Understanding the Changes in Positive and Negative Sentiments in the Discourse of the COVID-19 Pandemic in Alberta

by

Bennett Kuwan Tchoh

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Department of Digital Humanities

University of Alberta

Abstract

The severe acute respiratory syndrome caused by the SARS coronavirus was declared a pandemic in March 2020 by the World Health Organisation. Alberta as other provinces and countries had to face the challenge of fighting this syndrome that caused what is commonly known as the COVID-19 pandemic. The measures put in place by the government officials to fight the pandemic brought about big changes to the lives of Albertans. This study sought to explore the discourse that happened online about the COVID-19 pandemic in Alberta. One might expect that the discourse will change with changes in the gravity of the pandemic. But is this the case? Using tweets scraped from Twitter, the relationships between scores of positive and negative sentiments in these tweets and the number of patients in ICU, wave of the pandemic, mentions of oneself and people identified with (Us pronouns) and mentions of the other and people not identified with (Them pronouns) was explored using visualisations and statistical analysis tools. Three methods of measuring sentiments were used and the methods compared. Results showed that the discourse on the pandemic became less positive as we moved from the first to the fourth wave. A multiple linear regression showed that the best predictors of the sentiment score of the discourse was the wave of the pandemic and the use of Them pronouns. Comparison of the sentiment measuring methods suggested that the changes in the use of positive and negative words are not accompanied by a proportional change in the sentiment in the discourse. As expected, the number of patients in ICU correlated negatively with positive sentiment scores and positively negative sentiment scores. Unlike expected, the use of the Us and Them pronouns had similar correlations in magnitude and direction with sentiment scores. The findings from this study provide a better understanding of the changes in the sentiments of the discourse of the COVID-19 pandemic in Alberta and brings to light the importance of assessing the validity of measuring instruments and methods.

Keywords: sentiment analysis, COVID-19, Alberta, Twitter, ICU, personal pronouns.

Dedication

This thesis is dedicated to my mother Mrs. Emma Ndekum Kongran Kuwan who has always been there for me and wants what's best for me more than any other person.

Acknowledgement

Firstly, I want to thank God for seeing me through the master's program and helping me go through the different challenges I faced. Thank you to my Christian family and friends for all their support.

Thank you to my family for their prayers and support.

I want to thank my fiancé Bianca and her family who will soon to be my family for their support, encouragement, and prayers which they faithfully made for me.

A big thank you to my supervisor Dr. Geoffrey Rockwell for his support, encouragement, and patience throughout.

Thank you to the members of the supervisory committee Dr. Harvey Quamen and Dr. Thomas Barker. Their review and comments helped make the study better.

A big thank you to Nicola DiNicola for her assistance throughout the master's program. She went beyond her role as the graduate advisor for the Digital Humanities program to make sure I was doing well.

This study benefited from the CIFAR Catalyst Grant led by Daniel Baumgart on Guarding At-Risk Demographics with AI (GuARD-AI) in 2020. I had the opportunity to contribute to the grant's research as I was employed as a research assistant, and I used part of the data gathered under the grant in this study. Thank you to all my colleagues who helped collect the data for the grant's research project.

Table of contents

List of Tables

List of Figures

Introduction

In late 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) commonly known as COVID-19 was identified in Wuhan in China. It quickly spread to other countries and most countries in the world reported cases in the first quarter of 2020. By March 2020, it was declared a pandemic by the World Health Organization. The main symptoms of the infection were shortness of breath, tiredness, fever, and cough. The virus was first understood to be transmitted through contact with objects contaminated by droplets from the nose and mouth of infected people, but it was later understood that the virus could be transmitted by airborne droplets between people in close proximity (Jarvis, 2020). This led to the physical distancing measures implemented in most public spaces across the world to curb the spread of the virus. Most healthy individuals developed mild symptoms that resolved after a few days while the elderly, those with pre-existing medical conditions and those with a compromised immune system developed more serious symptoms. The majority of reported deaths occurred in this vulnerable group (Adab, Haroon, O'Hara, & Jordan, 2022).

In January 2020, the world watched in shock and amazement as Wuhan and other neighboring cities were put in lockdown creating scenes of ghost towns. As the virus spread exponentially, many countries imposed similar restrictions that limited the spread of the virus by limiting in-person interactions. The pandemic became the main topic of discussion as almost everyone's life was impacted by the social restrictions in place. This was particularly true for the province of Alberta which started implementing restrictions in March 2020 and declared a state of health emergency on the 17th of March 2020. With social gathering restrictions already in place and high use of online social media platforms, most interpersonal interactions moved online. Those that did not have the skills for online interaction had to learn fast.

Without much time to prepare, schools and universities had to quickly switch to online classes. Many people had to work from home. Zoom and online meeting services became household names as they were used by most workers that worked remotely. Popular brands of webcams were constantly out of stock and their prices were much higher than before.

Online influencers became even more influential as their audience increased with many people forced to stay at home and forced to turn to the internet for entertainment. With the intention of trying to make the best of the situation, many people tried to earn money through online activities and many people started an online career that way. The pandemic years are considered to be a boom period for influencers and online content creators (Gagliese, 2022).

Some companies like those in hospitality, travel, oil and gas were severely hit by the COVID-19 pandemic while others made record profits. Among those that made record profits were companies with online shopping websites like Amazon and Shopify, tech companies like Microsoft and Nvidia which produced computing and gaming hardware, and social media companies that own social networking services for example, Facebook and TikTok, online entertainment companies that owned Netflix and Prime Video (Novicio, 2021).

Canada and other regions in the world have experienced pandemics before but the COVID-19 pandemics brought about drastic changes in the life of almost all Canadians in a scale never experienced in their lifetime. The pandemic became the major topic of discussion online. Most people did not know what to expect and at the start it was both terrifying and intriguing to watch what was happening in the world and how their lives were affected. Government officials gave regular updates on the pandemic and on the measures they were implementing to fight it; the public reacted to these updates, and most people echoed these updates in an attempt to ensure that everyone was informed and to encourage compliance to fight the pandemic. As the pandemic evolved, people were encouraged to stay home to save lives and flatten the curve by preventing a surge in infections and hospitalisation that would further overwhelm the health sector. This was when it was understood that the virus would spread no matter what but the rate of spread had to be reduced. People were asked to postpone all travel plans. People were also encouraged to wear masks to reduce the spread. When vaccines became available, people were encouraged to get vaccinated. All these measures to fight the virus were substantially discussed online. There were varying opinions about the effectiveness and/or usefulness of these measures.

There was no cure to the virus and government and research laboratories reported that they were actively searching for cures and vaccines against the virus. Many people speculated on different home remedies and already available government approved drugs as potential treatments for the virus infection. A popular drug that received attention was Ivermectin that was known to treat parasitic diseases caused by worms in humans but was mostly used on reared animals. Most doctors trusted the advice of the medical governing body (for example the Center for Disease Control in the United States and the Public Health Agency of Canada in Canada) of their countries against the use of Ivermectin to treat COVID-19 while a few doctors promoted its use. Huge debates happened online about the effectiveness of this drug.

Very early on, the pandemic became politicised as the different political parties did not agree on the response the government officials had to the pandemic. The debates and clashes of opinions did not only happen in the parliamentary chambers and during news updates, but it was even more present online. Discussion about the pandemic happened across many media channels amongst which are academic publications, news reports, newspapers publications, etc. Individuals posted content online about the pandemic in many formats, including text, video, audio, live broadcasts, blog posts, etc., and most of these post generated responses which could be of the same format or could be in the form of a like and/or share. This created huge volumes of user generated data that can be used to get information about the COVID-19 pandemic.

Even before the pandemic, there was an exponential growth in online generated data that has been mined and used for research in many fields of study. There are more and more connected devices which have sensors that continuously record and save data. Companies and firms in every sector of activity (health, transport, agriculture, urban planning, etc.) utilize big data analytics to extract knowledge from these big data to achieve successes in their activities. In academia, big data generated from internet users' online activities has been recognised as a source of knowledge in many domains of study. Social sciences and the humanities are two of these domains of study that also employ big data to study human societies, human interactions and culture. The scale of big data makes it almost impossible for humans to directly analyze the data and so digital computing methods are used to analyse such big volumes of data which are usually themselves digitally born.

For researchers in the humanities who want to use big data, gathering and preparing this data requires that researchers have some computing skills. Interdisciplinary fields of study like Digital Humanities which is an intersection of the fields of computing, the human sciences, and the arts train students in the needed skills from different fields of study. Its focus is not merely the use of computational tools to gather and analyse textual data. Different forms of media are analysed, the aesthetics and art are explored with digital lenses and the implications of the use of digital methods are also analysed (Svensson, 2013).

The amount of content freely accessible online keeps increasing exponentially. All this content can be used as sources of data depending on the interests of the researcher. The text, audio, images, videos, layouts, that can be viewed in a browser in most cases can be downloaded and data extracted from them. Videos from YouTube can be downloaded, and the audio transcribed, the frames in the video file can be sampled and analysed to identify named entities, text from blogs can be downloaded, comments from posts and user information from social networking sites can be downloaded, etc. This can be done with or without specialized tools, but human intensive methods are very time consuming and hence in most cases customised programs are built or already available specialized program modules are used. The main limiting factor that would affect gathering this data would be legal use policies of the content creators and/or providers.

In recent years, with the rise in popularity of social media platforms, the content generated by users of the social media platforms have increasingly been used in research in many fields of study. Saxena (2013) lists the following seven characteristics of social media platforms:

1. The platform provides users with a webspace to upload their content. This webspace is usually free.
2. The webspace is given a unique web address that other users can use to visit the webpage.
3. The users are asked to build a profile by adding some details about themselves.
4. The platform has strong elements of connectivity which permits users to set up connections with others and upload content that is shared with connected users.
5. Content can be created and posted in real time and made accessible to other users.

6. Most social platforms have messaging functions which enable users to send private and public messages to each other.
7. The posted contents have timestamps that permit the viewers to know the timeline of events and locate posts in time.

These characteristics of social media platforms all constitute data that can be studied by researchers. One of the social media platforms that has generated much research interest is Twitter. It was popular during the pandemic, and it is used by many people in Canada.

Twitter is a social media platform used by many Albertans. Although Facebook is the most used social media platform in Canada by a big margin (Summerfield, 2023), public figures, politicians and government agencies are more inclined to use Twitter for their official communications. Twitter seems more than Facebook to be a platform for the communication or expression of ideas whereas Facebook seems to be more than Twitter to be a platform for multiple purposes amongst which are entertainment and self broadcast as an end in itself.

Twitter provides a wealth of data for any person interested in the study of the discourse on a particular topic in society. With the use of hashtags, users can easily search for posts on topics represented by the hashtag to learn about what is being said concerning the topic. For more large-scale text analysis projects, the Twitter API can be used to scrape and save tweets in bulk. To study the discourse on the COVID-19 pandemic in Alberta Twitter is a good source of data. The Premier of Alberta Jason Kenney and the Chief Medical Officer of Health Dr. Deena Hinshaw were both most active on Twitter and gave regular updates there.

Twitter was launched in July 2006 and has constantly grown in popularity. It can be described as a micro blogging social networking platform. Users post short texts called tweets with a character length limit which was previously 140 characters but currently set at 280 characters. According to Karger and Quan (2005), Twitter can be considered as microblogging because of the following three main characteristics,

1. there is a single author of the blog even though the topics covered in the posts might vary,
2. the posts are usually in the form of notes and not full large websites publications and,
3. the individual posts can easily be grouped together since they are usually in standardized formats like XML which are machine readable using RSS or in the case of Twitter, can be scraped as a JSON file with the Twitter API.

Twitter has evolved over the years and now allows users to post multimedia files such as pictures, videos, and sound files. Its networking ability allows users to follow other users. A followed user's posts appear in the Twitter feed of the followers. Generally, all tweets are public except when the user restricts the tweets to be visible only to their followers. Since 2014 an average of 500 million tweets are posted per day (Sayce, 2022). From the 2022 first quarter reports from Twitter, there are currently 237.8 million monetizable daily users (Dixon, 2022).

It is free to create a Twitter account and start posting tweets, following other users and being followed by other users. A popular feature of Twitter is the grouping of tweets by topic with the use of hashtags. This is done by prefixing one or more alphanumeric characters with the hash (#) sign in the tweeted text. This makes it possible for users to create and follow tweets on a specific topic. For example, the hashtags #abvote and #abpoli were popular during the elections in Alberta, while #edmontonoilers and #oilers are popular in tweets about the Edmonton-based hockey team.

Twitter is different from Facebook, the most popular social networking site in many ways (Rowe & Alani, 2014). Facebook is the number one social networking site (SNS), and it is the default SNS that people use to connect (Dixon, 2022b). Facebook is about giving users many options to broadcast themselves and to interact with each other. Users can post albums with videos and images, they can start an audio or video live broadcast in which they invite others to join and interact with other viewers, they can shop in the Facebook marketplace, they can date using Facebook dating, etc. Facebook wants to bring many aspects of real life into its platform. A step in that direction was the recent launch of the Metaverse. A metaverse is an online virtual environment which allows people to interact in it. Mark Zukerberg describes the metaverse as a place to connect, work, play, learn, and shop (Coursera, 2023). Twitter is more about making users communicate more efficiently. Users' posts are limited to 280 characters hence they have to be succinct and straight to the point to get their points across efficiently. Other users can interact with the tweet by liking it, retweeting it to their own followers, replying to the tweet and tagging other people to the tweet. Twitter is sometimes described as a battle ground of ideas and attention (Garrett, 2016). Usually, the main goal is to recruit adherents to the opinions expressed. It is the preferred networking platform for politicians and people that want to express their political views (Blackman, 2022). It was the favourite method of communication by Donald Trump, the former president of the United States before his account was banned in 2021.

The great amount of user-generated content available on Twitter has made Twitter attract interest as a valuable source of research data. The interest in Twitter as a data source has grown in the last decade. Researchers and policy makers have used data from Twitter for their different objectives and in many fields of study (Karami, Lundy, Webb, & Dwivedi, 2020).

Canada, like other countries was, faced with the challenge of dealing with the COVID-19 pandemic. The governments at the federal and provincial level implemented policies to curb the spread of the virus. With the decentralized system of government in Canada, the provincial governments were the ones that implemented the different policies like masking, lockdowns, and vaccine mandates which had the most impact on the life of the Canadians in that province. The three main top government officials that managed the pandemic in Alberta were the Premier Jason Kenney, the Minister of Health Tyler Shandro and the Chief Medical Officer of Health Dr. Deena Hinshaw. During the pandemic the government implemented many policies among which were varying levels of banning of group gatherings, closure of certain types of businesses, mask requirements in public spaces, mandatory isolation of contacts of diagnosed cases, and vaccine mandates. These policies were not welcomed by everyone and even the lifting of the policies was met with opposition by some groups of individuals. Many Albertans took to social media to voice their opinions on the pandemic and many pandemic hashtags were popular in the province.

An exploration of the tweets of Albertans on the COVID-19 pandemic would provide a view into how Albertans lived and felt about the pandemic. This study sought to explore this discourse that was going on online on the COVID-19 pandemic in Alberta. Twitter was used as the source of the discourse in Alberta. Through sentiment analysis, one aim was to observe how the sentiments on the pandemic changed during the different waves of the pandemic. The study also sought to compare methods of measuring positive and negative sentiments of tweets and to understand how these sentiments varied with official statistics on the number of people hospitalized and in ICU during the pandemic. Three methods of measuring the sentiment of the text will be compared. Another aim was to measure and explore how sentiment scores varied with the use by Twitter users of pronouns that represent themselves and groups they identify with (first person pronouns singular and plural) and the use of pronouns that represent other people and groups the other people identify with (second and third person pronouns singular and plural).

The sections that follow will be a literature review of studies from different areas of research, followed by a discussion of the ethics of doing research with data scraped from online sources. Then the methodology and results of the analysis will be presented. The study will end with the conclusion which will address the limitations of the study and provide recommendations for future studies.

1 Literature review

1.1 History of Twitter research

Since the publication of the first Twitter related papers in 2007 (one year after the official launch of Twitter) interest in Twitter has grown exponentially. Tens of thousands of papers have been published (Karami, Lundy, Webb, & Dwivedi, 2020). Very early on, the potential for use of Twitter as a research tool became apparent to many researchers. Ovadia (2009) studied the potential of using Twitter as a research tool. He noted that Twitter is a better option compared to sites like Google Trends because Twitter permits the tracking of the subjects being discussed and the researcher could access exactly what was being said by scrapping the full text of the tweet.

Williams, Terras, and Warwick (2013) published a study in which they reviewed all papers published on Twitter from 2007 to the end of 2011. This effectively covers all papers on Twitter since its creation up until 2011. They searched SCOPUS and Web of science for all published papers with the word 'Twitter' in the title, abstract and list of keywords. The bulk of their study was based on 575 papers that focussed on Twitter and microblogging. They found that three papers were published in 2007, eight in 2008 and 36 in 2009. There was a more significant increase in 2010 where 210 papers were published, and in 2011, 320 papers were published. They grouped the papers into 4 categories related to the focus of the papers. **Message** (the study focused on the content of the tweets), **user** (the study focused on Twitter users), **technology** (the study focused on the hardware and software used to access Twitter or its API) and **concept** (reviews and conceptualisation of Twitter and its uses). Most studies focused on the message followed by the user. One limitation of this study was that the authors relied only on the information in the abstracts to draw conclusions on the study so more detailed information about the study was left out and there is some chance of misclassification of the studies.

More recently, Karami, Lundy, Webb and Dwivedi (2020) did a review of studies published between 2006 and 2019 that had the word Twitter in their title or abstract. They used computational methods to mine the papers and abstracts and to analyze the data. They used Latent Dirichlet Allocation to identify major topics covered in Twitter research and tracked the topics over time. Of the 40 topics they considered, 29 had significant trends over time. Examples of topics with significant trends are sentiment analysis, stock market, content analysis and politics which had a positive significant trend.

Williams, Terras and Warwick (2013) used a more human-centered approach to gather the papers used in their study; on the other hand, Karami, Lundy, Webb and Dewivedi (2020) employed computational methods to find the papers used in their study. Both methods have both advantages and disadvantages. For example, Williams et al. were able to filter out papers that had the word Twitter in them but that were not really about Twitter while Karami et al. were not able to filter out the non-Twitter-related studies because of the greater number of papers they had. Of the 40 topics that they identified, they had to remove two topics from their analysis because they were not related to Twitter. The keywords of the topics were general terms used in paper abstracts.

As previously mentioned, researchers from many fields of studies have used Twitter in their research. We will now look at a few of these fields of studies and consider the way they used data from Twitter.

1.1.1 Twitter data and politics

According to Reeher (2004) the internet can be considered to have been used to make a major impact on elections in the US from the year 2000. The year 2000 was an election year in the United States and saw the candidates rely on the new offerings of the internet to gain votes. But without social networking sites (SNS) and the still low use of the internet by the public, the impact of the internet was still quite small as campaign webpages did not receive high traffic. In the election year of 2004 the number of internet users had greatly increased. Surveys by the Pew Research Centre coordinated by Rainie, Cornfield and Horrigan (2005) showed that the percentage of Americans who went online for election related activity rose from 18% to 29% during the 2000 and 2004 US presidential elections. Seventy-five million Americans went online for various campaign related activities (discussions over email, to get information and news about candidates, to donate to campaigns, etc.). The 2008 presidential election is considered to be the first time that candidates relied heavily on the internet for communication, networking, fundraising and organization. Barack Obama used many SNS and his victory according to many reporters was a sign that all future campaigns would have to rely heavily on social media. News reports called the 2008 election the Twitter and Facebook election. Barack Obama was well ahead in number of followers in comparison to his opponent John McCain by at least a 5-to-1 margin on most SNS (Pew Research Centre, 2008). Even before the presidential elections, Barack Obama, who was not a very known senator, used the internet and social media to gain the nomination as the democratic party presidential nominee. Many academic papers (for example Bimber, 2014; Goodman, Wennerstrom, & Springgate, 2011; Luck, Beaton, & Moffatt, 2010) have been published on Barack Obama's use of SNS for his campaign and most believe his use of SNS is one of the reasons why he won (Aaker & Chang, 2009). Barack Obama continued using Twitter during his presidency to push his political agenda and at one point he was the most followed person on Twitter. He is currently the second most followed person with 132.2 million followers, just behind Elon Musk who has 146.6 million followers.

