Department of Computer and Communication Systems Engineering

Faculty of Engineering

Universiti Putra Malaysia

43400 UPM Serdang Selangor

FINAL PROJECT REPORT

SENTIMENT ANALYSIS ON TWITTER: COVID-19 IN MALAYSIA

Course : ECC4306 Artificial Intelligence

Semester : 2 (2020/2021)

Lecturer : Prof. Madya Dr. Syamsiah bt. Mashohor

Students' Name : Muhammad Hazim Izzat Bin Zamri (194203)

Alif Zakuan Bin Yuslaimi (193258)

TABLE OF CONTENTS

1.0 BACKGROUND

Sentiment analysis is one of the most common Natural Language Processing (NLP) application. With sentiment analysis, understanding what people think about a subject by analyzing their posted tweets is no longer complicated.

NLP is an artificial intelligence topic which deals with computer-human interaction in natural language. NLP's ultimate goal is to read, decipher, interpret and grasp the human languages in a useful way. Most NLP techniques use machine learning to derive significance from human language.

In this project, we will be going to use Tweepy, which is an easy-to-use Python library for accessing the Twitter API. However, we will need to have a Twitter developer account and sample codes to do this analysis.

Tweepy provides access to the well documented Twitter API. With Tweepy, it is possible to get any object and use any method that the official Twitter API offers.

The aim of this project is to analyze what people think about the Covid-19 situation in Malaysia.

2.0 LITERATURE REVIEW

In this section several approaches on twitter sentimental analysis are reviewed. Opinion mining and sentiment analysis are advancing at a breakneck pace. The purpose of this mechanism is to extract text from any sentence. This data is extracted from a variety of social media platforms, including Twitter, LinkedIn, and Instagram. This issue can be resolved using a machine learning algorithm.

2.1 Algorithm Accuracy

In the first paper, Nirag T. Bhatt [1] conduct a survey on using Machine Learning technique to perform sentiment analysis. Based on the study, machine learning algorithm being compared to get result on which algorithm perform better in different features. The accuracy on each algorithm being tested. Other than that, algorithm such as Naïve Bayes, Logistic Regression, SVM Algorithm, Random Forest and Decision Tree are being compared to measure the accuracy each of the algorithm [2].

Method	TF-IDF Feature	Accuracy (%)
SVM	Yes	**84.04**
	No	83.23
Logistic Regression	Yes	83.15
	No	82.74
Multinomial Naïve Bayes	Yes	82.13
	No	82.08

Figure 1 Traditional machine learning algorithms [3]

2.2 Twitter Sentimental Analysis Approaches

O. Adwan et. al. [4] stated in the survey article, emerging Twitter Sentimental Analysis (TSA) approaches that combine machine learning and lexicon-based approaches with cognitive science, semantic web, and big data theories and technologies. Example of TSA approaches are big data platform, semantic Web technologies, based on cognitive science theories and Social Network Analysis (SNA) metrics.

These techniques also being discussed in article [5]. There are supervised Machine Learning, Ensemble and Lexicon-based methods approaches. All his methods has been discussed thoroughly including the features, algorithm being used and the outcomes.

2.3 Data Collection through using Tweepy library

S. Cahyaningtyas et. al. [3] using Tweepy library to extract all relevant data in twitter. According to the article, 115,931 raw data being obtained using tweepy provided by Python. Besides having tweepy as the library, to get or extract the data from twitter, the user needs to have twitter API. In order to get the API, the user must create twitter account and request for the API thus twitter will provide consumer key, consumer secret and access token [1].

2.4 Sentiment Classification

After determining if a sentence is opinionated, we must determine the statement's polarity. In sentiment analysis, there are three classification such as positive sentiments, negative sentiments and neutral sentiments. Depending on the sentiment analysis application, the subtasks of opinion holder extraction and object feature extraction may be considered unnecessary.

