

# Sentiment Analysis on Citizen's Emotion in Malaysia during the COVID-19 Pandemic: Impact due to Lockdown

Siti Nur A'dilah Hasnor<sup>1</sup>, Nahdatul Akma Ahmad<sup>2</sup>,  
Nursyahidah Alias<sup>3</sup>, Itaza Afiani Mohtar<sup>4</sup>, Nik Ruslawati Nik  
Mustapa<sup>5</sup>

College of Computing, Informatics and Media, Universiti Teknologi MARA, Perak Branch,  
Tapah Campus, 35400 Tapah Road, Perak, Malaysia  
Corresponding Author Email: nahdatul@uitm.edu.my

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v13-i5/16848>

DOI:10.6007/IJARBSS/v13-i5/16848

**Published Date:** 17 May 2023

## Abstract

People utilize social media platforms such as Twitter for social networking and to communicate their thoughts, feelings, and ideas with others. During the COVID-19 pandemic, Malaysians took to social media to express their feelings. This research aims to understand Malaysian Twitter users' discussions and psychological reactions to the COVID-19 pandemic. Supervised Machine learning technique was used to analyze 1,645 Tweets (written both in Malay and English) related to COVID-19 between 1st January 2020 and 7th June 2021. The tweets were pre-processed and then classified using Support Vector Machine (SVM) algorithm for the sentiment analysis. The classification results matched 6 basic emotions from Paul Ekman Model which are anger, fear, happy, love, sadness, and surprise. From the sentiment analysis, it was found that anger towards people who violated the Standard Operating Procedure (SOP) during Movement Control Order (MCO), political issues, and the issue of non-essential companies and factories disregarding the order to cease operation during MCO is dominant. This research may aid health professionals and researchers in better understanding Malaysian citizens' possible reactions during massive health crisis.

**Keywords:** Sentiment Analysis, COVID-19, Pandemic, Emotion, Mental Health

## Introduction

COVID-19 is said to originate in Wuhan, Hubei Province at the start of December 2019 when several patients reported severe respiratory infections. These patients had a background of working in the wholesale fish and seafood market, also known as wet markets (Huang et al., 2020). On March 2020, the World Health Organization (WHO) declared the COVID-19 as a global pandemic. Although this kind of respiratory disease had already caused both widespread death and cases of infections in Wuhan since 2019, the health crisis did not

show its true dimensions in Malaysia up until March 2020 where there was a significant jump in active cases in Malaysia due to the four-day religious gathering held at the Seri Petaling Mosque, Kuala Lumpur. This cluster became one of the largest threats to the pandemic situation in Malaysia for nearly four months. On 19th May 2020, according to the Director General of Health, Ministry of Health, Malaysia, 48% of Malaysia's COVID-19 cases were associated with the religious cluster, bringing the number of cases involved up to 3,347 cases (TheSunDaily, 2020). There onwards Malaysia became the country in Southeast Asia with the highest COVID-19 cases (Rahim, 2020). On 15th March 2020, due to the significant jump in active cases, the Prime Minister of Malaysia announced the decision of the federal government to implement the Movement Control Order (MCO) effective from 18th March until 31st March 2020 (Bunyan, 2020). The period has been extended four times as additional two-week "phases" over the course of two months as new cases continued to climb.

During the MCO period, Malaysians depended on virtual media to find, disseminate and share information from official sources such as the government, health bodies, news companies, as well as social media. Furthermore, social media is important in the lives of many people who are impacted by the COVID-19 pandemic. Social media has provided both opportunities and obstacles in health crisis communication, breaking the dominance of traditional news media (Lyu, 2012). Social media platforms also offer real-time information that includes many forms of content, such as text, audio and video. Twitter, one of the most well-known social media platforms, has been a key source of information sharing and self-documentation (Liu et al., 2010). Twitter has been one of the mediums for millions of people to discuss their thoughts on various issues. People have used Twitter to interact, express and distribute information regarding crises in the past, such as cyclones Soriano et al (2016), ebola Van Lent et al (2017), floods (Nair et al (2017), or Zika (Fu et al., 2016). As Malaysians have been battling COVID-19 since 2020 and the population is under lockdown, the significance of Twitter has escalated more than ever. Emotional distress has been linked to a wide range of public health issues, including an increase in alcohol use and drug addiction (Ashton et al., 2017), eating disorders (Litwin et al., 2017), anxiety and insomnia (Kirwan et al., 2017) among others. Furthermore, people who have encountered strong negative emotions are more inclined to believe anger-inducing rumours (Na et al., 2018) that in turns strengthen the emotional consequences through feedback. In this context, new research has offered methods to minimise population stress during social isolation, with promising outcomes (Pizzoli et al., 2020). However, to the best of the authors' knowledge, no comprehensive research on the emotional impact of COVID-19 has been published.

