

**Reddit vs Facebook: Comparing sentiment towards
tightening of COVID-19 restrictions in Singapore across
social media platforms**

Nicole Neo

Quantitative Methods in Social Science

Master of Arts Thesis, 2022

Columbia University

Table of Contents

Abstract	3
Introduction.....	4
Methodology	10
Results.....	24
Discussion	31
Conclusion	38
References.....	39
Appendices.....	42

Abstract

2021 was a turbulent year for Singapore. As COVID-19 infection numbers waxed and waned, the Singapore government responded with repeated tightening and loosening of restrictions. These disruptive yet protective measures generated heated discussions among Singaporeans. This study examines some of the online discourse to determine how sentiment towards the tightening of COVID-19 restrictions in Singapore varies across Reddit and Facebook. Using state-of-the-art lexicon-based sentiment analysis algorithms like VADER and TextBlob to analyze Reddit and Facebook comments, this study found a negligible difference in overall sentiment across these two platforms. Reddit and Facebook users have nuanced attitudes – they share both positive and negative sentiment about the tightening of restrictions. Compared to gold standard human expert scoring, VADER and TextBlob have a sentiment analysis accuracy level of 55% and 41% respectively. These findings are valuable given the limited literature about the sentiment towards COVID-19 restrictions in Singapore and the application of sentiment analysis algorithms to Singlish text.

Introduction

Background

2021 started on a good note for Singaporeans. With the beginning of a re-opening phase that loosened restrictions on group gatherings, and COVID-19 cases remaining at extremely low levels, it seemed as though the pandemic was coming to an end. However, that optimism was quickly shattered on May 4, 2021, when the government reimposed 'Phase 2' restrictions that limited gatherings to groups of 5 and mandated the closing of indoor gyms and fitness studios. What followed was a repetitive cycle of tightening and loosening of restrictions over the next 6 months as the government struggled to cope with new waves of infection caused by the Delta and Omicron variants. The frequently changing restrictions caused huge debates among Singaporeans, especially on social media platforms like Facebook and Reddit. Such discussion was often polarized, with some Singaporeans applauding the government's cautiousness amidst rising infection and hospitalization rates, while others lamented the continued cap on social activities despite high vaccination rates.

A survey by the Institute of Policy Studies conducted in 2021 revealed that sentiment towards COVID-19 restrictions in Singapore could be divided along age lines, as age had a statistically significant effect on Singaporeans' attitude towards the restrictions. Younger Singaporeans, whose mental and social wellbeing have suffered greater disruption during the pandemic compared to their older counterparts, were more willing to support the easing of restrictions. Older Singaporeans, who are more susceptible to infection and severe illness from COVID-19, were comparatively less keen on easing restrictions.¹ Given that the demographics of Singaporean users of Reddit and Facebook also differ by age (55.8% of Reddit users were

¹ Mathew Mathews et al., "Living with COVID-19 in Singapore: Attitudes, challenges and the way ahead" (Working Paper No. 43, Institute of Policy Studies, December 2021), 53-56, 67.

aged 18-34 compared to only 43.8% of Facebook users being in that age group as of December 2020)², it is plausible that the sentiment of Reddit and Facebook users towards COVID-19 restrictions will also differ.

This thesis therefore seeks to investigate if such sentiment differs across Reddit and Facebook. These platforms were chosen for their suitability to the research topic, as they both contain discussion-based features allowing for continued interaction through longer-form text compared to other platforms like Instagram (more visual media focused) and Twitter (shorter-form text interaction). This study will use state-of-the-art sentiment analysis algorithms to score the sentiment of Reddit and Facebook comments, and then determine if there is a difference in sentiment across the two platforms. The findings of this study will provide important insights into Singapore's COVID-19 policy and new learning points for sentiment analysis methods and the Natural Language Processing (NLP) field in general.

Previous work

There is currently no study that compares sentiment towards COVID-19 restrictions across Reddit and Facebook, but there have been similar studies of other COVID-19 related topics. While these studies may not directly address my research question, they nonetheless provide good understanding of existing methodology and of how sentiment differs across other platforms and towards other COVID-19 topics.

The only cross-platform study of sentiment towards COVID-19 related topics compares sentiment towards COVID-19 vaccines across Facebook and Twitter in the United Kingdom

² Simon Kemp, "Digital in Singapore: All the Statistics You Need in 2021", DataReportal, February 10, 2021, <https://datareportal.com/reports/digital-2021-singapore>.

and the United States. It found that Twitter sentiment was more negative than Facebook's, with the proportion of negative sentiment being almost twice that on Twitter compared to Facebook.³ The study also suggested that the difference in sentiment could reflect differences in user demographics across the two platforms,⁴ which gives us reason to expect that Reddit and Facebook sentiment towards COVID-19 restrictions could also differ as a result of varying user demographics. For sentiment analysis, this study used an ensemble model of lexicon-based methods VADER and TextBlob and a pre-trained Bidirectional Encoder Representations from Transformers (BERT) model. TextBlob had marginally better performance than VADER, and the lexicon-based methods had higher accuracy in detecting positive sentiment while BERT had higher accuracy in detecting neutral and negative sentiment.⁵

Other studies of sentiment towards COVID-19 related topics examined only a single social media platform, but involved relevant sentiment analysis techniques. One such study analyzed the sentiment of comments on COVID-19 related Facebook posts made by Public Health Authorities (PHAs) of Singapore, the UK, and the US. It found that while a majority of the comments on each page were negative, Singapore's PHA still had the most favorable sentiment in its page comments, followed by the US and the UK.⁶ A temporal analysis of the changes in sentiment over time also revealed that from mid-February to early March 2020, the comments became more positive and supportive of the PHAs' responses to the pandemic.⁷ This

³ Amir Hussain et al., "Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study," *Journal of Medical Internet Research* 23, no. 4 (2021): 5-6, accessed September 28 2021, <https://www.jmir.org/2021/4/e26627/>.

⁴ Hussain et al., "Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines", 6.

⁵ *Ibid.*, 2.

⁶ Aravind Sesagiri Raamkumar, Soon Guan Tan and Hwee Lin Wee, "Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison," *Journal of Medical Internet Research* 22, no. 5 (2020): 6, accessed September 27, 2021, <https://www.jmir.org/2020/5/e19334/PDF>.

⁷ Raamkumar, Tan and Wee, "Measuring the Outreach Efforts of Public Health," 6-8.

study hence reveals the value of conducting temporal analysis to generate additional insights from the data.

There is also some literature on the application of sentiment analysis methods to text containing Singaporean English (hereafter ‘Singlish’). Singlish is an English-based creole spoken by Singaporeans that contains elements of different languages such as the Hokkien, Teochew and Cantonese dialects spoken by the Chinese, Bahasa Melayu spoken by the Malays, and Tamil spoken by the Indians. Singlish, being an informal language, is not currently recognized by algorithms like VADER and TextBlob. This could be problematic as Singaporeans tend to use a lot of Singlish on Reddit and Facebook, and Singlish contains many sentiment-bearing expressions that would otherwise go undetected by the algorithms. For example, among the top Singlish tokens in my comment data are expressions like ‘heng’ (fortunately), ‘haiz’ (sigh), ‘ccb’ (a vulgarity) and ‘jialat’ (seriously bad). The only solution attempted for lexicon-based sentiment analysis thus far has been a manually constructed Singlish dictionary of 1024 terms from the Dictionary of Singlish and Singapore English⁸, the Oxford Singlish Dictionary⁹ and a Wikipedia list of Singlish vocabulary¹⁰. However, the researchers who used this method cautioned against it since Singlish is not a full language but just a variant of English, and there are many Singlish expressions that cannot be translated exactly or easily into English.¹¹

⁸ Jack Lee, “A Dictionary of Singlish and Singapore English,” accessed June 20, 2022, <http://www.mysmu.edu/faculty/jacklee/>.

⁹ Colin Goh and Y.Y. Woo, eds., *The Oxford Singlish Dictionary* (Singapore: Angsana Books, 2001).

¹⁰ Wikipedia contributors, “Singlish vocabulary,” *Wikipedia, The Free Encyclopedia*, accessed June 20, 2022, https://en.wikipedia.org/w/index.php?title=Singlish_vocabulary&oldid=1102023299.

¹¹ Siaw Ling Lo et al., “A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection,” *Knowledge-Based Systems* 105 (2016): 245, accessed June 23, 2022, <https://doi.org/10.1016/j.knosys.2016.04.024>.

Considering the gaps in the literature, this thesis can add value by being the first quantitative study that examines and compares sentiment across Reddit and Facebook about the tightening of COVID-19 restrictions in Singapore. The results of the study can provide a deeper understanding of how platforms with different user demographics may (or may not) present different sentiment towards the same topic. This study also serves as a test bed for NLP practices as it applies and evaluates the performance of state-of-the-art sentiment analysis algorithms on Singlish text and out-of-training domains. VADER was initially trained on tweets, opinion editorial articles, technological product reviews and movie reviews data,¹² but not on Reddit and Facebook data.

Scope of this thesis

For this study, the geographical scope was limited to Singapore. While it was challenging to ensure that the comment data exclusively discussed Singapore's COVID-19 restrictions, I tried to increase the relevance of the data to Singapore as much as possible by extracting comments from the r/singapore subreddit and Singapore-based Facebook pages.

Secondly, this study primarily examined sentiment towards the tightening of COVID-19 restrictions in Singapore, and not the loosening of restrictions. This was to ensure interpretability of the sentiment analysis results. By assuming that comments were responding to posts about the tightening of restrictions, positive-scoring comments could be interpreted as supporting tighter restrictions while negative-scoring comments could be interpreted as opposing them. Conversely, if we assumed comments to be about restrictions in general, it would become less clear whether a positive comment meant support for or opposition to the

¹² C.J. Hutto and Eric Gilbert, "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 no. 1 (2014): 222, accessed July 2, 2022, <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>.

restrictions. It was difficult to ensure that all comments were about the tightening of restrictions, but this was done on a best-efforts basis by only extracting comments from posts about the tightening of restrictions.

Thirdly, the temporal scope of this thesis was set as May to October 2021 since restrictions changed the most in this period. This period also saw the greatest debate about the restrictions since restrictions were extended for a long period despite improving hospitalization and vaccination rates. After November 2021, restrictions were gradually loosened and were only maintained but not re-tightened during the Omicron wave. A timeline of the restrictions from May to October 2021 can be found in Appendix A.

Lastly, this study only examined the quantitative differences in sentiment between Reddit and Facebook comments and did not investigate any underlying factors that might be responsible for this difference. While this study has cited other work discussing age as a possible determiner of differences in sentiment, this only served to explain the motivation behind the research question and was not meant to shape the interpretation of the results.

Methodology

Research design

The research question for this thesis is: How does sentiment towards the tightening of COVID-19 restrictions in Singapore vary across Reddit and Facebook? The hypothesis is that sentiment on Reddit and Facebook differ, with Reddit sentiment being more negative than Facebook sentiment. In this study, the target population is all Reddit and Facebook content (posts, comments, etc.) about the tightening of COVID-19 restrictions in Singapore. However, obtaining a representative random sample of this population is highly challenging. Social media platforms often restrict web scraping of data due to privacy issues, hindering the identification or development of full sampling frames needed to obtain a representative sample.¹³ For instance, Facebook only allows scraping of public data from groups and pages, and it also strictly limits the type and amount of data that can be scraped. As such, it would be difficult to obtain a probability-based sample of this population where every member of the population (i.e. every related post or comment on Reddit or Facebook) has an equal, non-null probability of being sampled. This study therefore uses convenience sampling, where the sample is obtained based on the availability and accessibility of data, and all members of the population hence have an unequal probability of being sampled.