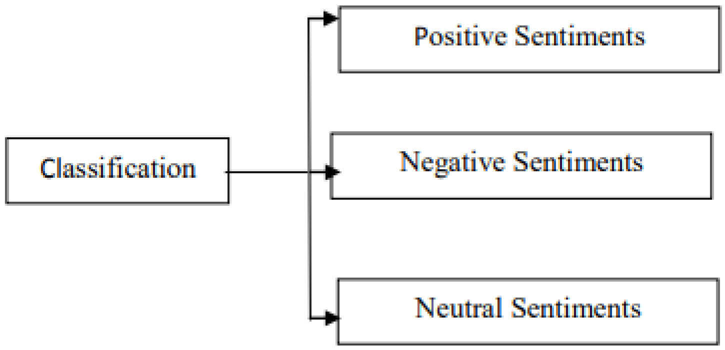


Figure 2 Sentiment Categorization [2]

P. Pandey et. al. [2] stated that sentiment analysis very important in helping with product analysis. Any statement that has a greater number of positive words it will be considered as a positive tweet.

2.5 Summary

Based on the reviewed literature, several studies have been conducted to test the hypothesis of the chosen performance capability and design emphasis. All this reference material sometimes increases the information about Twitter sentiment analysis and its test results that prove it useful.

To sum up, this project will focus on the classification of positive, negative and neutral sentiment regarding Covid-19 in Malaysia. Besides, an improvement on the system can be implemented compared to previous research.

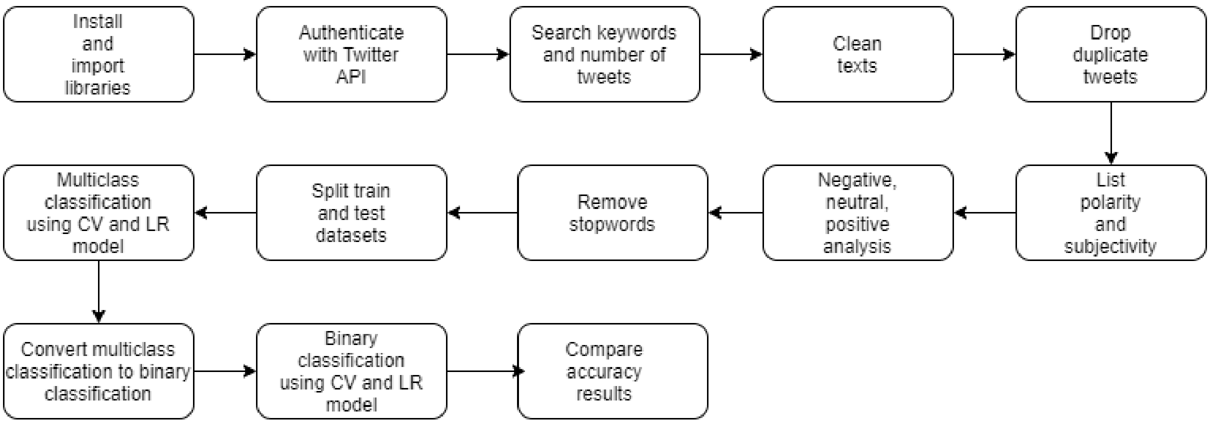
3.0 PROBLEM STATEMENT

1. Manually extract a sentence, read and analyze tweet by tweet and organize them into understandable format.
2. Difficulty to detect sentiment with inappropriate English.

4.0 OBJECTIVES

1. To Implement analysis using an algorithm to derive the sentiment distribution of Covid-19 in Malaysia.
2. To determine the positive, negative or neutral sentiments towards subject of interest.

5.0 METHODOLOGY



The first step is to install and import the required libraries. Before carrying out the analysis, we would need to install Textblob and Tweepy libraries using !pip install command on the terminal. the libraries will be imported for this sentiment analysis project. A Twitter Developer Account is required for authentication. Getting Twitter developer account usually takes a day or two, or sometimes more, for the application to be reviewed by Twitter.

After the authentication for Twitter API is successful, we will use Tweepy to get text and use Textblob to calculate positive, negative, neutral, polarity and compound parameters from the text.

The next step is to obtain Tweets with a keyword or a hashtag. For example, the scenario is that, the user should type keyword or hashtag (for example, lockdown malaysia) and type how many tweets that they want to fetch and analyse. After getting a specified amount tweets about “lockdown malaysia”, we can have a look at the number of tweets with different sentiments.

