A sentiment analysis on COVID-19 tweets
-
Research Report

John Yoo
Tom Willis
Yilu Liu

Date of submission
02/06/2020

Table of contents

1 Research Question

Under the current COVID-19 situation, our team initially drafted a research question: How does the sentiment surrounding COVID-19 differ over time and between locations, or to pose the question more broadly, we are conducting a sentiment analysis to discover public opinion specifically surrounding COVID-19 and to discover how the pandemic impacts peoples' mental health. To analyse the sentiments, we decided to conduct a quantitative research which implements sentiment analysis technique from natural language processing to determine the sentiment scores for Tweets discussing COVID-19. To answer this question, we chose to train a neural network model using SVM classification on COVID-19 Tweets posted from March to April 2020. The model takes a tweet as input, analyses its wording and labels it as negative, neutral or positive with a sentiment score.

Regardless, our research question has been narrowed down and refined during the training process, as we came to realise our model's training accuracy has reached a bottleneck of ~70.69% and we think this indicates any results we obtain from our model might not represent an accurate enough view on the sentiments surrounding COVID-19. Therefore, our current research question focuses on improving our neural network model first using different feature extraction techniques to provide high-performing sentiment predictions on Tweets related to COVID-19.

2 Introduction

There exist quite a few reasons why we consider the question, and research into the field of pandemics, is both interesting and important.

2.1 Background

Several reports have shown that when the outbreak hits the globe, most governments have their top priorities as flattening the curve to ensure hospitals have enough beds for patients and as saving the economy [1]. Consequently, people's mental health took a back seat to physical health. Common consequences of disease outbreaks include stress, anxiety, depression, anger and uncertainty with estimates of between 25% and 33% of the general public [2]. A recent US research shows that with lockdowns, non-essential-industry closures, travel restrictions and other social distancing dictates, levels of mental illness are eight times higher than usual among adults over the world [3]. Obviously, Coronavirus itself, with the social and financial disruption it has caused, is affecting the way people live, work, and communicate and it's also making a significant impact on our mental health.

2.2 Social media and mental health

As another impact COVID-19 has done, global in-home social media consumption has spiked during the pandemic [4], resulting in COVID-19 topics taking over different social media giants, such as Twitter. Since early March, COVID-related trends on Twitter keep taking up most of the Twitter's major pages including "# Explore" and "What's happening", with millions of users twittering about COVID-19 and sharing their experiences, thoughts and concerns surrounding the novel virus. This situation has offered ample text samples for our research and it also indicates underlying needs of analysing the reason behind the social media usage boost.

Moreover, there's a direct relation between social media and mental health. People with mental illness are increasingly turning to popular social media for multiple purposes, e.g. to share their feelings and to seek advice [5]. During the current situation, the need of seeking help from the community has increased and this has been reflected by the increase of numbers of such Tweets. However, limited research exists exploring social media utilisation by individuals experiencing different levels of mental illness, despite the fact that social media could serve as a significant part of mental health analysis platform [6], we hope our research could serve as a prompt for this open thread as to offer insights to other researchers.

2.3 Aim of sentiment analysis

As an important information-gathering technique, sentiment analysis could help us analyse the pandemic situation to offer ways to mitigate its impacts. It addresses a current issue that has yet to be analysed to a great depth or have lessons learnt from. It is a global issue affecting many parts of life across the globe, meaning answers could provide value to a diverse and wide reaching audience. The sentiment could be an indicator to population mental health and wellbeing, which can be plotted against time and analysed against events occurring or other trends in order to gain value.

Different governments could use such scores, in conjunction with other scores, as a rating of their pandemic response at different times. Analysing how the public trusts the government's response, or how effective the public finds government policy. This retrospection could be on, and provide value to, the current policy, pandemic mechanisms in place, what governments could do better, and what the people think their government could do better [7].

Using technology to harvest public opinion is also an important and growing field. O'Connor and Conway claimed that social media has been used extensively in marketing for sentiment analysis and negative-emotion language on Twitter has been shown to correlate well with influential events [8]. It is a field that is still evolving and improving and research in this field would be helpful.