Therefore, this research objectives are as following

- To investigate Malaysians' emotions expressed through social media during the COVID-19 pandemic lockdown.
- To highlight regarding emotional distress and experience of Malaysian's during lockdown.

### **Materials and Methods**

The following are the detailed activities that were used in this study.

#### **a. Data Extraction**

Data were gathered from Twitter beginning in January 2020 until end of July 2021 utilising the developer's Twitter API. A python module called snsrape is used to extract the tweets that will enable the extraction of older tweets with the help of tweepy module. These

data comprised of written tweets related to the COVID-19 pandemic in Malaysia. Hashtags or keywords such as “Coronavirus”, “corona”, “stay home”, “stay safe”, “#KitaJagaKita”, “lockdown”, “covid-19” and “COVID-19” were used to run the query. In order to obtain specific Malaysian Twitter user’s tweets, the geolocation was implemented during filtration of extraction where the coordinates of Kuala Lumpur are used within 400 miles radius from that coordinate. Aside from that, the following features were also collected for each tweet message which were

- Each message-level tweet (full text)
- Function features of hashtags, username and datetime

A total of 2,154 tweets were extracted. After the removal of duplicates and unnecessary tweets, 1,645 tweets were saved in CSV files as the dataset for this study. The data collection method for this study complied with Twitter’s terms of Service and Developer’s Agreement and Policy.

### **b. Pre-processing the Data**

It is a method of cleaning up the data, determining the variable to be used and converting it into a proper structure to make it more suitable for analysis onward. This step is the most critical step of the whole process. Data pre-processing or data cleaning is required to ensure the quality of the data. Hence, before feeding the prepared tweets into the developed classifier model, various filtering technique were used to clean the data which are:

- Removing the hashtags symbols and their content (e.g., #COVID-19), @users, and URLs from the text message because they did not contribute to the study analysis.
- Remove special character, punctuation, and numbers from the dataset because they did not aid in the detection of profane remarks.
- Remove the HTML character. For an example &amp;, &lt;, &reg and etc.
- Remove any newlines.
- Split any attached words. For an example, ‘soterrified’ was converted to ‘so terrified’.
- Converting the tweets into lower case.

The cleaning process was done via built-in python package called re, which can be used to work with regular expression. The data was then placed in a database. This study implemented MySQL as its desired Database Management System (DBMS). The database stores the extraction of tweets details, according to date time, the author’s username and the text itself for easier access.

### **c. Training Model**

During this stage, the sentiment analysis model was trained before moving to the analysis phase to improve its performance for the better result ahead. A training dataset was obtained from Kaggle and Github which consisted of two languages: English and Bahasa Malaysia. A total of 175,809 data were used to train the model. In this study, a dataset was used to train the model using the supervised machine learning technique which is Support Vector Machine (SVM) algorithm. Previous researchers have concluded that SVM algorithm does well for most of the extraction features compared to Naïve Bayes and Maximum Entropy algorithms (Pang & Lee, 2008). A python module called scikit-learn was used to train the model. Before the training process began, the data was split using `train_test_split()` function and then vectorised using Tf-idf method to convert the string data into numerical data. The data was then trained using `fit()` function with SVM classifier. Training of a model is necessary

in order to learn the different patterns, rules, and features. The trained model was then saved as pickle file for future use.

#### d. Test Model

Once the model has been trained on a given dataset, it is ready to be tested. A test dataset was provided to the model to check the accuracy of the model. The confusion matrix, precision, recall and F1-score for the model was also determined.

#### e. Analysis

Next, the model classified each tweet into six basic emotions according to Paul Ekman Model (Ekman, 2005) which are anger, fear, happy, love, sadness, and surprise. For each given tweet, this technique returned one emotion from the six categories. The extracted data were then predicted using a predict() function to predict the emotion of each tweet. Once the scoring of the tweets was done on the basis of emotions, a word cloud was created for each emotion to analyse all related words from a particular emotion.