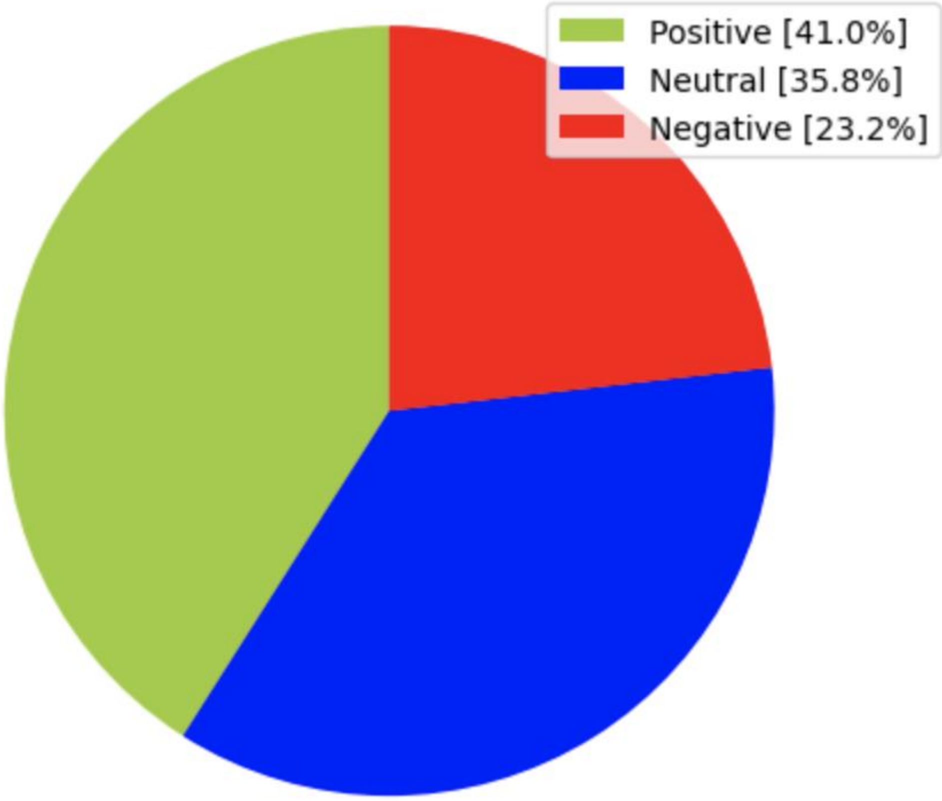


Figure 3 Pie chart example of sentiments

The Tweets will have to be cleaned to perform the sentiment analysis. When we have a look at the Tweet list we can see some duplicated tweets, and so we will need to drop duplicates records using *drop_duplicates* function. The texts can be cleaned by using lambda function and clean RT, link, punctuation characters and finally convert to lowercase.

Finally, we can perform the sentiment analysis. First, a new data frame (*tw_list*) and a new feature (*text*) will be created. The cleaned texts can be used to calculate polarity, subjectivity, sentiment, negative, positive, neutral and compound parameters. For all calculated parameters, new features were created to the data frame. The data frame can be split into 3 groups based on sentiment. For this project, 3 new data frames will be created (*tw_list_negative, tw_list_positive, tw_list_neutral*) and import from original *tw_list* data frame

A chart can be created by using a number of sentiment tweets. A word cloud can also be created using the tweets, so we can visually observe which words most used in these tweets.

Figure 4 Word cloud example

Applying count vectorizer provides the capability to preprocess the text data prior to generating the vector representation making it a highly flexible feature representation module for text.

Applying stemmer provides the root of words. Which means words that come from the same root can be eliminated, such as; connect, connection, connected, connections, connects comes from “connect”. By applying the stemmer function, all these words are considered the same.

We then can finally analyse sentiment using Tweets and we can observe which words are the most used and which words used together.

6.0 DATASETS

Data will be extract from Twitter using API. From there, Twitter will grant our group as developer account after finish reviewing the details. From there, consumer key, consumer secret and access token will be given to request authentication.