3 Team

3.1 Teamwork

Work was distributed between team members in a voluntary way. After working on the project during our weekly tutorial, the team would get together and individually volunteer what they were wanted to work towards for the next week. Or if required, sooner. The end of tutorials gave us an opportunity to peer review each other's work, as one or two of us would later give a short presentation to the class. This required us to have an understanding of each person's contributions.

Tom found investigating current research methods and finding most comfortable so he decided to focus his efforts there.

Yilu has majorly devoted her time on researching different quantitative research methods for conducting sentiment analysis; collecting and organising various datasets of COVID-19 tweets then compare each datasets' attributes to find the best dataset as input for the research model; and preprocessing data to form a new dataset for the model.

John implemented the classification model, conducted analysis on it and focused his time on researching sentiment and semantic features to extract from the dataset.

Activity	**Members**
Brainstorming of ideas	Tom, Yilu
John joins the team	
Research of different methods	Everyone individually researched but compiled together
Decide on our research question	Together
Decide on research method	Together
Current state of the art	Tom
Gathering Twitter data	Yilu, John
Model and other magic	Yilu, John

3.2 State of the art summary

Before conducting research it is important to gain a current view of the research field you are attempting to expand. Our question lends itself to many fields, the most direct field being the sentiment surrounding the COVID-19 pandemic and other pandemics. As this is close to our own question, we would like to understand and expand upon the questions currently being asked and answered surrounding COVID-19 public sentiment, while also discovering other interesting questions we hadn't thought to ask.

There are already papers analysing the COVID-19 pandemic. Including one [9], analysed in depth later, that examines what information and insight a research team was able to gleam, by mining data from a platform similar to Twitter.

Another paper [10] looked at India's lockdown and conducted a sentiment analysis on Twitter, much like we have. They found that India's population on twitter were overall positive and trusting of the government's use of a lock-down to stifle the spread of COVID-19 .

Looking into the past now to see what we can learn from research on past outbreaks. During the 2009 pandemic of H1N1, a paper [11] used twitter to track levels of disease activity and public concern in the U.S. The researchers tracked the use of keywords to indicate the user's concern on a topic, and plotted this over time. Interestingly, this study found that the number of tweets about H1N1 does not correlate with the actual number of cases, showing a relation where the talk of H1N1, and the concern, drops as the number of cases increases. This could indicate that the public's concern surrounding H1N1 is low.

The H1N1 outbreak was also known colloquially as Swine Flu. Another Twitter analysis [12] tracked the use of the two terms. The colloquial Swine Flu, and the World Health Organisation official H1N1. They took a random sample of 25 tweets per hour, ignoring retweets to prevent popular posts from being prominent in the sample, and found that the use of the term H1N1 increased over time, while the term Swine Flu decreased.

3.2 Individual research paper analysis

Building the state of the art in sentiment analysis of tweets [13], summary by John Yoo

This paper has relevance to our research question as it implements a state of the art sentiment classifier using an SVM model, on the same tangent to what we have implemented for our research question.

Introduction: The purpose of the research study was to construct 2 state of the art sentiment analysis SVM classifiers to predict the sentiment of messages in tweets and SMS and to predict the sentiment of a term within a message. The model created was submitted in the Sentiment Analysis competition organised by the Conference on Semantic Evaluation Exercises Wilson et al. [13] and placed first in the tweet sentiment classification category.

Methodology: Two large word sentiment association lexicons were created for the purposes of the study, one from tweets with sentiment word hashtags and the other from tweets with emoticons. Using these two lexicons alongside a variety of semantic and sentiment features, an SVM model was trained on 9912 annotated tweets and tested on 3813 unseen tweets to achieve an F-score of 69.02 on the tweet set and 68.46 on the SMS set.

Pre-processing was conducted on the tweets, all URLs and user-ids were normalised to a default placeholder, each tweet was tokenised and part of speech tagged. A feature vector was created for each tweet using the following features:

Word n-grams: presence of n-grams of tokens and character n-grams. Presence of all capital character words. The number of occurrences of each part of speech tag and the number of hashtags used.

3 manually constructed sets of sentiment lexicons were used to score each token using a sentiment/emotion equation the features implemented were total count, total score, max score and score of the last token.