With convenience sampling, this study is rendered descriptive rather than inferential; any insights cannot be generalized to a population larger than the sample.¹⁴ When random sampling cannot be achieved, there is no basis for using inferential tools like hypothesis testing, which requires random sampling to generate inductive inferences. Even if p-values were

¹³ Marcus E. Berzofsky et al., "Probability-Based Samples on Twitter: Methodology and Application," *Survey Practice* 11 no. 2 (2018), accessed August 3, 2022, <https://doi.org/10.29115/SP-2018-0033>.

¹⁴ Marcus E. Berzofsky et al., "Probability-Based Samples on Twitter: Methodology and Application."

obtained they would be uninterpretable.¹⁵ Therefore, no hypothesis testing was conducted in this study and only descriptive statistics were used to validate the hypothesis.

The primary quantity of interest in this study is the difference in the central tendency of sentiment polarity scores of Reddit and Facebook comments. Both mean and median measures were of interest as the distribution of the samples could be skewed or could contain outliers, in which case the median would be a more useful measure of comparison than the mean. The sentiment polarity scores are determined by sentiment analysis algorithms and range from -1 (extremely negative) to 1 (extremely positive).

The research process for this study is shown in Figure 1 below and comprises the following steps:

1. Data extraction: Using Application Processing Interfaces (APIs) to extract the required data from Reddit and Facebook
2. Obtaining word frequencies: Tokenizing the comment data and taking word frequency counts of the tokens to identify tokens requiring more pre-processing.
3. Text pre-processing: Treating abbreviations and Singlish phrases, then cleaning and tokenizing text to prepare the comment data for sentiment analysis.
4. Sentiment analysis: Applying the VADER and TextBlob sentiment analysis algorithms to score the sentiment polarity of each comment.
5. Exploratory data analysis: Generating and visualizing summary statistics of the polarity scores, then comparing sample mean/median scores for Reddit and Facebook and validating the hypothesis.

¹⁵ Norbert Hirschauer et al., “Can p-values be meaningfully interpreted without random sampling?” *Statistics Surveys* 14 (2020): 72, accessed August 3, 2022, <https://doi.org/10.1214/20-SS129>.

6. Temporal analysis: Plotting sentiment polarity scores over time to understand how Reddit and Facebook sentiment changes over time.
7. Model evaluation: Generating word clouds of VADER and TextBlob output and comparing VADER and TextBlob scores to gold standard human expert scores to evaluate algorithm performance.

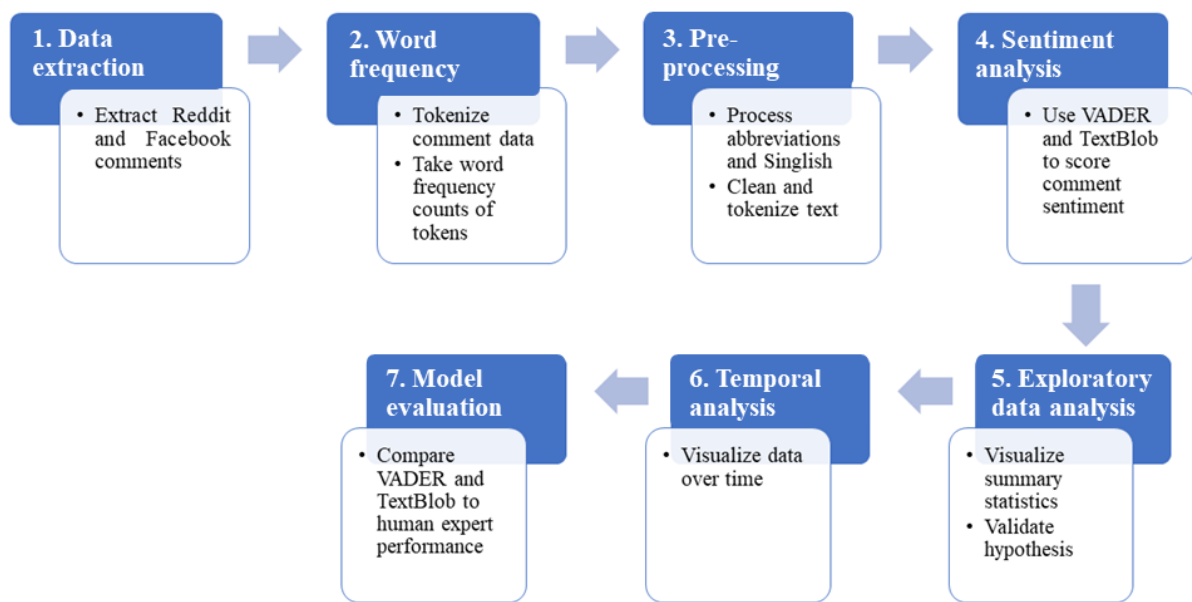


Figure 1: Flowchart of research process

Data extraction

To ensure better interpretability of results when comparing the descriptive statistics of the Reddit and Facebook samples, the samples should bear similar characteristics. This was achieved by sampling comments on Reddit and Facebook posts with identical content about the Singapore government's tightening of COVID-19 restrictions.

A source of content about Singapore's COVID-19 restrictions common to both platforms is online articles by Channel NewsAsia (CNA) and Straits Times (ST), two primary news channels of Singapore. When Singapore's Ministry of Health (MOH) announces new

COVID-19 restrictions, CNA and ST report on these announcements and post links to these articles on their respective official Facebook pages. Simultaneously, Redditors on the r/Singapore subreddit will repost the link to the same CNA and ST articles without adding any further commentary in the post. Hence, it would be adequate to compare the comments on such Reddit and Facebook posts as they contain the same content.

The Pushshift API¹⁶ was used to identify the relevant Reddit posts from which to scrape comments. These included posts from May to October 2021 that contained keywords like ‘covid’, ‘phase’, ‘measures’ and ‘suspended’, and which also contained links to CNA and ST articles. From the results, posts with articles that only announced the tightening (and not loosening) of COVID-19 restrictions in Singapore were selected. Articles about travel restrictions, border restrictions, vaccinations, and other topics beyond the scope of this thesis were excluded. Following this, the corresponding posts from the official CNA and ST Facebook pages were then manually identified. The final selection of 34 Reddit posts and 28 Facebook posts can be found in Appendix B. There are more Reddit posts than Facebook posts because some articles were posted only once on Facebook but multiple times on Reddit.

Comments from these posts were then scraped using the PushShift API and the facebook-scraper Python package¹⁷. The final dataset contains a total of 15 719 comments – 9112 from Reddit and 6607 from Facebook. These comments span the time period from May 4, 2021 when ‘Phase 2’ restrictions reducing group gathering sizes from 8 to 5 were introduced,

¹⁶ The Pushshift API provides capabilities to query an archive of all Reddit comments and submissions (posts). Full documentation can be found at <https://github.com/pushshift/api>.

¹⁷ facebook-scraper is a Python package that enables scraping of comments, reactions and other data from public Facebook pages without an API key. Full documentation can be found at <https://github.com/kevinzg/facebook-scraper>.

to October 20, 2021 when ‘Stabilization Phase’ restrictions limiting dine-in to groups of 2 were extended for another month.

Obtaining word frequencies

Obtaining word frequencies in a text is a valuable step that helps to identify key words in the text data, since the most frequently occurring words could contain the main idea or topic of the text. The top-occurring words could also be domain-specific words or filler words so frequently repeated in the text that they become less meaningful or informational – such words could then be classified as stop words to be used in later stages of analysis. Additionally, given the relevance of the comment data to Singapore and COVID-19 restrictions in Singapore, some words in the data could be policy-related abbreviations or Singlish words that need further explaining or translating before the data is ready for sentiment analysis. Taking word frequency counts helps to identify the abbreviations and phrases that appear most often in the data.

To ensure that only the most meaningful words are counted, the comment data needs to be adequately pre-processed. Using the regex library in Python,¹⁸ hyperlinks, digits and punctuation were removed from the data. The text data was then converted to lowercase to ensure that words with different capitalizations (e.g. ‘NO’ and ‘no’) would be counted as the same word. Stop words – common words in the English language that have little informational meaning (e.g. ‘can’, ‘be’, ‘I’) – were also removed using the Natural Language Toolkit (NLTK) package in Python.¹⁹

¹⁸ regex is a Python package that enables replacement of substrings of text within a textual body using pattern matching. More documentation can be found at <https://pypi.org/project/regex/>.

¹⁹ NLTK is a package that provides comprehensive functionalities for NLP in Python; it supports text classification, tokenization, stemming, parts-of-speech tagging, etc. Full documentation can be found at <https://www.nltk.org/api/nltk.html>.

After cleaning the text data, tokenization was performed using NLTK's word tokenizer function. Tokenization is the process of dividing a larger body of text into smaller parts called tokens. It goes beyond separating text by whitespace into words, even separating contractions (e.g. splitting 'it's' into 'it' and 's') and hashtags.²⁰ A total of 209 387 tokens were obtained from the text and the top 20 most frequently occurring tokens are listed in Figure 2 below.

Top 20 tokens (English)	
Token	Frequency
people	2230
covid	1503
like	1417
still	1249
cases	1188
vaccinated	1165
get	1159
go	1107
one	991
even	961
time	888
us	853
need	842
think	828
back	824
singapore	822
also	810
going	767
open	731
government	698

Figure 2: Top 20 most occurring tokens in the comment corpus and their frequency counts.

Within the list are some domain-specific key words like 'covid', 'cases', 'vaccinated', 'singapore' and 'government'. These key words show that the extracted comments do discuss the COVID-19 situation and the government's responses to the pandemic in Singapore, hence

²⁰ Daniel Jurafsky and James H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Third Edition draft)* (Harlow: Pearson Publishing, 2020), 2.

reflecting the relevance of the extracted data to the research question. The list also contains less meaningful words such as ‘us’ and ‘also’ that could be considered as additional stop words should we need to produce more meaningful output later.

Next, using NLTK’s English dictionary as a filter, ‘non-English’ tokens were identified. Tokens outside of NLTK’s English dictionary are likely to be Singlish expressions, Internet slang or abbreviations. Such ‘non-English’ tokens and abbreviations need to be replaced with English words, otherwise they may not be captured during sentiment analysis by the lexicon-based sentiment analyzers that are designed to recognize English words. The top 20 non-English tokens, their meaning, and their frequency counts are listed in Figure 3 below.

Top 20 tokens (non-English)		
Token	Meaning	Frequency
cb	Circuit Breaker (the name of Singapore’s lockdown measures in April-June 2020), or ‘Chee Bye’ (a Singlish vulgarity that means cunt and is equivalent in severity to the F word)	453
ktv	Karaoke	382
govt	Government	372
lol	Laugh out loud	297
ppl	People	238
gov	Government	223
etc	Et cetera	220
wfh	Work From Home	166
hbl	Home-Based Learning	155
others	Others	143
moh	Ministry of Health	138
pls	Please	138
mrt	Mass Rapid Transit (the name of Singapore’s subway system)	134
mtf	Ministry Task Force	112
ndp	National Day Parade	103
hari	Day in Malay, likely part of the phrase Hari Raya which is another name for the Eid festival	102
vax	Vaccine	100
mmtf	Multi-Ministry Task Force	99
oyk	Ong Ye Kung (Singapore’s Minister of Health)	99
reddit	Reddit	85

Figure 3: Top 20 most occurring non-English tokens in the comment data and their frequency counts.

