

RESEARCH

Open Access

# Data mining twitter for COVID-19 sentiments concerning college online education



Daniel Brandon<sup>1\*</sup>

## Abstract

In the last decade there has been a large increase in corporate and public reliance on social media for information, rather than on the traditional news and information sources such as print and broadcast media. People freely express their views, moods, activities, likes/dislikes on social media about diverse topics. Rather than surveys and other structured data gathering methods, text data mining is now commonly used by businesses to go through their unstructured text in the form of emails, blogs, tweets, likes, etc. to find out how their customers feel about their company and their products/services. This paper reports upon a study using Twitter (recently renamed to “X”) data to determine if meaningful and actionable information could be gained from such social media data in regard to pandemic issues and how that information compares to a traditional survey. In early 2020, the COVID-19 pandemic hit and forced colleges to move classes to an online format. While there is considerable literature in regard to using social media to communicate geo-political issues and in particular pandemics, there is not a study using social media to explore public sentiment in regard to COVID’s forcing online education upon the public. In this study, text data mining was used to gain some insight into the feeling of Twitter users in regard to the effect of COVID-19 and the switch to online education in colleges. This study found that Twitter data mining did produce actionable information similar to the traditional survey, and the study is important since its results may influence organizations to explore the use of Twitter (and possibly other social media) to obtain people’s sentiments instead of (or in addition to) traditional surveys and other traditional means of gathering such information. This paper demonstrates both the process of text data mining social media and its application to current real-world issues.

**Keywords** Business analytics, COVID-19, Coronavirus, Data mining, Online learning, Sentiment analysis, Text analytics, Text mining, Twitter

## Introduction

In March of 2020, COVID-19 forced colleges to shut down their physical campuses and transition classes to an online format. While there is considerable literature in regard to using social media to communicate issues regarding COVID and such pandemics, there is not a study using social media to explore public sentiment in regard to COVID’s forcing online education upon the

public. This study is relevant and important since its results may influence organizations to explore the use of Twitter (and possibly other social media) to obtain people’s sentiments instead of (or in addition to) traditional surveys and other traditional means of gathering such information. For our study here, the main research question is:

- Can valuable and definitive sentiment information be obtained from Twitter comparable to what might be obtained from a traditional survey?

\*Correspondence:

Daniel Brandon  
dbrandon@cbu.edu

<sup>1</sup> School of Business, Christian Brothers University, Memphis, TN 38104, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Our other research questions are:

- Are people's sentiments positive or negative in regard to moving to online education due to COVID?
- Did people's sentiments change from the time schools ceased physical operation and moved to online education to the time when the next semester started in the late summer of 2020?

In the general field of text data analytics, there is an interesting area called "sentiment analysis." This analysis is mostly used by companies to go through their emails, blogs, Facebook posts and likes, Twitter tweets, etc. to find out how their customers feel about their company and their products/services. It is also broadly used in business to analyze supply chain operations to find ways to improve practices.

As the COVID outbreak progressed, traditional communication media were employed first to survey the spread such as print, email, telephone, and broadcast based communication [18]. However, in the last decade there has been a large increase in corporate and public reliance on social media for information, rather than on the traditional news and information sources such as print and broadcast media. On social media people freely express their views, moods, activities, likes/dislikes on social media about diverse business, political and social phenomena.

Today many businesses also use social media to promote their products and services and share company information. Users of products and services share their experiences and reviews over social media thus providing a large and rich pool of information in the form of unstructured text. As a result, this pool of information has become a valuable source for conducting research.

In this research paper, the focus is on Twitter (recently renamed to "X") data. Twitter data in the form of tweets have been widely used in text data mining applications for a variety of social, political, and business purposes [3, 26, 30]. Twitter tweets have also been used in healthcare and medical businesses and specifically for crisis situations including use in pandemics [4, 7, 12, 17, 20, 22, 23, 34].

### Twitter service

Twitter service (or the Twitter App) is a social media platform that permits authenticated users to compose and send brief messages known as "tweets." Users can also receive tweets from other users. A tweet is a text string that can contain up to 280 characters, along with images, videos, and other multimedia content. This provides a quick and easy way to share thoughts, updates, news, and information. Users can follow other users and

can also like, retweet (repost), or reply to tweets. Users can also use hashtags to categorize their tweets and make them more discoverable to other users.

Twitter dates back to 2006, when Jack Dorsey, a student at NYU (New York University), envisioned a new internet app to send brief messages to others. He shared his idea with coworkers at the podcasting company Odeo, Evan Williams and Biz Stone, and together they launched Twitter [10]. The founding purpose of Twitter was to create a microblogging application where users could post short messages, or "tweets," which were originally limited to 140 characters. The platform was initially launched as a simple way for people to keep in touch with friends and family, but it quickly gained popularity and evolved into a powerful communication tool for businesses and other organizations as well as individuals.

Twitter is run through its website or mobile app and is free to use. The company makes money through advertising and through the sale of data to businesses and other organizations. In the early years, Twitter struggled to monetize its platform; but after several years, it became one of the world's most used social media networks, with millions of global users. Twitter's real-time nature and ease of use made it an ideal platform for live events and breaking news. It was used extensively during major events such as the Arab Spring and the London Riots, and has since become a key tool for journalists, politicians, and activists around the world.