Figure 5 shows the data that been extracted from Twitter using Twitter authentication. A total of 5000 tweets of data being analyze for this project purpose.

1	user	location	tweets
2	M_Wayne_Brown	Bangkok, Thailand	RT @BangkokPostNews: Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the g
3	JoshLee75167117	Damansara, Selangor	@thevibesnews @malaysia_covid I really wish for Sabah gain another thousands cluster forthcoming. What a dumb decision
4	TerryDieHeiden	Hell ï¸â€ðŸ”¥	@farysafiqahs In relation to covid ke.. kalau covid tu alot of other countries are handling it better than Malaysiaâ€¦ https://t.co/PC
5	COVID_19Disease		#Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the governmentâ€¦ https://t
6	bsindia	India	#Malaysia will extend its current restrictions until the Covid-19 pandemic is further contained, Prime Ministerâ€¦ https://t.co/glGYB
7	BahangD	Malacca, Malaysia	RT @fmtoday: The Delta variant, which was detected in Malaysia recently, has caused a surge in cases in Kudus within a week. #FM
8	FiqhTabayyun	Kuantan, Pahang	RT @AlfonsoTheOne: Adopt â€˜jab first, register laterâ€™ approach to flatten Malaysiaâ€™s Covid-19 curve, govt toldðŸ'ðŸ'ðŸ'ð
9	AlfonsoTheOne		Adopt â€˜jab first, register laterâ€™ approach to flatten Malaysiaâ€™s Covid-19 curve, govt toldðŸ'ðŸ'ðŸ'ðŸ'ðŸ'ðŸ'ðŸ'ðŸ'ðŸ'ðŸ' http
10	fmtoday	Malaysia	The Delta variant, which was detected in Malaysia recently, has caused a surge in cases in Kudus within a week.â€¦ https://t.co/HGI
11	BangkokPostNews	Thailand	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government hâ€¦ https://
12	drmsdah	Kolkata	RT @MoHFW_INDIA: #LargestVaccineDrive #Unite2FightCorona
13	yimingwuzere	Malaysia	RT @Nicholas_Gomez: Struck off cosmetic doc as COVID advisor!? @MARiiMalaysia @MITIMalaysia can you clarify please?
14	saroki19	Malaysia	Selangor has the highest covid-19 cases. Logically you should allocate more vaccine to selangor. You put out the maâ€¦ https://t.cc
15	ravinsingh8	Pendatang	@drkalai61 In 2019, there were 173,746 deaths in Malaysia or 476 a day. So yesterday 60 Covid deaths equals 13% extâ€¦ https://t.
16	wawaggmu	Malaysia	Malaysia (based on total population)
17	kyodo_english	Tokyo, Japan	Malaysia extends COVID-19 lockdown indefinitely
18	ajimmonn	Alor Setar	RT @Nicholas_Gomez: Struck off cosmetic doc as COVID advisor!? @MARiiMalaysia @MITIMalaysia can you clarify please?
19	asiaonecom	Singapore	As Malaysia's Covid-19 vaccination drive revs up, sceptics threaten to spoil the party https://t.co/oxMUvZFKwo https://t.co/iCPuoL
20	yeshaallya	Malaysia	Not to defend this government, but itâ€™s literally there if you would just Google it
21	syahiirimran	Malaysia	I just realised that next sem is my last year already. LIKE WTF IS HAPPENING RN. Iâ€™M NOT PAYING THE UNI AND GRADUATâ€¦ h
22	SAMLgk1		@malaysia_covid @syahredzan Is it true the vaccination center in POLIMAS Jitra is closed for 6 days due to vaccine stock shortage?
23	ejatghani	N-N-M	RT @drmusanordin: This 4K figure is indefensible.
24	Rob23527077	Malaysia	@malaysia_covid Bro. Is ok to me since closed for few days. I rather they finished it fast then keep in storage.
25	MyMedicalRsrc	Malaysia	Health minister says almost 7.4 million Covid-19 vaccine doses administered as of Sunday | Malay Mail https://t.co/n15VdnRORK
26	tortpace		@malaysia_covid Last week, nurse in a clinic kesihatan said they stopped inoculation as their vaccines were resentâ€¦ https://t.co/
27	ernamh	Kuala Lumpur, Malaysia	@brandon_pat91 Maybe Google. It's already in the papers what they're doing or at least trying ðŸ™„ Sabah has to deferâ€¦ https://
28	LaiKH5	Singapore	@malaysia_covid Letâ€™s see what will KJ said again,
29	zhuge8f	Malaysia	RT @azmi_radarsaham: World Bank has cut its forecast on Malaysia's GDP from 6.9% to 4.5% due to resurgence in Covid 19 infecti
30	MoJamesApted	Thailand - Madagascar - Egypt	RT @NikkeiAsia: CHINA'S COVID-19 VACCINES | Indonesia orders 58.5% of its vaccines from Chinese pharmaceuticals while Malays
31	AmitDeb52112501	Tripura, India	RT @BJYMNABADAL: #LargestVaccineDrive #Unite2FightCorona