More than half (55%) of the population of voting age in the United States used the internet for political purposes. Seventy-four percent of internet users reported going online to get information, communicate with others and share information on politics during the 2008 election campaign (Smith, 2009). But was the election outcome decided by the internet campaign? Budak (2010) found that getting information about the candidates from online sources (including Twitter) did not predict political engagement in the form of voting. In his study, the participants that had Twitter as their main source of election information was quite low; about 22 out of 2254 participants. But that does not mean that the information gotten from online sources did not influence the vote of those that voted. For the 2004 elections, Rainie et al. found that 52% of those that used the internet for political purposes reported that the information they got helped them decide who to vote for. Contrary to the Budak's (2010) finding, Ranie et al. found that 23% of those that used the internet for political purposes reported that the use of the internet encouraged them to vote. In more recent years the number of internet and Twitter users has increased. In 2022, Twitter reported 237.8 million daily active users. This was the highest number of daily active users ever reported (Dixon, 2022c). (It should also be noted that the use of other SNS has also increased and many other SNS have more users than Twitter). In 2019, 42% of a sample of adult Twitter users reported that they use Twitter for political discussions (Hughes & Wojcik, 2019).

Twitter data has been used as primary data for political research like projections of election outcomes, testing the public opinions, measuring the response to campaigns, etc. With 237.8

million daily active users, reading, sharing, and creating tweets, it is understandable why Twitter data has been widely used in research.

The usual election prediction method has been the use of polls, but these can be very expensive to conduct. Many studies published in the last decade have explored the use of data from Twitter to predict election outcomes. This is usually done by predicting the sentiment towards the candidates in the tweets scraped for this purpose. Liu, Yao, Guo and Wei (2021) used a combination of sentiment analysis and economic growth to predict election results at a county level. Their model had an accuracy rate of 81%. Their study was based on the 2016 US presidential election. In their methodology, they used the Twitter ID of the users to calculate the sentiment for each user towards Donald Trump and Hillary Clinton, the two presidential candidates. Based on the score, they decided on which candidate the user was more likely to vote for. Amador, Collignon-Delmar, Benoit and Matsuo (2017) used millions of tweets to predict the vote in Britain on whether Britain should leave the European Union (the Brexit vote). They did a correlation between different poll results (internet polls, telephone polls, likely-voter polls, etc.) and classified tweets (leave or remain). Results showed a good correlation between internet polls and Twitter data. The correlation between internet polls and telephone polls was poor and did not reach significance. Their conclusion suggested that Twitter data was better than telephone polls to predict election outcomes and could be used in place of internet polls due to their high correlation. Some past research on election outcome prediction using data from Twitter used models that rely solely on data from Twitter and while others have included other independent variables like economic data, data from other social networking platforms, Google trends (Kassraie, Modirshanechi, & Aghajan, 2017). As will be seen in the reviews of other areas of interest in which Twitter data was used, the type of data used in combination with data from Twitter depends on the areas of research and usually improves the prediction model.

1.1.2 Twitter and the financial market

In today's world, the financial market is a fast-paced environment with people buying and selling stocks, digital currency, foreign currencies, products and services right from the comfort of their homes or on the move on mobile devices. The financial market is a dynamic environment with the prices of the commodities and equities constantly changing under multiple influences. Buyers have to choose when best to buy and sellers when best to sell. Predicting the future value of goods, services and equities will help the stakeholders make maximum gains. Different types of internet user data have been used to predict market behavior. For example, Perlin, Caldeira, Santos and Pontuschka (2016) used the frequency of search queries to determine market volatility and to test market strategy based on them. They found that an increase of searches with the word 'stock' predicted volatility and negative returns in the near future. Interestingly, they also found that an increase in stock returns also predicted an increase in searches with the word stock. Albahli et al. (2022) used machine learning to train a model with tweets and financial data of ten stocks from Google. They used the StockSentiWordNet model which is specifically designed for sentiment analysis related to the stock market. The tweet text was tokenized into bigrams which made sentiment detection more efficient for example 'not good' was evaluated correctly as negative sentiment but if single words were considered as tokens, the presence of the word good would have increased the chances of the tweet to be wrongly attributed a positive sentiment. Their two final test models performed well with a prediction accuracy of 81.4% and 86.06%. Akbiyik et al. (2021) used data from 30 million Bitcoin related tweets to test the prediction of tweets on bitcoin's next day volatility. They used a multilingual sentiment analysis tool called Valence Aware Dictionary for Sentiment Reasoning abbreviated as VADER (Hutto & Gilbert, 2014) to attribute sentiment values to the tweets. The study found that the semantic content of the tweets (sentiment scores gotten from VADER) was not as predictive as the metadata about the author of the tweet (for example number of followers). The most predictive model was a combination of user information, tweet sentiment and volume of tweets. This might be so because of the higher volatility of cryptocurrency compared to other traded commodities (like foreign currency trading) in the stock market (Baur & Dimpfl, 2021). Unlike fiat currency, cryptocurrency does not have a regulatory board that controls its production and distribution, so its value is very dependent on demand and supply and very influenced by the hype of the media. The stock market is also similar to cryptocurrency in the sense that it can be very volatile in its reactions to media hype. Influential people like Elon Musk and Donald Trump's tweets on cryptocurrencies and companies in the stock market have had a significant effect on their value. Brans and Scholtens (2020) studied the effects of Donald Trump's tweets mentioning a company on the value of their shares. They found that negative tweets were followed by a reduction in their market value while on the other hand positive tweets did not have a significant effect on their market value. Tweets by Elon Musk on cryptocurrency have been observed to have a positive effect on the value of the currency. That was particularly true in the case for the Dogecoin (Ante, 2023). Tafti, Zotti and Jank (2016) studied the changes in volume of trade of popular NASDAQ 100 stocks with variations of mentions of the stocks on Twitter. They used a regression linear prediction model. The indicators from Twitter improved the accuracy of the model by only 0.2%. This showed the difficulty of using predictions from Twitter text for analysis with the aim of monetary gain. But the researchers suggested the use of more advanced machine learning models. Nti, Adeoya and Weyori (2020) compared the prediction of stock market prices in Ghana's stock market from a variety of text data sources (Google Trend, Twitter, forum posts, web news) using an Artificial Neural Network. They observed a prediction accuracy that ranged from 41% to 60.5% for individual sources. In a model that included all the sources, their prediction accuracy reached 77%.

1.1.3 Twitter data and health

Public health experts continuously monitor the population for any outbreaks of diseases. Most state health departments monitor and keep reports of medical visits and are able to detect abnormal patterns that might signal disease outbreaks or other situations of interest. Early detection of disease outbreaks is essential to start the responses as early as possible to limit the spread and increase the chances of containment. Other health issues like mental health are also of concern and so surveillance of emotional changes in the population might be needed to monitor mental wellbeing. With so many adults active online, Twitter might be a low-cost alternative for disease monitoring and health research. There might even be some advantages of using Twitter to monitor diseases instead of other sources of information or data from Twitter could be used together with other data sources like the records of medical visits, weather reports, etc.

Twitter offers the possibility of real time observation of the population since Twitter users will usually tweet on their current experiences. Researchers and government officials could monitor identified keywords that inform on the health of the population. That could be particularly effective for mental health which is becoming more and more prevalent and was quite bad during the COVID-19 pandemic (Gobbi et al., 2020; Jolly, Batchelder & Baweja, 2020).

The Google Flu Trends (GFT) is a popular service offered online by Google that started in 2008 with data from 2004 onwards. Google tracked queries related to flu from 25 countries and presented the fluctuations in queries graphically. It is still available online at https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac_ although the data has not been updated since 2015. They compared the number of queries with the baseline level for each country and it was used for early detection and forecasting of flu outbreaks. Predictions of flu outbreaks were shown to have very high accuracy in predicting doctor visits prior to 2010 (Ginsberg et al., 2009) but the predictions were found to be inaccurate between 2011 and 2013 (Lazer, Kennedy, King, & Vespignani, 2014). Preis and Moat (2014) used GFT data from 2010 to 2013 and historical flu data from medical sources during this period and found that training their model with both data sources improved the prediction ability compared to either source used alone.

Zhang, Lyu et al. (2021) did a study in which they searched COVID-19 related tweets and from there used regular expressions to identify users with depression. They searched for phrases like 'my depression', 'I have/developed/got/suffer(ed) from mild/severe depression', etc. to identify users with depression. Their goal was to identify the regular language use of individuals with depression, so they excluded the tweets used to identify the users with depression by regular expression and trained a model with 200 other tweets from these users. They also ran personality tests, sentiment analysis, demographic identification, eight psychological category analyses from the LIWC software, user networking (engagement) assessment and used the scores for each user in their model. Their model had an accuracy of 78.9%. They found that features that favoured the classification into the depression group were, usage of first-person pronouns, neuroticism, talking about biological processes like sleep or eating, exhibiting sadness. Weidener and Li (2014) did a study in which they mapped mentions of healthy food (fruits and vegetables) and unhealthy foods (deserts, cakes, cookies) on the US map to verify if there were more mentions of unhealthy food in areas identified as food deserts (regions that lack access to healthy foods) by the US Department of Agriculture. Their findings suggest that healthy food is more likely to be found in areas that aren't food deserts. They also found a higher proportion of tweets about unhealthy food in these food deserts.

Paul and Dredze (2012) developed the Ailment Topic Aspect Model (ATAM) which is similar to LDA (topic modeling). The model discovered health-related topics which could be used to identify different ailments. The output of ATAM was compared to LDA and ATAM detected more unique ailments (14) while LDA detected significantly less (10). ATAM was tested for its effectiveness in flu surveillance using tweets by calculating the number of tweets assigned to ATAM's flu ailment per week divided by the total number of tweets. A Pearson correlation was run with CDC flu data and a correlation of 0.93 was observed, suggesting that the flu and other diseases can be monitored through Twitter.

Nduwayezu et al. (2019) compared the volume of tweets from Nigeria on malaria and precipitation data at three time point offsets (same week, one week and two weeks). They found significant correlations between tweet volume and precipitation reports from all the offset points. The highest correlation was found with the two-week offset (0.75) followed by that from the one week offset (0.69) and the least was from precipitation from the same week (0.55) i.e., no offset.

Sinnenberg et al (2017) did a systematic review of academic papers that contained the word Twitter and a variety of medical terms in the abstract or title. Some of the findings were that, 56% did content analysis to identify themes in the tweets, 26% were monitoring mentions of diseases, 15% did sentiment analysis, and 5% were predicting prevalence of diseases.

Studies that have used data from Twitter abound in literature. The three areas of interest and examples of studies in those areas mentioned above are only a small sample of the wealth of published studies. These studies have used a variety of methods to analyse the data scraped from Twitter. Some of the most frequently used methods (or processes) of data pre-processing in preparation for analysis and analysis proper will be presented next.

1.2 Methods of analysis used in with Twitter data

There has been a variety of methods used to analyze data from Twitter ranging from the simple analysis like frequency comparisons (number of tweets on a subject, counting the number of mentions, etc.) to more complex, blackbox methods like machine learning models.

But before the data is analysed, the data usually has to be prepared for analysis. This will transform the data into a format which can be put through analysis methods and/or tools to test the study hypothesis. This transformation is usually made up of two main steps which are data cleaning and preprocessing.

1.2.1 Cleaning

In the cleaning stage, part of the data that is not useful for the study is removed. Most Twitter scrappers will return a lot of metadata with each tweet. According to Twitter, each tweet has about 100 types of associated metadata, for example time the tweet was created, ID of user, gender of user, whether it is a retweet, region from where the tweet was tweeted, etc. Most Twitter scrapers return a JSON file which contains much of this metadata as attributes. Depending on the goal of the study, the researcher will remove some of these attributes from the data and use some of the attributes to filter out some of the data. For example, it is common practice for retweets to be filtered out so the researcher can do so by dropping all tweets that have a Tweet ID for the 'is a Retweet of' attribute. It is also common for Twitter user handles, i.e. groups of characters that start with the @ sign and URLs to be removed from the text of each tweet. Usually, the tweet text goes through sentiment analysis and user handles and URL links are not very useful for sentiment analysis. The tweets can be further cleaned by removing symbols, numbers, foreign words, hashtags and emoticons.

Stopwords are usually removed because they are so commonly used that they are of low importance to most studies. Removing them helps give more focus to the important words of the corpus. It also reduces the size of the corpus and hence reduces the training time for machine learning models. Most language processing tool kits like NLTK have a stopword list of very common English words. It is common for researchers to identify other words particular to the corpus that can be added to the stopword list.

The cleaning process is very dependent on the study objectives so it varies widely in literature. For example, studies might want to investigate the types of URL links that are most shared about a particular topic so they will not clean out the links in the tweets. Other studies might want to focus more on the network of users or their popularity and so their username handles will not be cleaned out of the tweet text. An interesting study (Park, Baek, & Cha, 2014) analyzed the use of emoticons—and the punctuation signs that represent emotions—on Twitter by people from different cultural backgrounds and hence emoticons would not be removed from the tweet text. Some sentiment analysis programs—for example LIWC—even take into consideration the sentiment expressed in the emoticon when performing sentiment analysis.

1.2.2 Preprocessing

When the text has been cleaned, it is then preprocessed to prepare it for analysis. The first step of preprocessing usually involves putting all the text in lowercase (for example, Liu, Yao, Guo and Wei, 2021) to ensure that all the words used with different capitalisation are not considered to be different words during analysis. Depending on the objective of the study, other preprocessing steps might involve the following.

1.2.2.1 Grouping tweets

By user: In some cases, the opinions of individuals and not of the corpus as a whole is required. This can be used in cases like election prediction to determine the opinion of individual voters based on their tweets (for example Liu, Yao, Guo and Wei (2021) mentioned above).

By geographical location: Usually researchers restrict their scrape of Twitter to specific regions. In certain cases for example when using a corpus of data scraped by someone else or when creating different groups from the data scraped for more detailed analysis, they have to group their data from Twitter based on the geographical location of the user (for example Liu, Yao, Guo and Wei, (2021), Weidener and Li (2014) discussed above).

By time periods: Sometimes, researchers might want to compare different time periods and so they group the scraped tweets by time periods. For example, Chaudhry et al (2021) compared the sentiments from scraped tweets before, during and after the 2020 US presidential election. They also compared the 2016 and 2020 elections.

More than one type of grouping can be used together. For example, if a researcher wants to know how users' sentiments change over time periods, the tweets can be grouped by user and then the group of each user is further grouped into the different time periods.

Normalization: This usually involves lemmatization and stemmatization of the text. This is usually done to reduce the number of unique words that have the same general meaning. There are both advantages and disadvantages of this process and its usage will depend on the study being done. Some advantages include faster analysis and limiting the possibility for lexicon based models to miss some words. One disadvantage of this is that unique usage of words in the corpus and the unique meaning they bring is lost.

All these stages of data cleaning and preprocessing will modify the text and might have a significant impact on the results of the study hence affecting the study's validity, confidence and the applicability of the findings. They can also be a source of bias that would affect the applicability of the findings.

1.2.3 Analysis methods

Below are some of the methods that can be used to analyse the text to get some insights to test the research questions.

1.2.3.1 Volume analysis

This type of analysis usually involves the comparison of grouped tweets. A researcher could for example scrape tweets from Twitter and later group the tweets by user, time period, hashtag and later compare the volume of each group. The volume could be measured in terms of actual text volume or in terms of number of tweets. Sprenger, Tumasjan, Sandner, & Welpe (2014) in their study of the relationship between Twitter and the financial trading market found a positive correlation with 99% confidence interval between message volume (number of daily tweets) and trade volume (number of shares traded). Volume analysis is principally concerned with the number of tweets. The next methods presented are methods of analysis that are used to analyse the content of the tweets.

1.2.3.2 TF-IDF

TF-IDF stands for term frequency-inverse document frequency. It is used to determine the importance of words in a document which is part of a corpus of documents. TF-IDF is calculated by multiplying two metrics:

TF-IDF = TF*IDF

TF is the frequency of the word in the document. The relative frequency of the word is sometimes used because the different documents that make-up the corpus are usually not of the same size**. IDF** is the quotient obtained by dividing the number of documents in the corpus by the number of documents in which the word is found. Sometimes the natural logarithm of this number is used instead. The words with high TF-IDF in each document are the words that differentiate the document from the other documents and can be considered as an indication of what the subject matter of that document is. A use case scenario of TF-IDF could be the comparison of two different waves of the pandemic to see what words are peculiar (characteristic) to each wave. It is also widely used in pre-analysis as a method of vectorisation. Here, the TF-IDF value of each word for each document is calculated and then submitted to advanced machine learning methods. In the study which had as aim to compare the efficiency of different machine learning models in categorising tweets, Saeidi, da S. Sousa, Milios, Zeh, and Berton (2020) found that TF-IDF training vectors outperformed Word2Vec in classifying tweets.

1.2.3.3 Topic Modeling

Topic modelling in text analysis is a statistical technique used to identify topics or themes from a text document or from a group of text documents. It works by finding recurrent clusters of words in the document(s) provided. It is an unsupervised method because a labelled dataset is not provided to the topic modelling algorithm. It is a text mining technique that helps to identify the main themes of documents. A common topic modelling method used is Latent Dirichlet Allocation. Its output is usually of two types:

1. topic to term distributions which is made up of a list of topics derived from the whole corpus. The words in each topic each have value which represents their importance to the topic.
2. document to topic distributions (topics assigned to every document and how much of the topic is in the document).

In a study of data from Twitter on the COVID-19 pandemic, grouped data (by wave of the pandemic, location, gender of Twitter user) could be run through a topic modelling algorithm to identify the topics associated with each group of tweets. LDA is considered to be a type of machine learning approach (Blei, Ng, & Jordan, 2003)—but a machine learning approach to data analysis will be presented on its own later in this discussion of methods. Wesslen (2018) provides a more in-depth description and review of topic modelling use in social sciences and presents potential futures uses for causal inferences.

1.2.3.4 Principal Component Analysis

Principal component analysis (PCA) is a technique of data analysis usually performed on data with many dimensions. Dimensions can be considered as variables and the aim of PCA is to reduce the number of dimensions while keeping as much information of the dataset as possible. In the case of weekly text files gathered from Twitter, the dimensions are the words in the weekly text file. Each word could be seen as a vector which is made up of values which are the property of the word in all the documents. This property could be its raw frequency, relative frequency, TF-IDF score, a logarithmic transformation of its frequency, its position in the document, etc. The aim of PCA is to identify a few principal components by grouping the dimensions together in a way that each principal component represents as much as possible the variance of all the words (dimensions) of its constituents. PCA is similar to topic modelling in that the dimensions that make up a principal component are usually words that make up a single theme or concept expressed in the text. PCA is sometimes considered to be a method of topic modelling. For example, Chung and Pennebaker made use of PCA in their topic modelling method which they called the meaning extraction method (MEM) (Boyd, 2017; Chung & Pennebaker, 2008). The results of the PCA analysis are unusually presented graphically in a two-dimensional space which means that most of the other possible dimensions are not considered—efficiently. The amount of variance represented by these two first dimensions should always be given much importance when discussing the results. Words that cluster together represent words with similar use in the text, but isolated words that have high loading in the dimension are also very important as they show high contribution to the dimension and hence should be given importance in the discussions. PCA is also used in many different fields of study (for example oceanography, machine vision, and weather analysis) and can be adapted to the different types of datasets by methods described in more detail by Jolliffe and Cadima (2016).

1.2.3.5 Linear Regression

Regression analysis is a statistical tool that shows the relationship between a dependent variable and one or more independent variables. Linear regression assumes there is a linear relationship between the variables. It can be used as a way to predict (or estimate) the value of the dependent variable. In a situation where there are many independent variables, regression can determine which of the variables better or best predict the dependent variable. In simple linear regression, there is one independent variable and one dependent variable. In multiple linear regressions, there are two or more independent variables and one dependent variable.

The R squared value shows how much of the dependent variable is explained by the independent variable(s). Another output of regression analysis is the beta coefficient. It represents the magnitude and direction of the amount of change that a change in one unit of a given variable will bring about in the change in the dependent variable when all other variables are constant in a multiple regression model.

In simple linear regression, the R squared value is calculated by finding a regression line with a slope such that the sum of the square of the distance of the data points to the line (Sfit) is the least. Next the average of the dependent variable is calculated and the sum of the squares of the difference of the points of the dependent variable (Smean) with the mean is calculated.

R squared = (Smean - Sfit) / Smean

The process is the same for multiple linear regression with an adjustment made to account for the increased number of independent variables. In multiple regression, the model calculates R squared with and without each of the variables; if the change of R squared is big and significant, then the new variable can be included in the regression model.