Over the years, Twitter has added several features to enhance the user experience, including photographs, videos, and live streaming. The company has also acquired several other companies to expand its capabilities, such as Periscope for live streaming, and Vine for short-form video. The use of Twitter has expanded in recent years to become a vital tool for more applications including business marketing.

Today many public figures, celebrities, and politicians have a presence on Twitter, and it has become an important tool for communication and information dissemination in many different fields. However, in recent years, Twitter has faced criticism for its handling of misinformation, hate speech, and political propaganda on its platform. The company has responded by implementing new policies and tools to address these issues, but the challenge of moderating a platform used by hundreds of millions of people around the world remains a difficult task. In September 2022 Twitter's shareholders voted to accept a purchase offer of about \$40 billion from Elon Musk, and the company is now reorganizing and has been renamed to "X."

### Tweets vs text messages

Twitter tweets and text messages are often thought of interchangeably. Both are both forms of brief digital

communication, but they differ in some key characteristics. Short message service (SMS), commonly called texting, and Twitter are both methods for digitally sending short messages. SMS is a cellular service, whereas Twitter is an internet service. Twitter tweets could originally contain up to 140 characters now 280, while SMS messages can have 160.

Twitter tweets are public messages that can be read by anyone who has access to the internet, regardless of whether they have a Twitter account or not. These messages can now include text, images, videos, and links.

Text messages, on the other hand, are private, short messages commonly sent between two people typically through a mobile phone. These messages are typically used for more personal and informal communication. Unlike Twitter, text messages are not public and can only be seen by the sender and recipient(s). Commonly, texts only go privately to a single recipient (but groups can be set up), while Twitter tweets can be viewed by the public.

As with SMS, one can use Twitter to send private tweets, but Twitter is mainly focused on social sharing. By default, anybody can see one's tweets. Users who want to "follow" someone will automatically see that person's messages on their Twitter page, and those who follow you will see your tweets on their Twitter page. Unlike many other social media sites, one can follow anyone you want on Twitter, even without their permission.

#### **Tweet format and content**

There are about 30 fields of information inside of each tweet, and the tweet is contained in the data portion of the TCP/IP internet packet. Overall, the content and purpose of a Twitter tweet can vary greatly, but it is typically a short and concise message. Format features include:

1. Character limit: A Twitter tweet can have a maximum of 280 characters, including spaces and punctuation. This limit was increased from 140 characters in 2017.
2. Image and video: A Twitter tweet can also include an image or video, which will be displayed within the tweet.
3. URL (Universal Resource Locator): A Twitter tweet can include a URL link, which will be shortened using Twitter's URL shortening service. Shortened links permit one to include long URLs in a message while staying under the character limit of the tweet. The link shortening process uses information such as link click frequency to estimate relevance. This process also guards against malicious sites that distribute malware.
4. Hashtags: Hashtags are keywords preceded by the "#" symbol that help categorize tweets and make them

discoverable by others. Twitter users can include hashtags in their tweets to increase the visibility of their message.

5. Mentions: Twitter users can mention other users in their tweets by including the "@" symbol followed by the username. Mentioning another user will notify them of the tweet and increase the visibility of the tweet to their followers.

Content features in addition to the text also include emoticons and emoji which users employ to add an emotional or visual element to their tweets. Retweets and quotes allow users to also resend and quote tweets from other users to share other people's content with their followers.

Hashtags are now a key feature of Twitter that were first introduced in Twitter in 2007 and have since become a popular feature on the platform. They can be used to participate in conversations and events, follow trending topics, and amplify the reach of your tweets. They are commonly used to create a Twitter chat or to organize an event. Twitter hashtags are keywords or phrases that are preceded by the hash symbol (#) and are used to categorize and organize content on the platform. When a user includes a hashtag in a tweet, it makes the tweet readily discoverable to those who search for that particular hashtag. Hashtags can also be used to add context to one's tweets or to express a particular sentiment or emotion. To create a hashtag, all one needs to do is include the hash symbol (#) before a word or phrase in the tweet. Hashtags can contain letters, numbers, and underscores, but not spaces or special characters. Hashtags should be short and easy to remember and relevant and specific to the main topic. Using hashtags that are already popular will increase the reach of the tweet.

#### **Methods**

The Twitter API (Application Programming Interface) is a set of tools and specifications which allow programmers to interact with Twitter's platform and access its data. With the Twitter API, developers can create applications that can perform actions such as posting tweets, retrieving user information, or searching for specific tweets or topics. The API also provides access to real-time streaming data, allowing developers to monitor live tweets and trends as they occur. The Twitter API is commonly not only used by developers to build social media management tools, analytics platforms, and other applications that incorporate Twitter data, but by researchers to study twitter content. Fields of information available through the Twitter API include the text of the tweet plus: date and time tweet sent, latitude and longitude of sending devices, origin place information, type of device

used, information about sending entity, retweet count, favorite count, and hashtags.

To use the Twitter API, one first needs to create a Twitter account. Today that has to be a developer account. Once a developer account is approved, one creates a new Twitter App and generates the required API keys and access tokens. One will need to choose the appropriate programming language (typically R or Python) and library that will be used to make the API requests. Next one uses the chosen library functions to send requests to the Twitter API, passing in the API keys and access tokens. This process is shown in the block diagram in Fig. 1.

**Text data mining**

Text data mining deals with helping people and computers understand the “meaning” of a text document. Text data mining is commonly used in the business world by companies to go through their unstructured text to find out how their customers feel about their company and their products/services. In contrast, structured data are found in spreadsheets, relational databases and data warehouses, whereas unstructured text is the form in the form of emails, blogs, tweets, likes, etc.