Other features implemented were: numbers of contiguous sequences of exclamation/question marks, whether the last token contained an exclamation or question mark. Number of elongated words. Presence or absence of 1000 token clusters representing tweet content, and the total count of negated contexts.

Results:

Table 2: Message-level task: The macro-averaged F-scores obtained on the test sets with one of the feature groups removed. The number in the brackets is the difference with the all features score.

Experiment	**Tweets**	**SMS**
All features	69.02	68.46
All - lexicons	60.42 (-8.60)	59.73 (-8.73)
All - ngrams	61.77 (-7.25)	67.27 (-1.19)
All - word ngrams	64.64 (-4.38)	66.56 (-1.9)
All - char. ngrams	67.10 (-1.92)	68.94 (0.48)
All - negation	67.20 (-1.82)	66.22 (-2.24)
All - POS	68.38 (-0.64)	67.07 (-1.39)
All - clusters	69.01 (-0.01)	68.10 (-0.36)
All - encodings	69.16 (0.14)	68.28 (-0.18)

The most influential feature was the sentiment lexicon features with n-grams being the second most influential. Removing the sentiment encoding features such as hashtags, emoticons and elongated words had no impact on the performance.

The second classifier trained on term sentiment within a message utilised a similar feature set. Additional features added were: Upper case count, count of stop-words, position of the term and term splitting: whether the term contained a hashtag made up of multiple words.

An SVM classifier was trained on 8891 annotated terms in tweets and applied to 4435 tweets in the test set and 2334 terms in unseen SMS messages as well.

The model obtained an F-score of 88.93 on the tweet set and 88 on the SMS set. As opposed to the sentiment message classifier, n-gram features were the most influential with sentiment lexicon features being the next most useful.

Table 3: Term-level Task: The F-scores obtained on the test sets with one of the feature groups removed. The number in the brackets is the difference with the all features score.

Experiment	Tweets	SMS
All features	89.10	88.34
All - ngrams	83.86 (-5.24)	80.49 (-7.85)
All - word ngrams	88.38 (-0.72)	87.37 (-0.97)
All - char. ngrams	89.01 (-0.09)	87.31 (-1.03)
All - lexicons	85.15 (-3.95)	83.70 (-4.64)
All - negation	88.38 (-0.72)	86.77 (-1.57)
All - stopwords	89.17 (0.07)	88.30 (-0.04)
All - encodings	89.16 (0.06)	88.39 (0.05)

Conclusion: In this research paper, 2 SVM classifiers for the classification for message level and term level sentiment were trained. The conclusion was reached that from both sentiment message and sentiment term classifiers a variety of features based on surface form and lexical categories can produce a high performing model. It was found that the sentiment lexicon features along with n-gram features were the most influential features overall.

Top Concerns of Tweeters During the COVID-19 Pandemic: Inforveillance Study by Yilu Liu

One of the team members, Yilu, has also discovered a study [14] conducted last month that used unigrams and bigrams to analyse word frequencies and then implemented Latent Dirichlet allocation for topic modelling to identify the topics of tweets related to COVID-19. Moreover, they have as well performed sentiment analysis to calculate the interaction rate per topic. As a result, they have found twitter users interests: origin of the virus, how to not get infected, its impact on people, countries and economy. For the sentiment analysis part, the research outcome showed that the mean sentiment was positive for 10 topics and 2 topics are negative.

Thus, the research team reached a conclusion that social media provides an opportunity to communicate and share health information in public and there's a need for a more proactive and agile public health presence on social media to combat the spread of fake news. So despite using similar methodologies like our team, the research has a slightly different initiative and their conclusion therefore focuses on proving the necessity of including social media as a portal of health information communication for the general public and monitoring social media. Instead, we wish to conduct sentiment analysis to explore the relationship between peoples' sentiments revealed through their Tweets and the pandemic situation itself to see if the pandemic has done any visible impact on our mental health.