#### f. System Architecture

The system architecture provides an overall overview of the system. As shown in Figure 1, the data can be obtained by extracting tweets from Twitter using its Application Programming Interface (API) provided by the developer. The collected tweets are stored inside the database. The database storage used is MySQL. Before feeding the tweets into a sentiment classifier, the tweets were pre-processed to obtain a more accurate result. Pre-processing included tokenization and cleaning the tweets from unwanted hashtags, links, usernames, stop words, and noise, before finally converting to lower case. The sentiment classifier will classify the results which are the emotions corresponding to the tweets; happy, surprise, love, sadness, fear, and anger, and store it into the database. The results were then visualized using the Dash framework and displayed on a web application that can later be accessed by other parties. The sentiment classifier was developed by training the Support Vector Machine (SVM) algorithm with a labelled Twitter dataset obtained from the internet.

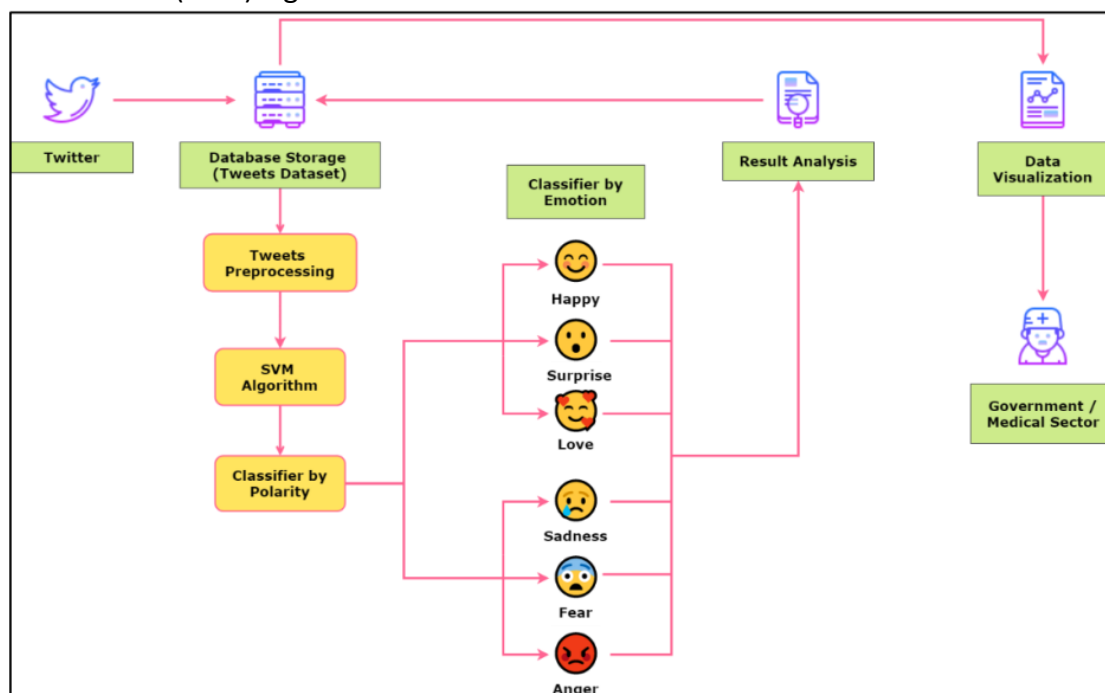


Fig. 1 System Architecture

## Results and Discussion

### A. Analysis of Malaysian Citizens' Emotion during COVID-19

Malaysian citizens' emotion during COVID-19 have direct correlation with their mental health and current emotional state. In addition to that, they started to depend more on social media to express their thoughts, therefore, it has become essential to monitor citizens mental health state on social media. Thus, this research utilized tweets shared on Twitter that contained the word 'covid19', 'coronavirus', 'COVID-19', 'stay home' and 'lockdown' to determine the emotions associated with the tweet.

For this research, a total of 1645 tweets produced from the query 'covid19', 'coronavirus', 'COVID-19', 'stay home' and 'lockdown' were extracted directly from Twitter. The tweets extracted are dated from 26th January 2020 to 7th June 2021. They are then cleaned from any unwanted details such as the user's Twitter handle, hashtag, and links. The tweets are also removed from words that contain less than three letters and stop words which are words that are used frequently but do not contribute to the content of the tweets. These data were analyzed by predicting the emotion contained in the tweets by using the sentiment analysis classifier developed. The emotion predicted is then used to visualize the emotion of citizen where the number of tweets for each emotion is counted and displayed according to dates.