Text mining can discover valuable, non-trivial, and previously unknown information from large collections of unstructured data. It involves various techniques such as machine learning and natural language processing (NLP) to extract meaningful insights and patterns from volumes of text data. Some common applications of text data mining include sentiment analysis, emotional analysis, topic modeling, named entity recognition, document classification, and information extraction. The overall goal of text data mining is to automatically identify and extract relevant information from unstructured text data, making it easier for businesses, organizations, and researchers to make data-driven decisions.

**Twitter data mining**

Twitter data mining refers to the process of collecting, cleaning, transforming, and analyzing data specifically from Twitter. The data collected from Twitter can be in the form of tweets, hashtags, mentions, and other relevant information. Twitter data mining is commonly used for various purposes such as:

1. Sentiment analysis: to determine the sentiment of tweets and understand public opinions about a particular topic, product, or event.
2. Emotional analysis to discover specific emotions embedded in the tweets
3. Marketing research: to gather insights about customer preferences, buying behavior, and product reviews.
4. Trend analysis: to identify trending topics and understand what people are talking about on Twitter.
5. Event analysis: to track events as they happen and gather real-time insights about them.
6. Social network analysis: to identify influencers and study the spread of information on Twitter.

The insights obtained from Twitter data mining are used by businesses, organizations, and governments to make informed decisions and commonly to develop targeted marketing strategies.

**Data mining and public health**

Data mining is a valuable tool for the public health sector and in particular for the study of disease causes, disease spread, and pandemics as it allows researchers to analyze large amounts of unstructured and structured data and extract meaningful insights from it [24]. During a pandemic, large amounts of structured data are generated from various sources, including medical records, surveillance systems, surveys, and demographic data. A number of studies have used structured data in online health databases such as PubMed or Medline to study health related issues [6, 8, 27]. The use of these structured studies is most appropriate for historical events rather than pandemic situations in real time. Traditional surveys via telephone, email, or web forms also provide structured medical data [28]. Today much unstructured data are also typically available in the form of social media posts, and this is available in real time. Pandemic data mining can help researchers to:

1. Track the spread of the disease: Data mining algorithms can be used to map the spread of the disease over time and space. This information can help researchers understand the dynamics of the disease and the factors that influence its spread.



Fig. 1 Twitter API process block diagram

2. Identify risk factors: By analyzing demographic data, medical records, and other relevant information, data mining can help researchers identify risk factors for severe disease outcomes. This information can be used to guide targeted interventions and prioritize resources.
3. Predict disease outbreaks: Predictive models based on data mining algorithms can be used to forecast the spread of the disease and predict future outbreaks. This information can be used to inform public health response efforts and prepare for future pandemics.
4. Evaluate the effectiveness of interventions: Data mining can also be used to evaluate the effectiveness of interventions, such as vaccination campaigns and public health measures. This information can be used to refine and improve public health responses in real-time.

Overall, pandemic data mining has become a very important tool for the study of public health situations as it allows researchers to make sense of large amounts of complex data and extract meaningful insights that can better guide response efforts.

#### COVID-19 in 2020

In early 2020, COVID-19 began to spread around the world causing crises involving both physical and mental health which overwhelmed existing health care systems [1, 2]. In March of 2020, COVID-19 forced colleges to shut down their physical campuses and transition classes to an online format. In the weeks that followed many writers, educators, and researchers studied the situation to determine the sentiment of students, parents and other college stakeholders in regard to a variety of related issues including the virus, the actions of colleges, the mental health of students, and to the effectiveness and desirability of online education [9, 11, 13, 15, 21, 25, 36].

One prominent study was the Simpson Scarborough National Survey [25] which involved many senior level school students who were set to enroll in a traditional college or university in Fall 2020. It also included in the study many current college students already enrolled in a traditional college or university. Key findings in the survey were:

- Many of the current college students criticized the quality of the online education and most said it was worse than traditional in-person instruction
- About one tenth of the students planning to attend college now would instead choose a two-year community college, start an online college instead, or simply defer college all together—take a “gap year”

Early in the COVID pandemic, social media became involved in several manners. Various social media platforms were used to warn people of the impending dangers. Later social media platforms were used to educate people into ways to mitigate the physical and mental health problems caused by the virus [29]. As evermore people practiced isolation and social distancing due to COVID, social media were used more not only for business but also personal communication [35]. As a result of COVID, global investments in pandemic preparedness dramatically increased including various forms of modern communication including the internet and social media [14]. In contrast, for this study social media is being used to study the sentiments of people in regard to the virus and its effect upon education.

In this study, a data mining analysis of Twitter tweets obtained via the Twitter API will be used to study COVID-19 online college sentiment as opposed to the traditional survey. While the data gathered here were back in 2020, the final report preparation and publication were delayed to limitations placed by the ongoing COVID pandemic on resources and collaboration.