However, there are a lot of quantitative research methods implemented in their research that are worthing looking at. For instance, they had some useful data preprocessing techniques, which includes removing non-English tweets, normalising usernames, lemmatising different spellings of the same words. The first two of the above rules have also been implemented by our team since input data prepossessing is an essential task for informal texts and could ensure more efficient sentiment analysis [15]. In addition, they have also taken an interesting path analysing the processed tweets using unigram and bigrams, which we have also implemented as a feature, and their approach served as a good reference to us. Needless to say, we have also compared with their sentiment analysis method as we were using the same Python textblob library for validating our model performance in the beginning of our research. That being said, our focus is on the sentiment analysis side instead of Latent Dirichlet Allocation leveraging, we then abandoned their sentiment analysis method in order to include more complicated features.

Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China by Tom Willis

The team member, Tom, looked at a study [9] conducted in early 2020 that used social media to analyse public opinion of COVID-19 in China. The study specifically looks at the social media platform Sina-Weibo, which is comparable to Twitter. The researchers pursued a few different methods of mining and analysing the Weibo text. The methods include the analysis of topics, a spatial analysis, and a time series analysis.

The researchers put a lot of thought into their time series analysis on the number of messages each day. They immediately identify that the amount of messages has low and high peaks every day. They also made an effort to eliminate the seasonal effects on the data. They found that the number of messages per day seemed to rise in line with the trend of confirmed cases of COVID-19.

The researchers also analysed the spatial distribution of the analysed Weibo text messages. Their results showed that the texts were concentrated in the east-central areas of the country, with a high density of COVID-19 related texts from COVID-19 hotspots, Wuhan, Beijing, and Shanghai. The researchers performed a Spearman correlation analysis to define whether this correlation is statistically significant. The results include a Spearman correlation coefficient of

0.84 and the value (p = 0.00 < 0.01), indicating that this is a significant positive correlation with confidence.

The researchers created a topic extraction and classification model based on the random forest algorithm and a latent Dirichlet allocation model. They used this model to identify topics and subtopics related to COVID-19 in the Weibo text messages. These there are seven main topics including " 'events notification','popularization of prevention and treatment', 'government response', 'personal response', 'opinion and sentiments', 'seeking help', and 'making donations' " [9]. The topics of personal response, opinion and sentiments, and seeking help were further divided into 13 sub-topics so that this information is more specific to what the topics refer to.

4 Implementation

4.1 Initial stages

To answer our research question on uncovering public opinion surrounding COVID-19 we decided to train a classifier and use the classifier to predict the sentiment of unannotated COVID-19 related tweets in order to gather information, extract trends and patterns regarding its impact on the social media platform. We chose to do this as with this information we would be able to determine various factors to how the pandemic has affected us. In particular we would be able to extract sentiment from specific locations and specific timeframes in order to construct a view on how developments of the pandemic and government responses have been received.

In our initial stages of research we focused on gathering information from various related research papers in the field. Our first objective was to determine how studies regarding past pandemics were conducted, and this gave us an insight into the importance of such a study and the outcomes we could expect. In particular Signorini et al [16] used twitter to track levels of disease activity and public concern in the U.S during the H1N1 pandemic, this paper motivated us to use a classification model in our research study.

4.2 Choosing a model and dataset

At this point we had decided to use a classification model to predict the sentiment of COVID19 tweets. Go and L.Huang [17] found that SVM performed the best in classification on tweets, outperforming Naïve Bayes and Maxentropy models, we opted to use SVM for our research question as it has been proven to be the most effective on such text categorisation tasks and especially robust on large feature spaces which was important for our specific task.

Following this, we collated a number of potential datasets to use for our model. As candidates we had Stanford Universities dataset of 4 million positive and negative annotated tweets, Kaggle annotated COVID19 datasets and live covid19 tweets collated by Lamsal [18]. We made the decision to use the Kaggle annotated COVID19 tweets as we firstly required COVID19 tweets for our classifier, and secondly the annotations allowed use of the tweets directly for our training and testing datasets.

The Kaggle dataset we chose contained annotated tweets ranging from 12/03/2020 to 30/04/2020. The sentiments were one of three possibilities, either: positive, neutral or negative, we chose a random subset of 4000 from each in our dataset.

In the first stages of implementing our SVM model we initially trained the model on the whole tweets in our dataset and obtained a classification accuracy of ~63% on a test set of 1000 unseen tweets. Our goal now was to improve upon its predictive performance.