Among the top 20 non-English tokens are a mix of Singlish words, Internet slang and abbreviations for various government policies and institutions in Singapore. These tokens have a rather high frequency count and contain meaningful domain-specific information or sentiment, so it is important to replace them with an equivalent word or phrase in English so that the sentiment analyzers can capture them. A dictionary was constructed for this purpose on a best-efforts basis. Tokens were only replaced if they could easily be replaced with an English word or short phrase, and tokens with a frequency count of less than 10 were not replaced. Some exceptions were made for the following reasons:

- The phrase ‘Hari Raya’ refers specifically to the Islamic holiday Eid al-Fitr that does not have an equivalent name in English. The phrase was therefore not translated.
- The abbreviation ‘ICU’ stands for Intensive Care Unit, and the word care contains positive sentiment that would be captured by sentiment analyzers, even if ‘ICU’ does not bear positive sentiment. To prevent confusion, ‘ICU’ was not expanded.
- ‘CB’ was used in the comment data both as a reference to the lockdown measures and as a vulgarity bearing highly negative sentiment. As it would be impossible to systematically determine which meaning this token had in each comment, it was removed entirely from the data altogether to prevent confusion.
- Common Internet slang expressions like ‘lol’, ‘wtf’, and ‘lmao’ were not replaced as one of the sentiment analyzers used in this study, VADER, is able to recognize and capture sentiment in these expressions.

Text pre-processing

Unlike other common NLP tasks, text pre-processing is not necessary for sentiment analysis. Many conventional pre-processing steps such as the removal of digits, punctuation, and emoticons, the removal of stop words, and the stemming or lemmatization of words, could

potentially hamper the accuracy of sentiment analysis and hence should not always be adopted.²¹ This is because punctuation (especially exclamation and question marks), emoticons (e.g. :D or :-O), capitalization (e.g. ‘GREAT’ vs ‘great’), and stop words (which include degree modifiers or negation words like ‘not’, ‘very’ and ‘but’) all affect the polarity and intensity of sentiment in the text, so removing these features will erode these nuances and change the sentiment. Lemmatization also ignores sentiment intensity in words with the same base roots (e.g. ‘good’ and ‘better’).²² The same sentence can receive drastically different scores with and without pre-processing, so pre-processing should be applied only if absolutely necessary.²³

Sentiment analyzers are also designed to handle some of these text features, hence reducing the need for pre-processing. Both TextBlob and VADER can process negation words like ‘but’ and ‘not’,²⁴ while VADER accounts for punctuation, capitalization, degree modifiers, Western-style emoticons, sentiment-related acronyms (e.g. ‘lol’, ‘wtf’), and common sentiment-bearing Internet slang (e.g. ‘nah’ and ‘meh’).²⁵ Therefore, to preserve as many sentiment-bearing features as possible, the comment data was minimally pre-processed. Hyperlinks, hashtags and line breaks were removed to increase machine readability, while Singlish text and abbreviations were replaced using the manually constructed dictionary. No punctuation, digits or stop words were removed, no lemmatization or stemming was applied, and capitalization was also preserved.

²¹ Min Zhou, “StopWords and Lexicon Normalization in Sentiment Analysis,” July 22 2019, accessed July 7, 2022, <https://medium.com/data-science-blogs/stopwords-and-lexicon-normalization-for-sentiment-analysis-f9f10f0d4108>.

²² Min Zhou, “StopWords and Lexicon Normalization in Sentiment Analysis,” July 22 2019.

²³ Holly Emblem, “Are you Scared, VADER? Understanding how NLP Pre-Processing Impacts VADER Scoring,” July 18 2021, accessed July 7, 2022, <https://towardsdatascience.com/are-you-scared-vader-understanding-how-nlp-pre-processing-impacts-vader-scoring-4f4edadbc91d>.

²⁴ Min Zhou, “StopWords and Lexicon Normalization in Sentiment Analysis,” July 22 2019.

²⁵ C.J. Hutto and Eric Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” 221.

Sentiment analysis

After some minimal pre-processing, sentiment analysis was conducted on the comment data. There are two types of sentiment analysis – supervised and unsupervised. Supervised sentiment analysis involves training machine learning algorithms on data that is pre-annotated with sentiment labels (e.g. ‘positive’, ‘negative’ or ‘neutral’), so that the algorithm can classify the sentiment of new text data. Whereas unsupervised sentiment analysis uses sentiment lexicons (lists of words and phrases that are positive or negative)²⁶ to obtain the sentiment polarity of individual parts of the text, then aggregates the scores into an overall score that determines the entire text’s polarity.²⁷ Unlike supervised sentiment analysis, unsupervised sentiment analysis does not require pre-labelled data. As the comment data in this study has not been manually annotated with sentiment labels, unsupervised lexicon-based sentiment analysis would be more suitable.

Among the many lexicon-based sentiment analysis tools used in NLP studies, SentiWordNet, VADER and TextBlob are more commonly used. SentiWordNet comprises 147 306 synsets, each of which has a positive, negative or neutral score on a scale of 0 to 1. However, there are key limitations to SentiWordNet. Most of the synsets also do not have positive or negative polarity, greatly reducing the utility of SentiWordNet in producing output with meaningful sentiment scores. Moreover, SentiWordNet has not been designed to account for sentiment-bearing features in the context of social media text.²⁸ The limited utility of SentiWordNet is reflected in its poorer performance relative to other sentiment analysis

²⁶ Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan, “A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis,” *Asian Journal of Computer Science* 8 no. S2 (2019): 2, accessed July 2, 2022, <https://doi.org/10.51983/ajcst-2019.8.S2.2037>.

²⁷ Alexander Hogenboom et al., “Exploiting emoticons in sentiment analysis,” (paper presented at the 28th Annual Association for Computing Machinery Symposium on Applied Computing, March 2013), <https://doi.org/10.1145/2480362.2480498>.

²⁸ C.J. Hutto and Eric Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” 218.

algorithms. In a study that compared the performance of NLTK (which contains SentiWordNet resources), TextBlob, and VADER on Rotten Tomatoes movie review data, NLTK had the lowest F1 score of 57%, while TextBlob and VADER had much higher scores of 79% and 82% respectively.²⁹ With this in mind, I chose to only use VADER and TextBlob for this study.

VADER and TextBlob have similar scoring scales. TextBlob returns scores for both subjectivity (i.e. whether something is an opinion or a fact) and polarity (i.e. positivity/negativity and intensity of sentiment),³⁰ but for the purposes of this study only the polarity score will be used and the subjectivity score will be ignored. TextBlob's polarity score ranges from -1 to 1, where -1 indicates extreme negativity and 1 indicates extreme positivity.³¹ VADER returns a normalized, weighted compound score; it is a sum of valence scores of each sentiment-bearing feature in the text as given by the lexicon, with that sum normalized to also be between -1 and 1.³²

However, VADER and TextBlob differ in the types of sentiment-bearing features they consider in scoring. According to TextBlob documentation, TextBlob primarily uses the WordNet lexicon and does not consider the valence of non-textual features like emoticons or punctuation.³³ Contrastingly, VADER incorporates the sentiment of punctuation, capitalization, Western-style emoticons, acronyms (e.g. 'lol', 'wtf'), and common Internet slang (e.g. 'nah' and 'meh').³⁴ These characteristics make VADER particularly accustomed to sentiment in

²⁹ Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis," 5.

³⁰ Siaw Ling Lo et al., "A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection," 236.

³¹ Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis," 4.

³² *Ibid.*

³³ Full documentation for TextBlob can be found at <https://github.com/sloria/TextBlob> and <https://textblob.readthedocs.io/en/dev/>.

³⁴ C.J. Hutto and Eric Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," 221.

online microblog-type contexts,³⁵ and possibly explains its superior performance to TextBlob in the aforementioned study by Bonta et al..³⁶ Despite the differences in TextBlob and VADER, both algorithms will be used for this study. These two are currently the strongest performing sentiment analysis algorithms in the field, and there has been limited literature promoting the use of one over the other.

Exploratory data analysis

After the comments were scored using VADER and TextBlob, exploratory data analysis was conducted to determine if there was a difference in mean and median scores between Reddit and Facebook comments.

Summary statistics (mean, median, maximum, median values, etc.) were obtained using the `describe()` function in Python's pandas package. Boxplots and histograms were also generated from the data for easy visual comparison of these statistics. The visualizations show the range, distribution and shape of the data, providing greater insight beyond just the differences in central tendencies across the samples.

Temporal analysis

Temporal analysis was also conducted to gain a more granular understanding of the data. For this, the comment data was grouped by date and platform and the mean VADER or TextBlob score was computed for each group. This data was then used to generate line plots of Reddit and Facebook comment sentiment scores over time, with sentiment score on the y-

³⁵ *Ibid.*, 225.

³⁶ Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan, "A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis," 5.

axis, date on the x-axis, and a different colored line for Reddit and Facebook. A total of two plots – one for VADER scores and one for TextBlob scores – were generated.

Model evaluation

To assess the accuracy of the reported results, it is important to evaluate VADER and TextBlob's accuracy at detecting sentiment polarity. A common method of evaluation is to compare the output of both algorithms using word clouds. Word clouds display words according to their frequency in the text, with higher-frequency words appearing bigger and bolder in the word cloud. By generating word clouds of the positively and negatively scored text from each algorithm, and observing whether the word clouds rightfully contain positively and negatively associated words, we can roughly observe the accuracy of the algorithm.

To generate the word clouds, the comment data was first split into four sets – comments with positive VADER scores, negative VADER scores, positive TextBlob scores, and negative TextBlob scores. Comments with a strictly neutral score of 0 were not used for this purpose. To ensure that the word clouds contained meaningful words, the four sets of data underwent some pre-processing. Digits, punctuation and English stop words from NLTK were removed from the text. The text was also converted to lowercase so that the same words with different capitalization would not be differentiated. Common domain-specific phrases (e.g. 'covid', 'lockdown', 'vaccine', 'ministry of health') and other high frequency tokens with no sentiment were removed since they would appear very frequently in the text, but would not be helpful in assessing the accuracy of the sentiment analysis algorithms.

As word clouds are a rudimentary evaluation tool that only looks at word frequencies, I will also compare the algorithm-scored output against output from my manual scoring for a

more comprehensive evaluation of accuracy. As a native English speaker, I have the linguistic ability to accurately determine the overall sentiment expressed in complex, multi-sentence comments; as a native Singlish speaker, I would also be able to accurately detect sentiment expressed in Singlish. For this task, my scoring therefore provides a suitable gold standard against which to compare algorithm output.

As human scoring requires significantly more time than machine scoring, I will only manually score a sample of 100 randomly selected comments from the data. To ensure consistency with the algorithm scoring, I will maintain the same scoring scale of -1 to 1. Since VADER and TextBlob score on a continuous scale, I aim to do so as closely as possible by scoring on four degrees of negativity (Slightly Negative = -0.25, Moderately Negative = -0.5, Very Negative = -0.75, Extremely Negative = -1.0) and four degrees of positivity (Slightly Positive = 0.25, Moderately Positive = 0.5, Very Positive = 0.75, Extremely Positive = 1.0). Comments with mixed or indifferent sentiment will be scored as 0. I will score each comment twice in two different time periods and return the average of both scores. A selection of the manually scored comments can be found in Appendix C.

Results

Results of exploratory data analysis

The answer to the research question of how sentiment towards tightening of COVID-19 restrictions varies across Reddit and Facebook can be found by examining the distribution of the polarity scores of Reddit and Facebook comments. Figure 4 displays summary statistics of the VADER scores, while Figure 5 displays summary statistics of the TextBlob scores.

Looking at Figure 4, the mean VADER polarity score is positive (> 0) for both platforms, but the mean Reddit score (0.0344) is marginally less positive than that of Facebook (0.0466). Standard deviations are similarly high for both samples (0.4711 for Reddit and 0.4443 for Facebook), indicating a great dispersion of data points with respect to the mean. The median score for both platforms is 0, a neutral score.