#### Sentiment analysis

Sentiment analysis, often called opinion mining, is a subset of NLP that focuses on determining the sentiment expressed in a piece of text, such as a social media post, a product review, or a news article. The goal of sentiment analysis is to classify a given text as having a positive, negative, or neutral sentiment. There are several techniques used in sentiment analysis, including:

1. Rule-based approaches: In this method, a set of rules and lexicons (lists of words with associated sentiment scores) are used to classify the sentiment of a text. For example, words like "awesome" and "fantastic" may be associated with a positive sentiment, while words like "terrible" and "disgusting" may be associated with a negative sentiment.
2. Machine learning approaches: In this method, machine learning algorithms, such as support vectors, decision trees and forests, and neural networks, are trained on a large annotated dataset to identify patterns and make predictions about the sentiment of new texts.
3. Hybrid approaches: This method combines rule-based and machine learning-based approaches to leverage the strengths of both methods and improve the accuracy of sentiment analysis.

Sentiment analysis has numerous applications, including social media monitoring, customer feedback analysis, and opinion mining. It can be used by businesses to gain

insight into their customers' opinions and preferences, by governments to track public sentiment about political issues, and by researchers to study the spread of information and ideas online.

It is important to note that sentiment analysis is a challenging task, and the accuracy of sentiment analysis can be influenced by factors such as the complexity of the text, the use of sarcasm and irony, and the cultural context of the text.

### Sentiment and emotion lexicons

A *sentiment lexicon*, also known as a sentiment dictionary or sentiment word list, is a list of words and phrases that are associated with a particular sentiment, such as positive, negative, or neutral. Sentiment analysis lexicons are used in rule-based natural language processing to perform sentiment analysis, which involves determining the overall sentiment expressed in a piece of text.

A typical sentiment analysis lexicon contains a list of words and phrases, along with their associated sentiment scores. The sentiment scores can range from  $-1$  (strongly negative) to  $1$  (strongly positive), or be assigned to one of several categories, such as positive, negative, and neutral.

Sentiment analysis lexicons can be constructed in several ways, including manual annotation, using crowd-sourced data, or using machine learning algorithms. The accuracy of a sentiment analysis lexicon strongly affects the quality of the results of sentiment analysis, so it is important to choose a lexicon that is well-suited for the specific task and language being used.

A text data mining *emotion lexicon*, also known as an emotion dictionary or emotion word list, is a list of words and phrases that are associated with specific emotions, such as joy, anger, sadness, and fear. Emotion lexicons are used in rule-based text data mining to perform emotion detection, which involves determining the emotions expressed in a piece of text.

A typical emotion lexicon contains a list of words and phrases, along with their associated emotion scores. For example, words like "excited" and "joyful" may be associated with the emotion of joy, while words like "angry" and "frustrated" may be associated with the emotion of anger. Like sentiment lexicons, emotion lexicons can be constructed in several ways, including manual annotation, using crowd-sourced data, or using machine learning algorithms.

### Results and discussion

In order to access and select Twitter tweets, one needs to write a program to call the Twitter API function and specify the search parameters. The search parameters for this study include the language, start date, end date, and keywords. Many methods, tools, and languages have

been used for unstructured text analysis. Commonly used languages are R and Python, and many tools are available for each language in the form of open source packages (libraries).

For this research, an R program script file was created to access the Twitter API. The choice of R over Python is just the author's preference as both languages have available libraries for the Twitter interface. Set (vector) processing and SQL program level functions are directly available in the R language.

The R program was run weekly between May and August of 2020. The first run in May represents the early sentiments when the pandemic became widespread in the USA and most colleges switched to online mode, and the last run in August represents the sentiments just before college started back up after the summer break. Thus the study provided not only comparative sentiment for those two critical points in time but also the full longitudinal analysis.

Each run of the R program downloaded the last six weeks of recent English language tweets containing the words "covid," "online," and "college." The R rTweet package was used. Search terms and other non-relevant terms (i.e., http, https, etc.) were stripped out.

The R tidyverse package was used for data cleaning and text processing. Tidyverse consists of open source R packages completed by Hadley Wickham [31–33]. The group of packages provide for an underlying design philosophy, grammar, and data structures for text processing, so called tidy data. Tweets were converted to tidy format, and the text was further converted to lowercase, punctuation was removed, and stop words (a, an, the, etc.) were also removed.

Next the count of each word was found, and a plot of word count in ascending order was plotted (via R ggplot function). Next a "word cloud" drawn using the R wordcloud package. Then a sentiment analysis was performed for both positive/negative sentiments and for standard emotions.

There are several commonly used lexicons for discovering the opinions and/or emotions expressed in the text. The tidytext package provides access to several lexicons, including:

- AFINN by Finn Arup Nielsen [19]
- BING by Bing Liu [5]
- NRC by Saif Mohammad and Peter Turney [16]

These lexicons provide a correlation between words and positive/negative opinions and/or specific emotions as anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. NRC groups words into these specific emotion categories, whereas BING places words into positive and

negative categories. AFINN associates words with a score between negative five and plus five for negative versus positive sentiment.

For the sentiment data analysis, one needs to first combine the data table containing the count for each word with the data table from a lexicon containing the sentiment score for each word. BING was used for positive/negative sentiment analysis, and the R inner join function was used to combine the word counts with the BING lexicon sentiments. The next step in the data analysis is to sort and group together the resulting joined table entries by sentiment. The final step is to then add up the entries for each sentiment to get a total sentiment score. In R one can perform these data analysis steps via a traditional nested “for loop” or use the R vector manipulation functions. However, a more direct set processing method is also available using SQL (Structured Query Language) as is commonly done in business applications with relational database tables. The R sqldf package was used so that SQL statements could be used to manipulate the R tidydata data frames, here using the SUM and GROUP\_BY functions; a code example is:

```
bing<-get_sentiments("bing")
count.words.bing<-inner_join(count.words, bing)
head(count.words.bing)
library(sqldf)
y<-count.words.bing
w<-sqldf("select sentiment, sum(n) as s from y group by sentiment")
```

NRC was used for emotional analysis by doing an inner join of our word counts with the NRC lexicon. The R sqldf package was again used so that SQL statements could be used to manipulate the R tidydata dataframes, again using the SUM and GROUP\_BY functions; a code example is:

```
nrc<-get_sentiments("nrc")
count.words.nrc<-inner_join(count.words, nrc)
z<-count.words.nrc
x<-sqldf("select sentiment, sum(n) as s from z where sentiment!='positive' and sentiment!='negative' group by sentiment")
```

The R qplot function was then used to visualize the emotional analysis.