After predicting the sentiment of each tweet using the developed sentiment classifier, it was found that among all the accumulated tweets, the tweets for anger have the highest percentage with 57.8% which is 951 tweets. This is followed by happy with 33.6%, sadness with 4.56%, love with 2.68%, fear with 1.09% and lastly surprise with 0.24% as shown in Figure 2. Based on the high number of tweets for anger, it can be assumed that the Malaysians' mental wellbeing are not in a good condition during COVID-19 pandemic.

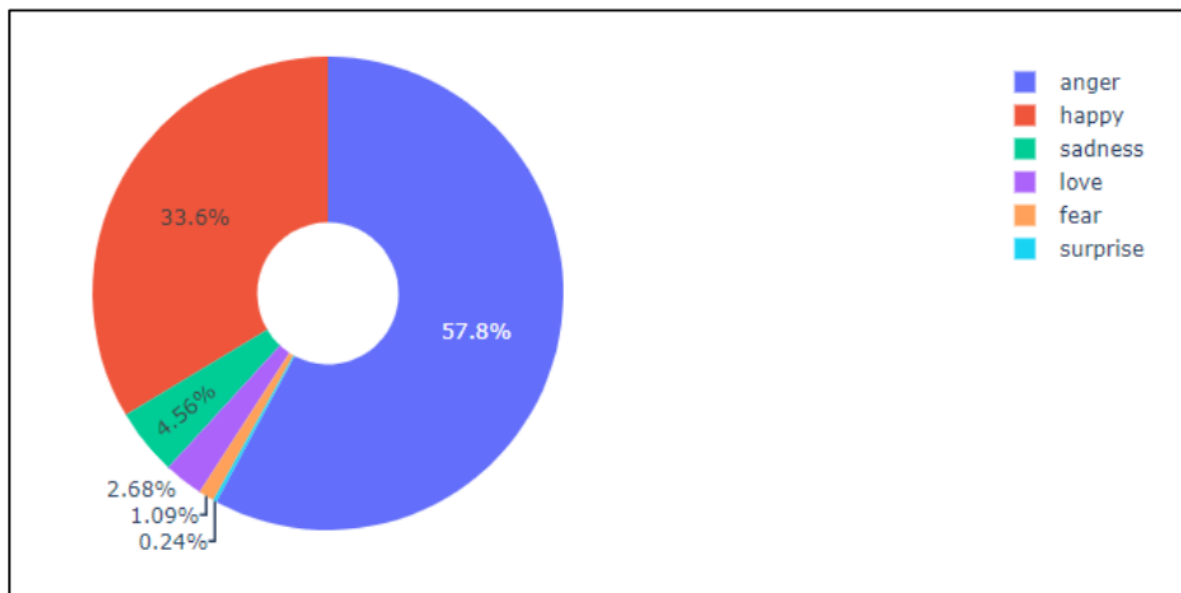


Fig. 2 Percentage of Tweets Emotion Distribution using Pie Chart

The tweets predicted as anger can be interpreted as a strong feeling of annoyance, displeasure, or hostility of citizens of Malaysia towards COVID-19 outbreak. In the dataset that has been extracted from Twitter, the dataset for anger class contains a lot of overlapping

words such as “stay”, “home”, “please”, “lockdown”, and “covid”. Due to these overlapping words, it is predicted that Malaysians are angry towards COVID-19 as the words that most frequent appear in word cloud as in Figure 3 below. This class indicates a negative outcome where it can be interpreted as the citizen being annoyed and displeased with the COVID-19 outbreak.