Regression can be used to find out what variables predict the measured data extracted from Twitter texts. But in a complex world like the one we live in, the relationship between variables is not always linear and hence linearity of the data which is one of the assumptions of this analysis can also be its limitation. Jain, Tripathi, Dwivedi, and Saxena (2018) used multiple regression to predict the price of two cryptocurrencies. They grouped tweets by sentiment and counted the number of tweets in each sentiment group. These numbers served as input to the multiple regression model. They found R squared scores of 44% and 50% for the two currencies, which is quite low for predictions for monetary gains. Regression analysis is best done by computer software as they are very efficient and fast in finding the regression line or plane. Gordon (2015) discusses in more detail the use of regression analysis in the social sciences.

1.2.3.6 Sentiment analysis

Sentiment analysis here can be defined as the use of computational natural language processing (NLP) tools to identify and measure the affects expressed in electronic text. Most of the studies mentioned in the literature review above (for example Albahli et al., 2022; Liu, Yao, Guo and Wei, 2021; Zhang et al., 2021) that use the content of tweets to predict or explain the dependent variable do so through sentiment analysis. Sentiment analysis is a very popular NLP method. It is used not only on tweets but also on text from a variety of sources including customer reviews of products, newspaper articles, books and novels, etc. Although there are many human emotions (happiness, sadness, depression, confusion, anger, love, hate, disgust, etc), most studies just consider positive and negative emotions. The positivity and negativity of a text can be considered as separate constructs and measured independently of each text, or they can be considered as opposite extremes of the same construct and hence one can be measured and the other inferred from the measurement. Although it is expected that the values of positivity and negativity if measured on text in the same corpus should have a very strong negative correlation, this is not always the case. Although the correlations between them are usually negative, varying strengths of this correlation have been found, suggesting that positivity and negativity are not entirely opposite sides of the same construct (An, Ji, Marks, & Zhang, 2017). Humans are complex beings who can feel many emotions at the same time. Studies have reported individuals reporting being sad and happy at the same time or about the same situation (Larsen, McGraw, & Cacioppo, 2001). But the varying strengths of correlation between positivity and negativity observed can be explained by many factors, amongst which are the validity, and sensitivity of the measuring instruments.

Most tools used to perform sentiment analysis are lexicon-based and work with an algorithm to compute the sentiment score of the text based on the words from the lexicon matched in the text. The effectiveness of the tools is based on the content of the lexicon dictionary and the algorithm. The lexicon might be simple and just contain two lists of words: a list of positive words and a list of negative words with no sentiment valence score for each word. It might be accompanied by an equally simple algorithm which simply counts the number of words from the text that were matched in the lexicon lists and gives an overall classification of the text as either positive or negative based on the raw frequency of the positive and negative words matched. In most cases, the words in the lexicon have different sentiment valence scores and the algorithm computes the final sentiment score of the text by some calculation (Bhattacharya, Sarkar, Kole, & Jana, 2022). Algorithms vary in their level of complexity. Some algorithms are able to take into consideration the context of use of the words to determine for example if a word that can be considered negative is used in a positive context. A popular and more advanced lexicon-based tool is VADER. VADER has been specially designed for sentiment analysis of text scraped from social media (Hutto & Gilbert, 2014). VADER is part of the NLTK toolkit (Bird, Loper, & Klein (2009).

For sentiment analysis on grouped tweets, the researcher has to decide not only on the tool to use, but also on how to score the groups. For example, should all the tweets be merged into a single text file and the text file passed all at once through sentiment analysis? If that is the route chosen by the researcher, the researcher has to make sure the tool used is suited for sentiment analysis on large texts. If the researcher decides to pass the tweets one at a time through sentiment analysis, how to give a total score to the group of tweets is another decision to be taken by the researcher—that comes after the researcher should have chosen a sentiment analysis tool that has been proven in literature to work on short texts like tweets.

To determine if the group has an overall positive or negative sentiment, the researcher could:

1. Find the average sentiment score for the group by adding the sentiment score for each tweet and dividing the sum by the number of tweets. The researcher will then check in what sentiment range the score falls and attribute that sentiment to the group.
2. Another option will be to compare the number of negative and positive tweets and the sentiment with the most tweets will be considered to be the overall sentiment of the group (the modal sentiment). For a given group of tweets, if the negative tweets are fewer but very negative while the positive tweets are more numerous than the negative tweets but not very positive, the average score of the group might be in the negative range while the modal score will classify the group as positive. Hence two researchers who use these two methods might come to different conclusions about the same group.
3. Another option is to find the sentiment score of each tweet and sum all the calculated scores of the group of tweets and use this sum as the overall sentiment of the group.
4. Another option will be to give separate scores for positive sentiment and negative sentiment for each group. There are many ways to score each sentiment. One option is to analyse each tweet and classify it as positive or negative, then find the total number of tweets for each sentiment and report the score of each sentiment as a fraction of the total

number of tweets. Another option is to analyse each tweet and obtain a sentiment score and classify the tweet into positive and negative, then compute the average score for the positive and negative sentiment. Another option will simply be to report raw frequency scores of the number of positive and negative tweets in the group.

Above are some of the common and straightforward methods that the researcher might use for sentiment analysis of tweets. It is important to note that all methods have advantages and disadvantages. It can be noticed from the above discussion that there are many decisions to be made by the researcher in the process of sentiment analysis of the tweets. These decisions in most cases influence the outcome of the study and hence the objectivity of the research and the possible bias that ensues should be taken into consideration in all studies both by the researcher and the reader of the research publication.

There are also more complex methods for the analysis of tweets available for research as we will see next.

1.2.3.7 Machine learning approach

Machine learning can simply be defined as the use of computational algorithms to learn the data representations of the dataset (Bi, Goodman, Kaminsky, & Lessler, 2019). There are mainly two types, the supervised and the unsupervised learning models.

In **supervised learning**, the machine learning algorithm is trained on a labeled set of training data and the machine then takes the knowledge from the training data and applies it to the new data. There are two broad subtypes of supervised machine learning algorithms.

1. There are classifiers where the output is a distinct label. A use example is when the researcher wants to extract tweets of a particular topic or to classify tweets according to sentiments. The efficiency of the classifier depends very much on the training dataset provided by the researcher.
2. Regression algorithms, where the output is a continuous value. A use example is when the algorithm is used on a tweet to determine the sentiment score (usually a value in the range of -1 to 1) and not simply the sentiment type (positive, negative, or neutral). A training set made of tweets with associated scores is used to train the model.

Usually, the labelled dataset is created by independent labellers and a high percentage of agreement between the labellers is sought. The same process can be used for sentiment analysis of tweets.

In **unsupervised machine learning**, the algorithm is not given labeled training data. The full dataset is given to the algorithm for it to discover hidden patterns in the data that might be interesting to the researcher. Some of the main functions of unsupervised machine learning are:

- Associations, where the algorithm tries to find relationships between certain aspects of parts of the data,
- Clustering, this is the process of grouping similar aspects of the data together, for example, entries that have similar characteristics.
- Dimension reduction, which is similar to PCA where the algorithm removes data (features, variables) that demonstrates less variance hence making the dataset easier to understand.

Both supervised and unsupervised machine learning have many more variations. There is also semi-supervised learning which is a middle ground of both methods. Machine learning will not be used in this study and further discussion of the different types is beyond the score of this work.

1.3 Ethics of research using scraped data from Twitter

Although Twitter is generally a public platform, there are several ethical issues to be considered when doing research on data scraped from there —and from other online sources although the focus will be on Twitter.

The main ethical concern is the issue of getting consent from the users to use their published information and content. Should consent be sought from the user? If yes, how should we go about that? For large scale studies in which millions or even billions of tweets are gathered using the Twitter API, it is almost impossible to seek consent, but for studies with a smaller set of Twitter users, it might be possible to get consent (Lomborg & Bechmann, 2014). Most users make public posts on Twitter. Twitter scrapers can be set to scrape all tweets from users and bringing together those tweets can reveal identifying personal information about the user. Personal information like health, sexuality, finances, religion, political affiliation, gender, and age could be revealed when all of the user's posts on Twitter are brought together. In qualitative research, publishing information from Twitter in academic reports—even while not naming the individual account—might bring unnecessary or unwanted exposure to the user. Quoting users' posts from Twitter can make them identifiable considering that if the quote is entered in today's powerful search engines, it is very likely that the author could be identified. Twitter users are usually unaware that Twitter sells their data and might have an incorrect perception of the audience of their Tweets. This shows that they might have a false sense of privacy and security (Bernstein, Bakshy, Burke, & Karrer, 2013; Proferes, 2017). A survey in 2021 (Pew Research Center, 2021) found that 35% of adult Twitter users in the US reported their account being private or were unsure. Of these users, only 17% had private accounts.

Consent is required for all studies involving human subjects for all universities that receive federal funding in the US; hence, these universities must have an Institutional Review Board (IRB). In Canada, the Natural Sciences and Engineering Research Council (NSERC), the Canadian Institutes of Health Research (CIHR) and the Social Sciences and Humanities Research Council (SSHRC) are the three main funding agencies and together they make up the Tri-Council. The Tri-Council agencies are the main channels through which the government funds research. The funding agencies produced the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans –TCPS2 (2022) (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada, 2022a) to ensure the safety and well-being of human subjects in research. They later produced the Does Research Using Social Media Platforms Require Research Ethics Board Review? (Canadian Institutes of Health Research, et al., 2022b) to help research ethics boards to review the research involving human subjects with information collected on social media platforms.

Although there isn't a consensus among IRBs, many IRBs in the United States don't consider research with public data like that from Twitter to have human subjects based on the Federal regulations code hence most research that use data from Twitter don't report clearance from IRBs (Fiesler & Proferes, 2018). In Canada, the Tri-Council considers that research obtained from human-produced data available on social media is research with human subjects and requires clearance from IRBs with the exception, that the participants have no reasonable expectation of privacy and the information collected be in the public domain. In Mikal, Hurst and Conway's (2016) study, 26 Twitter users with and without a history of depressive symptoms discussed their views on the use of Twitter data for health monitoring in a focus group discussion. They found out that the users were generally aware of the public nature of Twitter and that their posts on Twitter could be used against them someday and so they were more mindful of what they said there. But they also expect what they say on Twitter to be lost in the mass of other tweets and retweets (theirs and others') and do not have an understanding of how much information could be gotten with the use of the Twitter API. They reported a general positive opinion for the use of publicly available data from Twitter at the population level provided that the data is properly anonymized.

In a meta study done by Takats et al. (2022) on published papers on health that used Twitter as a source of data between 2006 and 2019, they found that only a small percentage of the papers reported ethical clearance from IRBs. They identified 367 papers and of these, only 119 (32%) had ethical clearance from IRB. They also reported that 62 of these papers had identifying information of Twitter users and direct quoting of tweets. From the 367 papers, 131 papers reported attempts to anonymise the data they collected.

Most studies in which consent is obtained from the participants, the participants are usually promised anonymity. The Tri-Council agencies (Canadian Institutes of Health Research, et al., 2022a) requires the researcher to explicitly say how (or if) they will ensure anonymity. This precision and many other details about the study must be presented to the potential participant before they are asked to give their consent. Whether consent is given verbally or in writing, the researcher is legally bound by the anonymity terms of the informed consent form. The absence of this binding agreement between the researcher and the participants in most research with data scraped from Twitter and the fact that the data scraped from Twitter is already public data leads to many researchers not putting in adequate effort to anonymise data. In the hyper-networked world in which we live today, data that is not anonymized pose a greater risk because from a tweet quoted verbatim in a published study, the individual behind the account that made the tweet could easily be identified and his full name gotten. With the individual's name now known, other information like their geographical location (State and county) can be deduced from the posts they make. This additional information will in turn make it easier for more information about the individual (like their address), to be gotten by refining the search for the person by name and location on white pages websites. This then exposes the individual and their families and friends to more direct harm from trolls who are ever so present.

As previously mentioned, Twitter is recognised as the number one social media platform on which people share opinions and debate on important issues affecting society. People post on Twitter for their opinion to be heard and to be seen. It is common for online news articles to include iframes of posts from Twitter as a sample of user opinion on the subject covered. Twitter makes sharing tweets very easy and provides several options for sharing public tweets. The easiest method is to retweet a tweet. This is a core functionality in Twitter where people both in support or against a tweet can share it to their followers and comment on it. One of the other several methods of sharing tweets is sharing to other social media platforms, for example, users can share a tweet to their Instagram story (Twitter, n.d.a).

One might ask why researchers need to ask for consent from Twitter users if everyone is allowed to share public tweets as much as they want?

Academic researchers are held to a high standard of practice and respect for privacy. They understand that some users are not familiar with the risk of a tweet going viral particularly on a polarising subject. The amount of backlash that the user might receive might be too much to bear and so impact the individual negatively. This backlash can be in the form of bullying, insults, threats (even life threats), harassment, loss of employment, isolation, cancelling, etc. Ethical review and informed consent were implemented to avoid causing any harm to participants and to inform the participants of the potential risk associated with the study. This harm can be physical, mental, psychological, etc. According to Ess (2007) ethical questioning as concerns internet research should take into consideration the possibility of harm to the participants and the vulnerabilities of these participants. Although the Tri-Council agencies say research ethics board clearance is not necessary when there is no reasonable expectation of privacy and the information is in the public domain, Tri-Council also states that it is still the researcher's responsibility to identify potential risk that their research my cause to participants and work with the research ethics board to control the potential risk to participants or groups (Canadian Institutes of Health Research, et al., 2022b).

Twitter users who want to remain anonymous use aliases but mentioning only virtual identities in studies does not preclude causing harm to those mentioned because harm does not occur only when the real identity of the person behind the Twitter account is known. In virtual environments in cyberspace, many people interact with virtual identities or aliases. Even when the virtual identity is an avatar which is very different from the real identity or it is just a handle composed of e-text, people still tend to feel pain when their virtual identity is attacked. One of the earliest examples of such an occurrence was a user accusing another of rape in 1993 in an online virtual community (Burgess et al., 1998). This implies that research data should be used in such a way that the original account that posted the tweets even though it might be a virtual identity is not identifiable, particularly in the discussion of sensitive issues.

This raises another issue in the academic world. Good academic practice will require that the original author should be credited, i.e., the source of what is said has to be stated. According to the Twitter terms, the user retains copyright of what they tweet as can be seen in the text below obtained from Twitter's Terms of Service page.

> You retain your rights to any Content you submit, post or display on or through the Services. What's yours is yours — you own your Content (and your incorporated audio, photos and videos are considered part of the Content). (Twitter, n.d.b)

But crediting the author of a post without the user's consent exposes the user to unsolicited attention (whether the attention is desired by the user or not). In the case where the user would appreciate the attention from their post, not identifying (or citing) them could be considered as not giving due credit and it could be a source of pain or harm to the user if they feel exploited and treated unfairly. In the case where the user would not have wanted exposure, citing the user will bring unwanted exposure that might be harmful to the user.

From the previous discussions, it seems that the main ethical issues arise when tweets have to be quoted verbatim (or paraphrased). The proposed solution of anonymising the tweet is a double-edged sword because it can be considered as good practise because it protects the user from the possibility of pain from unsolicited attention but also implies not crediting the original source and not bringing recognition to the original user in the case where the attention is desired. The solution to this will be to ask for consent from the user. In most cases, only a few tweets from a few users are quoted in a study. Even for studies with millions of tweets, just a few tweets are quoted to prove a point and it is very doable for the researcher to seek consent from the users who posted the tweets. The consent might be for the researcher to publish identifiable information or for the researcher to use the tweets of the users without giving them credit. In both cases, the possibility of harm and risk are well explained to the user(s) (participant(s)) and they are also informed of ways to address negative outcomes—as required by IRBs.

Studies—like this one—in which no part of the text of the tweets will be quoted and the tweets are simply collected, quantitatively analysed and the results published do not need consent and even ethical clearance because of the extremely low possibility of causing harm to those whose tweets are included in the study. Additionally, this study does not meet the criteria of the University of Alberta (University of Alberta, n.d.) to necessitate an ethics approval so ethics approval was not requested. It also does not meet the criteria of the Tri-Council agencies to require ethics approval.

The Tri-Council funding agencies require that the usage of data from the social media platform does not bridge any of the terms of use of the social media platform. In 2014, Twitter added in its privacy policy that data from the platform could be used for academic research (Fiesler & Proferes, 2018). Gold (2020) reviewed Twitter's use policy as pertains to research and produced a guide to help researchers and ethics reviewers to adhere to Twitter's use policies. Twitter doesn't permit scraping so gathering a database of tweets should only be done through its API whose use comes with provisions. Gold mentions the fact that the users are informed when signing up for a Twitter account that their content will be used for research and this implies implicit consent to researchers. The users are also aware that they can make their content private at any time, at which point their use in research is no longer permitted. This is why Twitter permits sharing of tweets by the sharing of tweet ids and user ids. This shared material can then be 're-hydrated' to get the tweet text through services provided by Twitter. In so doing Twitter ensures that only tweets that haven't been deleted and only those that are still public are available to be used in the research.

Another restriction of the use of Twitter is the use of data from Twitter to identify users and their activity off Twitter. Twitter also prohibits the use of data from Twitter to derive or infer information about the user's health, finances, political affiliation, ethnicity, beliefs, sexual orientation, trade union membership and guilt about a crime. This policy would likely make research like that of Zhang et al. (2021) whose aim was to identify users with depression in bridge of use policy. But Twitter also permits research that calculates aggregates from its contents (counting the number of tweets reporting depression symptoms) if user identifiers are not stored and/or used. Gold recognises many gray areas as pertains to using data from Twitter for academic research and recommends getting legal advice for research that can be considered high risk of being in bridge of use policy.

Another ethical issue raised by the use of data scraped from Twitter is that of validity of the academic research findings obtained from Twitter data. The problem of validity occurs at many stages, from data collection to data analysis and discussion.

Prior to the change to paid access in April 2023, academic researchers only had access to 1% of the tweet traffic on Twitter. Currently the entry level paid access to the Twitter API gives access to even less tweets, about 0.3% (Stokel-Walker, 2023). It is unclear how the different topics discussed over Twitter are represented in these 0.3% of tweets. This raises the issue of bad sampling. The researchers might now have an unrepresentative sample which does not permit generalisation on the population under study. Although the number of tweets for a popular hashtag might be very great despite the limited access, the fact that the sample is such a small percentage of the population could be problematic, especially for hashtags that aren't very popular.

Twitter’s hitherto free API gave developers the ability to create bots that could post a high volume of tweets. These spam tweets could seriously bias research leading to misleading conclusions and recommendations from published research. Currently these free bots are restricted to 1500 posts per month (50 a day) which somewhat limits the ability of bot creators to saturate certain conversations with spam tweets. Spam does not only come from bots but can also come directly from humans too. Individuals might decide to repeatedly retweet and post certain opinions on a conversation. The researcher has to come up with ways to avoid scraping those tweets or ways to identify and remove unwanted tweets from their datasets. Bias does not only originate from ill intentioned individuals but could also come from the scraping process itself. The scrapers could pick up tweets that aren’t part of the topic under investigation. These could be because of inefficient scraper configuration, poor choice of keywords, keywords that are used in other topics, etc. The researcher must find ways to identify and remove these tweets. Some studies have used trained classifiers to filter out unwanted tweets while others have used NLP methods like matching known spam words to identify spam. Most of these issues cannot be completely addressed by the researcher despite their best efforts. It is the researcher’s ethical requirement to list all known and possible limits to the validity of their study.

The data used in this study was scrapped using three popular hashtags, #abhealth, #albertadoctors, and #covid19ab. Although these hashtags were among the most popular hashtags used, they represented only a small proportion of the hashtags of the tweets that were posted on the COVID-19 pandemic in Alberta. The following hashtags were very popular in the corpus of tweets gathered and it is very likely that they occurred in other tweets that did not have any of the three hashtags used to gather tweets for this study., #covidab, #yeghealth, @CMOH_Alberta, #albertadoctors, #FireKenney, #ABpoli, #FiretheUCP. Some of these popular hashtags like #FireKenney and #FiretheUCP express specific opinions which are most likely negative critique of the handling of the pandemic by government officials. Scrapping them might be a source of bias and so we stuck to the neutral and popular hashtags.

1.4 The COVID-19 pandemic on Twitter

From the third quarter of the year 2019, the world has been experiencing a global health crisis from the COVID-19 pandemic. This is the first modern day global pandemic that affected the lives of virtually everybody in modern societies.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) commonly known as COVID-19 was first identified in Wuhan in China. It spread to other countries and most countries in the world reported cases in the first quarter of 2020. By March 2020, it was declared a pandemic by the world health organization (Ghebreyesus, 2020).