Figure 5 also shows the mean TextBlob score for Reddit (0.0475) being slightly less positive than Facebook's (0.0589). Standard deviations of TextBlob scores (0.2465 for Reddit and 0.2478 for Facebook) are similar across both platforms, but smaller than that of the VADER scores. Here the median score for both platforms is also zero.

<u>VADER scores</u>		
	Reddit (n = 9112)	Facebook (n = 6607)
Mean	0.0344	0.0466
Standard Deviation	0.4711	0.4443
Median	0.0000	0.0000
Min	-0.9959	-0.9968
Max	0.9988	0.9971

Figure 4: Summary statistics of VADER polarity scores of Reddit and Facebook comment subsamples. Scores are shown to four significant figures.

	<u>TextBlob scores</u>	
	Reddit (n = 9112)	Facebook (n = 6607)
Mean (SD)	0.0475	0.0589
Standard Deviation	0.2465	0.2478
Median	0.0000	0.0000
Min	-1.0000	-1.0000
Max	1.0000	1.0000

Figure 5: Summary statistics of TextBlob polarity scores of Reddit and Facebook comment subsamples. Scores are shown to four significant figures.

The distribution of these samples can also be seen from visualizations like boxplots and histograms. Figure 6 depicts side-by-side boxplots of Reddit and Facebook polarity scores. In a boxplot, the box marks out the interquartile range (IQR) of values between the first quartile and third quartile (25th to 75th percentiles); the bottom of the box indicates the first quartile (Q1) value, the solid black line in the box indicates the median (Q2) value, and the top of the box indicates the third quartile (Q3) value. The dotted red line shows the mean value. The whiskers indicate the values within 1.5 times of the IQR from the lower and upper quartiles (e.g. $Q1 - 1.5 \times IQR$ for the lower whisker and $Q3 + 1.5 \times IQR$ for the upper whisker); any values outside the whiskers are deemed as outliers.³⁷ Figure 6 shows that for both VADER and TextBlob scores, the mean Reddit comment score is barely lower than the mean Facebook score. The IQR of Reddit scores is slightly larger than that of Facebook's, indicating a greater spread in the data. Generally, VADER scores also have greater spread than TextBlob scores – the VADER score boxplots have much longer whiskers and no outlier points compared to TextBlob boxplots.

³⁷ Michael Galarnyk, "Understanding Boxplots," September 12, 2018, accessed August 21, 2022, <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>.

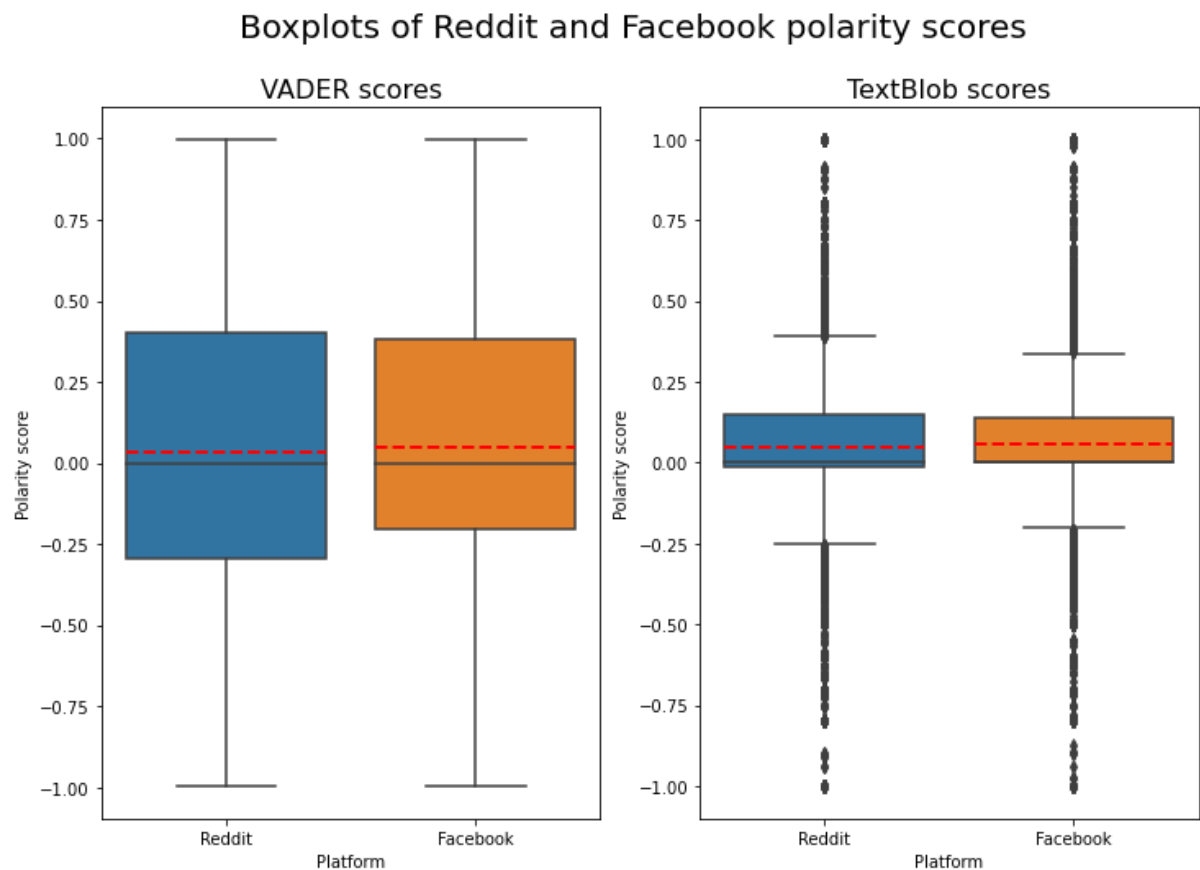


Figure 6: Boxplots of VADER and TextBlob polarity scores of Reddit and Facebook comment samples

Figure 7 contains four histograms showing the shape of the four subsamples. The first row shows the distribution of VADER scores for the Reddit and Facebook subsamples, while the second row shows the distribution of TextBlob scores. All subsamples have a fairly symmetrical distribution. This is notable because when there is no skew in the data, the mean is a more useful indicator of central tendency compared to the median, which is at 0 for all subsamples. The histograms show the mode to be 0 across all subsamples, as seen from the zero-inflation with a huge proportion of values concentrated at 0. The histograms also show a larger spread of VADER scores compared to TextBlob scores, echoing what was shown in Figure 6. This reflects that VADER scores are more varied than TextBlob scores, so VADER could be a more nuanced and sophisticated sentiment analysis model than TextBlob.

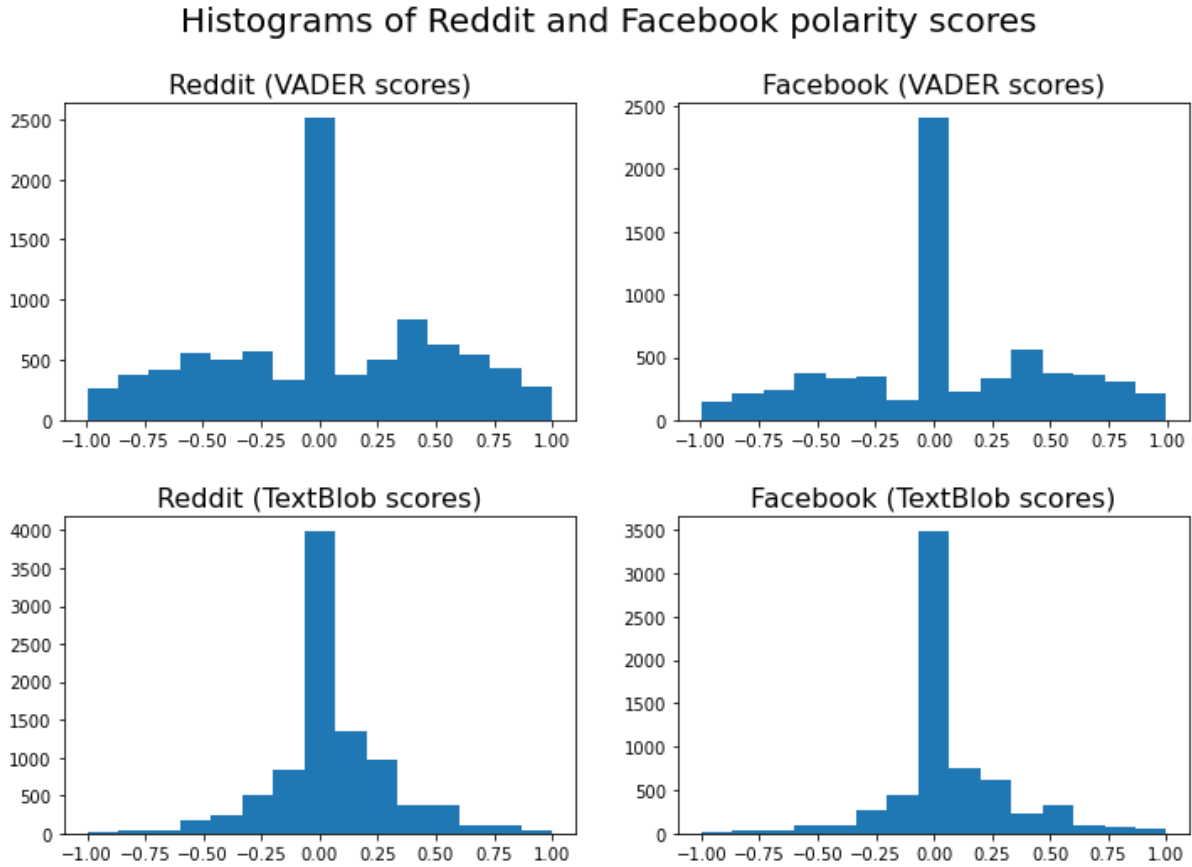


Figure 7: Histograms of VADER and TextBlob sentiment polarity scores of Reddit and Facebook comment samples

From the summary tables and visualizations, we have sufficient information to determine the validity of the hypothesis that sentiment on Reddit and Facebook differ, with Reddit sentiment being more negative than Facebook sentiment. Based on the mean score, which is the more useful indicator of central tendency here compared to the median, we can conclude that Reddit and Facebook comments are not very different in sentiment. The mean Reddit sentiment score is only about 0.01 less than the mean Facebook sentiment score, and on a sentiment polarity scale of -1 to 1, this difference is negligible. Having said that, this conclusion applies only to this sample of data which was obtained conveniently and not through random sampling, so the findings above are merely descriptive and not inferential. They cannot be generalized to the wider population of all Reddit and Facebook content about COVID-19 restrictions in Singapore without further research and supporting evidence.

Results of temporal analysis

Figure 8 shows how the mean polarity scores of Reddit and Facebook comments change over time from May to October 2021. The blue line represents Reddit scores while the orange line represents Facebook scores.



Figure 8: Line plots of VADER and TextBlob sentiment polarity scores of Reddit and Facebook comment samples across time from May to October 2021

While the line plots of VADER scores have greater amplitude than those of TextBlob scores, the general trend is that both Reddit and Facebook scores vary across time and no one platform has a consistently higher or lower score than the other. Facebook and Reddit lines frequently crossed each other – in the second half of May 2021, Facebook sentiment was slightly more positive than Reddit sentiment, but in September 2021, Reddit sentiment was often more positive than Facebook sentiment.