**Word clouds**

The chart in Fig. 2 was created to itemize the top words appearing in the tweets:

However, a "Word Cloud" is perhaps a more effective and visually appealing graphic utilizing text size for visual emphasis of words weighted by their frequency of occurrence. It is easy to portray prominent words. Here the cloud shows the 100 most common words with the font size related to the word frequency. The initial word cloud is shown in Fig. 3.

**Initial sentiment and emotion analysis**

The initial May 6 2020 sentiment analysis was 44.9% negative and 55.1% positive. It was surprising to see more positive than negative considering the apparent more negative press on the subject. The initial emotional analysis is shown in Fig. 4, and it was surprising to see the trust emotion so high.

**Longitudinal analysis**

The R script was rerun weekly in 2020 between May 6 and August 10. Each time the last six weeks of tweets

are analyzed. The number of positive and negative items were entered into Excel each week, and then plotted versus time producing the moving average plot in Fig. 5. As can be seen, attitudes were becoming more negative as the summer progressed.

**Final sentiment and emotion analysis**

Shown in Fig. 6 is the final emotional analysis on August 10, 2020. It is now 65% negative and only 35%

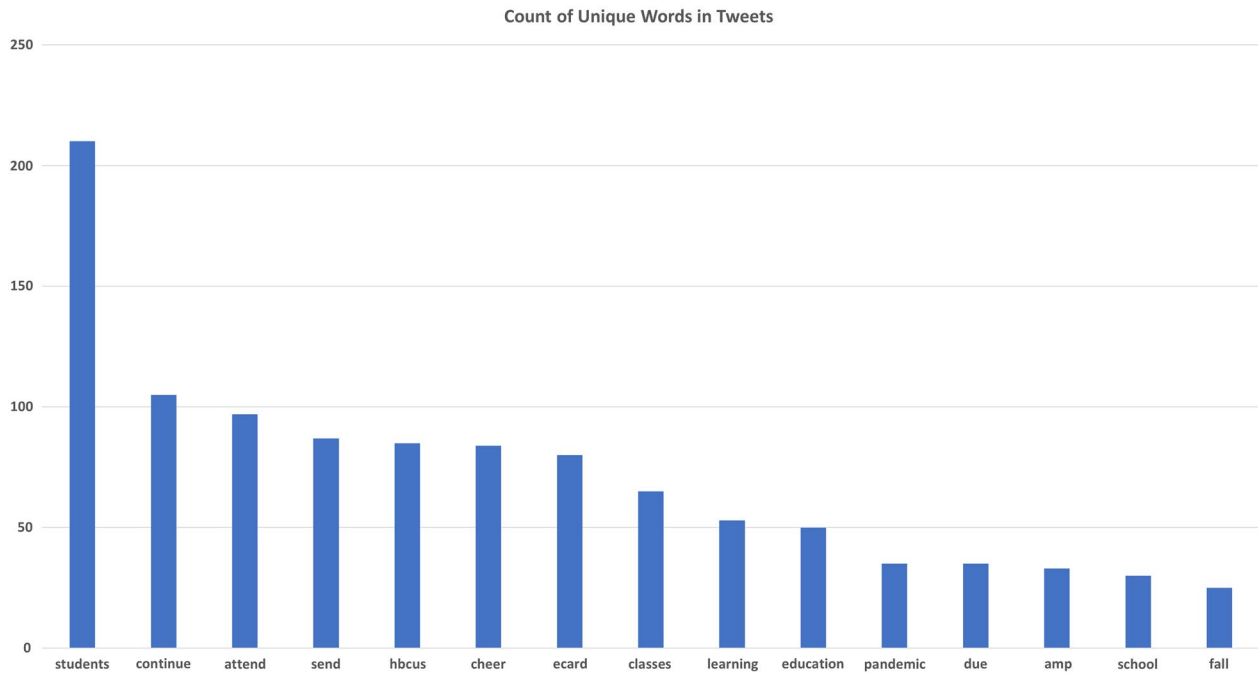


Fig. 2 Word counts



Fig. 3 Word cloud

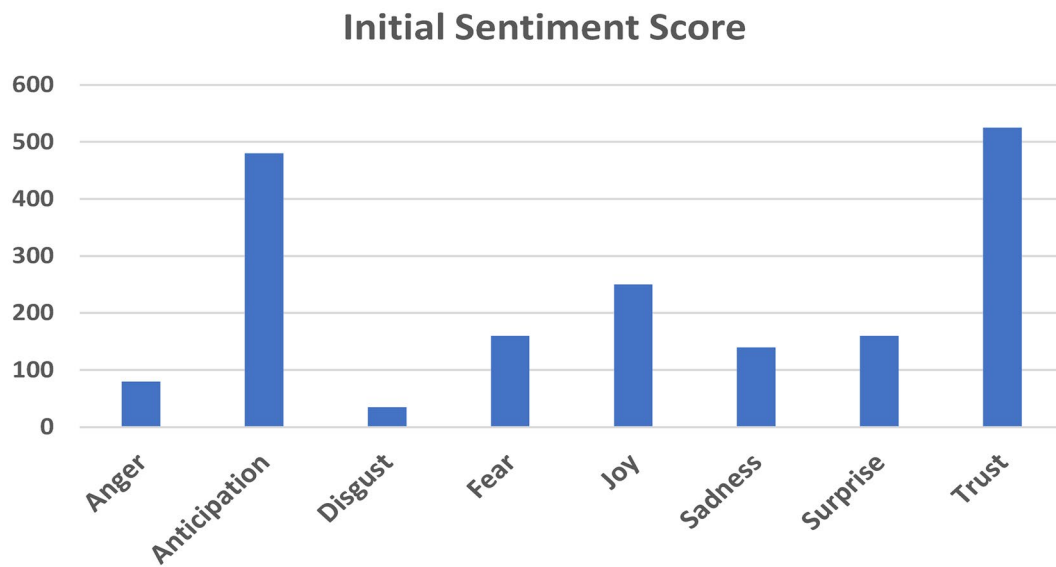