Research on the COVID-19 pandemic using data scraped from Twitter spanned many areas of interest and different geographical locations. Mental health was a domain of interest as the COVID-19 caused governments to implement lockdowns, and people were isolated from loved ones. Talbot, Charron, and Konkle (2020) analyzed a small sample of tweets about the experience of pregnancy during the pandemic. Using VADER, they found that there was a general negative emotion associated with the tweets with expectant mothers reporting stress and anxiety. They complained of isolation, sleep related issues, and depression. Zhang, et al. (2021)

developed a pipeline to be used to identify Twitter users with depression by using identified keywords. They trained a machine learning classifying model to identify users with depression from the general public. Then they implemented it to be used to measure changes in depression in the general public of Twitter users. Their classifier had an accuracy of 78.9%. Koh and Liew (2022) scraped tweets by users with keywords loneliness and COVID-19 over a 60-day period starting from May1st 2020. They used machine learning (topic-modeling) and identified three themes: effects of loneliness on the community, social distance and loneliness and lastly the effects of loneliness on mental health. They compared the trend of these three themes in North America, Europe, and other regions. They reported temporal as well and some geographical variations of the three themes. For example, the topic of the effects of loneliness on mental health increased in the time period of the study.

Saha, Torous, Caine and De Choudhury (2020) used a quasi-experimental design to check the change of discourse related to mental illness resulting from the COVID-19 pandemic by comparing a corpus of the 1% of all tweets available through the Twitter API over the exact time period before and during the pandemic (March 24 to May 24th in 2019 and 2020). They used existing machine learning trained classifiers to identify tweets expressing mental health symptoms. They found that during the pandemic, such tweets increased by 14% while tweets relating to expressions of support increased by 5%. This increase was short lived and later a decline in reports of both symptomatic expressions and support was observed.

In a pandemic like the COVID-19 pandemic, it is important to know the opinion of the public on the measures that the government implements. This can also serve as a measure of the effectiveness of the policies. For this purpose, Tsai and Wang (2021) did a sentiment analysis for each day on tweets collected for three policies, the stay-at-home policy, the social distancing policy and the mask wearing policy. They tracked the change in attitude towards these policies over time. The sentiments were generally negative with a few peaks which could be explained by the major stories happening. For example, they explained peaks (which actually means an increase in positive sentiments although the graph's curve remained in the negative sentiment range) in early March to be due to the announcement of the Centre for Disease Control and prevention of the recommendation to stay at home. This suggests that there was a strong positive reaction to this order despite the overall negative sentiment caused by the pandemic in general.

Many studies that have used textual data have tried to find correlations with the discourse going on online and official metrics of the pandemic. Yan, Law, Nguyen, Cheung and Kong (2021) scraped reddit posts about three Canadian cities and among other analyses, they tested the correlation between number of comments and new daily infections. They tracked comments to posts on reddit that provided regular updates on the COVID-19 pandemic in three Canadian big cities, Vancouver, Toronto, and Calgary. They found that for Vancouver, the number of comments had a positive correlation with the number of cases. In their study on the COVID-19 discourse around masking in the US, Al-Ramahi, Elnoshokaty, El-Gayar, Nasralah, and Wahbeh (2021) first sought to find the main topics associated with anti-mask discourse and then tested the correlation of the volume of anti mask discussion with the daily COVID-19 cases. They found that the highest correlation occurred at a nine-day lag of the number of cases. This correlation was 0.77 but it should be noted that the correlation at no lag was just slightly lower. A similar study was done by Chen, Zhong and Pang (2021) in which they compared tweet volume with number of daily cases and found that the correlation was only significant for the early stages of the pandemic. Many other studies have explored the correlation between COVID-19 discourse happening online and COVID-19 statistics measured in the population. For example, Zhang, Yi, Chen & He, (2021) considered the correlation between sentiment scores, Twitter activity and new cases in four cities in the US and four cities in Canada. They ran many Pearson correlations at time lags ranging from -30 days to +30 days between number of positive cases and number of tweets, retweets, likes, replies and sentiment scores. Interestingly, they found that the peak values of the correlation between number of cases and sentiment scores, retweets, likes, and replies occurred at the no lag time point. The peak correlation between number of cases and number of tweets was at a +1-day lag. This suggests that activity on Twitter is very reactive to the COVID-19 metrics.

1.5 The current study

Alberta is one of the 10 provinces in Canada. It is considered to be the most conservative province in Canada. Like other provinces, it was severely hit by the pandemic. As of June 2023, there have been more that 633,000 diagnosed cases, more than 10 million doses of vaccine administered, and 5,769 deaths. Official statistics show that 91.1% of Albertans aged 12 and above have received at least one dose of vaccine (Government of Alberta, 2023).

Comparing other provinces and territories with data from March 2023 (The New York Times, 2023), Alberta reported 5,622 total deaths, which was the third highest number of deaths in Canada. It was behind Quebec with 18,160 deaths and Ontario with 16,234 deaths. These provinces are also more populated than Alberta. The death rate per 100,000 people was 222 for Quebec, 121 for Ontario and 138 for Alberta. The provinces with the least deaths were Nova Scotia (794) Newfoundland and Labrador (318) and Prince Edward Island (93). They had a death rate per 100,000 people of 86, 61 and 65 respectively. A complete list of the number of deaths and death rate for all provinces and territories can be found in appendix A.

For this study, Twitter was chosen as the source of the discourse. Much of the discourse around the pandemic happened online. The ease of online communication and/or expression these days makes it one of the preferred mediums of communication by most people. Twitter is also very popular among Canadians in general and Albertans in particular. A recent report states that 40% of adults in Canada have a Twitter account, with 65% of those with a Twitter account under the age of 25. There is a greater tendency for Canadians with a household income of at least $100,000 and a postgraduate degree to have a Twitter account. In Canada, Twitter has a 13.78% share of the SNS visits. Nearly eight million Canadians use Twitter (Bush, 2023). With so many active users, Twitter is a valuable source of research data.

It is common practice in recent years for people to go online to react to social events by posting about it or sharing posts made by another person to their own community or network. Brett (2011), in his study on why people share information online states that sharing information is an innate part of being human and linked to human evolution. The tendency to share information did not happen because of technology, but technology has made it easier. Online communication has the potential to reach a greater audience. One Twitter post can be read by many followers located in different geographical regions. Unlike live speech that just exists in the moment, the online posts are permanently available and can be read by people at different times. There were restrictions of in person gatherings during the COVID-19 pandemic that favoured online communications even further.

Technology has made it very easy and effective to share information such that sharing information online seems to be the first choice for people. This means that Twitter can also be used as an early indicator for a variety of phenomena and so can be used as a real time source of information from society. The tendency to share information online in real time has been encouraged—and/or has been exploited—by broadcasting agencies such that it is common to see different types of TV programs add Twitter hashtags to encourage online discussion of the program. For example, reality TV like *The Voice* and *American Idol* in the US encourage online discussion by adding the special hashtags to their broadcasts. Twitter being a microblogging platform with word character limit, users can quickly tweet their opinions and thoughts about topics as the topic unfolds. The user doesn't have to compose a whole article or voluminous blog entry. Communication on Twitter is characterised by the 3Vs (volume, velocity and variety) of big data. It is a great source of large volumes of information that is constantly being added to at a very fast rate and covers a great range of topics.

Twitter offers a kind of fly on the wall observation of society. Other ways of data collection like questionnaires, interviews, focus group discussions, and experimental testing are artificial in the sense that participants know they are being observed and even unconsciously, this affects their behavior and responses which might become more of a performance. In this sense, data from Twitter is more true and more representative of the participants' (Twitter users') true self. It is now left to the researcher to find ways to collect and use this data as objectively as possible by selecting valid variables and their indicators.

The main variables of this study were the number of patients in ICU, the wave of the pandemic, the use of pronouns referring to oneself and the ingroup (the group the person identifies with) and the use of pronouns referring to another person and the outgroup (the group the person does not identify with), and sentiment scores measured using three methods.

Like many of the studies discussed in the literature review, this study explores how sentiment scores from the discourse varied with the gravity of the pandemic. Unlike most other studies, the number of patients in ICU was used as a measure of the gravity of the pandemic. Most studies explored sentiment variations in the discourse with respect to the number of cases (Mansoor, Gurumurthy, & Prasad, 2020) and deaths (Ragothaman, & Huang, 2021). Patients in ICU are those patients that are critically ill and are usually on life support as their body is shutting down due to the effects of the COVID-19 virus. It is different from diagnosed cases because although the case number might be high, a very low percentage of these cases end up in hospital. It is also different from the number of deaths because being in ICU during the pandemic is an active situation in which the patient is fighting for their life and the overwhelmed and overworked staff is doing their best to provide care for the patient. And there was also the additional fear in the province of Alberta of running out of ICU beds and hence more fear in people as they thought they might not have beds if they or any person they care about ended up needing intensive care. Being in ICU is being in a stage of much suffering, with no effective treatment for COVID-19 and with very high chances of death. Patients in ICU better represents the dire situations brought about by the pandemic because it is a more compelling source of concern for the public because of the prospect of themselves or another person being left to die without proper medical care because of insufficient ICU beds.

In studies in the literature review, sentiment analysis was done with methods that range in complexity from simple frequency counts to machine learning. One of the aims of this study was to explore and compare different ways of measuring and/or computing sentiment scores. Many studies in literature have compared sentiment analysis methods. Some have focused on comparing lexicon-based models with machine learning while others have compared different lexicon-based tools. Bonta, Kumaresh, and Janardhan (2019) in their study of movie reviews compared SentiWordNet which is part of the NLTK Toolkit, Textblob and VADER. These three programs process text differently and have different lexicons. They found VADER to have the highest accuracy (77%) followed by Textblob (74) and NLTK's SentiWordNet (62%). For this study, three methods of measuring the sentiment scores were used.

The simple measuring method (SMM). Before the analysis is run, all the tweets are combined into a single text file. This method is the simplest lexicon-based sentiment analysis method, in which each word is independently identified as positive, negative, and neutral. The sum of the words in each category is calculated and the ratio between the categories is calculated.

The complex measuring method (CMM). This method uses a complex algorithm—which takes into consideration punctuation amplifiers, degree switches, negation prefixes, etc.—to attribute a single sentiment score to each tweet. Based on cut-off scores, the tweet is classified as positive, negative or neutral. The ratio of the number of tweets in each category is used as the output score.

The average score method (ASM). This method used the same sentiment analysis algorithm as the CMM but the tweets are not classified into sentiment groups. The average score of the group of tweets is calculated. Only one score, the average weekly score (AWS) is the output for the group of tweets.

As discussed in the sentiment analysis section of the literature review, each of these methods produce objective results but their use has different implications and validity depending on the study objectives, but the choice of one method over another is under the subjective control of the researcher. In this study, the methods used were compared and the relationship between the methods was explored. The advantages and disadvantages of the methods, and what each method brings to the analysis was discussed.

Another aim of the study was to explore how language use referring to oneself and the group identified with and language use referring to another person or a group not identified with varied with the sentiment scores calculated from the discourse on the COVID-19 related text scraped from Twitter. The COVID-19 pandemic brought about polarization in the population. There were mainly two groups that can generally be described as, those that were for and those that were against the restrictive measures put in place to fight the pandemic. This divide became even more visible with the implementation of vaccine mandates which targeted the unvaccinated. The two groups had violent interactions on online platforms (Wu, Zhao, Lu, & Chen, 2022). Othering language is usually accompanied by negative opinions about the out group (the group not identified with or not belonging to) and positive opinions about the in group (the group identified with or belonging to) (Burnap & Williams, 2016). The use of first-person pronouns could be considered to be an indicator of talking about the ingroup while the use of second and third person pronouns could be considered to refer to the outgroup. Exploring how the use of these two sets of pronouns vary with the sentiment scores of the text could be telling of how Twitter users express sentiment when talking about the ingroup and the outgroup.

It was also sought to identify which of the main variable(s) (Number of patients in ICU, Wave of pandemic, the use of pronouns) best predicted the change in the sentiment expressed on Twitter on the pandemic as measured by the AWS—which was the main dependent variable of the study.

1.5.1 Hypotheses

To guide the study, the following hypotheses were emitted.

The SMM, the CMM and the ASM are all methods of measuring the sentiment of text. Despite the difference between these methods, there is still much overlap between the methods. Because of this overlap, it is hypothesized that,

1. There will be a positive correlation between sentiment scores derived from SMM and sentiment scores derived from the CMM.

The ASM and the CMM use VADER's complex analysis algorithm although the way the scores are computed vary. Because of this it is hypothesized that,

2. For the same sentiments, the correlation between the AWS and sentiment scores measured with the CMM will be greater than the correlation between AWS and sentiment scores measured with the SMM.

The number of patients in ICU is an indicator of the gravity of the pandemic. As the number of patients in ICU rose, so did the risk of more restrictions being instated and the risk of the province running out of ICU beds. It was expected that this reflected in the discussions on Twitter and so it was hypothesized that,

3. There will be a significant correlation between the number of patients in the ICU and sentiment scores. The significant correlation between the variables will be as follows:
    - 3.1. There will be a negative correlation between patients in ICU and positive sentiment.
    - 3.2. There will be a positive correlation between negative sentiments and the number of patients in ICU.
    - 3.3. There will be a stronger correlation between the number of patients in ICU and the sentiment scores derived from the CMM compared to the sentiment scores derived from SMM.

4. There will be a negative correlation between the neutral sentiment and the number of patients in ICU.
5. There will be a negative correlation between the number of patients in ICU and AWS.

Burnap and William (2016) reported that othering language is usually accompanied by negative opinions about the othered persons. It is expected that this would reflect in conversions with mentions of second and third person pronouns. It was hypothesized that,

6. The use of the Us pronouns will correlate positively with positive sentiments and negatively with negative sentiment while the use of Them pronouns will correlate positively with negative sentiments and negatively with positive sentiments.

Pandemic fatigue (Haktanir et al., 2022), lockdown fatigue (Goldstein, Yeyati, & Sartorio, 2021) were experienced by most people during the pandemic as restrictive measures were imposed and lifted and reimposed during the waves of the pandemic. It was expected that as new restrictions were imposed during each wave, the discourse on Twitter would become increasingly negative as the population experiences the new restrictions more and more negatively.

7. The discourse during each subsequent wave will be significantly less positive than that of the previous wave.

Hate speech proliferates during crises (Persak, 2023; UN, 2020) and this hate speech is usually directed towards other people and groups. The number of patients in ICU is an objective indicator of the gravity of the pandemic. It was expected that these two variables could be the best predictors of the sentiment in the discourse on Twitter. It was hypothesized that,

8. In a multiple regression model with the Us and Them pronouns, patients in ICU and wave of pandemic, the best predictor of AWS will be the number of patients in ICU and the use of the Them pronouns.

2 Methodology and Results

For this study on the discourse of the COVID-19 pandemic in the province of Alberta, several methods of data collection, preprocessing and data analysis were used. The tools and the procedures of the data collection and analysis will be reported in this chapter. The results from the analysis will also be reported.

2.1 Data

The raw data from which the variable scores were extracted consisted of tweets scraped directly from Twitter and official statistics on the number of patients in ICU obtained from the government of Alberta's official website.

2.1.1 Data Collection

The data was gathered within a parent research project titled Guarding At-Risk Demographics with AI (GuARD-AI) started in 2020 and led by Dr. Daniel Baumgart. The project was funded by a CIFAR Catalyst Grant. The project was run in collaboration with Dr. Geoffrey Rockwell and his research team. With the aim of capturing the discourse of the COVID-19 pandemic in Alberta, data was collected from many sources. To gather the discourse that happened in news reports, the text of 2959 **news articles** was gathered between the 11th of May 2020 and the 31st of October 2021. The Premier of Alberta and the Chief Medical Officer of Health (CMOH) gave regular **COVID-19 updates** on the state of the pandemic in the province of Alberta. The texts of these updates were gathered as part of the discourse of the pandemic in Alberta. Between the 1st of March 2020 and the 31st of March 2022, we gathered 200 updates from the Premier and 249 updates from the CMOH. To get the discourse produced by the population, t**weets** were scraped from Twitter. Twarc (Summers, 2020) was used to scrape Twitter via the Twitter API. Twarc is a command-line-based module that runs in python. The scraper was set to run every 12 hours. Its output was a JSON file which contained the scraped tweets. We used three hashtags to scrape tweets from Twitter: #abhealth, #albertadoctors, and #covid19ab. We started scraping tweets on March 17, March 18 and April 21, 2020, for #albertahealth, #albertadoctors and #covid19ab respectively. We planned to scrape Twitter until October 31st, 2021, but we experienced a malfunction of our scraper and the last tweet for all the hashtags was scrapped on September 21st, 2021. All the data from the parent project is available for free at https://borealisdata.ca/dataverse/covid-discourse.

For this study, only the data scraped from Twitter was because only the text that was produced by the general Albertan public was needed for this study. The tweets were grouped by weeks that started on Mondays.

Number of patients in ICU: The government of Alberta keeps updated records of the number of patients in ICU on their website (https://www.alberta.ca/stats/covid-19-alberta-statistics.htm#data-export). This data can readily be downloaded as a CSV file with each row corresponding to a day in the COVID-19 pandemic. The data starts on the 6th of March 2020, and it is constantly being updated.

2.1.2 Preprocessing and Data Extraction

Because scraping with the three hashtags did not start at the same time, 5 weeks at the beginning which did not have tweets from all the hashtags were removed. The last week starting on September 20th 2021 was also removed because it had only two days of tweets. The final data used in this study covered a period of 74 weeks. The first week starts on the 20th of April 2020 and the last week starts on the 13th of September 2021.

The data scraped from Twitter had both tweets and retweets. There were many more retweets than tweets in the scraped tweets. The scraped tweets were grouped into weeks that started on Mondays in the period during which the scrapper was run. Each group of tweets that represented a week in the pandemic was converted into a CSV file in which each row had one tweet.

The weekly CSV files were loaded into a pandas dataframe for preprocessing. All retweets were filtered out. Tweets from popular users get retweeted a lot by other users that agree or disagree with the content of the tweet and hence might overshadow the opinion of the majority of users. Next, all the links and hashtag and mentions of Twitter accounts were removed. This was done by running a Python script that looped through each word of each tweet and removed words that started with '#', 'http' and '@'. Unlike in most other studies, stopwords weren't removed because they made up part of the words of the lexicon categories (for example personal pronouns) to be used as variables in this study. From each CSV file, a separate weekly text file was produced by combining all the text from all the tweets. These weekly text files were to be used in part of the next step which was to extract measurements of the study variables from the weekified data.

2.1.3 Tools and data extraction

LIWC22 (Boyd, Ashokkumar, Seraj, & Pennebaker, 2022) and VADER (Hutto & Gilbert, 2014) were used to get quantitative data for the analysis to test the hypotheses.

LIWC stands for Linguistic Inquiry and Word Count. It is a dictionary-based text analysis program that quantifies word use. Its dictionary is made up of about 6400 words, word stems and emoticons. Each word can belong to one or several word categories.

For each text file, LIWC outputs 4 categories of score, summary language variables (example of variables include word count, words per sentence, emotional tone), linguistic dimensions (example of variables include first person pronoun, 3rd person pronoun, prepositions), other grammar (example of variables include common verbs, quantifiers) and psychological processes (example of variables include anxiety, sadness, family, death). For this study, LIWC was used to extract measurements of first, second and third person pronouns—both singular and plural—from the weekly text files. The result output from LIWC was a CSV file with five columns that had the titles ‘i’, ‘we’, ‘you’, ‘shehe’ and ‘they’. The columns of ‘i’ and ‘we’ were combined by calculating their mean to make a single column representing the ‘Us’ variable which measures the use of first person singular and plural pronouns. The ‘you’, ‘shehe’ and ‘they’ columns were combined by calculating their mean to make a single column representing the ‘Them’ variable which measures the use of second and third person singular and plural pronouns. This was done to get a score for the mentions of oneself and one’s group and the score for the mentions of the other and the other’s group respectively.

VADER (Valence Aware Dictionary and sEntiment Reasoner) was used to get sentiment scores from the weekly CSV data loaded into Python and from the weekly text files. VADER has a lexicon and a processing algorithm that is specifically designed for—short—microblogging texts. This makes it one of the best options for sentiment analysis for data from Twitter (Hutto & Gilbert, 2014). Its creators were aided in their design of the lexicon by studying other well-established lexicons like LIWC’s. VADER can recognise happy emoticons (for example, :D, :), happy faces emojis and happy sentiment acronyms (for example lol which stands for laughing out loud) as positive emotions. When a text is analysed using VADER, the output is a dictionary object with four key and value pairs. The keys are pos, neg, neu, and compound which represent the ratio of text in the positive category, the ratio of text in the negative category, the ratio of text in the neutral category, and a general sentiment score of the text normalised to fall between -1 and 1. The pos, neg and neu scores are obtained from a simple word count of the words in the text belonging to each category. The compound score on the other hand is obtained by using a more intelligent algorithm that understands the context of the usage of words and considers the valence of each word before a general score is attributed to the text.