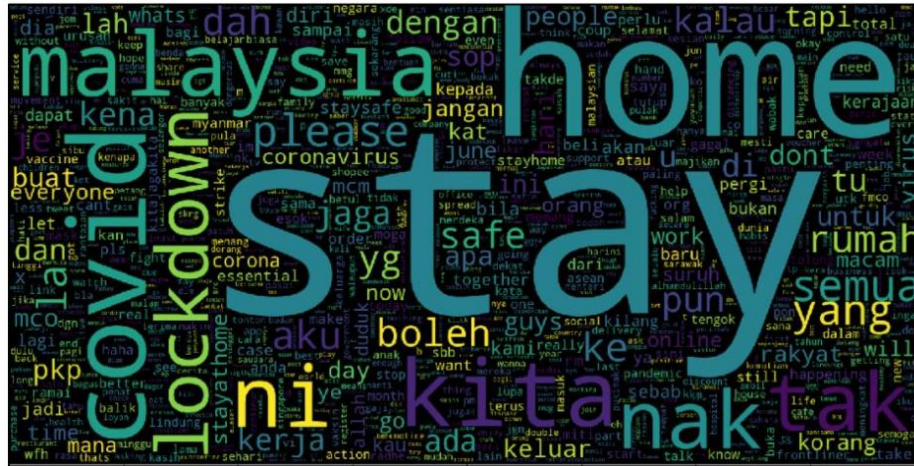


Fig. 3 Word Cloud of anger class datasets

Figures 4 and 5 demonstrate an example of anger tweets from the extracted dataset. News of individuals violating Standard Operating Procedure (SOP) during Movement Control Order (MCO) became one of the factors that caused anger among Malaysians. Many tweets from extracted datasets were about Malaysians complaining about SOP violations by individuals. Photos posted on social media, such as defying the SOP, uploaded by the perpetrators themselves provoked anger among Malaysians, resulting to the highest percentage of anger sentiment throughout this analysis.

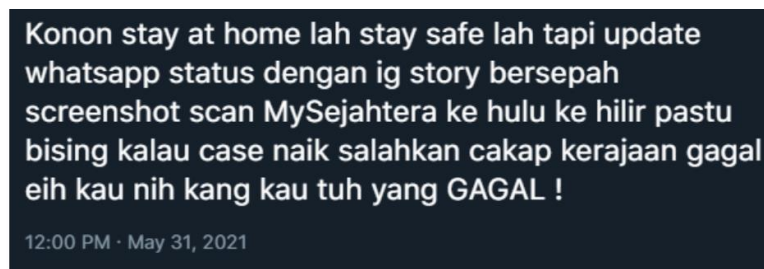


Fig. 4 Example of Anger Tweets from Extracted Dataset

(Translation: Pretending to ‘stay at home’, ‘stay safe’, but WhatsApp updates, IG stories show screenshots of MySejahtera check-ins. Then you have the insolence to blame the government for rising cases. Eih, you’re the one to blame.)

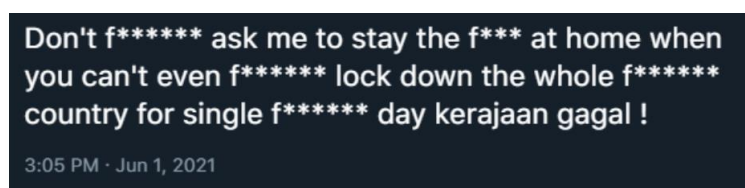




Fig. 7 Word Cloud of happy class datasets

During MCO, Malaysians were ordered to stay at home. Being confined and having all previous routines jeopardized, Malaysians began to be interested in buying online since many e-commerce websites, such as Shopee and Lazada, have provided a lot of shocking sales, vouchers, and special sales day every month during MCO period. For instance, Shopee would have a sales day on 7.7, 8.8, 10.10, and 12.12 of every month, and people began to share what they bought on Shopee, causing it to go viral on social media. This sale or promotion may indirectly be therapy to those who are bored at home. This might contribute to the high percentage of happy sentiment across this analysis.

### B. Analysis of Malaysian Citizens' Emotion Based on Group of Month from 2020 to 2021

The first three months of 2020, when COVID-19 outbreak started in Malaysia, most people in Malaysia are annoyed and angry. They consequently vent their feelings on social media. Figure 8 depicted that from January 2020 until March 2020, the percentage of anger emotion are the highest than any other emotions. As previously stated, the Tabligh cluster was the first large cluster that triggered a dramatic increase in COVID-19 cases in Malaysia. Amongst them were individuals refusing to practice self-quarantine because they feel that COVID-19 is not a threat, thus exposing other individuals to the virus. This might be due to the lack of awareness of the dangers of the COVID-19 virus during that time. Perhaps, this might be the reason why the anger sentiment is the highest than other sentiments during that duration.