We improved our model in two steps, preprocessing and feature extraction.

4.3 Preprocessing and feature extraction

Basing our preprocessing steps loosely on Mohammad et al [19], we performed data preprocessing on each of the 12000 tweets. For the purposes of our research question we only considered English tweets. All URLs, user-ids and hashtags were removed from each tweet along with all unnecessary punctuation, these tokens were removed as they are not useful in determining the sentiment of a tweet. We also attempted to correct all spelling errors within each tweet and we removed stop words; which are a set of commonly used words that do not help in extracting sentiment.

With our data preprocessing done we were ready to extract features from our tweets to train our SVM model. From pang et al [20] we adopted the unigram and bigram absence and presence feature and extended upon this. We implemented 3 additional categories of features. The second feature category we implemented utilised lexicon dictionaries, we adopted a dictionary of 6800 positive and negative labelled words to create multiple features. We implemented the sum of sentiment scores for positive and negative words in each tweet, a frequency score for positive and negative words and the sentiment of the last positive or negative word in each tweet.

The third feature category we focused on was negation. From our research on relevant studies we discovered that negation was an important semantic aspect in classifying bodies of text, however, difficult to implement. We proposed the following implementation idea to detect negation within a tweet for our study: identify the negating term within the sentence and determine its negating influence on the tweet, by this we mean to determine where the negation ends. For example in the sentence "Can't find any toilet paper anywhere. Found this pair of slippers for sale though" the word "Can't" is the negating term and its negation ends with the full stop before "Found". We implemented our own list of stop punctuation to determine when a negation's range of influence would end the list includes colons, semi-colons, full stops, exclamation marks and question marks. A negation tag "_NEG" was appended onto each negated word token, in this way we would not lose any information on exactly which part of the sentence was negated when we eventually created bigrams and unigrams from the tweet.

The last feature category we implemented for our model was part of speech tags. Part of speech tags are labels assigned to each word in a body of text to indicate its part of speech and often also other grammatical categories such as its tense, if its plural or singular, etc.

We imported and used the NLTK library to tag each word within each tweet, and then similar to the negation feature we appended this tag to the word as a way to bind the information of the tag and the word into one token.

4.4 Testing and obtaining results

We separated our dataset into 10000 tweets in the training set and 2000 in the test set. Using the features we were able to train the SVM classifier on the set of 10000 annotated tweets. We performed cross validation and tuned the hyperparameters for the SVM model, choosing the RBF kernel and the value C = 0.005. Applying the test set of 2000 unseen tweets to the classifier and testing our model on this test set returned a classification accuracy of 70.69%.

We also experimented on repeating the same classification process but with the removal of one feature group at a time in order to find the most influential features for the COVID-19 dataset. With this insight we hypothesised that we would be able to further improve our model. We found that the most influential feature was the set of lexicon dictionary features. The second most influential features were the unigram and bigrams feature sets. Interestingly we found that removing the negation feature actually aided in the predictive performance of the classifier.

AS our model had a relatively low accuracy performance especially for the neutral class classification we came to the conclusion that any results we obtained from predicting using the model would not represent an accurate enough view on the sentiment surrounding COVID19 and that our results would not be useful. Thus we instead opted to focus our efforts on analysing the faults with our model first and attempt to make improvements.

For our implementation of the SVM model used to classify sentiment on COVID19 related tweets we found 2 major flaws that we hypothesised to have influenced the relatively low predictive performance. The first flaw can be seen in the low neutral class true positive predictive percentage, our model had a 50% accuracy on classifying neutral tweets. A major reason for this flaw could potentially be seen in our implementation of the lexicon dictionary feature. For this feature we overlooked neutral words and only focused on determining positive and negative sentiments, together with the fact that this feature was the most influential in our model we can see that improving upon this feature could help in the prediction of neutral class tweets. The second flaw we identified after testing our model was the detrimental negation feature. We found that removing the negation feature actually aided in our models ability to classify the sentiment of the tweets. After a review of our feature implementation and further research into the topic we found that our negation feature had not covered all aspects of negation in the english language. Xiang et al [21] explained the differences in explicit and implicit negation, and levels of semantic representation, from this paper we were able to identify that our negation feature was lacking in many aspects. Words such as “fails” in the sentence “it utterly fails to solve the main problem” were not caught by our feature set to be a negation, there were many such examples in our dataset that were not caught.