Three methods were used to get sentiment scores from each week in the study sample. The simple measuring method (SMM), the complex measuring method (CMM), and the average score method (ASM) were used.

To get the sentiment scores using the simple measuring method (SMM) the weekly text files were passed though VADER and for each file, the pos, neg and neu scores were extracted. This gave the ratio of positive, negative and neutral words in that week of text. These variables were named Pos_lex, Neg_lex and Neu_lex respectively. The suffix 'lex' was used to indicate that the scores were obtained from simple lexicon-based word count analysis.

To get sentiment scores using the complex measuring method (CMM), VADER was used on each tweet of the weekly CSV file and the compound score was used to determine if the tweet was positive, negative or neutral. The compound score ranged from -1 to 1. If the score was less than -0.05, then the tweet was classified as negative, if the score was between -0.05 and 0.05, then the tweet was classified as neutral and if the score was greater than 0.05 then the tweet was classified as positive (Hutto & Gilbert, 2014). For each weekly CSV file, three scores —positive, negative and neutral—were calculated. These were obtained by dividing the total number of tweets in each category by the number of tweets for that week. The variables were named Pos, Neg and Neu. Hence, the sum of the scores of the Pos, Neg and Neu variables for each week was 1 since the scores were ratios of the number of tweets in each category.

Finally, the average score method (ASM) was used to get the average weekly sentiment score (AWS). This was calculated by dividing the sum of the compound scores for each tweet by the total number of tweets in the week. Because VADER does not work well for large text files, the compound scores for each week could not be gotten from the weekly text file (created by combining all the tweets of the week into a single big text file) and so it had to be calculated from the sentiment score from each tweet. When the large weekly text files were passed to VADER for sentiment analysis, the compound scores produced by VADER were extreme values very close to 1 and -1.

Wave of pandemic. The COVD-19 pandemic was characterised by waves which can be described as periods of high transmission and cases. Some factors that can influence the length of a wave of the pandemic are restrictive measures which usually reduce transmissions and bring down the number of identified cases and the emergence of new variants which are usually more contagious and increase the number of cases. There are many criteria that can be used to determine the periods of the waves (Ayala et al., 2021) but for this study, visual inspection was used to estimate the start and end date of each wave. The number of diagnosed cases histogram (can be seen in appendix B) obtained from the government of Alberta's website (https://www.alberta.ca/stats/covid-19-alberta-statistics.htm#total-cases) was used to estimate the start of each wave. The boundary between each wave was the point in the histogram between two peaks where there was the least number of cases. The second wave was considered to start on September 28, 2020. The third wave was considered to start on February 22, 2021. The fourth wave was considered to start on July 12, 2021.

2.2 Analysis and Results

To test the hypotheses emitted and to explore the data, visualisations and statistical analysis tools were used. For exploration by visualisation, a word cloud and a line graph were used. The statistical analysis tools used were correlation, ANOVA and multiple linear regression.

2.2.1 Descriptives statistics of the study variables

The descriptive statistics of the study variables consisting of the minimum value, maximum value, the mean, the standard deviation, and the significance score for the Shapiro-Wilk test for normality are presented in table 1 below.

Table 1

Descriptive statistics of the main study variables.

Variable	Min value	Max value	mean	Standard deviation	Normality Shapiro–Wilk test
US	.91	1.79	1.37	.18	.608
THEM	.59	1.15	.84	.13	.467
ICU	4.86	215.57	59.91	54.21	**.000**
POS	.31	.50	.39	.04	**.032**
NEG	.02	.37	.28	.04	.940
NEU	.25	.39	.33	.03	.087
TOT_AVG	-.04	.16	.06	.05	.561
POS_LEX	.11	.15	.13	.01	.474
NEG_LEX	.07	.14	.1	.02	.923
NEU_LEX	.73	.80	.77	.02	.054

2.2.2 Word cloud

Voyant-Tools (Sinclair & Rockwell, 2016) was used to create a word cloud of the first 205 most frequently used words in the cleaned text from Twitter. Voyant Tools (VT) is a suite of tools available for free online at www.voyant-tools.org for the analysis and visualisation of texts. The tool Cirrus was used to create the word cloud. The size of each word in the word cloud is a function of the frequency of the word in the text. The higher the frequency of the word in the text, the bigger the word in the word cloud. The color and orientation of the words are randomly attributed by the tool. The word cloud can be seen in figure 1 below. It should be noted that Voyant-Tools' default stopword list was applied on the text before the visualization was produced. The corpus of Twitter tweets can be found here https://voyant-tools.org/?corpus=6887fe68d8785dd29c736a8ba93cc73c.

Figure 1

The first 205 most frequent words in a word cloud

Note: The colour and position of the words in the word cloud are randomly attributed

It should be noted that the hashtags and the mentions of Twitter user accounts and URLs were not included in the text used to produce the word cloud. The word cloud suggests that the gathered tweets are effectively about the COVID-19 pandemic in Alberta. The high frequency words include Alberta, cases, covid, vaccine, health, Kenney, government. Table 2 below shows the first 140 words with highest frequency.

Table 2

First 140 high frequency words

Rank	Term	Count	Rank	Term	Count	Rank	Term	Count	Rank	Term
1	alberta	41539	36	hospital	7479	71	live	4732	106	community
2	covid	37597	37	apply	7337	72	way	4724	107	vaccines
3	cases	30569	38	total	7325	73	week	4708	108	thank
4	health	25384	39	school	7064	74	kids	4690	109	wave
5	people	21174	40	responsibilities	6877	75	measures	4657	110	protect
6	new	20977	41	duties	6847	76	doctors	4591	111	death
7	just	15286	42	family	6809	77	doing	4558	112	watch
8	care	14980	43	itâ™s	6756	78	support	4557	113	reported
9	albertans	14052	44	work	6696	79	say	4510	114	virus
10	kenney	14013	45	icu	6491	80	medical	4431	115	read
11	today	12872	46	jason	6447	81	year	4430	116	physician
12	like	11948	47	dr	6372	82	long	4386	117	shandro
13	vaccine	11763	48	good	6333	83	i'm	4325	118	look
14	time	11660	49	mask	6180	84	staff	4319	119	great
15	need	11052	50	numbers	6123	85	really	4308	120	dose
16	canada	10860	51	want	5971	86	safe	4295	121	nurses
17	pandemic	10561	52	ab	5945	87	plan	4287	122	second
18	public	9720	53	medicine	5874	88	stop	4277	123	come
19	government	9579	54	workers	5774	89	data	4254	124	risk
20	day	9406	55	premier	5606	90	alberta's	4149	125	south
21	active	9114	56	make	5606	91	weeks	4088	126	number
22	calgary	9048	57	services	5599	92	general	4019	127	contact
23	edmonton	9046	58	schools	5496	93	lives	4014	128	we're
24	deaths	8842	59	think	5467	94	iâ™m	4002	129	high
25	zone	8780	60	healthcare	5390	95	getting	3998	130	needs
26	know	8734	61	help	5386	96	testing	3981	131	hey
27	province	8375	62	vaccinated	5300	97	tests	3908	132	daily
28	it's	8085	63	open	5227	98	got	3904	133	can't
29	update	8054	64	case	5075	99	doses	3853	134	vaccination
30	going	8034	65	said	5029	100	news	3826	135	patients
31	ucp	7957	66	days	4926	101	better	3736	136	city
32	restrictions	7922	67	home	4910	102	stay	3681	137	months
33	hinshaw	7845	68	masks	4896	103	minister	3649	138	test
34	says	7610	69	donâ™t	4877	104	spread	3630	139	response
35	right	7487	70	ahs	4829	105	variant	3629	140	rate

The main government officials that provided updates were the Chief Medical Officer of Health of the province (CMOH) Dr. Deena Hinshaw, the Premier Jason Kenney and the Minister of Health Tyler Shandro. The CMOH gave more updates than the Premier but, comparing the size of their last names it was noticed that the Premier (Kenney) was more discussed in the tweets than the CMOH (Hinshaw). The Premier's first name (Jason) is also bigger than the CMOH's first name (Deena). Both the first and last name of the Minister of Health (Tyler Shandro) had less mentions than the first and last names of the CMOH respectively. The Terms tool on VT shows the frequency of words used in the corpus. Table 3 below shows the raw frequency of the first and last names of the government officials. The Terms tool confirms the observations from the word cloud. It also suggests that the public preferred using the government official's last names in their discussions. The Premier had the highest mentions, followed by the CMOH and then the Minister of Health.

Table 3

Frequency of mentions of the names of the Premier, Minister of Health and the Chief Medical Officer of Health for the province of Alberta

Government Official	First name	Last name	Total
The Premiere	6447(6524)	14013(15854)	20460(22378)
CMOH	2371(2398)	7845(8302)	10216(10700)
Minister of Health	1004(1021)	3230(3604)	4234(4625)

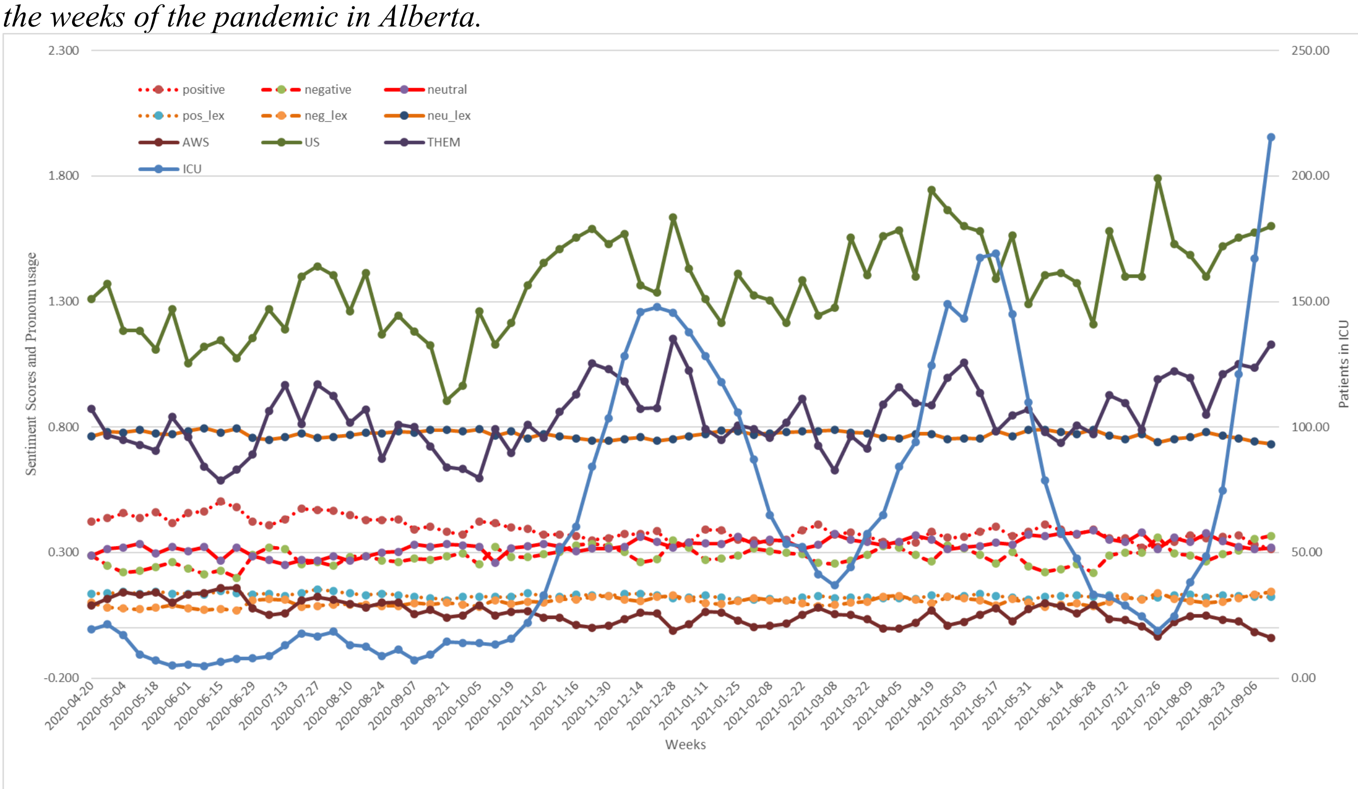
Note. Numbers in brackets refer to the frequency of the names with suffixes—for example, apostrophe s.

2.2.3 Line Graph

A graph is a method of visualisation representing data from two variables. It is usually ordered according to the variable on the x-axis value. Usually, the independent variable is on the x-axis while the dependent variable is on the y-axis. The points on the graph are joined by a line and the graph can be used to visualize the trend in the data that might not be obvious by just looking at a list of numbers. In this study, all the quantitative data was loaded into excel to create a line graph to show the variation of the study variables over the weeks of the pandemic. A two vertical axes graph was created because of the difference in range of values of the study variables. Figure 2 below shows the line graph with weeks (time) on the horizontal axis; number of patients in ICU tracked on the right vertical axis; and Us and Them pronouns and sentiment scores tracked on the left vertical axis.

Figure 2

Changes in Us and Them pronouns, sentiment scores, and the number of patients in ICU over the weeks of the pandemic in Alberta.



From Figure 2 above we notice that Us and the Them pronouns have very similar variation patterns suggesting their use follows the same pattern. The AWS line despite its oscillation shows a gradual downward slope towards the right suggesting that the overall sentiment becomes less positive over time.

It is not very objective to use graphs to draw conclusions on the relationship between variables since most graphs are subject to bias both in their design and their interpretation. In the design of the graph, the choice of the scale, choice of visual aspects (for example color and line texture) are some aspects of the graph that the researcher can implement consciously or unconsciously to favour a desired interpretation. More objective statistical procedures will be done next.

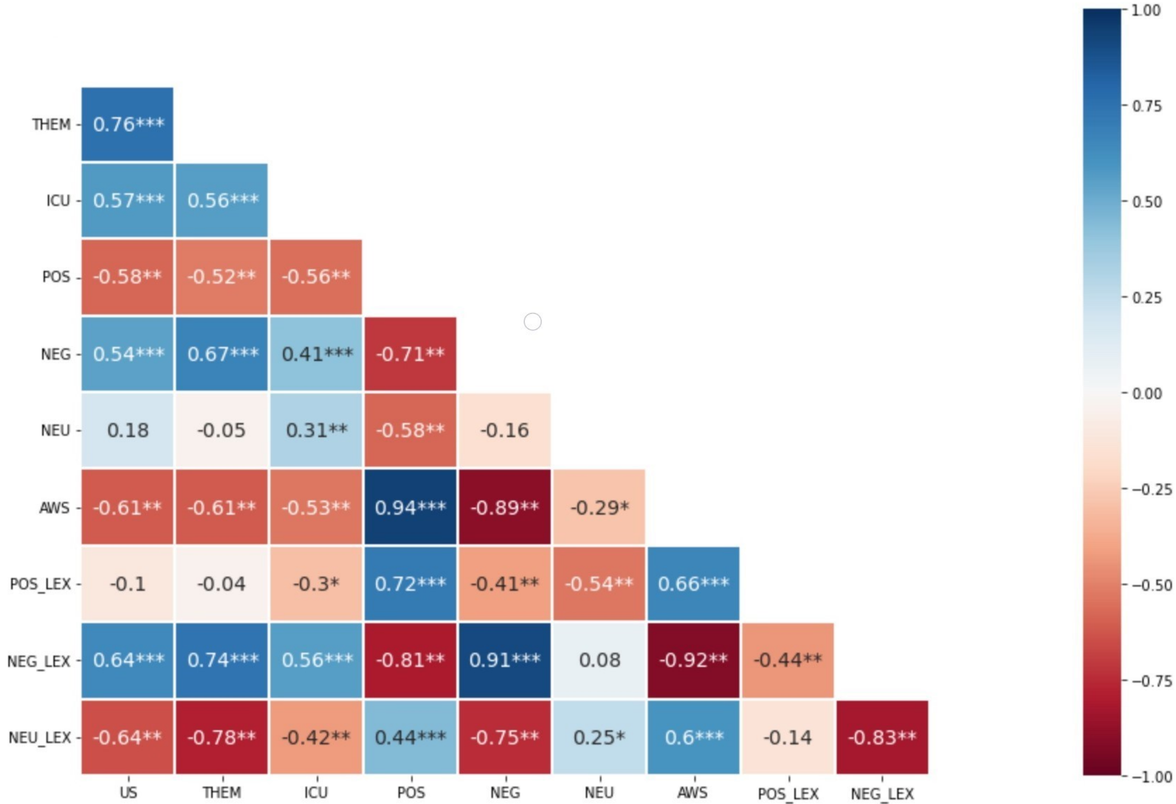
2.2.4 Correlation

A bivariate Pearson correlation was run with the main study variables to observe the relationship between them. The correlations between sentiment measures, Us and Them pronouns and number of patients in ICU were calculated. The results can be found in Figure 3 below.

From Figure 3, we notice that most of the variables correlate with each other in varying degrees and in different directions. The correlations mentioned below are those that permitted the test of this study's hypothesis.

Figure 3

Correlation between the study variables



Note. * Correlation if significant at .05 level, ** correlation significant at 0.01 level, *** correlation significant at 0.001 level

There was a significant and positive correlation between pairs of sentiment scores obtained by using the SMM and the CMM. The correlation between the positive sentiment measures was r= .72, n=74, p=0.001. The correlation between the negative sentiment measures was r=.91, n=74, p=0.001. The correlation between neutral sentiment measures was r=.25, n=74, p=0.05.

Positive sentiment measured with the CMM had a very strong correlation with AWS r=.94, n=74, p=0.001 while that measured with the SMM had a correlation of r=.66, n=74, p=0.001. The negative sentiment scores had similar correlations with AWS. The scores measured with the CMM had a correlation of r=-.89, n=74, p=0.01 while the score measured with the SMM had a correlation of r=-.92, n=74, p=0.01.

There was a significant negative correlation between measures of positive sentiment and the number of patients in ICU. The correlation between the number of patients in ICU and the CMM sentiment scores was r=-.56, n=74, p=0.01 while that with the SMM sentiment scores was r=-.3, n=74, p=0.05. So there was a stronger correlation between the number of patients in ICU and CMM positive sentiment scores. There was a significant positive correlation between measures of negative sentiment and number of patients in ICU. The correlation between the number of patients in ICU and CMM sentiment scores was r=.41, n=74, p=0.01 while that with the SMM sentiment scores was r=.56, n=74, p=0.01. So there was a stronger correlation with SMM sentiment scores for the negative sentiment.

Number of patients in ICU had a positive correlation with neutral sentiments measured by the CMM r=.31, n=74, p=0.01, and had a negative correlation with the neutral sentiment measured by the SMM r=-.42, n=74, p=0.01.

There was a negative correlation between the number of people in ICU and the average weekly score (AWS), r=-.53, n=74, p=0.01.

The use of the Us and Them pronouns had similar correlations (in direction and in magnitude) with the same sentiment measures. The Us and THEM pronounces had a negative correlation with CMM positive sentiment scores, r=-.58, n=74, p=0.01 and r=-.52, n=74, p=0.01 respectively. The US and Them pronouns had a positive correlation with CMM negative sentiment scores, r=.54, n=74, p=0.001 and r = .67, n=74, p=0.001 respectively. The US and Them pronouns had a negative correlation with the SMM positive sentiment scores but it did not reach statistical significance, r=-.1, n=74, p=.374 and r=-.04, n=74, p=.723 respectively. The Us and Them pronouns correlated positively with the SMM negative sentiment scores, r=.64, n=74, p=0.01 and r=.78, n=74, p=0.01 respectively.

2.2.5 ANOVA

A one-way ANOVA was run to compare the AWS during each wave of the pandemic. Descriptive statistics of the four groups (waves) can be found in table 4 below. To test if there was a difference in the overall sentiment in each wave, a one-way between groups ANOVA was performed. The Shapiro-Wilk test was used to ensure that each group did not significantly deviate from a normal distribution. Additionally the assumption of homogeneity of variance was satisfied based on Lavene's test, $F(3, 70)=.298$, $p=.827$.