There were 3 time points with bigger differences in the Reddit and Facebook polarity scores – May 18, September 12, and September 30. The comments from these dates were closely inspected to better understand the differences (see Appendix D for a selection of these comments). My observations were as follows:

- **May 18 (mean Reddit score = 0.0153, mean Facebook score = 0.1677):**

Comments were responding to articles announcing the suspension of in-person private tuition and enrichment classes for students aged 18 and below. Some Reddit comments were concerned about the negative impact on students' academic support and on tutors' incomes, while several Facebook comments supported the move for protecting students and called on the government to extend this measure to other settings like institutions of higher learning.

- **September 12 (mean Reddit score = 0.2895, mean Facebook score = 0.0318):**

Comments were responding to the suspension of in-person visits to residential care homes amid the rise in COVID-19 cases. Redditors were rather supportive of the measures as they recognized the need to protect the vulnerable. Facebook users expressed concern about the negative impact on the mental health of elderly in the care homes. Many Facebook comments also complained about the last-minute

nature of the announcement and the government's lack of transparency about the extent of infection in the care homes.

- **September 30 (mean Reddit score = -0.0107, mean Facebook score = -0.1624):**

Comments were responding to the move to home-based learning in universities amid rising COVID-19 cases. While the mean Reddit score was negative, it was less negative than that of Facebook's possibly due to a few positive comments on Reddit expressing an understanding of the need for these measures and the ability to adapt to the measures. Facebook comments mostly expressed disappointment about the imposition of restrictions despite rising vaccination rates; many also cast doubt on the effectiveness of vaccination in reducing transmission.

These comments show that Reddit and Facebook users are not monolithic in their attitudes towards COVID-19 restrictions in Singapore. Given that Reddit users tend to be younger than Facebook users,³⁸ we would expect Redditors to generally be less supportive of all restrictions compared to their Facebook counterparts. However, according to the comments above, Redditors were less supportive than Facebook users of measures impacting younger students, but more supportive about measures concerning vulnerable elderly in care homes. These nuanced, varied attitudes of Reddit and Facebook users could explain why on average there is barely a difference in sentiment expressed across both platforms, unlike what was initially hypothesized.

³⁸ Simon Kemp, "Digital in Singapore: All the Statistics You Need in 2021", DataReportal.

From the word clouds, VADER and TextBlob seem to generally capture positive and negative sentiments accurately. Most of the words in the positive score word clouds have positive sentiment, and vice versa with the negative word clouds. There are however some exceptions. For instance, VADER and TextBlob pick up on 'risk' for positively scored comments, while VADER picks up on 'good' and TextBlob picks up on 'hope' even for negatively scored comments.

Nonetheless, it is premature to conclude from the word clouds alone how accurate VADER and TextBlob are and which model of the two is more accurate, since word clouds are only a rudimentary tool that show the highest frequency words in each data without additional contextual information about the use of these words in the comments. All we can conclude from the word clouds is that VADER and TextBlob are both able to detect positive and negative sentiment-bearing words quite accurately. We need a more detailed inspection of the scoring output to fully compare and evaluate the performance of these two algorithms.

Comparing VADER and TextBlob scoring against gold standard human scoring provides a more comprehensive evaluation of the accuracy of these algorithms. There are two aspects to a polarity score – the orientation of the score (whether it is positive, negative or neutral) and the magnitude of the score (how positive or negative the score is). As orientation is the more important measure of accuracy (it does not matter how close the score magnitudes are if the orientation is already inaccurate), I compared how accurately the orientation of the VADER and TextBlob scores matched that of the human scores. As human scoring was only done on a subsample of 100 comments, we can only compute algorithm accuracy using these 100 comments.

Figure 10 compares the algorithm scores to gold standard human scores for five of the manually scored comments (see Appendix C for more manually scored comments). In this table, the ‘score_vader’ column is the VADER score, ‘score_tb’ is the TextBlob score, ‘score_human_avg’ is the human score (average score from two scoring sessions) and ‘orient_match_vader/tb’ is an indicator column with value = 1 for a match and value = 0 for a mismatch between the orientations of the human score and the algorithm score.³⁹

	text	score_vader	score_tb	score_human_avg	orient_match_vader	orient_match_tb
1	Amazing 4G leadership! Please call for an elections again this year because I am very sure a lot of people will wanna vote for you all and give you the well deserving mandate! So proud to have smart leaders like you all!!!	0.9645	0.6737	-0.3750	0	0
2	I'll give our stupid government full credit instead. Whatever happens in other countries is out of our control, but what's within our control is our response to what happens in other countries and our government had to fuck it up, all in the name of eCoNoMy	-0.7311	-0.2200	-0.5000	1	1
3	Go back to phase 2, but more flights are arriving! Song boh?	0.0000	0.3125	-0.5000	0	0
4	time for the "Containing Crowd Battle" to start, sayonara F&B businesses	-0.3818	0.0000	-0.5000	1	0
5	>own up to your mistakes People's Action Party? Admit that they are wrong? HAHA	0.4576	-0.0667	-0.5000	0	1

Figure 10: A table comparing the orientations of VADER and TextBlob scores against that of human scores for five comments. Scores are shown to four significant figures.

Summing the indicator values across all 100 comments, we obtain the total number of matches and mismatches across the manually scored subsample as seen in Figure 11. Dividing the number of matches over the subsample size of 100, we obtain an accuracy score of 55%

³⁹ Note: A match occurs when the human score and model score have the same orientation. If both scores are positive (polarity score > 0) or both are negative (polarity < 0) or both are neutral (polarity score = 0), there is a match, otherwise there is a mismatch.

for VADER and 41% for TextBlob. We can hence conclude that based on this subsample, VADER is a more accurate sentiment analyzer than TextBlob, but both do not have high accuracy relative to gold-standard human scoring. To assess the accuracy of VADER and TextBlob on the entire comment data, more time and resources would be needed to manually score the rest of the comments.

n = 100		Matches	Mismatches	Accuracy
VADER		55	45	55%
TextBlob		41	59	41%

Figure 11: Table showing the number of orientation matches between VADER/TextBlob and human scores and the accuracy of VADER and TextBlob.

Limitations

Despite best efforts to improve the validity and accuracy of this study by (i) extracting relevant data, (ii) using state-of-the-art sentiment analyzers, and (iii) processing the text more adequately for sentiment analysis, there are still significant limitations to this study that should be highlighted.

The first limitation of this study arises from the limited accuracy of the sentiment analysis algorithms which are still inadequate in many aspects compared to a human analyzer. There are several reasons why VADER and TextBlob were highly inaccurate in analyzing the comments in the data, unlike their high accuracies touted in other studies. Here, the algorithms were often unable to detect sarcasm that was otherwise easily detected by the human scorer using contextual understanding. For the first comment in Figure 12 below, VADER and TextBlob mistakenly scored the comment as positive (0.9645 and 0.6737 respectively), but the human scoring was negative (-0.375). While this comment contained several words with

positive sentiment like ‘amazing’, ‘proud’, and ‘deserving’, only the human scorer was able to detect the sarcastic use of these words given the contextual knowledge Singaporean citizens usually express overwhelming support for the Singaporean government in a sarcastic manner. More comments where VADER and TextBlob were unable to detect sarcasm can be found in Appendices C and D.

Additionally, many comments were long and complex – they spanned several sentences and often contained a mix of sentiment within the comment. Linguistic knowledge is required to accurately determine the overall dominant sentiment of such comments, and as this was something the human scorer possessed while VADER and TextBlob did not, the human scorer outperformed the algorithms. We can see this from the second comment in Figure 12. This comment was scored negatively by VADER (-0.9387) and TextBlob (-0.0955), but was manually scored as slightly positive (0.2500). The human scorer had the linguistic knowledge to recognize the user’s main message in the comment, which was that the elderly should consider getting vaccinated as the risk of the vaccine is much lower than expected even for elderly with chronic illnesses. Since the comment was meant to be encouraging, the overall sentiment polarity of the comment was scored positively. Whereas VADER and TextBlob, likely seizing on the negative sentiment-bearing words like ‘risk’, ‘allergic’, ‘painful’, and ‘worse’, scored the comment as a negative one.

Moreover, despite the efforts to translate Singlish phrases into English, some phrases remained untranslated so VADER and TextBlob were unable to accurately determine the sentiment of comments containing such phrases. For example, the third comment in Figure 12 contains the Hokkien dialect phrase ‘wa meng ti’, which means ‘I ask god’ when literally translated to English, and is a phrase used by Singaporeans to express exasperation at a

situation. Equipped with an understanding of Singlish, the human scorer was able to score this as moderately negative (-0.5000), while VADER and TextBlob both scored it as neutral since this phrase was absent from their sentiment lexicons.

	text	score_ vader	score_ tb	score_ human _avg
1	Amazing 4G leadership! Please call for an elections again this year because I am very sure a lot of people will wanna vote for you all and give you the well deserving mandate! So proud to have smart leaders like you all!!!	0.9645	0.6737	-0.3750
2	Have you asked her to consider re-assessing the risk? My mom is 62, diabetic, on dialysis and is allergic to several types of meds, and recently got the second dose with few complaints. We will eventually open up so we told her that it is for her own safety that she lowers the risk of getting the disease. Sneezing is painful for my mom because of her peritoneal dialysis residual pain every day, so we noted that COVID will make her feel worse than any vaccine in all likelihood. Nothing bad happened in the end. I understand that it can be riskier for some people and this is just my experience. We had also delayed the first vaccination till early June this year.	-0.9387	-0.0955	0.2500
3	Wa meng ti wa meng ti	0.0000	0.0000	-0.5000

Figure 12: A selection of comments where VADER and TextBlob were unable to accurately detect sentiment polarity. Scores are shown to four significant figures.

The second limitation of this study lies in the relevance of data used. The study sought to investigate sentiment towards the tightening of COVID-19 restrictions in Singapore on Reddit and Facebook, but it was difficult to ensure that the extracted comments pertained exactly to that. As seen from Appendices C and D, several comments still discussed other restrictions (e.g. border control measures), vaccination policies, and the pandemic situation abroad. Some comments were also responding to other comments in the thread instead of the article content. Therefore, the sentiment in the extracted comment data may not fully represent Reddit and Facebook sentiment towards the tightening of COVID-19 restrictions in Singapore.

The third limitation is due to sampling difficulties. As mentioned in the Methodology section, the target population of this study is all online content on Reddit and Facebook about the tightening of COVID-19 restrictions in Singapore. However, due to data extraction challenges I could only obtain a convenience sample instead of a random sample of this population. The inability to conduct random sampling meant that inferential methods like hypothesis testing could not be conducted. The findings of this study only apply to the sample data used and cannot be generalized to the population, hence limiting the utility of this study.

Conclusion

Although Reddit and Facebook users in Singapore are demographically different, this difference does not manifest in their sentiment towards COVID-19 restrictions in Singapore. The sentiment on both platforms fluctuated considerably as the restrictions changed from May to October 2021, but for the whole period the overall mean sentiment polarity score was negligibly different between the two platforms. This conclusion was reached after conducting sentiment analysis with VADER and TextBlob, two state-of-the-art lexicon-based algorithms. This study saw new applications of these algorithms to COVID-19 related Reddit and Facebook comments containing Singlish expressions, and while NLP methods were used to help the algorithms perform better on such data, the algorithms still did not perform very well. Notwithstanding its limitations, this study provides valuable contributions to the literature on NLP, social media sentiment analysis, and COVID-19 policy in Singapore. This is one of a few studies that apply NLP methods to Singlish text, compare sentiment across Reddit and Facebook, and examine online sentiment towards the tightening of COVID-19 restrictions in Singapore. Future extensions of this thesis which improve on the methodology for Singlish sentiment analysis, and which conduct this analysis on a more representative sample, would provide highly utile findings on this pertinent topic.