### Conclusions

In this paper, textual data mining has been used to gain insight into the feeling of Twitter users for the effect of COVID-19 on college education and the switch to online mode. Tweets were downloaded from Twitter using an R program written to select relevant English language tweets. The program was run weekly from May to August of 2020. Each week’s downloaded and mined data provided a viewport into users’ feelings via sentiment analysis. It was found that:

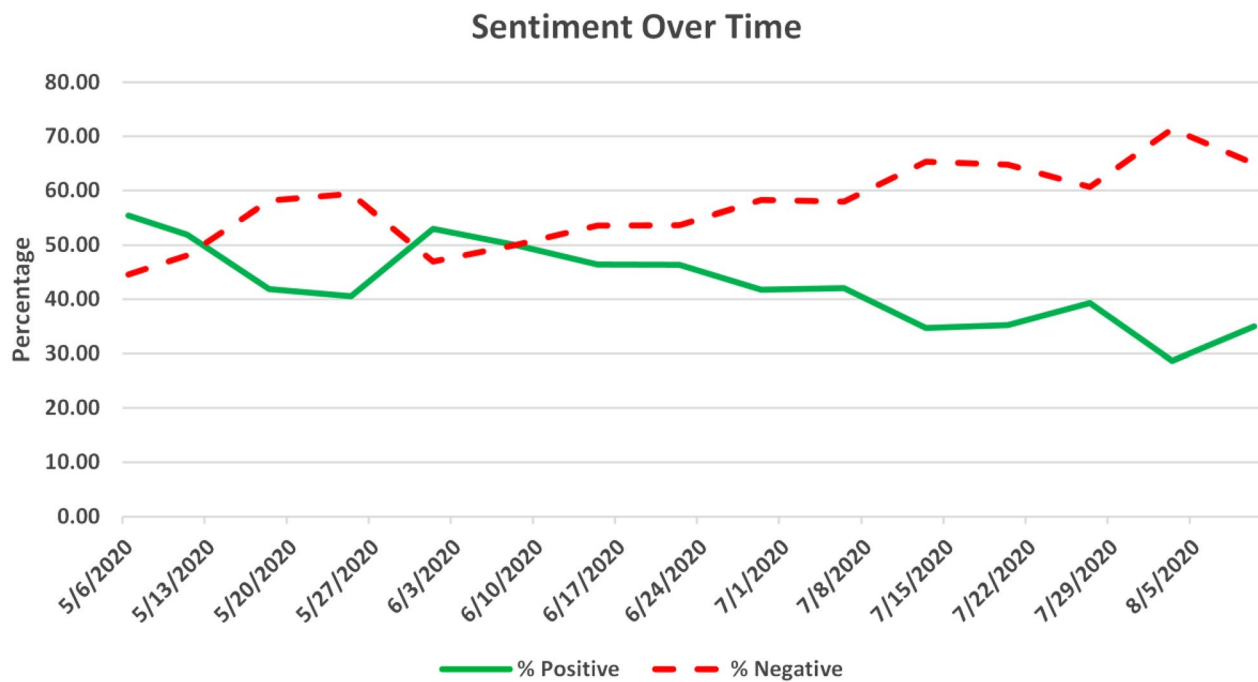
- The initial May 6 2020 sentiment analysis was 44.9% negative and 55.1% positive
- The initial emotional analysis showed strong for “trust” and “anticipation”
- The final August 10 2020 sentiment analysis was 65% negative and only 35% positive
- The final emotional analysis still showed strong for “trust” and “anticipation,” but “anger,” “fear,” and “disgust” were higher
- The longitudinal analysis over the summer of 2020 showed increasingly negative sentiment and reducing positive sentiment

positive. Trust and anticipation are still high, but there is more anger, fear, and disgust. This is consistent with surveys as the Simpson Scarborough National Survey.





**Fig. 4** Initial emotional scores



**Fig. 5** Sentiment over time

Our first research question was:

- Can valuable and definitive sentiment information be obtained from Twitter comparable to what might be obtained from a traditional survey?