There was a main effect of the independent variable (wave of pandemic), $F(3, 70)=27.449$, $\eta 2=.544$, $p<0.001$ which means that there was a significant difference in the average sentiment in the different waves of the pandemic in Alberta. It was found that 54.4% of the variance in the AWS was accounted for by the wave of the pandemic. A post-hoc Tukey's HSD test with homogenized group sizes was used to evaluate the difference in the means of the individual groups. There was a significant difference between the average sentiment in wave 1 (mean=.104) and wave 2 (mean=.036, $p<0.001$), wave 3 (mean=.050, $p<0.001$) and wave 4 (mean=.01, $p<0.001$). There was a significant difference between the average sentiment score in wave 3 (mean=.050) and wave 4 (.012, $p=0.015$). There was no significant difference in the overall sentiment between wave 2 (mean=.036) and wave 3 (mean=.050, $p=0.528$) and between wave 2 (mean=.036) and wave 4 (mean=.012, $p=0.197$). The results are presented graphically with the means plotted against the wave in Figure 4 below.

Table 4

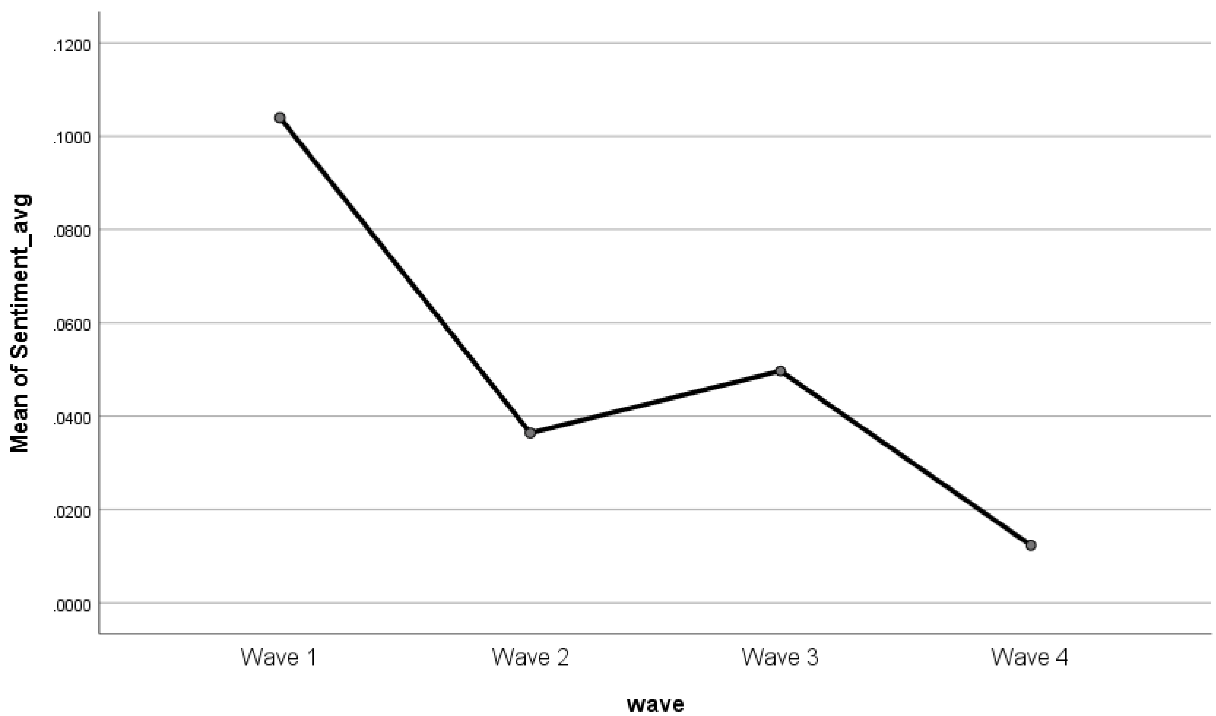
Table showing number of weeks and mean average weekly score AWS scores with standard deviation for each wave of the pandemic.

Group	N	Mean	Standard deviation
Wave 1	23	.104	.034
Wave 2	21	.036	.028
Wave3	20	.050	.031
Wave 4	10	.012	.045

Note. N is number of weeks in wave

Figure 4

Chronological change in the average weekly score (AWS) during each wave of the pandemic.



2.2.6 Multiple Regression

Multiple linear regression is a statistical technique that measures the prediction of the dependent variable by the independent variables. It follows the same principle as a simple linear regression in which there is only one dependent variable and one predictor variable. In simple linear regression, the prediction is done with the use of a straight line with a gradient such that the sum of the square of the distance of the data points to the line is least. In multiple linear regression, there are many predictor variables, and the prediction is done using a high dimensional object like a plane. Most computer softwares that calculate multiple linear regressions run multiple regression models with and without each predictor variable and compare the change in r squared and p (measure of significance) values to determine which variables are the best predictors and to determine if the addition of the predictor variables significantly increases the prediction of the model.

A multiple regression in which the wave of pandemic, the number of patients in ICU and the use of Us and Them pronouns were predictors (independent variable), and the AWS was the dependent variable was run to discover which independent variable(s) was/were the best predictor(s) of the AWS. Preliminary analyses were conducted to ensure no violations of assumptions of normality (the dependent variable is normally distributed), linearity (there is a linear relationship between the variables), multicollinearity (very high correlation between the predictor variables) and homoscesdesticity (the dependent variable has similar variance at different levels of the predictor variables).

The overall result showed that the model was significant $F(4, 69)= 20.775$, $R2=.546$, $p<0.001$. In this model, only the wave of pandemic and the Them pronouns significantly predicted the average weekly sentiment score. The results are summarized in table 5 below.

Table 5

Multiple regression result table showing the prediction of each variable on dependent variable (AWS)

Variable	B	SE B	β
wave	-0.017	.005	-.385*
ICU	.000	.000	-1.39
Us	-.008	.038	-.032
Them	-.12	.044	-.347**

Note. R^2=.546, *p<.001, **p=.009

The wave of the pandemic was the strongest predictor (β=-.385, p<.001). This shows that as new waves of the pandemic came about, the overall weekly sentiment became less positive. The Them pronoun also significantly predicted the overall weekly sentiment (β=-.347, p=.009). As the usage of the Them pronouns (pronouns referring to the other and the other's group) increased, the sentiment in the tweets became less positive. It should be noted that prediction does not imply causation. It should also be noted that this regression solution is only valid for this set of variables in this model.

3 Discussion

The general aim of this study was the exploration of the COVID-19 discourse in Alberta. Tweets on the COVID-19 pandemic were scraped from Twitter and grouped into weeks. These tweets were then submitted to sentiment analysis using the average score method (ASM) and an average weekly sentiment (AWS) score was obtained for each week. Two other sentiment analysis methods or procedures were used to get other sentiment scores. A simple lexicon-based method which was called simple measuring method (SMM) in this study and a complex measuring method (CMM) which is also lexicon-based but uses intelligent algorithms that identifies the context of use of words in the tweets. The other study variables were the average number of patients in ICU for the week, the wave of the pandemic, and the usage of Us pronouns (first person pronouns) and Them pronouns (second and third person pronouns). A number of hypotheses were proposed, and statistical analysis was done to test the hypotheses.

The first hypothesis stated that there will be a positive correlation between the same sentiments measured with the SMM and with the CMM. The result of the Pearson correlation confirmed this hypothesis. There was a strong positive correlation between measures of the positive and negative sentiment while the correlation between the measures of the neutral sentiment was significant but weak. The SMM of sentiment analysis was run by grouping all the tweets for the week into a single text file and passing the text file through VADER's simple lexicon sentiment analysis algorithm which identifies positive, negative, and neutral words and computes a ratio of their frequency as strength of these sentiments. On the other hand, the CMM was run by passing each tweet at a time through VADER's complex sentiment analysis algorithm and classifying the tweet as positive, negative or neutral if the compound score of the tweet was greater than .05, less than -.05 and in the range between -.05 to .05 respectively. The total number of tweets in each category was used to find the ratios which made up the score for each sentiment. It can be noticed that using the CMM, the range of scores for which a tweet is classified as neutral is quite small hence less tweets will be classified as neutral. Additionally, the neutral content (words) of tweets classified as positive or negative is ignored in the CMM since the tweets can only be classified in one sentiment group. On the other hand, in the SMM, the neutral content of the tweets are identified, quantified and are represented in the final neutral score. From Table 1 we see that the average neutral sentiment score gotten from the SMM is more than twice that of the average neutral score gotten from the CMM. It might not be surprising then that the neutral sentiment group had such low correlation considering the different ways in which the neutral sentiment scores were calculated.

As mentioned above, the sentiment score for each week was measured in three different ways, the SMM and the CMM just compared above and the ASM that produced the AWS obtained by summing the sentiment score from each tweet (obtained by using VADER’s complex algorithm) and dividing the sum by the number of tweets for that week. Of the three measurements of sentiment, the average weekly sentiment score (AWS) was considered to be the main sentiment score of the study against which the scores from the other methods are compared. The AWS also seems to be the most valid measure of the sentiment of the weekly groups of tweets considering that the value of the sentiment score of each tweet is taken into account and not the overall sentiment group of the tweet. Unlike the other methods, the ASM produces a single score that represents the average positivity score of the week. The next best method seems to be the CMM since it also used VADER’s complex algorithm and the least valid sentiment scores are those obtained from the SMM. This conjecture led to the second hypothesis which states that, if compared with the correlations with the scores obtained by the SMM, the correlations between the sentiment scores from the CMM will have a stronger correlation with the AWS and the correlation with the positive sentiment scores will be positive and that with the negative sentiment scores will be negative.

The results from the correlation analysis showed that, for positive sentiment scores the correlation with the score from the CMM was greater than that from the SMM. The difference between the two scores was 0.3. For negative sentiment scores, the scores from the SMM had a slightly stronger negative correlation. The difference between the two correlations was just .03. So the hypothesis was only partly verified. But overall, this study suggests that the CMM is a better method compared to the SMM for measuring the sentiment from a group of tweets considering that it had overall a stronger correlation—i.e. the sum of the absolute values of correlations will be greater for the scores gotten from the CMM.

Interestingly, for the neutral sentiment, the CMM scores had a weak negative correlation with the AWS while neutral SMM scores had a moderate positive correlation with AWS. This means that, as the AWS increases, less tweets are identified as neutral with the CMM and vice versa. This is expected considering that the increased—positive—polarization of tweets that drive an increase of the AWS likely also explains the fact that less tweets are falling in the small score range for the neutral tweets category. The positive correlation between neutral scores obtained from the SMM and AWS means that as the AWS score increases (increase in positivity), the percentage of neutral words in the tweets also increases. This suggests that more neutral words are used to express positivity and less neutral words are used to express negativity.

The third hypothesis states that there will be a negative relationship between the number of patients in ICU and positive sentiments and a positive relationship between the number of patients in ICU and negative sentiments. This hypothesis was true for sentiment scores obtained from both the CMM and the SMM. It was also hypothesized that the correlation between CMM sentiment scores and patients in ICU will be greater than the correlation between SMM scores and patients in ICU. The assumption behind this hypothesis was that compared to the SMM, the CMM was a more valid measurement method. This hypothesis was only valid for the positive sentiment score and not the negative sentiment score. This suggests that the SMM is a more sensitive measure of negative sentiment in tweets which occur in response to changes in the number of patients in ICU. On the other hand, the CMM is a more sensitive measure of changes in positive sentiment which occur as a result of changes in the number of patients in ICU. This also suggests that as the number of patients in ICU increases, there is a greater use of negative words even if they are not always used to express negative sentiments—because the number of tweets that are classified as negative increases at a slower rate. So when the number of patients in ICU increases, the writing styles among Twitter users becomes more negative (as measured by the SMM) than how the text actually conveys negative sentiment (as measured by the CMM). On the other hand, when the number of patients in ICU reduces, Twitter users convey more positive sentiment (as measured by the CMM) than how their writing style becomes more positive (as measured by the SMM). This might be explained by the fact that the COVID-19 pandemic is a phenomenon that is experienced negatively by Twitter users but they try to stay positive—considering that the AWS mostly stayed above zero and the mean was 0.06. When the conditions of the pandemic become better, Twitter users are quick to post more positive tweets but the change to the use of positive words does not occur at the same rate as they are cautiously positive. But when the conditions of the COVID-19 pandemic become worse, the Twitter users tend to use more negative words but they try not to be overly negative. They change to post more negative tweets but not as much as they change to use more negative words.

In hypothesis 2 above, the SMM negative sentiment scores had a stronger correlation with the AWS compared to the correlation of the CMM negative sentiment score with the AWS. In hypothesis 3.3, the SMM negative sentiment scores had a stronger correlation with the number of patients in ICU compared to the correlation between CMM negative sentiment scores and the number of patients in ICU. This seems to confirm the explanation proposed above that states that for this study the SMM which is a simple lexicon-based method for sentiment analysis is a better method for measuring negative sentiments while the CMM is a better method of measuring positive sentiment.

A better explanation of this finding might be that, as more people were admitted into ICU during the COVID-19 pandemic in Alberta, there was a general increase in the usage of negative words on Twitter but the use of those negative words was not only to express negative sentiment, but they were also used was to express neutral or even positive sentiments. As explained in VADER's GitHub page, two authors could express the same sentiment and/or opinion on the same topic but one author might use more negative words than the other. VADER's simple lexicon-based sentiment analysis will produce very different positive and negative sentiment scores for these authors. These scores are more reflective of the authors' writing styles than the authors' sentiments. So the Twitter users' writing style changes to a more negative writing style with an increase of patients in ICU and it was accompanied by a less proportionate increase in negative sentiment.

The fourth hypothesis states there will be a negative correlation between the neutral sentiments scores and the number of patients in ICU. It was expected that as the COVID-19 pandemic worsened, the conversations on Twitter would become more polarized, hence less neutral. The neutral scores measured with the CMM showed the opposite of what was hypothesized. It had a positive correlation with the number of patients in ICU. This means that as the number of patients in ICU increased, the number of tweets that were identified as neutral increased. On the other hand, as hypothesized, the neutral sentiment scores measured with the SMM had a negative correlation with the number of patients in ICU. This means that as the number of patients in ICU increased, the number of neutral words used in tweets decreased—and vice versa.

A possible explanation of the observed positive correlation between CMM scores for the neutral sentiment and the number of patients ICU is that, the rate at which Twitter users switched from posting positive tweets to posting neutral tweets as the number of patients in ICU increased was higher than the rate at which Twitter users switched from posting neutral tweets to posting negative tweets. But because the CMM uses VADER's complex algorithm to categorize tweets, the reduction of neutral word use as measured by the SMM mustn't be accompanied by the reduction of the overall number of tweets that are classified as neutral. This can be explained by the—better—explanation provided to the results of hypothesis 3.3 above. So this finding also suggests that the higher presence of polarized words (non neutral words) was accompanied by a change of usage of the polarised words.

Number of patients in ICU had different strengths of correlations with measures of positive and negative sentiment scores measured by the SMM and the CMM. The correlation of the positive sentiment measured with the CMM was greater than that measured with the SMM while the correlation of the negative sentiment measured with SMM was greater than that measured with the CMM. This means there were always variations in the use of words (positive, negative, and neutral) that were not transmitted in the perceived sentiment of the text as measured by the

CMM. Because the measures of positive, negative, and neutral sentiments were reported as proportions of a whole, it is this residual proportion of variation that appears strongly as opposite directions of correlation of the sentiment score of the neutral sentiment measured by the SMM and the CMM with the number of patients in ICU.

The fifth hypothesis stated that there will be a negative correlation between the number of patients in ICU and the AWS. There was a significant, moderate, and negative correlation between the number of patients in ICU and AWS. This means that as the number of patients in ICU increased, the average sentiment of the week became less positive and vice versa. This is expected considering that the increase in number of patients in ICU—which in itself means an increase of the number of people suffering and at high risk of dying, an increased strain on the medical system, medical services not available or delayed for other patients, risk of more public restrictions—is associated with many negative consequences which are discussed online. The polarized groups of opinions that formed during the COVID-19 pandemic were constantly clashing on issues like restrictions, lockdowns, and vaccine mandates. As the pandemic worsened as—indicated by the number of patients in ICU—these clashes normally are expected to be more frequent. Clashes are usually associated with negative language (Uyheng & Carley, 2021) which are reflected in the AWS. In the province of Alberta, increases in the number of patients in ICU led to a shortage of ICU beds. A high percentage of those in ICU were unvaccinated. In the politicised pandemic, this fueled more negative critique of the unvaccinated who were accused of overwhelming the medical personnel and using up medical resources that could have been used for other patients who had been waiting long for medical services like surgeries. Those in support of people’s choice to remain unvaccinated also responded to these critiques in negative terms. People who saw more restrictions and lockdowns as a solution to fight the pandemic clashed with those that were against those measures. People who thought that unvaccinated individuals should be put under more restrictions clashed with those that were of the opinion that these restrictions were unethical and discriminatory. These more frequent clashes caused the tweets scraped from Twitter to be less positive as the number of patients in ICU increased.

The sixth hypothesis stated that the use of the Us personal pronouns will correlate positively with positive sentiments and negatively with negative sentiment while the use of Them pronouns will correlate positively with negative sentiments and negatively with positive sentiments. It was hypothesized this way because it was expected that in the politicised COVID-19 environment, Twitter users would write more positive tweets when writing about themselves and a group to which they belong, while on the other hand, Twitter users would write more negative tweets when they write about others and the group the others belong to. It was observed that the frequency of use of Us and Them pronouns had similar correlation strengths and directions with positive and negative sentiment scores. They both correlated negatively with positive sentiment scores but their correlation did not reach significance for the SMM positive scores. They correlated positively with the negative sentiment scores. This means that, as the use of Us and Them pronouns increased, the tweets became more negative and vice versa. The use of the Us and Them pronouns also correlated negatively with the AWS (the average weekly score which is a single sentiment score for the week).

The observed negative correlation between the Us and Them pronouns use and AWS suggests that Twitter users tweeted more negatively when referring to someone even if that person was themself or the group to which they belonged. This can be understandable when we consider that negative tweets are not always directed at someone else. Negative tweets could be a user complaining about the negative effects of the pandemic on him/herself and it could also be a user stating her/his opinion for example, that the government officials are managing the pandemic poorly. Interestingly the use of the Us and Them pronouns correlated negatively with the score of the use of neutral words as measured with the SMM. This means that as the use of the Us and Them pronouns increased, the use of neutral words decreased. This suggests that as Twitter users used more personal pronouns, thereby referring to people, they expressed opinions about those who they were referring to and these opinions are not neutral and more likely negative than positive. It should also be noted that the use of Us and Them pronouns correlated positively with the number of patients in ICU. This suggests that as the COVID-19 pandemic made more people critically sick, Twitter users started using more personal pronouns as they also tweeted with more negative words.

The seventh hypothesis stated that as the pandemic evolved, the discourse around the pandemic in Alberta gradually became less positive. It was expected that in the four waves covered in the period considered in this study, the positive sentiment will significantly lessen from one wave to the other. Results from the between subjects ANOVA (independent group ANOVA) showed that there was a significant effect of the wave of the pandemic on the AWS. Post hoc Tukey's HSD test showed that—and as Figure 4 suggests—the discourse around the first wave was the most positive. It was significantly more positive than the discourse during all the other waves. The discourse became less positive during the second wave and stayed at approximately the same level during the third wave. Then during the fourth wave, the discourse became significantly less positive than during the first and third waves. This suggests that it might have been the case that at the start of the pandemic, most Albertans were experiencing their first pandemic and lockdown and although there were many people and groups against the lockdowns there was also a lot of general interpersonal support and encouragement with slogans like, we are in this together, stay home and save lives. The pandemic and the lockdowns were new and intriguing although it was also scary. Most people did not know what to expect. The first lockdown was hard and many people's mental health was affected (Saha, Torous, Caine, & De Choudhury, 2020). Restrictions started to be lifted in May 2020 and the state of emergency was lifted in mid June 2020 and people could get back to indoor and outdoor group activities and dine at restaurants which they had missed. This came at the right time as it was at the start of summer.

Cases were slowly rising in early autumn, and another state of emergency with new restrictions was finally declared on November 24, 2020. Gyms, indoor dining, hair stylists, fitness centres, bars, had to close. Albertans already knew how difficult the first lockdowns were and many expressed their negative feelings about the new lockdown and its effects online.

During the second wave, the government announced a four-step reopening plan that was based on the number of hospitalisations. Alberta started step 1 on February 8, 2021 and the second step started fully on March 8th, 2021. But when the cases and hospitalisation increased—signalling the third wave— there was a roll back into stage 1 on April 9, 2021. The difference in restrictive measures between the two steps was not very great and step two lasted just one month. Step 1 of the four-step reopening plan started after the peak of the second wave and this same plan was in place and continued into the third wave. During the presentation of the plan, it was clearly explained to the public that the movement forward from one step to another was based on the number of hospitalisations and that there was always the possibility of moving backwards. Albertans could anticipate the introduction of stricter measures during rollback to previous steps as the hospitalisations increased during the third wave.