References

Berzofsky, Marcus E., Tasseli McKay, Y. Patrick Hsieh and Amanda Smith. “Probability-Based Samples on Twitter: Methodology and Application.” *Survey Practice* 11 no. 2 (2018).

Accessed August 3, 2022. <https://doi.org/10.29115/SP-2018-0033>.

Bonta, Venkateswarlu, Nandhini Kumaresh and N. Janardhan. “A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis.” *Asian Journal of Computer Science* 8 no. S2 (2019): 1-6. Accessed July 2, 2022. <https://doi.org/10.51983/ajcst-2019.8.S2.2037>.

Emblem, Holly. “Are you Scared, VADER? Understanding how NLP Pre-Processing Impacts VADER Scoring.” July 18, 2021. Accessed July 7, 2022.

<https://towardsdatascience.com/are-you-scared-vader-understanding-how-nlp-pre-processing-impacts-vader-scoring-4f4edadbc91d>.

Galarnyk, Michael. “Understanding Boxplots.” September 12, 2018. Accessed August 21, 2022. <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>.

Goh, Colin and Y.Y. Woo, eds. *The Coford Singlish Dictionary*. Singapore: Angsana Books, 2001.

Hirschauer, Norbert, Sven Grüner, Oliver Mußhof, Claudia Becker and Antje Jantsch. “Can p-values be meaningfully interpreted without random sampling?” *Statistics Surveys* 14 (2020): 71-91. Accessed August 3, 2022. <https://doi.org/10.1214/20-SS129>.

Hogenboom, Alexander, Daniella Bal, Flavius Fransincar, Malissa Bal, Franciska de Jong, Uzey Kaymak. "Exploiting emoticons in sentiment analysis." Paper presented at the 28th Annual Association for Computing Machinery Symposium on Applied Computing, March 2013. <https://doi.org/10.1145/2480362.2480498>.

Hussain, Amir, Ahsen Tahir, Zain Hussain, Zakariya Sheik, Mandar Gogate, Kia Dashtipour, Azhar Ali and Aziz Sheikh. "Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study." *Journal of Medical Internet Research* 23, no. 4 (2021): 1-10. Accessed September 23, 2021. <https://www.jmir.org/2021/4/e26627/>.

Hutto, C.J. and Eric Gilbert. "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 no. 1 (2014): 216-225. Accessed July 2, 2022. <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>.

Jurafsky, Daniel and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Third Edition draft)*. Harlow: Pearson Publishing, 2020.

Kemp, Simon. "Digital in Singapore: All the Statistics You Need in 2021." DataReportal. February 10, 2021. <https://datareportal.com/reports/digital-2021-singapore>.

Lee, Jack. "A Dictionary of Singlish and Singapore English." Accessed June 20, 2022. <http://www.mysmu.edu/faculty/jacklee/>.

Lo, Siaw Ling, Erick Cambria, Raymond Chiong, David Cornforth. "A multilingual semi-supervised approach in deriving Singlish sentic patterns for polarity detection." *Knowledge-Based Systems* 105 (2016): 236-247. Accessed June 23, 2022.

<https://doi.org/10.1016/j.knosys.2016.04.024>.

Mathews, Mathew, Syafiq Suhaini, Mark Hou and Fiona Phoa. "Living with COVID-19 in Singapore: Attitudes, challenges and the way ahead." Working Paper No. 43, Institute of Policy Studies, 2021.

Raamkumar, Aravind Sesagiri, Soon Guan Tan and Hwee Lin Wee. "Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison." *Journal of Medical Internet Research* 22, no. 5 (2020): 1-12. Accessed September 27, 2021.

<https://www.jmir.org/2020/5/e19334/PDF>.

Wikipedia contributors. "Singlish vocabulary." *Wikipedia, The Free Encyclopedia*. Accessed June 20,

2022. https://en.wikipedia.org/w/index.php?title=Singlish_vocabulary&oldid=1102023299.

Zhou, Min. "StopWords and Lexicon Normalization in Sentiment Analysis." July 22 2019.

Accessed July 7, 2022. <https://medium.com/data-science-blogs/stopwords-and-lexicon-normalization-for-sentiment-analysis-f9f10f0d4108>.

Appendices

Appendix A – Timeline of COVID-19 restrictions, May – October 2021

Date announced (DD/MM/YY)	Restriction
4/5/2021	Phase 2 restrictions from 8 to 30 May <ul style="list-style-type: none"> Group gathering limits down from 8 to 5 people No more than 50% of employees to work in office, down from 75% Maximum capacity at events down from 750 to 250 Museum and public library capacity limits down from 65% to 50% Indoor gyms and fitness studios to close till 30 May
14/5/2021	Phase 2 (Heightened Alert) restrictions from 16 May to 13 June <ul style="list-style-type: none"> Group gathering limits down to 2 people, limited to 2 gatherings a day Work From Home to be the default work arrangement No dining-in, only takeaway and delivery to be offered Museums, public library, and tourist attraction limits down to 25% Events and live performances to be capped at 50 people No wedding receptions to be held
16/5/2021	Primary schools, secondary schools and Junior Colleges to move to full Home-Based Learning from 19 May
18/5/2021	In-person private tuition and enrichment classes for younger students aged 18 and below to be suspended from 19 May
10/6/2021	Phase 3 (Heightened Alert) restrictions in two stages <ul style="list-style-type: none"> From 14 June: Group gathering limits up from 2 to 5 people From 21 June: Possible resumption of dining-in
18/6/2021	Dining-in and mask-off indoor fitness activities to resume on 21 June in groups of up to 2 people
7/7/2021	Dining-in group limits up from 2 to 5 from 12 July; Work From Home to remain the default work arrangement
16/7/2021	Dining-in and mask-off indoor fitness activities back to groups of 2 for the unvaccinated and remaining at groups of 5 for the vaccinated
20/7/2021	Resumption of Phase 2 (Heightened Alert) restrictions from 22 July to 18 August <ul style="list-style-type: none"> Group gathering limits down to 2 people No dining-in, only takeaway and delivery to be offered
4/8/2021	All visits to hospital wards to be suspended from 5 to 18 August
6/8/2021	Dining-in back to groups of 5 for the vaccinated and groups of 2 for the unvaccinated and those dining outdoors from 10 August; Group gathering limits up from 2 to 5; 50% of employees allowed to return to the workplace from 19 August
18/9/2021	Primary schools and Special Education schools to move to Home-Based Learning from 27 September to 6 October
22/9/2021	All visits to hospital wards to be suspended from 24 September to 23 October
24/9/2021	Stabilisation Phase restrictions from 27 September to 24 October <ul style="list-style-type: none"> Group gathering limits down from 5 to 2, limited to one gathering a day Dining-in group limits down from 5 to 2 Work From Home to be the default work arrangement
20/10/2021	Stabilisation Phase restrictions extended to 21 November

Appendix B – Facebook and Reddit posts used for comment extraction

	Article title	Article URL	Reddit post link	Facebook post link	Date (DD/MM/YY)
1	Singapore to cut social gathering size from 8 to 5 amid rising Covid-19 cases; effective May 8-30	https://www.straitstimes.com/singapore/health/singapore-to-cut-social-gathering-size-from-8-to-5-amid-rising-covid-19-cases	https://www.reddit.com/r/singapore/comments/n4li0v/singapore_to_cut_social_gathering_size_from_8_to/	https://www.facebook.com/TheStraitsTimes/posts/10157877541112115	4/5/2021
2	Cap of 5 people for social gatherings, household visits to return as Singapore tightens COVID-19 measures	https://www.channelnewsasia.com/singapore/cap-of-5-people-social-gatherings-household-visits-covid-19-moh-1344191	https://www.reddit.com/r/singapore/comments/n4li5g/cap_of_5_people_for_social_gatherings_household/	https://www.facebook.com/ChannelNewsAsia/posts/10158274255327934	4/5/2021
3	Limit on employees who can return to workplace back at 50%; firms urged to adhere to tighter Covid-19 rules	https://www.straitstimes.com/singapore/limit-on-employees-who-can-return-to-workplace-back-at-50-as-covid-19-measures-are	https://www.reddit.com/r/singapore/comments/n52529/limit_on_employees_who_can_return_to_workplace/	https://www.facebook.com/TheStraitsTimes/posts/10157878977002115	4/5/2021
4	COVID-19: Indoor sports facilities to close temporarily, outdoor exercise classes to continue with reduced capacity	https://www.channelnewsasia.com/news/singapore/indoor-sports-facilities-close-outdoor-classes-allowed-covid-19-14754552	https://www.reddit.com/r/singapore/comments/n61hcv/covid19_indoor_sports_facilities_to_close/	https://www.facebook.com/ChannelNewsAsia/posts/10158278202967934	6/5/2021
5	Group sizes down from 5 to 2, dining-in suspended as Singapore tightens COVID-19 measures	https://www.channelnewsasia.com/news/singapore/covid-19-phase-2-dining-in-work-from-home-tightened-measures-1365476	https://www.reddit.com/r/singapore/comments/nc0vau/group_sizes_down_from_5_to_2_diningin_suspended/	https://www.facebook.com/ChannelNewsAsia/posts/10158295565437934	14/5/2021
6	Group sizes down from 5 to 2, dining-in suspended as Singapore tightens COVID-19 measures	https://www.channelnewsasia.com/news/singapore/covid-19-phase-2-dining-in-work-from-home-tightened-measures-14808382	https://www.reddit.com/r/singapore/comments/nc0veq/group_sizes_down_from_5_to_2_diningin_suspended/	https://www.facebook.com/ChannelNewsAsia/posts/10158295565437934	14/5/2021
7	No dining in, social gatherings capped at 2 people from May 16 as S'pore tightens Covid-19 rules	https://www.straitstimes.com/singapore/health/no-dining-in-social-gatherings-capped-at-2-people-from-may-16-as-spore-tightens	https://www.reddit.com/r/singapore/comments/nc0vwe/no_dining_in_social_gatherings_capped_at_2_people/	https://www.facebook.com/TheStraitsTimes/posts/10157899793777115	14/5/2021

