March to July. Over this 5-month period of change, the paper has analyzed and observed what major events caused these sentiment changes and how they would increase or decrease the rate of COVID -19 infection. One of the main goals of this research was to discuss why America had such a high rate of infection compared to New Zealand and analyze possible reasons for America's high rates of infection. Some of the main reasons for America's high rate of infection is due to the citizens and their Presidents lack of care in regards to the pandemic by not following the health and safety measures suggested by health professionals. Another reason for the high rate of infection is due to two main factors, the murder of George Floyd which sparked Black Lives Matter Riots and the Anti-Mask and Anti-Lockdown protests. Due to these main points and the fact that American states were unable to put proper lockdown procedures in place, can be credited for the high rate of infection in the country. In conclusion, when it comes to dealing with a pandemic, it is ideal to quickly close off the country's borders and initiate a nation-wide lockdown.

References

[1]   Shear MD, Mercosh S. (2020, April 17) Trump Encourages Protest Against   Governors   Who   Have   Imposed   Virus   Restrictions.   New   York   Times,   Updated   April   29,   2020   https://www.nytimes.com/2020/04/17/us/politics/trump-coronavirus-governors.html

[2]   Darliansyah A, Wandabwa HM, Naeem MA, Mirza F, Pears R. (2019) Long-Term Trends in Public Sentiment in Indian Demonetisation Policy. CCIS 2018;932:65-75

[3]   Watson KTrump probably can't impose a nation lockdown or quarantine. Here's why. CBS News 6 April 2020. https://www.cbsnews.com/news/national-lockdown-quarantine-president-powers/

[4]   Aratani L. How did face masks become a political issue in America? The Guardian 29 June 2020, https://www.theguardian.com/world/2020/jun/29/face-masks-us-politics-coronavirus

[5]   World Health Organization. (2020). Advice on the use of masks in the context of COVID -19: interim guidance, World Health Organization, 5 June 2020. https://apps.who.int/iris/handle/10665/332293. License: CC BY-NC-SA 3.0 IGO

[6]   CDC. Considerations for Wearing Masks. CDC 7 August 2020. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fprevent-getting-sick%2Fcloth-face-cover.html

[7]   McKelvey T. Coronavirus: Why are Americans so angry about masks? BBC 20 July 2020. https://www.bbc.com/news/world-us-canada-53477121

[8]   Safe Software. Data Validation. (2020). https://www.safe.com/

[9]   Mottl, D. Get Old Tweets 3. Pypi. 2019. https://pypi.org/project/GetOldTweets3/#description

[10] Matošević G, Bevanda V. Sentiment analysis of tweets about COVID-19 disease during pandemic, 2020 43rd International Convention on Information,   Communication and Electronic Technology (MIPRO)*,* Opatija, Croatia, 2020, pp. 1290-1295, doi: 10.23919/MIPRO48935.2020.9245176.

[11] Barkur, G, Vibha, Kamath GB. , (2020). 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.

[12] Davison   WP.   Public   Opinion.   Encyclopedia   Britannica   (2017,   March   17). https://www.britannica.com/topic/public-opinion