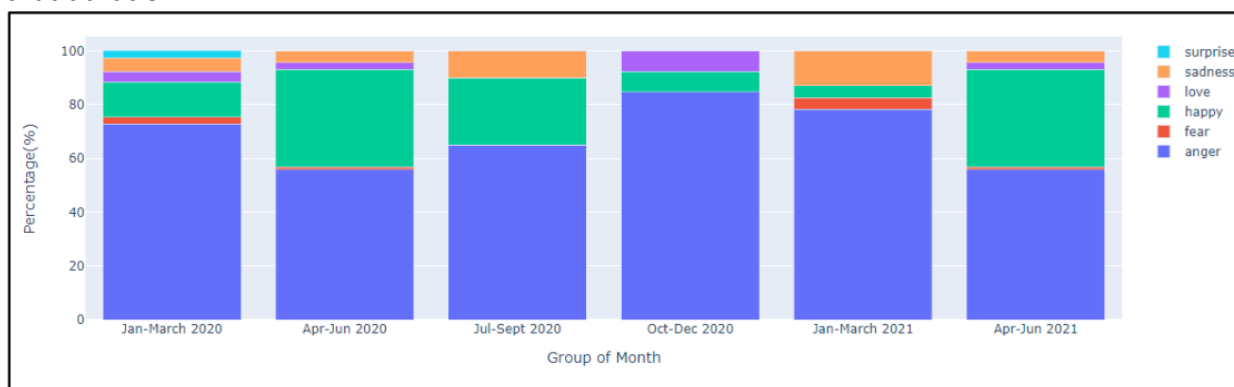


Fig. 8 Bar Chart of Malaysian Citizens' Emotions Based on Group of Months from 2020 to 2021

Figure 8 illustrates a bar chart of Malaysian citizens' emotions based on a group of months from 2020 to 2021. From October through December 2020, the anger emotion is predicted to have the highest percentage than in other months. This might be due to an issue which occurred during this period. On the 26<sup>th</sup> September 2020, the Sabah state elections were held. Although the Election Commission of Malaysia has established certain SOP, there are still individuals who do not follow the SOP, such as not maintaining physical distance and being in crowded areas. Since then, Sabah saw a rapid increase of COVID-19 cases. According to Berita Harian Online (2020), the government has admitted that the rise of COVID-19 cases in Malaysia is due to the Sabah state elections. This issue has provoked anger among Malaysians, resulting in the highest percentage rate of anger sentiment during this time period compared to other months.



### C. Sentiment Analysis Evaluation

The sentiment analysis developed during the development stage is evaluated in order to measure the effectiveness and performance of the classifier. The sentiment analysis is evaluated using standard sentiment analysis evaluation, which is accuracy, precision, recall, and F1-score with the help of the confusion matrix. The value of precision, recall and F1-score of a developed sentiment classifier is important in preparation for the calculation of the overall accuracy of the classifier. Precision and recall are improved calculations to measure a sentiment analysis classifier in comparison to basic accuracy calculation.

Confusion matrix or also known as the error matrix is commonly used to calculate and visualize a machine learning classifier model performance. It is a summarization of the prediction result gathered during the testing phase of the sentiment analysis development where the classifier is used to predict the sentiment of a test dataset where the true value is already known. By using the confusion matrix, it is easier to observe the number of correct and incorrect predictions for each class. Therefore, the confusion matrix can be used as a reference for the calculation of precision, recall, and F1-score of the developed sentiment analysis model. The size of the table depends on the number of classes used in the model. For example, a model with three classes produces a table with a size of 3x3. The sentiment classifier that was developed for this application has a confusion matrix with a size of 6x6 as this classifier model consists of six classes.