8	Only 2 visitors per household per day, no dining-in allowed: Covid-19 rules in S'pore from May 16	https://www.straitstimes.com/singapore/health/only-2-visitors-per-household-per-day-no-dining-in-allowed-covid-19-rules-in-spore	https://www.reddit.com/r/singapore/comments/nc0wni/only_2_visitors_per_household_per_day_no_diningin/	https://www.facebook.com/TheStraitsTimes/posts/10157899892712115	14/5/2021
9	Group sizes down from 5 to 2, dining-in suspended as Singapore tightens COVID-19 measures - CNA	https://www.channelnewsasia.com/news/singapore/covid-19-phase-2-dining-in-work-from-home-tightened-measures-14808382	https://www.reddit.com/r/singapore/comments/nc0y1s/group_sizes_down_from_5_to_2_diningin_suspended/	https://www.facebook.com/ChannelNewsAsia/posts/10158295565437934	14/5/2021
10	Fixed seating with one-metre spacing for recess, no intermingling, as MOE tightens measures to fight Covid-19	https://www.straitstimes.com/singapore/fixed-seating-with-one-metre-spacing-for-recess-no-intermingling-as-moe-tightens-measures	https://www.reddit.com/r/singapore/comments/ncctqd/fixed_seating_with_onemetre_spacing_for_recess_no/	https://www.facebook.com/TheStraitsTimes/posts/10157900823922115	15/5/2021
11	2-passenger limit for taxis, private-hire cars during tightened COVID-19 restrictions	https://www.channelnewsasia.com/news/singapore/2-passengers-taxis-private-hire-cars-grab-phase-2-covid-19-14820210	https://www.reddit.com/r/singapore/comments/ndfe1i/2passenger_limit_for_taxis_privathire_cars/	https://www.facebook.com/ChannelNewsAsia/posts/10158300015227934	16/5/2021
12	COVID-19: Primary, secondary schools and JCs to move to full home-based learning from May 19	https://www.channelnewsasia.com/news/singapore/covid-19-home-based-learning-primary-secondary-school-jc-14821222	https://www.reddit.com/r/singapore/comments/ndmusi/covid19_primary_secondary_schools_and_jcs_to_move/	https://www.facebook.com/ChannelNewsAsia/posts/10158300676777934	16/5/2021
13	COVID-19: Primary, secondary schools and JCs to move to full home-based learning from May 19	https://www.channelnewsasia.com/news/covid-19-home-based-learning-primary-secondary-school-jc-14821222?cid=FBcna&fbclid=IwAR2tl4G7S3Emfjssaq_Pi2Ofa5zYeb_LQ_6ShQylTMSMFQS9FYTQFvcIIo	https://www.reddit.com/r/singapore/comments/ndmv03/covid19_primary_secondary_schools_and_jcs_to_move/	https://www.facebook.com/ChannelNewsAsia/posts/10158300676777934	16/5/2021
14	S'pore schools to start full home-based learning from May 19 amid spike in Covid-19 cases	https://www.straitstimes.com/singapore/parenting-education/spore-schools-to-start-full-home-based-learning-from-may-19-amid-spike	https://www.reddit.com/r/singapore/comments/ndmuup/spore_schools_to_start_full_homebased_learning/	https://www.facebook.com/TheStraitsTimes/posts/10157904579472115	16/5/2021

15	All in-person private tuition and enrichment for students aged 18 and below to be suspended: MOE	https://www.channelnewsasia.com/news/singapore/in-person-tuition-enrichment-classes-suspended-covid-19-students-14833184	https://www.reddit.com/r/singapore/comments/nf7x2i/all_inperson_private_tuition_and_enrichment_for/	https://www.facebook.com/ChannelNewsAsia/posts/10158304639667934	18/5/2021
16	COVID-19 task force 'evaluating' timing and scope of reopening amid fresh outbreak: Lawrence Wong	https://www.channelnewsasia.com/news/singapore/covid-19-task-force-evaluating-the-timing-and-scope-reopening-15025980	https://www.reddit.com/r/singapore/comments/o102bk/covid19_task_force_evaluating_timing_and_scope_of/	https://www.facebook.com/ChannelNewsAsia/posts/10158367144152934	16/6/2021
17	COVID-19 task force 'evaluating' timing and scope of reopening amid fresh outbreak: Lawrence Wong	https://www.channelnewsasia.com/news/singapore/covid-19-task-force-evaluating-the-timing-and-scope-reopening-15025980?1	https://www.reddit.com/r/singapore/comments/o104h0/covid19_task_force_evaluating_timing_and_scope_of/	https://www.facebook.com/ChannelNewsAsia/posts/10158367144152934	16/6/2021
18	COVID-19 task force 'evaluating' timing and scope of reopening amid fresh outbreak: Lawrence Wong	https://www.channelnewsasia.com/news/singapore/covid-19-task-force-evaluating-the-timing-and-scope-reopening-15025980?cid=FBcna&fbclid=IwAR11Yxg0kKop1qNp3Ue4CUdhbH6P1fHcQ0gA0Gloa7DzvUFwRh5Mp3WDwIM	https://www.reddit.com/r/singapore/comments/o11sf4/covid19_task_force_evaluating_timing_and_scope_of/	https://www.facebook.com/ChannelNewsAsia/posts/10158367144152934	16/6/2021
19	COVID-19: Dining-in group size back to 2; groups of 5 only if all fully vaccinated	https://www.channelnewsasia.com/news/singapore/covid-19-how-many-can-dine-in-vaccinated-groups-of-5-2-15232846	https://www.reddit.com/r/singapore/comments/oldoyg/covid19_diningin_group_size_back_to_2_groups_of_5/	https://www.facebook.com/ChannelNewsAsia/posts/10158423307637934	16/7/2021
20	S'pore tightens Covid-19 rules as KTV cluster grows, pivoted night spots to suspend operations	https://www.straitstimes.com/singapore/health/covid-19-measures-to-be-tightened-over-ktv-cluster-differentiated-approach-for	https://www.reddit.com/r/singapore/comments/oldqa5/spore_tightens_covid19_rules_as_ktv_cluster_grows/	https://www.facebook.com/TheStraitsTimes/posts/10158023125857115	16/7/2021
21	Dine-in group size cut to 2 from July 19 as S'pore tightens Covid-19 rules; but up to 5 allowed at some F&B outlets, including fully vaccinated	https://www.straitstimes.com/singapore/health/dine-in-group-size-cut-to-2-from-july-19-as-spore-tightens-covid-19-rules-but-up-to	https://www.reddit.com/r/singapore/comments/oldqwa/dine_group_size_cut_to_2_from_july_19_as_spore/	https://www.facebook.com/TheStraitsTimes/posts/10158023136557115	16/7/2021

22	S'pore cuts dining-in cap to 2 as Covid-19 KTV cluster grows; up to 5 at some F&B outlets if fully vaccinated	https://www.straitstimes.com/singapore/health/dine-in-group-size-cut-to-2-from-july-19-as-spore-tightens-covid-19-rules-but-up-to	https://www.reddit.com/r/singapore/comments/oldr80/spore_cuts_diningin_cap_to_2_as_covid19_ktv/	https://www.facebook.com/TheStraitsTimes/posts/10158023136557115	16/7/2021
23	No dining in, social group sizes cut to 2 from July 22 as S'pore returns to phase 2	https://www.straitstimes.com/singapore/no-dining-in-social-group-sizes-cut-to-2-from-july-22-as-spore-returns-to-phase-2	https://www.reddit.com/r/singapore/comments/onx8xr/no_dining_in_social_group_sizes_cut_to_2_from/	https://www.facebook.com/TheStraitsTimes/posts/10158030653082115	20/7/2021
24	Visits to all hospital wards to cease for 2 weeks amid rise in COVID-19 community cases: MOH	https://www.channelnewsasia.com/news/singapore/visits-to-all-hospital-wards-to-cess-for-2-weeks-amid-rise-in-15378914	https://www.reddit.com/r/singapore/comments/oxn4p7/visits_to_all_hospital_wards_to_cess_for_2_weeks/	https://www.facebook.com/ChannelNewsAsia/posts/10158457835047934	4/8/2021
25	S'pore to maintain current Covid-19 measures, will rely on vaccination and testing to move ahead	https://www.straitstimes.com/singapore/health/spore-to-maintain-current-covid-19-measures-will-rely-on-vaccination-and-testing-to	https://www.reddit.com/r/singapore/comments/ph2utq/spore_to_maintain_current_covid19_measures_will/	https://www.facebook.com/TheStraitsTimes/posts/10158110812602115	3/9/2021
26	No easing of restrictions for now amid spike in COVID-19 cases: Lawrence Wong	https://www.channelnewsasia.com/singapore/covid-19-no-easing-restrictions-lawrence-wong-2154951	https://www.reddit.com/r/singapore/comments/ph2xt2/no_easing_of_restrictions_for_now_amid_spike_in/	https://www.facebook.com/ChannelNewsAsia/posts/10158509010257934	3/9/2021
27	MOH suspends in-person visits to residential care homes from Monday as COVID-19 cases rise among residents, employees	https://www.channelnewsasia.com/singapore/covid-19-old-age-homes-visits-suspended-2172411	https://www.reddit.com/r/singapore/comments/pmou64/moh_suspends_inperson_visits_to_residential_care/	https://www.facebook.com/ChannelNewsAsia/posts/10158522859952934	12/9/2021
28	Home-based learning for Pri 1 to 5 students amid rise in COVID-19 cases: MOE	https://www.channelnewsasia.com/singapore/home-based-learning-pri-1-5-students-amid-rise-covid-19-cases-moe-2186721	https://www.reddit.com/r/singapore/comments/pqhokh/homebased_learning_for_pri_1_to_5_students_amid/	https://www.facebook.com/ChannelNewsAsia/posts/10158532152462934	17/9/2021
29	No visitors to hospital wards allowed from Sept 24 to Oct 23, amid surge in Covid-19 cases	https://www.straitstimes.com/singapore/health/no-visitors-to-hospital-wards-allowed-from-sept-24-to-oct-23-amid-surge-in-covid-19	https://www.reddit.com/r/singapore/comments/pt3yxt/no_visitors_to_hospital_wards_allowed_from_sept/	https://www.facebook.com/TheStraitsTimes/posts/10158141394567115	22/9/2021

30	COVID-19: Dining-in group size limit at regular F&B outlets down to 2; social gatherings reduced	https://www.channelnewsasia.com/singapore/covid-19-dining-group-size-two-social-gatherings-reduced-moh-2199456	https://www.reddit.com/r/singapore/comments/puh01k/covid19_diningin_group_size_limit_at_regular_fb/	https://www.facebook.com/ChannelNewsAsia/posts/10158542015252934	24/9/2021
31	Work-from-home to be default from Monday as COVID-19 workplace measures tightened	https://www.channelnewsasia.com/singapore/work-home-default-covid-19-workplace-measures-2199566	https://www.reddit.com/r/singapore/comments/puh0ez/workfromhome_to_be_default_from_monday_as_covid19/	https://www.facebook.com/ChannelNewsAsia/posts/10158542018962934	24/9/2021
32	S'pore unis move classes online to reduce spread of Covid-19 as cases continue to rise	https://www.straitstimes.com/singapore/parenting-education/spore-unis-move-classes-online-to-reduce-spread-of-covid-19-as-cases	https://www.reddit.com/r/singapore/comments/pyfbzk/spore_unis_move_classes_online_to_reduce_spread/	https://www.facebook.com/TheStraitsTimes/posts/10158154314997115	30/9/2021
33	Stabilisation Phase extended to Nov 21; more time needed to stabilise COVID-19 situation: MOH	https://www.channelnewsasia.com/singapore/stabilisation-phase-restrictions-extended-nov-21-dining-social-gatherings-2256766	https://www.reddit.com/r/singapore/comments/qbyerz/stabilisation_phase_extended_to_nov_21_more_time/	https://www.facebook.com/ChannelNewsAsia/posts/10158582288427934	20/10/2021
34	Singapore's Covid-19 measures to be extended till Nov 21, to be reviewed at 2-week mark	https://www.straitstimes.com/singapore/health/singapores-covid-19-measures-to-be-extended-till-nov-21-to-be-reviewed-at-2-week	https://www.reddit.com/r/singapore/comments/qbyfn4/singapores_covid19_measures_to_be_extended_till/	https://www.facebook.com/TheStraitsTimes/posts/10158186209047115	20/10/2021