In the end we were unable to implement changes to our features to improve our model past our all features SVM model performance of 70.69%. For our research question we were able to propose a method to uncover public sentiment by training a classifier and using the classifier to obtain sentiment values for unannotated tweets relating to COVID-19. We were able to identify and implement key preprocessing methods, and features to train a SVM model, however, due to our models low performance we were unable to report accurate sentiment values. We proposed 2 faults that are potentially a major factor for this, and in a later section we propose 2 potential solutions to solving these faults.

5 Result and reflection

5.1 Outcome

As stated in section 4, with different features extracted, we trained the SVM classifier using the 10000 annotated tweets. We implemented cross validation to show how predictive the accuracy is depending on the sample tweets size and the number of predictor variables. We have also tuned our hyperparameters for our model in order to achieve best performance. Then we applied a test set of 2000 tweets to the model and gained varied different classification accuracy, which are displayed below:

Experiment	**Accuracy(%)**
All features	**70.69**
All - lexicons	63.86
All - lexicon score	67.56
All - lexicon freq	66.45
All - lexicon last word	68.67
All - word ngrams	65.34
All - negation	71.54
All - POS	68.77

Table 4. Prediction accuracy by feature experiments

Impressively, the table indicates that removing the negation feature would aid in the classification prediction performance. We as well discovered from the table that the most influential feature was the set of lexicon dictionary features, and the second most being the unigram and bigrams feature sets.

We have also generated the confusion matrix for our SVM model with all features with a True Positive Percentage demonstration, the positive, negative had a positive percentage of 82% and 76&, while the neutral class still returned a 50% accuracy.

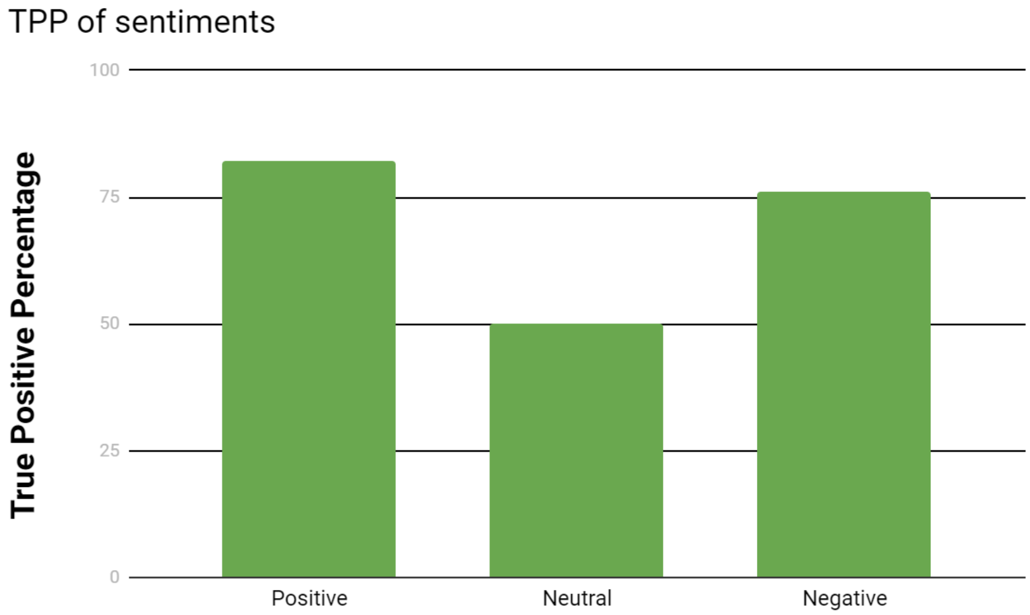

Figure 1. True Positive Percentage of *Positive*, *Neutral* and *Negative* classes

	Predicted: **Positive**	Predicted: **Neutral**	Predicted: **Negative**
Actual: **Positive**	**546**	84	36
Actual: **Neutral**	145	**334**	187
Actual: **Negative**	64	96	**506**

Table 5. Confusion matrix for *All features* model.