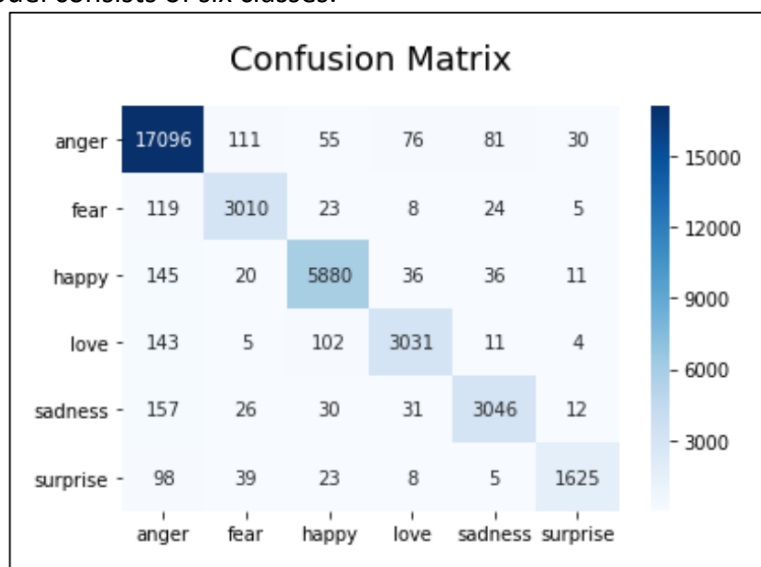


Fig. 9 The Confusion Matrix for Sentiment Analysis Model

The confusion matrix shown in Figure 9 visualizes the number of predictions for each class where each row represents the actual number of test case for each class and each column represent the predicted number for each class. Thus, each diagonal cell in the confusion matrix represents the true positive for each class where the classifier made correct predictions. In order for the sentiment analysis model to achieve high accuracy, these diagonal cells need to have a value that is higher than its adjacent cells. The sentiment analysis model for this project is tested using 35162 total cases which are split from the original dataset during the development phase. Based on Figure 9, each class has a high value of true positive with anger having the highest value of 17096. These show that the classifier model is going to have higher accuracy in predicting tweets with all sentiment and have a higher probability in predicting tweets with all sentiment as well.

**Conclusion**

Sentiment analysis is a computational and natural language processing-based approach for analysing people's sentiments, emotions, and attitudes in a given text (Beigi, Hu, Maciejewski, & Liu, 2016) and an essential method in social media research. This sentiment analysis work was based on a machine learning model for predicting Malaysians' emotions from Malay language and English tweets. This Twitter data analysis can be used to roughly indicate the public's perception and emotions during the COVID-19 pandemic. Public health agencies and the government should be aware that Twitter data may be used to examine public sentiment and knowledge levels regarding the COVID-19 pandemic. It is important to note that the levels of public emotions are dynamic, as presented in previous bar graphs. The findings also show that people use social media platforms to express dissatisfaction about current issues or express anger as well as exhibit negative emotions at various times.

From the Twitter data, it can be observed that people are trying to convey their dissatisfaction with government policies and with several issues that have arisen during the pandemic. For instance, how can the government respond to the discrepancy in SOP enforcement during MCO, to ensure public safety? This emotion analysis can help the government in identifying the needs of people throughout the pandemic period. For instance, online shopping has been found to reduce the stress of people who cannot leave the house due to MCO. Therefore, what initiatives can the government take? This study however, contains several limitations. Firstly, this study was conducted between 2020 and 2021, during the pandemic's early stages. Due to this, the overall picture of the pandemic is rather limited. Therefore, research that focuses on public sentiment following the aforementioned time may offer insightful data for comparison. Due to the abundance of information on COVID-19 on social media, it could be challenging to pinpoint the best sources of information. Only Twitter was used in this study to obtain data; future studies should also include mass media and other data sources in addition to social media data.

The current research makes a significant contribution to the existing knowledge by providing insights into Malaysians' sentiments and emotions regarding the COVID-19 pandemic. It uses a machine learning model to predict emotions from Malay language and English tweets, providing a snapshot of the public's emotions and perceptions during the pandemic. This work sheds light on the importance of social media data in examining public sentiment and knowledge levels during a health crisis. It highlights the significance of sentiment analysis as a computational approach to extract valuable information from vast amounts of social media data. This research could be used as a tool for public health agencies and the government to better understand the public's needs during the pandemic period. Furthermore, the findings reveal the need for the government to respond to the public's dissatisfaction and concerns regarding pandemic policies, such as discrepancies in standard operating procedure enforcement. Finally, the limitations of the study also suggest avenues for future research to explore public sentiment following the pandemic's early stages and to include other data sources in addition to social media data.

**Acknowledgement**

We would like to express our thanks to Universiti Teknologi MARA (UiTM) for providing the funding and facilities for this study.

**Corresponding Author**

Nahdatul Akma Ahmad

College of Computing, Informatics, and Media, Universiti Teknologi MARA, Perak Branch,  
Tapah Campus, 35400 Tapah Road, Perak, Malaysia  
Email: nahdatul@uitm.edu.my