Appendix C – Selection of 30 manually scored comments

Scores are shown to four significant figures

	text	score_vader	score_tb	score_human_avg
1	Amazing 4G leadership! Please call for an elections again this year because I am very sure a lot of people will wanna vote for you all and give you the well deserving mandate! So proud to have smart leaders like you all!!!	0.9645	0.6737	-0.3750
2	I'll give our stupid govt full credit instead. Whatever happens in other countries is out of our control, but what's within our control is our response to what happens in other countries and our govt had to fuck it up, all in the name of eCoNoMy	-0.7311	-0.2200	-0.5000
3	Go back to phase 2, but more flights are arriving! Song boh?	0.0000	0.3125	-0.5000
4	time for the "Containing Crowd Battle" to start, sayonara F&B businesses	-0.3818	0.0000	-0.5000
5	>own up to your mistakes PAP? Admit that they are wrong? HAHAAHAHAHAHA	0.4576	-0.0667	-0.5000
6	Wong Out	0.0000	0.0000	-0.3750
7	Please remember all this next time you vote.	0.3182	0.0000	-0.2500
8	Bro u said 3 idiots is favourite among your foreign uni people. Why do they relate to it? Is education same there as in India? I thought it was different	0.0000	-0.2313	0.0000
9	Wa meng ti wa meng ti	0.0000	0.0000	-0.5000
10	fully expecting a ton of posts in r/Sggonewild in the next month 🤔	0.4404	0.0000	0.3750
11	Yea, seems like a good idea, a [recent study]((but yet to be peer reviewed) in the UK suggests improved immunity when the 2nd dose of the Pfizer vaccine is taken after 12 weeks, so delaying it not only ensures more people can get vaccinated quickly, but could also improve immunity as well	0.9349	0.2556	0.7500
12	I'm sorry such a happy occasion for the both of you had to be postponed twice. It really sucks and your frustrations are perfectly justifiable! Hope the venue understands how this is completely out of your control and keeps the spot for you :)	0.8019	0.2286	0.2500
13	Too bad their chicken is dry asf(in my opinion)	-0.5423	-0.4556	-0.2500
14	Actually not true, dine in indoors is proven a higher risk activity both by the US and UK.	-0.5318	0.0375	-0.2500
15	awww this made me smile :')) ya hope this whole thing ends soon man. Take care ok!	0.8908	0.3812	0.7500
16	Crazy Rich Politicians, coming ~~soon~~ *never* to a theatre near you	0.296	-0.0417	-0.2500
17	Yea people keep bring up New York and California as an example, FYI, New York's 7 day average is 400+ close to 500, California is 900+ close to 1000 cases daily still. You take california aside since it's a huge state, NY only has 1.5x our total population roughly, and has 50x the number of	0.4784	-0.0444	-0.3750



















	cases DESPITE having 70% of adults having at least 1 dose. Like shit is that the level that is acceptable to people here all of a sudden? This sub would have called for Lawrence's head if those were the levels that we hit for no. of cases. But suddenly the goal posts shift because they are sian of this(who isn't?)			
18	You forgot gyms also. And ktvs and bars and karaokes	0.0000	0.0000	0.0000
19	Dear politics, guess what in 4 years we will still have from time to time 20 cases a day like the almighty flu. Most countries in the world are now open again with way lower vaccination rate but wait we are smarter than all of them. They have a plan we don't. Continue to speak fear about daily cases what are the figures on the critical cases ?	-0.4767	0.1250	-0.5000
20	This poor singapore..way to set back your economy by more	-0.4767	0.0333	-0.7500
21	thank you for all that you do!! :)	0.7256	0.5000	0.7500
22	Singpass—>Healthhub also shows vaccination status.	0.0000	0.0000	0.0000
23	Fuck the PAP	-0.2023	-0.1500	-0.5000
24	Totally defeats the point of pushing for vaccination ain't it...?	-0.3804	0.0000	-0.5000
25	Wtf I'm damn annoyed. I barely went out for these few months, and had plans to dine out with SO for our first anniversary this Friday. And things are ruined again :(-0.9337	-0.2583	-0.7500
26	Have you asked her to consider re-assessing the risk? My mom is 62, diabetic, on dialysis and is allergic to several types of meds, and recently got the second dose with few complaints. We will eventually open up so we told her that it is for her own safety that she lowers the risk of getting the disease. Sneezing is painful for my mom because of her peritoneal dialysis residual pain every day, so we noted that COVID will make her feel worse than any vaccine in all likelihood. Nothing bad happened in the end. I understand that it can be riskier for some people and this is just my experience. We had also delayed the first vaccination till early June this year.	-0.9387	-0.0955	0.2500
27	\> soggy prata *NOOOOOO*	0.0000	0.0000	-0.5000
28	With this kind of last minute knee jerk reaction it better be 2k+	0.1280	0.3667	-0.3750
29	You're being reactionary and small-minded like the government. There was **always** going to be another outbreak, regardless of its source, especially with the contagious Delta variant circulating. The question was: how will the Singapore G respond to it? Would they stick to their 'it's endemic, we have good vaccination rates, we must keep going' or would they panic, throw the curtains shut, and hide under the lockdown bed again? Now we know it's the latter.	-0.3866	0.2333	-0.5000
30	Why did you want to come back to this godforsaken country?	0.0772	-0.2000	-0.7500

Appendix D – Selection of comments from May 18, September 12 and September 30

Scores are shown to four significant figures

	text	score_vader	score_tb	platform
	May 12			
1	I think he's saying yes it's difficult for Ministry of Education to control, but once you get caught (someone sabotage you/contact tracing), it's implied that the Government will not come down lightly on you and you are proceeding at your own risk. Hopefully this is enough deterrence for private tutoring to stop	-0.7506	0.0574	Reddit
2	Hopefully parents can make arrangements for the kids to play with other kids in small groups. Lack of outside stimulus is terrible for kids.	-0.0772	-0.3438	Reddit
3	Tuition ctrs with 10 or more students have to register with Ministry of Education. But private teachers giving home tuition are not regulated currently. Even so, who would want to continue with home tuition in this climate.	0.5499	0.1667	Reddit
4	Tiresome, my tutee's parents (primary 4 and below) don't want online tutoring..... means no income at all for the next one month...	-0.2263	-0.0333	Reddit
5	Does this mean home schooling families no need to have lessons?	-0.2960	-0.3125	Reddit
6	Honestly, preschools should be closed as well. Childcare and studentcare to be open for parents who can't Work From Home. These young ones are told that they can't sing songs and play together with others... have you interacted with kids below 6yo? 🧑 Most childcare centres are pretty small. When they eat or during routine time, their masks are off and they'd definitely talk to each other. Many preschoolers have older siblings in primary or secondary school. I'm very worried that there'd be an outbreak in preschools coz can you imagine what these kiddos will be going through? Please don't wait until the bomb go jibaboom le then react. Like what these politicians have been doing. The older kids are suffering because of these politicians man 🤔	0.9164	0.1406	Facebook
7	All adults courses that are conducted in groups should be suspended as well. Stay safe	0.2633	0.5000	Facebook
8	Please make it mandatory for lifelong learning institutes to convert all classes to online classes too.. Apparently classes are still being conducted physically with multiple learners still coming in and out of the centres to attend their lessons in groups of more than 2 people. Though this round of viruses may be more prone towards children, that does not mean it does not affect the adults too. We should minimise all these risks now to prevent further clusters from happening. Singapore Ministry of Manpower Ministry of Health, Singapore	0.1531	0.1007	Facebook

9	I would say piano one on ones especially for the super young kids should still remain open physically especially this age group like 4 to 7 is too young to do zooming.	0.7693	0.0889	Facebook
10	Can you suspend the flights from India instead?	-0.3182	0.0000	Facebook
September 12				
1	As someone who works in a nursing home, I just feel so done with all these on-off-on-off restrictions. Sigh and again with the buying time for unvaccinated seniors to get vaccinated.... I have no words	-0.2732	0.0000	Reddit
2	Would a vaccine mandate help? What would you wish the government did as someone deep in the weeds?	0.6966	0.0000	Reddit
3	My nursing home has already achieved 90% vaccinated status.. the remaining 10% are mostly anti-vaccinated Not sure how much a vaccine mandate might help since these anti-vaccinated would rage about whether the government will take responsibility if they get side effects from the vaccine We've been encouraging them to get vaccinated for like a year... Feel more like a pushy salesman than a healthcare worker at this point	0.5373	0.2375	Reddit
4	I'm sure not all of them are anti-vaccinated right? Are there elderly at your nursing home who cannot be vaccinated due to medical reasons? What can be done to help this group of elderly?	0.6747	0.1652	Reddit
5	Do you think it would be feasible to allow visits if both resident and visitor are fully vaccinated? Blanket restrictions that apply as harshly to both vaccinated and unvaccinated are making less and less sense given the direction we have chosen to tackle the pandemic.	0.2263	-0.1778	Reddit
6	My heart goes out to dementia patients in homes. They may not understand the need to get vaccinated, and it may be difficult to explain to them, given the nature of dementia. For some, when they see a needle, they may also think that they are being harmed and may react -\-\ which could be violent -\-\ not all dementia patients of course, but it is tough for them and their caregivers bcoz they cannot understand what is going on.... 🤔 😞	-0.7964	-0.5630	Facebook
7	Is this the reason why you dont announce unlinked cases anymore? We as families deserves to know which facilities are affected so wheres the transparency?	-0.2406	0.0000	Facebook
8	No visitors , good temporary restriction. Be safe then sorry.	0.2960	0.2333	Facebook
9	Looks like many secrecy in infected announcements.. Maybe it's part of state secret.	-0.1779	0.0500	Facebook
10	Ministry of Health should only suspend visits to those residential care homes which were affected. Cannot treat old folks like prisoners!	-0.4678	0.0625	Facebook
September 30				
1	I feel reeally bad for students. 2 years online. University is also about socializing and learn to have face to face interactions with academics and fellow students. Stay strong. I really hope you have at least a year of REAL studies...	0.4576	-0.0333	Reddit

2	In my last year of university and honestly feels really tired that half of my time in university was spent online. There are perks and all but I really miss the social aspect of it too. When I looked back at the time spent in university, honestly pre-covid times had more significant memories for me. Post covid sems were a blur, and kinda boring from that point on.	0.4005	0.0417	Reddit
3	I can understand how tough a decision that must've been. It's an essential phase for him to be in school, but you also don't want to compromise his health and safety. I wish you the best of luck, hopefully things will get better soon so he can get back.	0.9791	0.2222	Reddit
4	I picked local you which is at least 1 year longer just so I can go for semester exchange and this? Wonderful.	0.6570	0.2333	Reddit
5	Dam sad, for other local universities at least you hall life to mingle. For SIM there's almost really no face 2 face interactions at all. I've not met anyone in university for the last 1 year except when there were 1-2 group projects. I wish I could connect with more people, but for the greater good this is part of the norm now.	0.7425	0.0750	Reddit
6	AGAIN, for the nth time, and I will continue to do so: Vaccination is not to make you IMMORTAL. Vaccination does not PREVENT infection or death. It's to REDUCE your potential of infection/ death.	-0.7706	0.0000	Facebook
7	How about secondary schools? We seem to have neglected this group.	-0.5267	-0.3000	Facebook
8	Epic failure because only relying on leaky vaccines. And hence....the number of hospitalised patients of 616 Fully vaccinated already outnumbered the 470 Not Fully vaccinated. So hospitals are swamped with Fully Vaccinated patients who can also land up needing O2 and hospital. Constantly spending time harping and bullying the unvaccinated these few months when they should have spent more time anticipating the explosion of cases And coming up with a game plan.	-0.7845	-0.0521	Facebook
9	Unis are institutions which, amongst other missions, are tasked to inspire cutting-edge tech and thinking. University administrators ought to, as far as possible, switch to online learning. Why make students fill lecture rooms for lectures which can just as effectively be given virtually?	0.7650	0.1437	Facebook
10	What's the long term effects of the 💉 ? 🧑🧒🧓 Why are more and more 💉 vaccinated individuals       being infected      at a higher rate than before vaccination 💉 ? Is the  weakening vaccinated       natural immune system? 🤔 🤔	-0.5913	0.3000	Facebook