The relaxation of restrictive measures between wave 2 and wave 3 wasn't like that which occurred between wave 1 and wave 2 which was much longer, and had very little restrictions left and it permitted Albertans to get back to an almost normal lifestyle. There was no big change in restrictions between wave 2 and 3 and Albertans experienced both waves as a single long period of high restrictive measures. This was reflected in the discourse online which showed no significant difference in positivity.

There was a significant reduction in positivity in the fourth wave. The reduction in positivity during the fourth wave can be explained by a combination of factors that were occurring at the same time and not just the return of some restrictive measures.

After the third wave vaccines were increasingly becoming available to the whole population and with the low number of cases, the government was eager to end all or most of the restrictions. The three step Open for Summer Plan was started on June 1st. The plan was to lift all restrictions by July 1 which is Canada day if 70% of Albertans received at least one vaccine dose. This goal was met and all restrictions (except some restrictions in long term care facilities) were lifted. According to Premier Kenney, the COVID-19 pandemic in Alberta had moved to an endemic status. But with the new Delta variant which was much more contagious, cases started rising. The government officials didn't seem concerned by the rising cases as they did not expect hospitalisations and ICU admissions to rise with rising cases. They gave fewer updates and Albertans became increasingly worried and this worry was reflected on Twitter. In late August and early September, the hospitalisation and ICU admissions had risen so much that there were shortages of ICU beds. More than 80% of patients admitted to ICU were unvaccinated. This led to the Premier to launch a controversial compensation program for vaccination. And on the 15th of September 2021, the Premier declared a state of emergency and new restrictions to reduce the spread of the virus to avoid overwhelming the health sector even more. He also announced a restriction exemption program that will restrict access to some public places to the unvaccinated adult population. All these new measures brought about clashes of opinions online between those that were for and against and this very likely contributed to bring down the positivity of the text on Twitter. The data considered in this study ends on the 19th of September 2021 and only captures the early days of these new measures.

The period between the first and second wave is similar to the period between the third and fourth wave. They were characterised by very few restrictions and they occurred in the good weather of summer. Experiencing new restrictions after a summer with almost no restrictions was difficult to accept and it reflected in the significant changes in positivity of the discourse on Twitter.

The eighth hypothesis stated that in a regression model with AWS as the dependent variable, the most significant predictors in the model will be the number of patients in ICU and the use of the Them pronouns. It was expected that as more people became critically ill during the pandemic and needed intensive care, the tweets posted will become less positive as measured by the AWS. It was also expected that because of the politicised and polarising nature of the pandemic, people will tend to post more negative tweets when they refer to others and other's groups. Results from the analysis only partly confirmed the hypothesis. In a multiple regression model run with these predictors, the only significant predictors of AWS were the wave of the pandemic and the use of the Them pronouns. They both negatively predicted the AWS. This means that, in trying to predict the AWS of tweets related to the pandemic in Alberta, after considering the predictive power of the wave of the pandemic and the use of Them pronouns, considering the predictive power of the number of patients in ICU and the use of the Us pronouns does not significantly improve prediction of the AWS. The results also say that subsequent waves of the pandemic brought about significant negative changes in the overall sentiment of the tweets posted. This is supported by the findings from the between subjects ANOVA. As hypothesized the Them pronouns significantly predicted the AWS. This shows that more mentions of other persons in tweets were generally associated with less positive sentiments being expressed—and vice versa. It should be noted that running simple regressions with each variable at a time with AWS will show significant negative relationships between the Us, Them and number of patients in ICU variables—as suggested from Figure 3 which shows the correlation between the study variables—but the beauty of multiple regression models is that it shows the best predictors for the dependent variable. The wave of the pandemic, although being a categorical variable with 4 categories, was the strongest predictor of the AWS during the pandemic. It suggests that Albertans experienced the four waves considered in this study differently. As previously discussed the pandemic waves were also different from each other; for example, the introduction of vaccines and vaccine mandates in the later waves, COVID-19 fatigue which was Albertans being increasingly just fed-up with the pandemic and were saddened by how their life was negatively affected.

The use of the Them pronouns was a better negative predictor of AWS than the use of the Us pronouns. This suggests that talking about the other and their group is more associated with negative speech on Twitter than talking about oneself and the groups belonged to. This finding gives some credence to the thought that led to the sixth hypothesis that stated that there will be a negative relationship between the use of Them pronouns and positive sentiments. It suggests that people are more negative when talking about someone or people that belong to a group they are not associated with.

Limitations

Despite the interesting findings of this study and serious efforts to be objective and scientific, this study was characterised by many limitations which will be disused next.

This study used only a very small sample of the discourse on the COVID-19 pandemic in Alberta. The data used in this study most likely represents less than one percent of the discourse of the COVID-19 pandemic in Alberta considering that only three hashtags were used. The discourse on the pandemic occurred on other social media platforms like Facebook, Instagram, Reddit, and on other media channels like Television news reports, radio programs and web articles.

The Twitter user demographic might not be a good representation of the demographics of the population of Alberta. According to Bush (2023) there is a greater tendency for individuals with a household income greater than $100,000 and with a college degree to have a Twitter account.

All sentiment measuring methods are not 100% accurate. VADER's accuracy was estimated at 88% in a study by Nair, Veena, & Vinayak (2021) that compared three sentiment analysis tools. This implies that there is a proportion of tweets that were miscategorized and this affected the results of the study.

In all studies designed and conducted by humans, there is always an element of researcher subjectivity and bias that is difficult to completely control despite the researcher's best efforts. Even unconsciously, the choice of methods of analysis, and the interpretation of results might have been biased towards the confirmation of the research hypothesis.

The Us and Them pronouns in this study were considered to represent Twitter users' mentions of oneself and the groups to which they identified with and mentions of the other and the groups they identify with respectively. But in reality, second and third person pronouns that make up the Them pronouns can be used when referring to someone a Twitter user identifies with and a group they identify with. For example, someone who has recovered from a COVID-19 infection after spending time in ICU can talk about patients currently in ICU using the third person pronouns. The person identifies with the patients in ICU but they are using the third person pronouns which is part of the Them pronouns which were considered in this study to be used to refer to people the Twitter user does not identify with.

Positive and negative emotions are both very broad groups of emotions. The negative emotion lexicon is made up of words that can be further classified into different sets for different emotions, verbs, nouns, etc. From the lexicon (available at https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vader_lexicon.txt) the negative words can be classified into diseases (flu, cancer phobia), fears (dread, fright, terrify), anger (livid, rage, fury) etc. It will be more interesting to study specific negative sentiments or emotions to explore how they vary during the pandemics.

Conclusion

This study was an exploration of the discourse of the COVID-19 pandemic in the province of Alberta. The main aim of the study was to explore how the discourse on the pandemic changed over time. Another aim of the study was to compare different methods of measuring sentiments from text. Three methods were compared here. The simple measuring method (SMM), the complex measuring method (CMM) and the average score method (ASM).

Tweets on the COVID-19 pandemic in Alberta were scraped from Twitter, divided into weekly groups, cleaned and quantitative data was gotten from the text of the tweets. The quantitative data were sentiment scores (positive, negative, and neutral sentiments scores) and the use of Us pronouns (first person pronouns singular and plural) and Them pronouns (second and third person pronouns singular and plural) for each week. Additionally, the weekly average of the number of people in ICU was obtained from Alberta's ministry of health's website. This data was visualized and quantitative data analysis methods were performed on the data extracted in order to test the study hypothesis.

A word cloud visualization of the data confirmed that the text gathered was really about the COVID-19 pandemic in Alberta. High frequency words were, Alberta, Covid, cases, health, and people. It was noticed that the Premier Jason Kenney was the most discussed government official. The second most discussed government official was Deena Hinshaw the Chief Medical Officer of Health, and the third most discussed government official was Tyler Shandro, the Minister of Health.

It was observed that there was a decrease in the positivity of the discourse as the waves of the pandemic unfolded. This decrease did not occur along all the successive waves of the pandemic. The significant decrease in positivity occurred when the next wave occurred after a summer during which almost all the restrictions were lifted.

A general correlation between the study variables was run to explore the relationship between them. There was a strong correlation between pairs of positive and negative sentiment cores measured using the SMM and the CMM but the correlation of the neutral sentiment was low. This means that the two methods were similar since all the correlations were positive and significant. But to understand the difference between the measuring methods, their correlations with the number of patients in ICU was compared. Results based on the magnitude of the correlation coefficients suggested that for measures of positive sentiment, the CMM had a stronger correlation while the SMM had a stronger correlation for negative sentiments. But further comparison of the positive and negative scores measured using the SMM and the CMM with the AWS (obtained from the average score measuring method) suggested that CMM was a better measure of sentiment analysis. But considering how the CMM and SMM scores are calculated, it can be said that CMM and SMM do not have the same level of validity as measures of sentiments of texts. SMM scores were based on raw categorisations of the individual words that make up the text while the CMM used VADER's complex analysis algorithms which uses rule-based enhancements that are sensitive to the context in which the words are used. This makes CMM to be a more valid measure of the actual sentiment of the text while SMM is less valid as a measure of the actual sentiment of the text but more valid as a measure of the writing styles i.e., the choice of positive, negative, or neutral words to express a given sentiment or opinion.

Since positive sentiments measured using the CMM had a stronger correlation with the number of patients in ICU than positive sentiment measured with the SMM, and the fact that negative sentiments measures by the SMM had a stronger correlation with the number of patients in ICU than negative sentiment measured with the CMM, the following postulate can be made about the COVID-19 discuss in Alberta.

- Starting from a hypothetical base level, as the number of patients in ICU increases (indicating a worsening of the pandemic crisis) Twitter users tend to use more negative words but these words are not used to express purely negative sentiments, hence the negative sentiments measured by the CMM has a lower correlation with number of patients in ICU. On the other hand, as the number of patients in ICU reduces, Twitter users tend to use much fewer negative words compared to how what they say reduces in negativity.

- Now for the positive sentiment, starting from a hypothetical base level, as the number of patients in ICU increases (indicating a worsening of the pandemic crisis) Twitter users tend to express much less positive sentiments than how less their use of positive words reduces. When the number of patients in ICU reduces, Twitter users express more positive sentiment than how their use of positive words increases.

In general, we can say that this pattern of variation of sentiment scores measured with the CMM and the SMM with the number of patients in ICU represent Twitter users' attempt to be positive and not overly negative amid the negative phenomenon which was the pandemic.

Unexpectedly, the use of Us pronouns and Them pronouns had similar correlations (in magnitude and direction but particularly in direction) with positive and negative sentiments scores. This shows that in a situation like a pandemic in which everyone is affected, people express similar types of sentiments when talking about themselves and others. But interestingly, a multiple regression model, revealed that mentions of the other is more predictive of the average sentiment than mentions of oneself and a group identified with.

This study showed that the discourse on the pandemic became less positive as the pandemic evolved. Comparing different methods of measuring sentiment scores permitted the comparison of changes in actual text sentiment and writing style that suggested that Twitter users in general tried to remain somewhat positive. It was found that mentions of the other was more predictive of sentiment scores in the discourse although it had similar use variation to mentions of oneself and the group to which they belong.

Recommendations and Directions for Future Studies

This study was a general exploration of the discourse of the COVID-19 pandemic in Alberta. From the findings of the study some cautious recommendations for practice can be made. Some of the limitations mentioned above can be addressed in future studies and the study itself could be modified for more interesting results.

Researchers who do sentiment analysis as part of their studies have to make informed decisions on which sentiment analysis tool to use and how to compute sentiment score depending on the objective of their study. Simple lexicon-based sentiment analysis tools are less reliable when used to measure sentiments. They have better reliability when used to measure the use of positive, negative, and neutral words and so their use is recommended in such situations.

Although the use of Us and Them pronouns had similar correlation patterns with sentiment scores, the use of Them pronouns was a better predictor of the sentiment scores. In future studies with methods like machine learning, priority can be given to measures of mentions of others and the group to which they belong in training models.

This study can be redone with a sample that is a better representation of the population of Alberta. Data from many social media platforms could be used. The sentiment scores from the different SNS could be compared to have a better understanding of how the pandemic might be discussed differently on different social media platforms. This will also lead to a bigger and more representative sample of the discourse of the pandemic in Alberta.

Positive and negative emotions are very broad categories through which the content of text can be analysed. Content analysis programs like LIWC have different dimensions of content analysis. One such dimension is psychological processes which has subcategories that measure specific aspects of the text, for example anger, family, and death. Knowing how the discourse about these categories vary during the pandemic will provide more interesting information than simply knowing how the positivity and negativity in the text varies over time. Another way of achieving this is by running topic modelling to identify the topics discussed in the text and tracking the topics over time. This has the advantage of not being lexicon based and so the content analysis process is not limited to fixed lexicons that might not be adapted to the text being studied.

It will be interesting to know with high levels of confidence the sentiment score of tweets when the Twitter user is talking about themself or someone or a group they identify with and when they are talking about another person and a group with whom they do not identify with. The study could be made more specific such that two or more groups of individuals could be identified and the content of the tweets when referring to a particular group is analysed. But care should be taken not to break Twitter's use policy that forbids making inference about certain opinions of Twitter users based on their posts on Twitter.

In this study, only one official medical metric of the COVID-19 pandemic was used. Future studies could run a model in which other medical metrics like number of vaccines administered, number of hospitalisations, number of deaths, number of new cases, or even other statistics that don't seem directly related to the pandemic, like prices of basic goods, in a model to explore how/if they predict the online discuss of the COVID-19 pandemic. Different hierarchical regression models could be used with these metrics to see the one that fits best.

In this study the SMM and the CMM of sentiment analysis were compared. The comparison led to the postulate that during the pandemic, when the pandemic worsened (increase in the number of patients in ICU) the people on Twitter used more negative words but they tried not to be overly negative. Future studies could explore different ways to verify this postulate. One way would be to identify and Track Twitter users over time to see if their tweets change in a way that confirms the postulate. Another way could be to create a sample of tweets from when the number of patients in ICU are at their highest level and when they are at their lowest levels, then identify negative words from VADER’s lexicon and explore how they are used in those two samples.

Reference

Aaker, C., Chang, V. (2009). Obama and the power of social media and Technology. *Stanford Business*. https://www.gsb.stanford.edu/faculty-research/case-studies/obama-power-social-media-technology

Adab, P., Haroon, S., O'Hara, M. E., & Jordan, R. E. (2022). Comorbidities and covid-19. *bmj, 377*. http://dx.doi.org/10.1136/bmj.o1431

Akbiyik, M. E., Erkul, M., Kaempf, K., Vasiliauskaite, V., & Antulov-Fantulin, N. (2021, February 27-March 3). *Ask" Who", Not" What": Bitcoin Volatility Forecasting with Twitter Data* [Conference presentation]. Sixteenth ACM International Conference on Web Search and Data Mining (WSDM '23), Singapore, Singapore. https://doi.org/10.48550/arXiv.2110.14317

Albahli, S., Irtaza, A., Nazir, T., Mehmood, A., Alkhalifah, A., & Albattah, W. (2022). A Machine Learning Method for Prediction of Stock Market Using Real-Time Twitter Data. *Electronics, 11*, 1-20. https://doi.org/10.3390/electronics11203414

Al-Ramahi, M., Elnoshokaty, A., El-Gayar, O., Nasralah, T., & Wahbeh, A. (2021). Public discourse against masks in the COVID-19 era: Infodemiology study of Twitter data. JMIR Public Health and Surveillance, 7(4). https://doi.org/10.2196/26780

Amador, J. C., Collignon-Delmar, S., Benoit, K., & Matsuo, A. (2017). Predicting the Brexit vote by tracking and classifying public opinion using twitter data. Statistics, Politics and Policy, 8(1), 85-104. https://doi.org/ 10.1515/spp-2017-0006

An, S., Ji, L. J., Marks, M., & Zhang, Z. (2017). Two sides of emotion: Exploring positivity and negativity in six basic emotions across cultures. *Frontiers in psychology, 8*. https://doi.org/10.3389/fpsyg.2017.00610

Ante, L. (2023). How Elon Musk's twitter activity moves cryptocurrency markets. *Technological Forecasting and Social Change, 186*. https://doi.org/10.1016/j.techfore.2022.122112

Ayala, A., Villalobos Dintrans, P., Elorrieta, F., Castillo, C., Vargas, C., & Maddaleno, M. (2021). Identification of COVID-19 waves: Considerations for research and policy. *International Journal of Environmental Research and Public Health, 18*(21). https://doi.org/ 10.3390/ijerph182111058

Baur, D. G., & Dimpfl, T. (2021). The volatility of Bitcoin and its role as a medium of exchange and a store of value. *Empirical Economics, 61*(5), 2663-2683. https://doi.org/10.1007/s00181-020-01990-5

Bernstein, M. S., Bakshy, E., Burke, M., & Karrer, B. (2013). Quantifying the invisible audience in social networks. *Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI 2013*, 21–30. https://doi.org/10.1145/2470654.2470658

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American journal of epidemiology, 188*(12), 2222-2239. https://doi.org/10.1093/aje/kwz189

Bimber, B. (2014). Digital media in the Obama campaigns of 2008 and 2012: Adaptation to the personalized political communication environment. *Journal of information technology & Politics, 11*(2), 130-150. https://doi.org/10.1080/19331681.2014.895691

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* O'Reilly Media, Inc.

Blackman, M. “What’s in a name? Twitter is an ideological battleground over the control of language.” *Opinionista*. Daily Maverick, 9 Nov. 2022, https://www.dailymaverick.co.za/opinionista/2022-11-09-whats-in-a-name-twitter-is-an-ideological-battleground-over-the-control-of-language/. Accessed 5 May 2023.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning research, 3*(2003), 993-1022.

Bonta, V., Kumaresh, N., & Janardhan, N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology, 8*(S2), 1-6. https://doi.org/ 10.51983/ajcst-2019.8.S2.2037

Boyd, R. (2017). Psychological text analysis in the digital humanities. In S. Hai-Jew (Ed.), *Data Analytics in Digital Humanities* (pp. 161–189). Springer International Publishing. https://doi.org/10.1007/978-3-319-54499-1_7

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*. Austin, TX: University of Texas at Austin. https://www.liwc.app/

Brans, H., & Scholtens, B. (2020). Under his thumb the effect of president Donald Trump’s Twitter messages on the US stock market. *PloS one, 15*(3). https://doi.org/10.1371/journal.pone.0229931

Brett, B. (2011). *The psychology of sharing: why do people share online.* New York Times Customer Insight Group. https://web.archive.org/web/20160410111238/http://nytmarketing.whsites.net/mediakit/pos/POS_PUBLIC0819.php

Budak, A. (2010). *Facebook, Twitter and Barack Obama: New media and the 2008 presidential elections*. Unpublished master's thesis]. Georgetown University.