## References

- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... & Cheng, Z. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, 395(10223), 497-506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
- TheSunDaily. (2020). 48% of nation's Covid-19 cases linked to Sri Petaling tabligh event. *The Sun Daily*. <https://www.thesundaily.my/local/48-of-nation-s-covid-19-cases-linked-to-sri-petaling-tabligh-event-HM2428393>
- Rahim, R. N. (2020). Covid-19: Malaysia tertinggi di Asia Tenggara [Covid-19: Malaysia highest in Southeast Asia]. *Harian Metro*. <https://www.hmetro.com.my/utama/2020/03/554843/covid-19-malaysia-tertinggi-di-asia-tenggara>
- Bunyan, J. (2020). PM: Malaysia under movement control order from Wed until March 31, all shops closed except for essential services. *Malaymail*. <https://www.malaymail.com/news/malaysia/2020/03/16/pm-malaysia-in-lockdown-from-wed-until-march-31-all-shops-closed-except-for/1847204>
- Lyu, J. C. (2012). How young Chinese depend on the media during public health crises? A comparative perspective. *Public Relations Review*, 38(5), 799–806. <https://doi.org/10.1016/j.pubrev.2012.07.006>
- Liu, I. L. B., Cheung, C. M. K., & Lee, M. K. O. (2010). Understanding twitter usage: What drive people continue to tweet. In PACIS 2010 - 14th Pacific Asia Conf. Inf. Syst. (pp. 928–939).
- Soriano, C. R., Roldan, M. D. G., Cheng, C., & Oco, N. (2016). Social media and civic engagement during calamities: the case of Twitter use during typhoon Yolanda. *Philippine Political Science Journal*, 37(1), 6–25. <https://doi.org/10.1080/01154451.2016.1146486>
- Van Lent, L. G. G., Sungur, H., Kunneman, F. A., Van De Velde, B., & Das, E. (2017). Too far to care? Measuring public attention and fear for Ebola using Twitter. *Journal of Medical Internet Research*, 19(6). <https://doi.org/10.2196/jmir.7219>
- Nair, M. R., Ramya, G. R., & Sivakumar, P. B. (2017). Usage and analysis of Twitter during 2015 Chennai flood towards disaster management. *Procedia Computer Science*, 115, 350–358. <https://doi.org/10.1016/j.procs.2017.09.089>
- Fu, K.-W., Liang, H., Saroha, N., Tse, Z. T. H., Ip, P., & Fung, I. C.-H. (2016). How people react to Zika virus outbreaks on Twitter? A computational content analysis. *American Journal of Infection Control*, 44(12), 1700–1702. <https://doi.org/10.1016/j.ajic.2016.04.253>
- Ashton, K., Bellis, M. A., Davies, A. R., Hughes, K., & Winstock, A. (2017). Do emotions related to alcohol consumption differ by alcohol type? An international cross-sectional survey of emotions associated with alcohol consumption and influence on drink choice in different settings. *BMJ Open*, 7(10), e016089. <https://doi.org/10.1136/bmjopen-2017-016089>
- Litwin, R., Goldbacher, E. M., Cardaciotto, L., & Gambrel, L. E. (2017). Negative emotions and emotional eating: The mediating role of experiential avoidance. *Eating and Weight Disorders - Studies on Anorexia, Bulimia and Obesity*, 22(1), 97–104.

<https://doi.org/10.1007/s40519-016-0301-9>

- Kirwan, M., Pickett, S. M., & Jarrett, N. L. (2017). Emotion regulation as a moderator between anxiety symptoms and insomnia symptom severity. *Psychiatry Research*, 254, 40–47. <https://doi.org/10.1016/j.psychres.2017.04.028>
- Na, K., Garrett, R. K., & Slater, M. D. (2018). Rumor acceptance during public health crises: Testing the emotional congruence hypothesis. *Journal of Health Communication*, 23(8), 791–799. <https://doi.org/10.1080/10810730.2018.1527877>
- Pizzoli, S. M. F., Marzorati, C., Mazzoni, D., & Pravettoni, G. (2020). An Internet-Based Intervention to Alleviate Stress During Social Isolation With Guided Relaxation and Meditation: Protocol for a Randomized Controlled Trial. *JMIR Research Protocols*, 9(6), e19236. <https://doi.org/10.2196/19236>
- Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/1500000011>
- Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An Overview of Sentiment Analysis in Social Media and Its Applications in Disaster Relief. In *Sentiment Analysis and Ontology Engineering* (pp. 313–340). Springer International Publishing. [https://doi.org/10.1007/978-3-319-30319-2\\_13](https://doi.org/10.1007/978-3-319-30319-2_13)
- Ekman, P. (2005). Basic Emotions. In *Handbook of Cognition and Emotion* (pp. 45–60). John Wiley & Sons, Ltd.