The Twitter text mining analysis provided very similar overall results as the traditional survey, but at a much lower cost and time. The survey can, however, provide more detailed information. The next research questions were:

### Final Sentiment Score

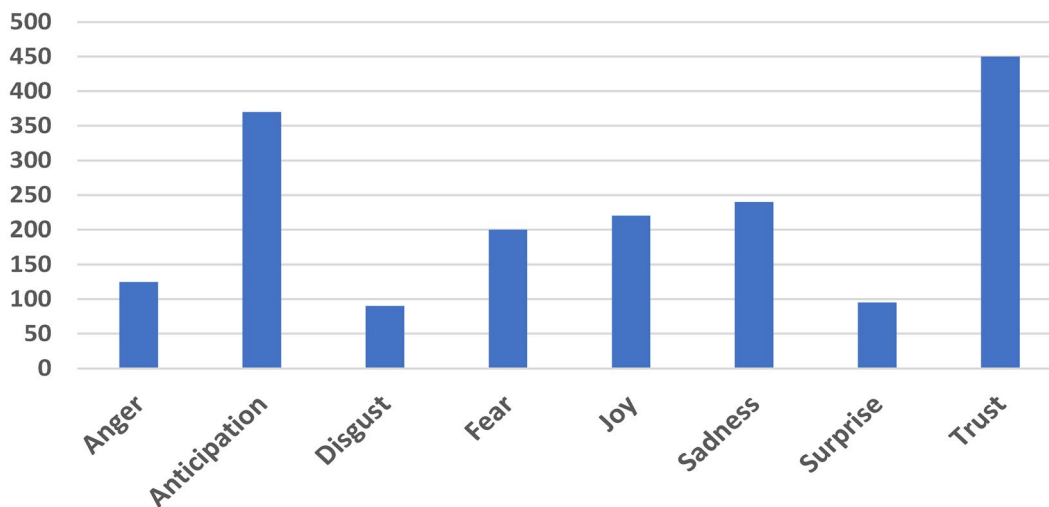


Fig. 6 Final emotional scores

- Are people’s sentiments positive or negative in regard to moving to online education due to COVID?
- Did people’s sentiments change from the time schools ceased physical operation and moved to online education to the time when the next semester started in the late summer of 2020?

It was found that the public’s sentiment about COVID’s forcing schools to online education was mixed, but somewhat more negative. Over the time between the initial school closings and the start of the fall term, the public’s sentiments became more negative. Thus, this paper has demonstrated both the process of text data mining and its application to current real-world issues.

**Abbreviations**

API	Application programming interface
IP	Internet protocol
NLP	Natural language processing
NYU	New York university
SMS	Short message service
SQL	Structured query language
TCP	Transmission control protocol
URL	Universal resource locator

**Acknowledgements**

Not applicable.

**Author contributions**

Not applicable, all contributions are from lone author.

**Funding**

The author did not receive support from any organization for this work. The author has no relevant financial or non-financial interests to disclose.

**Availability of data and materials**

The datasets generated during and/or analyzed during the current study are available to a Twitter user via the Twitter API. The R code used for this research is available from the author on reasonable request.

**Declarations**

**Ethical approval and consent to participate**

Not applicable, there were no animal nor human participants.

**Consent for publication**

Not applicable.

**Competing interests**

The author declares that they have no competing interests.

Received: 26 July 2023 Accepted: 27 November 2023

Published online: 12 December 2023

**References**

1. Abbas J (2020) The impact of coronavirus (SARS-CoV2) epidemic on individuals mental health: the protective measures of Pakistan in managing and sustaining transmissible disease. *Psychiatr Danub* 32(3–4):472–477. <https://doi.org/10.24869/psyd.2020.472>
2. Abbas J (2021) Crisis management, transnational healthcare challenges and opportunities: the intersection of COVID-19 pandemic and global mental health. *Res Glob* 3:100037. <https://doi.org/10.1016/j.resglo.2021.100037>
3. Ahmad N, Siddique J (2017) Personality assessment using Twitter tweets. *Procedia Comput Sci* 112:1964–1973