Burgess. J., Jahr. M., Keljo. J., Schroeder J., Sweitzer. W. (1998). *Controlling the Virtual World: Governance of On-Line Communities*. Stanford University. https://cs.stanford.edu/people/eroberts/cs201/projects/1998-99/controlling-the-virtualworld/index.html

Burnap, P., & Williams, M. L. (2016). Us and them: identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data science, 5*, 1-15. https://doi.org/ 10.1140/epjds/s13688-016-0072-6

Bush, O. (2023). *Twitter Statistics in Canada*. Made in CA. https://madeinca.ca/twitter-statistics-canada/

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council. (2022a). *Does Research Using Social Media Platforms Require Research Ethics Board Review? Guidance to support the implementation of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans –TCPS 2 (2022)*. Canada. Retrieved from https://ethics.gc.ca/eng/reb-cer_social-sociaux.html

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council. (2022b). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – TCPS 2*. Canada. Retrieved from https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2022.html

Chaudhry, H. N., Javed, Y., Kulsoom, F., Mehmood, Z., Khan, Z. I., Shoaib, U., & Janjua, S. H. (2021). Sentiment analysis of before and after elections: Twitter data of US election 2020. *Electronics, 10*(17). https://doi.org/10.3390/electronics10172082

Chen, N., Zhong, Z., & Pang, J. (2021). An exploratory study of COVID-19 information on twitter in the greater region. *Big Data and Cognitive Computing, 5*(1). https://doi.org/10.3390/bdcc5010005

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. *Journal of research in personality, 42*(1), 96-132. https://doi.org/10.1016/j.jrp.2007.04.006

Coursera. (2023). *What Is Facebook Metaverse?* Coursera. https://www.coursera.org/articles/facebook-metaverse

Dixon, S. (2022). *Twitter - Statistics & Facts*. Statistica. Retrieved from https://www.statista.com/topics/737/twitter/#dossierKeyfigures

Dixon, S. (2022b). *Most popular social networks worldwide as of January 2022, ranked by number of monthly active users*. Statistica. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Dixon, S. (2022c). *Number of monetizable daily active Twitter users (mDAU) worldwide from 1st quarter 2017 to 2nd quarter 2022*. Statistica. https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide/

Ess, C. (2007). Internet research ethics. In A. Joinson., K. Y. A. McKenna., T. Postmes & U. Reips (Eds), *The Oxford handbook of Internet psychology* (pp. 487-502). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199561803.013.0031

Fiesler, C., & Proferes, N. (2018). "Participant" perceptions of Twitter research ethics. *Social Media+ Society, 4*(1), 2056305118763366. https://doi.org/10.1177/2056305118763366

Gagliese, J. (2022, July 8). *The Rise Of The Influencer: Predictions For Ways They'll Change The World*. Forbes. Retrieved from https://www.forbes.com/sites/theyec/2022/07/08/the-rise-of-the-influencer-predictions-for-ways-theyll-change-the-world/?sh=6e3006b643a7

Garrett, S. (2016). *The digital battleground: The political pulpit to political profile*. Augustana College. https://digitalcommons.augustana.edu/cgi/viewcontent.cgi?article=1073&context=celebrationoflearning

Ghebreyesus, T. A. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. World Health Organisation. Retrieved from https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012-1014. https://doi.org/ 10.1038/nature07634

Gobbi, S., Płomecka, M. B., Ashraf, Z., Radziński, P., Neckels, R., Lazzeri, S., ... & Jawaid, A. (2020). Worsening of preexisting psychiatric conditions during the COVID-19 pandemic. *Frontiers in Psychiatry, 11*. https//doi.org/ 10.3389/fpsyt.2020.581426

Goldstein, P., Levy Yeyati, E., & Sartorio, L. (2021). Lockdown fatigue: The diminishing effects of quarantines on the spread of COVID-19. *CID Working Paper Series*. https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37369329

Goodman, J., Wennerstrom, A., & Springgate, B. F. (2011). Participatory and social media to engage youth: from the Obama campaign to public health practice. *Ethnicity & disease, summer 21*(3 0 1).

Gold, N. (2020). Using Twitter data in research: Guidance for researchers and ethics reviewers. University College London. Retrieved from https://www.ucl.ac.uk/research-ethics/sites/research_ethics/files/using_twitter_in_research_v1.1_final.docx

Gordon, R. A. (2015). *Regression Analysis for the Social Sciences*. Routledge.

Government of Alberta. (2023, June). *COVID-19 Alberta statistics*. Alberta. https://www.alberta.ca/stats/covid-19-alberta-statistics.htm#highlights

Haktanir, A., Can, N., Seki, T., Kurnaz, M. F., & Dilmaç, B. (2022). Do we experience pandemic fatigue? current state, predictors, and prevention. *Current Psychology*, *41*(10), 7314-7325. https://doi.org/10.1007/s12144-021-02397-w

Hughes, A., & Wojcik, S. (2019, August 2). *10 facts about Amercians and Twitter*. Pew Research Center. https://web.archive.org/web/20190807070838/https://www.pewresearch.org/fact-tank/2019/08/02/10-facts-about-americans-and-twitter/

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. https://doi.org/10.1609/icwsm.v8i1.14550

Jain, A., Tripathi, S., Dwivedi, H. D., & Saxena, P. (2018, August 1). Forecasting price of cryptocurrencies using tweets sentiment analysis. *Proceedings in 2018 eleventh international conference on contemporary computing (IC3), India.* (pp. 1-7). https://doi.org/ 10.1109/IC3.2018.8530659

Jarvis, M. C. (2020). Aerosol transmission of SARS-CoV-2: physical principles and implications. *Frontiers in Public Health, 8*. https://doi.org/10.3389/fpubh.2020.590041

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci, A374*. https://doi.org/10.1098/rsta.2015.0202

Jolly, T. S., Batchelder, E., & Baweja, R. (2020, September 17). *Mental health crisis secondary to COVID-19–related stress: a case series from a child and adolescent inpatient unit.* Psychiatrist. https://www.psychiatrist.com/pcc/covid-19/covid-19-related-stress-in-children/

Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and research: A systematic literature review through text mining. *IEEE Access*, *4*. https://doi.org/10.1109/ACCESS.2020.2983656

Karger, D.R. and Quan, D. (2005), "What would it mean to blog on the semantic web?." *Journal of Web Semantics, 3*(2-3), 147-157. https://doi.org/10.1016/j.websem.2005.06.002

Kassraie, P., Modirshanechi, A., & Aghajan, H. K. (2017). Election Vote Share Prediction using a Sentiment-based Fusion of Twitter Data with Google Trends and Online Polls. *Proceedings in the 6th International Conference on Data Science, Technology and Applications (DATA), 2017*, 363-370. https://doi.org/10.5220/0006484303630370

Koh, J. X., & Liew, T. M. (2022). How loneliness is talked about in social media during COVID-19 pandemic: Text mining of 4,492 Twitter feeds. *Journal of psychiatric research, 145*(6), 317-324. https://doi.org/10.1016/j.jpsychires.2020.11.015

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science, 343*(6176), 1203-1205. https://doi.org/ 10.1126/science.1248506

Larsen, J. T., McGraw, A. P., and Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *J. Pers. Soc. Psychol. 81*, 684–696. doi: 10.1037/0022-3514.81.4.684

Liu, R., Yao, X., Guo, C., & Wei, X. (2021). Can we forecast presidential election using twitter data? An integrative modelling approach. *Annals of GIS, 27*(1), 43-56. https://doi.org/10.1080/19475683.2020.1829704

Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media*. The Information Society, 30*(4), 256-265. https://doi.org/10.1080/01972243.2014.915276

Luck, E., Beaton, J., & Moffatt, J. (2010, September 9-12). *The social media (r)evolution: Obama's political campaign*. [Paper presentation]. Global Marketing Conference, Okura, Tokyo. https://core.ac.uk/download/pdf/10903115.pdf

Mansoor, M., Gurumurthy, K., & Prasad, V. R. (2020). Global sentiment analysis of COVID-19 tweets over time. *The Research Gate Journal of Computer Science and Engineering*.

Mikal, J., Hurst, S. & Conway, M. (2016) Ethical issues in using Twitter for population-level depression monitoring: a qualitative study. *BMC Med Ethics, 17*. https://doi.org/10.1186/s12910-016-0105-5

Nair, A. J., Veena, G., & Vinayak, A. (2021). Comparative study of twitter sentiment on covid-19 tweets. *Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021)*, 1773-1778. https://doi.org/10.1109/ICCMC51019.2021.9418320

Nduwayezu, M., Satyabrata, A., Han, S. Y., Kim, J. E., Kim, H., Park, J., & Hwang, W. J. (2019). Malaria Epidemic Prediction Model by Using Twitter Data and Precipitation Volume in Nigeria. *Journal of Korea Multimedia Society, 22*(5), 588-600. https://doi.org/10.9717/kmms.2019.22.5.588

Novicio, T. (2021, March 8). 15 Companies That Benefited The Most From The Pandemic. *Yahoo Finance*. https://finance.yahoo.com/news/15-companies-benefited-most-pandemic-101257486.html

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence From Ghana. *Applied Computer Systems, 25*(1), 33-42. https://doi.org/10.2478/acss-2020-0004

Ovadia, S. (2009). Exploring the potential of Twitter as a research tool. *Behavioral & Social Sciences Librarian, 28*(4), 202-205. https://doi.org/ 10.1080/01639260903280888

Park, J., Baek, Y. M., & Cha, M. (2014). Cross-cultural comparison of nonverbal cues in emoticons on Twitter: Evidence from big data analysis. *Journal of Communication, 64*(2), 333–354. https://doi.org/10.1111/jcom.12086

Paul, M. J., & Dredze, M. (2012). A model for mining public health topics from Twitter. *Johns Hopkins University*. https://www.cs.jhu.edu/~mdredze/publications/2011.tech.twitter_health.pdf

Perlin, M. S., Caldeira, J. F., Santos, A. A., & Pontuschka, M. (2017). Can we predict the financial markets based on Google's search queries?. *Journal of Forecasting, 36*(4), 454-467. http://dx.doi.org/10.2139/ssrn.2525909

Peršak, N. (2023). Hate Speech in Times of the COVID-19 Pandemic: The Social Costs and Legal Implications of a Health Crisis. *Medicine, Law & Society*, *16*(2). 227–246. https://doi.org/10.18690/mls.16.2.227-246.2023

Pew Research Centre. (2008, September 15). McCain vs. Obama on the Web. *Pew Research Center*. https://www.pewresearch.org/journalism/2008/09/15/mccain-vs-obama-on-the-web/

Pew Research Center. (2021, November 15). The Behaviors and Attitudes of U.S. Adults on Twitter. *Pew Research Center*. https://www.pewresearch.org/internet/2021/11/15/the-behaviors-and-attitudes-of-u-s-adults-on-twitter/

Preis, T., & Moat, H. S. (2014). Adaptive nowcasting of influenza outbreaks using Google searches. *Royal Society Open Science, 1*(2), 140095. https://doi.org/10.1098/rsos.140095

Proferes, N. (2017). Information flow solipsism in an exploratory study of beliefs about Twitter. *Social Media + Society*, 3(1). https://doi.org/10.1177/2056305117698493

Ragothaman, A., & Huang, C. (2021). Sentiment Analysis on COVID-19 Twitter Data. *International Journal of Computer Theory and Engineering, 13*(4), 100-107. https://doi.org/10.7763/IJCTE.2021.V13.1297 /

Rainie, L., Cornfield, M., and Horrigan, J. (2005). The Internet and Campaign 2004. *Pew Internet & American Life Project*. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2005/PIP_2004_Campaign.pdf.pdf

Reeher, G. (2004). Information and American Democracy: Technology in the Evolution of Political Power.[Book Review]. *Perspectives on Politics, 2*(2), 366-367. https://doi.org/10.1017/S1537592704380977

Rowe, M., & Alani, H. (2014, June). Mining and comparing engagement dynamics across multiple social media platforms. *Proceedings of the 2014 ACM conference on Web science*, 229-238. https://doi.org/10.1145/2615569.2615677

Saeidi, M., da S. Sousa, S. B., Milios, E., Zeh, N., & Berton, L. (2020). Categorizing online harassment on Twitter. *Proceedings in Machine Learning and Knowledge Discovery in Databases: International Workshops of ECML PKDD 2019*, 283-297. https://doi.org/10.1007/978-3-030-43887-6_22

Saha, K., Torous, J., Caine, E. D., & De Choudhury, M. (2020). Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. *Journal of Medical Internet Research, 22*(11). https://doi.org/ 10.2196/22600

Saxena, S. (2013, August 11). Key characteristics of Social Media. *Easymedia*. https://www.easymedia.in/7-key-characteristics-of-social-media/

Sayce, D. (2022). The Number of tweets per day in 2022. *David Sayce*. https://www.dsayce.com/social-media/tweets-day/

Sinclair, S. & Rockwell, G. (2016). Voyant: see through your text. *Voyant Tools*. http://voyant-tools.org/

Sinnenberg, L., Buttenheim, A. M., Padrez, K., Mancheno, C., Ungar, L., & Merchant, R. M. (2017). Twitter as a tool for health research: a systematic review. American journal of public health, 107(1), e1-e8. https://doi.org/10.2105%2FAJPH.2016.303512

Smith, A. (2009). *The Internet's role in campaign 2008 (Vol. 15)*. Pew Internet & American Life Project. https://www.pewresearch.org/internet/wp-content/uploads/sites/9/media/Files/Reports/2009/The_Internets_Role_in_Campaign_2008.pdf

Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. European Financial Management, 20(5), 926-957.

Stokel-Walker, C. (2023, March 10th). Twitter's $42,000-per-Month API Prices Out Nearly Everyone. *WIRED*. https://www.wired.co.uk/article/twitter-data-api-prices-out-nearly-everyone

Summerfield, P. (2023, March 16). The platforms and apps with the most reach in Canada. *Media in Canada*. https://mediaincanada.com/2023/03/16/the-platforms-and-apps-with-the-most-reach-in-canada/

Summers, E. (2020). twarc (Version 2). *GitHub repository*. https://github.com/DocNow/twarc

Svensson, P. (2013). Humanities computing as digital humanities. In M. Terras, J. Nyhan, & E. Vanhoutte (Eds.), *Defining Digital Humanities* (pp. 159-186). Routledge.

Tafti, A., Zotti, R., & Jank, W. (2016). Real-time diffusion of information on Twitter and the financial markets. *PloS one, 11*(8). https://doi.org/10.1371/journal.pone.0159226

Takats, C., Kwan, A., Wormer, R., Goldman, D., Jones, H. E., & Romero, D. (2022). Ethical and methodological considerations of twitter data for public health research: systematic review. *Journal of Medical Internet Research, 24*(11). https://doi.org/10.2196/40380

Talbot, J., Charron, V., & Konkle, A.T. (2021). Feeling the Void: Lack of Support for Isolation and Sleep Difficulties in Pregnant Women during the COVID-19 Pandemic Revealed by Twitter Data Analysis. *Int. J. Environ. Res. Public Health, 18*(2). https://doi.org/10.3390/ijerph18020393

The New York Times. (2023). Tracking Coronavirus in Canada: Latest Map and Case Count. *The New York Times*. https://www.nytimes.com/interactive/2021/world/canada-covid-cases.html

Tsai, M. H., & Wang, Y. (2021). Analyzing Twitter data to evaluate people's attitudes towards public health policies and events in the era of COVID-19. *International Journal of Environmental Research and Public Health, 18*(12). https:/doi.org/10.3390/ijerph18126272

Twitter. (n.d.a). How to share a Tweet on Instagram Stories. *Twitter*. https://help.twitter.com/en/using-twitter/instagram-stories-share

Twitter. (n.d.b). Terms of Service. *Twitter*. https://twitter.com/en/tos

United Nations. (2020). A Pandemic of Hate. *Impact and Prevention*. Https://Www.Un.Org/En/Hate-speech/Impact-and-prevention/A-pandemic-of-hate

University of Alberta. (n.d). Do I need research ethics approval?. *University of Alberta, Research and Innovation*. https://www.ualberta.ca/research/services/research-ethics/do-i-need-research-ethics-approval.html

Uyheng, J., & Carley, K. M. (2021). Characterizing network dynamics of online hate communities around the COVID-19 pandemic. *Applied Network Science, 6*(1). https://doi.org/10.1007/s41109-021-00362-x

Wesslen, R. (2018). Computer-assisted text analysis for social science: Topic models and beyond. *arXiv*. https://doi.org/10.48550/arXiv.1803.11045

Widener, M. J., & Li, W. (2014). Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography, 54*(2), 189-197. https://doi.org/10.1016/j.apgeog.2014.07.017

Williams, S. A., Terras, M. M., & Warwick, C. (2013). What do people study when they study Twitter? Classifying Twitter related academic papers. *Journal of Documentation, 69*(3), 384-410. https:/doi.org/10.1108/JD-03-2012-0027

Wu, X. K., Zhao, T. F., Lu, L., & Chen, W. N. (2022). Predicting the hate: A GSTM model based on Covid-19 hate speech datasets. *Information Processing & Management, 59*(4). https://doi.org/10.1016/j.ipm.2022.102998

Yan, C., Law, M., Nguyen, S., Cheung, J., & Kong, J. (2021). Comparing public sentiment toward COVID-19 vaccines across Canadian cities: analysis of comments on reddit. *Journal of Medical Internet Research, 23*(9). https:/doi.org/10.2196/32685

Zhang, Y., Lyu, H., Liu, Y., Zhang, X., Wang, Y., & Luo, J. (2021). Monitoring depression trends on twitter during the COVID-19 pandemic: observational study. *JMIR Infodemiology, 1*(1). https://doi.org/10.2196/26769

Zhang, Q., Yi, G. Y., Chen, L. P., & He, W. (2021). Text mining and sentiment analysis of COVID-19 tweets. *arXiv*. https://doi.org/10.48550/arXiv.2106.15354

Appendix

Appendix A: Summary of Data on Deaths in Provinces

Province of Territory	Total Deaths	Deaths per 100,000	Total Cases
Canada	51,720	138	4,617,095
Prince Edward Island	93	65	56,523
Northwest Territories	22	53	11,511
Quebec	18,160	222	1,320,325
Alberta	5,622	138	629,269
Nova Scotia	794	86	140,793
Saskatchewan	1,890	172	153,651
Yukon	32	89	4,989
Manitoba	2,464	193	154,712
Ontario	16,234	121	1,601,325
New Brunswick	834	112	88,866
Newfoundland and Labrador	318	61	54,757
Nunavut	7	19	3,531
British Columbia	5,249	113	396,817

Data from https://www.nytimes.com/interactive/2021/world/canada-covid-cases.html

Appendix B: Histogram of Diagnosed cases in Alberta

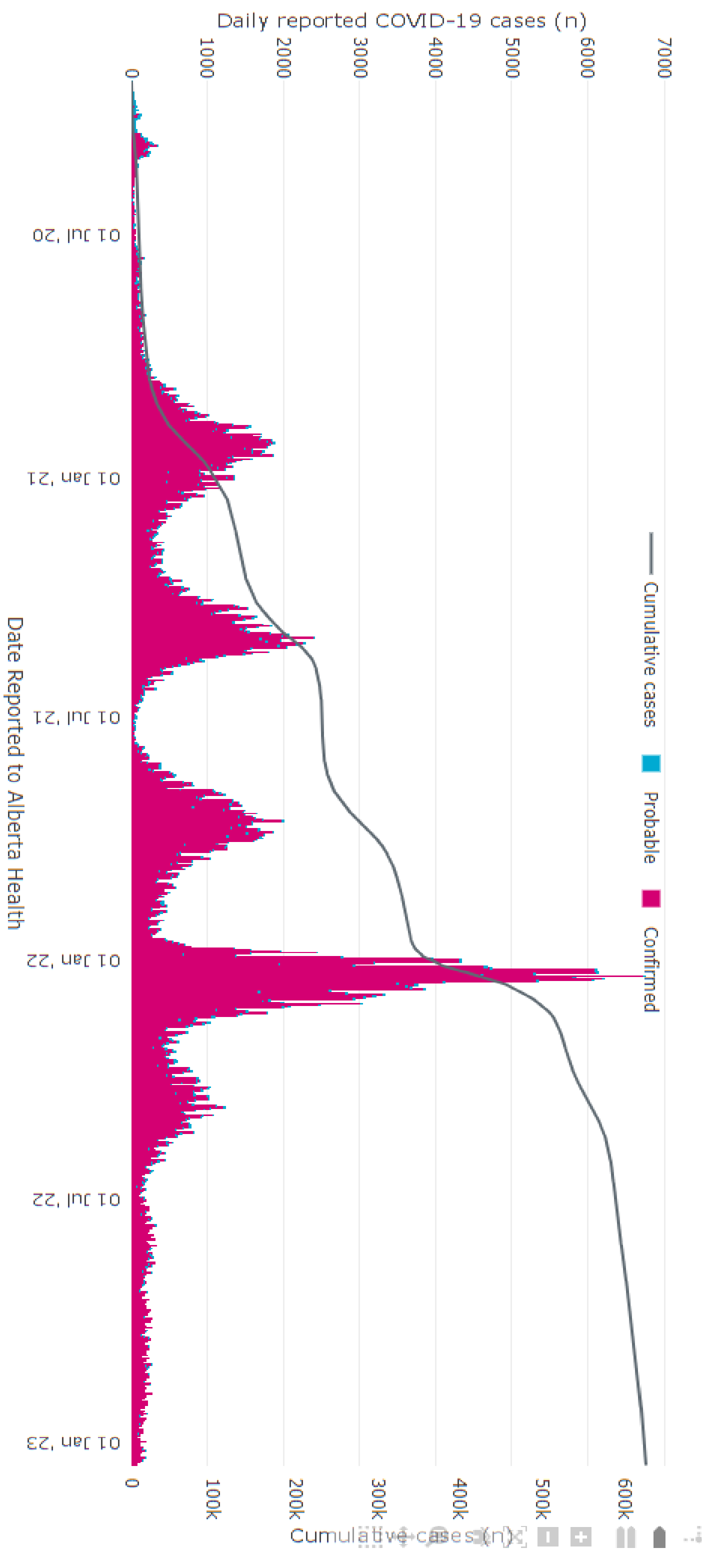


Figure 1: COVID-19 cases in Alberta by day and case status. Probable cases include cases where the lab confirmation is pending. Cases are under investigation and numbers may fluctuate as cases are resolved. Data included up to end of day January 16, 2023.