From the result we concluded, we decided we would not conduct further analysis on predicting sentiments surrounding COVID-19 as the model is not able to generate highly accurate predictions on the view of sentiment studies with COVID-19 tweets until the performance is improved. As previously mentioned, because of the existing accuracy problem, we are not able to answer our original question so we changed the scope of our research.

5.2 Future Work

As mentioned in our report of what we did to answer our research question, we were unable to completely fulfill the outcomes we set when we proposed our question initially. A classifier was implemented but we were unable to obtain accurate results. We identified 2 flaws: a low neutral class prediction score and a detrimental negation feature. Here we propose 2 potential solutions in order to solve these issues.

Peddiniti et al [22] proposed the idea that the same sentence or phrase could have different meanings and thus a different sentiment in different domains. An example of this is the word "unpredictable" which is positive in the context of movies, however, negative if used in most other contexts. This is the general idea of domain dependence, and we hypothesised that with a lexicon dictionary of words relating specifically to COVID-19 with all three classes of sentiment we would be able to improve on the neutral tweet classification accuracy and thus the performance of the model as a whole.

The second flaw was the detrimental negation feature, we identified that there our feature implementation did not catch all forms of negation, we propose that by implementing a context extractor using an algorithm such as logistic regression we could cover the aspects of negation we were unable to in our implementation. From Xiang et al [21] we would be able to focus on covering both implicit and explicit negations and assign levels of sentiment to both.

With these two additions a model with a much higher performance could be trained and using this we would be able to fulfill our original research question. Applying sets of unannotated tweets, pertaining to specific timeframes and locations we could uncover trends and discover valuable information in how COVID-19 has affected public sentiment and how this is changing over time.

5.3 Project Summary

One aspect worth reflecting on is the choice of research methods we implemented. As there are various research methods covered in the lectures and we acquire barely no previous research experience, we in the beginning tried to look into as many of the methods as possible, and as a result we allocated a great deal of extra time and effort studying different methodologies. But in the end, due to changes of our research question and scope, only a small portion of the methods we investigated actually got implemented by us. Perhaps next time when we have experience deciding the implementation of our next research, with better time management skills, we would determine our research approach in a more effective manner and we would be able to save some time that could be used on conducting other analysis.

In hindsight there are two principal points we would have changed: 1) we should have dedicated more time to the initial stages of our research in order to determine the feasibility of our model from an earlier stage. We realised at a late stage that our models performance was too low to use in sentiment analysis of unannotated tweets. As a result of this, we were unable to fulfil our initial research questions goal. 2) We should have searched more extensively for features to extract from our training dataset. While we did implement key semantic and sentiment features it was in the end not sufficient enough to obtain a high predictive performance.

6 Reference

[1]"Government support and the COVID-19 pandemic", OECD, 2020. [Online]. Available: http://www.oecd.org/coronavirus/policy-responses/government-support-and-the-covid-19-pandemic-cb8ca170/. [Accessed: 02- Jun- 2020].

[2]"Mental Health Ramifications of COVID-19: The Australian context", Black Dog Institute, 2020. [Online]. Available: https://www.blackdoginstitute.org.au/wp-content/uploads/2020/04/20200319_covid19-evidence-and-reccomendations.pdf. [Accessed: 02- Jun- 2020].

[3]J. Twenge and T. E. Joiner, “Mental distress among U.S. adults during the COVID-19 pandemic”, 07-May-2020. [Online]. Available: psyarxiv.com/wc8ud. [Accessed: 02- Jun- 2020].

[4]S. Fischer, "Social media use spikes during pandemic", Axios, 2020. [Online]. Available: https://www.axios.com/social-media-overuse-spikes-in-coronavirus-pandemic-764b384d-a0ee-4787-bd19-7e7297f6d6ec.html. [Accessed: 02- Jun- 2020].

[5]J. Naslund, K. Aschbrenner, L. Marsch and S. Bartels, "The future of mental health care: peer-to-peer support and social media", Epidemiology and Psychiatric Sciences, vol. 25, no. 2, pp. 113-122, 2016. Available: 10.1017/s2045796015001067.