4. Ahmed W, Bath P, Demartini G (2017) Using Twitter as a data source: an overview of ethical, legal, and methodological challenges. *Adv Res Ethics Integr* 2:79–107
5. Bing L (2020) <https://juliasilge.github.io/tidytext/reference/sentiments.html>. Last Accessed June 1 2020
6. Farzadfar F, Naghavi M, Sepanlou SG, SaeediMoghaddam S, Dangel WJ, Davis Weaver N, Larijani B (2022) Health system performance in Iran: a systematic analysis for the global burden of disease study 2019. *The Lancet* 399(10335):1625–1645. [https://doi.org/10.1016/S0140-6736\(21\)02751-3](https://doi.org/10.1016/S0140-6736(21)02751-3)
7. Fung ICH, Yin J, Pressley KD, Duke CH, Mo C, Liang H, Fu KW, Tse ZTH, Hou SI (2019) Pedagogical demonstration of twitter data analysis: a case study of world AIDS day, 2014. *Data* 4:84
8. Hafeez A, Dangel WJ, Ostroff SM, Kiani AG, Glenn SD, Abbas J, Mokdad AH (2023) The state of health in Pakistan and its provinces and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Global Health* 11(2):e229–e243. [https://doi.org/10.1016/S2214-109X\(22\)00497-1](https://doi.org/10.1016/S2214-109X(22)00497-1)
9. Hess A (2020) How the class of 2020 became the class of COVID-19. CNBC
10. Hosch WL (2009) Twitter microblogging service. *Britannica*, <https://www.britannica.com/topic/Twitter>
11. Kelderman E (2020) Spurred by Coronavirus, Some Colleges Rush to Move Online. *Chronicle of Higher Education*
12. Kim EHJ, Jeong YK, Kim Y, Kang KY, Song M (2016) Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *J Inf Sci* 42:763–781
13. Marcus J (2020) New york times, Will the Coronavirus Forever Alter the College Experience?
14. Micah AE, Bhangdia K, Cogswell IE, Lasher D, Lidral-Porter B, Maddison ER, Dieleman JL (2023) Global investments in pandemic preparedness and COVID-19: development assistance and domestic spending on health between 1990 and 2026. *Lancet Global Health* 11(3):e385–e413. [https://doi.org/10.1016/S2214-109X\(23\)00007-4](https://doi.org/10.1016/S2214-109X(23)00007-4)
15. McCauley A (2020) How COVID-19 Could Shift The College Business Model: 'It's Hard To Go Back'. *Forbes*
16. Mohammad S, Peter T (2020) [https://emilhitfeldt.github.io/textdata/reference/lexicon\\_nrc.html](https://emilhitfeldt.github.io/textdata/reference/lexicon_nrc.html), Last Accessed June 2020
17. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, Brownstein JS (2014) A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *J Med Internet Res* 16:e236
18. NeJhaddadgar N, Ziapour A, Zakkipour G, Abolfathi M, Shabani M (2020) Effectiveness of telephone-based screening and triage during COVID-19 outbreak in the promoted primary healthcare system: a case study in Ardabil province, Iran. *Z Gesundh Wiss*, PP. 1–6.
19. Nielson F (2020) <https://github.com/fnielsen>, Last Accessed June 2020
20. Reyes-Menendez A, Saura JR, Alvarez-Alonso C (2018) Understanding WorldEnvironmentDay user opinions in Twitter: a topic-based sentiment analysis approach. *Int J Environ Res Public Health* 15:2537
21. Rolla News, Government Technology (2020) Researchers Study Social Media to Track COVID-19 Sentiments
22. Samuel J, Md Nawaz Ali GG, Md. Mokhlesur Rahman, Esawi E, Yana S (2020) COVID-19 public sentiment insights and machine learning for tweets classification. <https://www.preprints.org/manuscript/202005.0015/v1>, Last Accessed June 2020
23. Samuel J, Garvey M, Kashyap R (2019) That message went viral?! exploratory analytics and sentiment, analysis into the propagation of tweets. In: *Annual Proceedings of Northeast Decision Sciences Institute (NEDSI), Conference, USA*
24. Schmidt CA, Cromwell EA, Hill E, Donkers KM, Schipp MF, Johnson KB, Hay SI (2022) The prevalence of onchocerciasis in Africa and Yemen, 2000–2018: a geospatial analysis. *BMC Med* 20(1):293. <https://doi.org/10.1186/s12916-022-02486-y>
25. Simpson (2020) Higher ed and Covid-19, simpson scarborough national survey. [simpsonscarborough.com](https://simpsonscarborough.com), Last Accessed June 2020
26. Skoric MM, Liu J, Jaidka K (2020) Electoral and public opinion forecasts with social media data: a meta-analysis. *Information* 11:187
27. Shoib S, GaitanBuitrago JET, Shuja KH, Aqeel M, de Filippis R, Abbas J, Arafat SMY (2022) Suicidal behavior sociocultural factors in developing countries during COVID-19. *Encephale* 48(1):78–82. <https://doi.org/10.1016/j.encep.2021.06.011>
28. Soroush A, Ziapour A, Abbas J, Jahanbin I, Andayeshgar B, Moradi F, Cheraghpouran E (2021) Effects of group logotherapy training on self-esteem, communication skills, and impact of event scale-revised (IES-R) in older adults. *Ageing Int* 47(4):758–778. <https://doi.org/10.1007/s12126-021-09458-2>
29. Wang D, Su Z, Ziapour A (2021) The role of social media in the advent of COVID-19 pandemic: crisis management, mental health challenges and implications. *Risk Manag Healthc Policy* 14:1917–1932
30. Wang Z, Ye X, Tsou MH (2016) Spatial, temporal, and content analysis of Twitter for wildfire hazards. *Nat Hazards* 83:523–540
31. Wickham (2020) <https://joss.theoj.org/papers/https://doi.org/10.21105/joss.01686>, Last Accessed June 2020
32. Wickham (2020) Tidyverse. [www.tidyverse.org](http://www.tidyverse.org). Last Accessed June 2020
33. Wickham H, Grolemund G (2017) *R for data science*, O'Reilly
34. Ye X, Li S, Yang X, Qin C (2016) Use of social media for the detection and analysis of infectious diseases in China. *ISPRS Int J Geo Inf* 5:156
35. Yu S, Draghici A, Negulescu OH, Ain NU (2022) Social media application as a new paradigm for business communication: the role of COVID-19 knowledge, social distancing, and preventive attitudes. *Front Psychol* 13:903082
36. Young J (2020) Scenes from college classes forced online By COVID-19, <https://www.edsurge.com/news/2020-03-26-scenes-from-college-classes-forced-online-by-covid-19>, Last Accessed June 2020

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---