[6]A. Shepherd, C. Sanders, M. Doyle and J. Shaw, "Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation", BMC Psychiatry, vol. 15, no. 1, 2015. Available: 10.1186/s12888-015-0408-y.

[7]R. Arunachalam and S. Sarkar, "The New Eye of Government: Citizen Sentiment Analysis in Social Media", Semanticscholar.org, 2013. [Online]. Available: https://www.semanticscholar.org/paper/The-New-Eye-of-Government%3A-Citizen-Sentiment-in-Arunachalam-Sarkar/422b44f254f540817ca615cbb24d793f9be7bebf. [Accessed: 02- Jun- 2020].

[8]M. Conway and D. O'Connor, "Social media, big data, and mental health: current advances and ethical implications", Current Opinion in Psychology, vol. 9, pp. 77-82, 2016. Available: 10.1016/j.copsyc.2016.01.004 [Accessed: 02 - Jun - 2020].

[9] X. Han et al, "Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China," International Journal of Environmental Research and Public Health, vol. 17, (8), pp. 2788, 2020. [Online]. Available: https://search-proquest-com.virtual.anu.edu.au/docview/2393189872?pq-origsite=summon. [Accessed: 02- Jun- 2020]

[10] A. Signorini, A. M. Segre and P. M. Polgreen, "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic," PLoS One, vol. 6,

(5), 2011. [Online]. Available: https://search-proquest-com.virtual.anu.edu.au/docview/1295071593?pq-origsite=summon. [Accessed: 02- Jun- 2020]

[11] G. Barkur, Vibha, G. B. Kamathx, "Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India," 2020. [Online]. Available: https://www-sciencedirect-com.virtual.anu.edu.au/science/article/pii/S1876201820302008. [Accessed: 02- Jun- 2020]

[12] C. Chew, G. Eysenbach, "Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak", 2010. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0014118#pone.0014118-Picard1. [Accessed: 02- Jun- 2020]

[13] T. Wilson, Z. Kozareva, P. Nakov, S. Rosenthal, V. Stoyanov, and A. Ritter. 2013. SemEval-2013 task 2: Sentiment analysis in twitter. In Proceedings of the International Workshop on Semantic Evaluation, SemEval '13, Atlanta, Georgia, USA, June. [Accessed: 01 - Jun - 2020]

[14]A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi and Z. Shah, "Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study", Journal of Medical Internet Research, vol. 22, no. 4, p. e19016, 2020. Available: 10.2196/19016 [Accessed: 02 - Jun - 2020].

[15]I. Latha, G. Varma and A. Govardhan, "Preprocessing the Informal Text for efficient Sentiment Analysis", International Journal of Emerging Trends & Technology in Computer Science, vol. 1, no. 2, pp. 58-61, 2012. [Accessed: 02 - Jun - 2020].

[16] A. Signorini, A. M. Segre and P. M. Polgreen, "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic," *PLoS One,* vol. 6, *(5),* 2011. [Accessed: 01 - Jun - 2020]

[17] Go, R. Bhayani, L.Huang. "Twitter Sentiment ClassificationUsing Distant Supervision". Stanford University, Technical Paper,2009 [Accessed: 01 - Jun - 2020]

[18] Rabindra Lamsal, "Coronavirus (COVID-19) Tweets Dataset", IEEE Dataport, 2020. [Online]. Available: http://dx.doi.org/10.21227/781w-ef42. [Accessed: 01 - Jun - 2020]

[19] Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu, "NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets". National Research Council Canada, 2013 [Accessed: 01 - Jun - 2020]

[20] Pang, B.and Lee, L. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts". 42nd Meeting of the Association for Computational Linguistics[C] (ACL-04). 2004, 271-278. [Accessed: 01 - Jun - 2020]

[21] Ming Xiang, Julian Grove, Anastasia Giannakidou. "Explicit and implicit negation, negative polarity, and levels of semantic representation". Linguistics Department, University of Chicago, 2014. [Accessed: 01 - Jun - 2020]

[22] V. M. K. Peddinti and P. Chintalapoodi, "Domain adaptation in sentiment analysis of twitter," in Analyzing Microtext Workshop, AAAI, 2011. [Accessed: 01 - Jun - 2020]