Figure 5 Data frame with features such as user, location and tweets

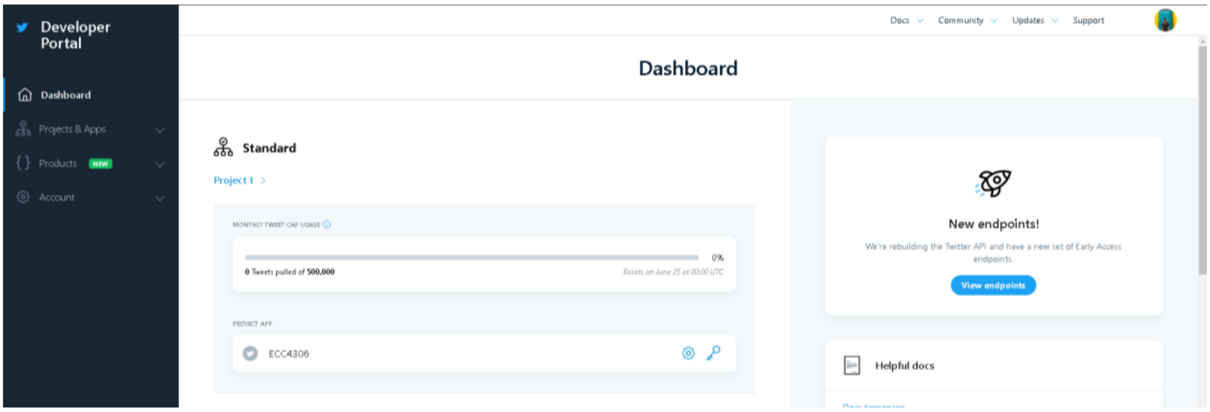


Figure 6 Dashboard of Developer Account

7.0 RESULTS AND DISCUSSION

	user	location	tweets
0	M_Wayne_Brown	Bangkok, Thailand	RT @BangkokPostNews: Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government has…
1	JoshLee75167117	Damansara, Selangor	@thevibesnews @malaysia_covid I really wish for Sabah gain another thousands cluster forthcoming. What a dumb decision
2	TerryDieHeiden	Hell 🔥	@farysafiqahs In relation to covid ke.. kalau covid tu alot of other countries are handling it better than Malaysia… https://t.co/POS8mUwRb1
3	COVID_19Disease	NaN	#Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government… https://t.co/RhgcRfwdBK
4	bsindia	India	#Malaysia will extend its current restrictions until the Covid-19 pandemic is further contained, Prime Minister… https://t.co/glGYB6fp9M
...	...	...	...
4995	awan_biru89	Kuala Lumpur	RT @Soya_Cincau: Malaysia will allow COVID-19 vaccines to be sold via the private market soon. Khairy said, "If you want to go fast, you pa…
4996	syedalaqib	NaN	RT @malaymail: JKJAV: Covid-19 vaccine administration above 250,000-mark for three days in a row https://t.co/rt6pxYvklH
4997	broz6969	NaN	RT @malaysiakini: MP SPEAKS | Malaysia will never win the war against Covid-19 if… https://t.co/115usvcOUl
4998	jasze36	Malaysia	RT @malaymail: JKJAV: Covid-19 vaccine administration above 250,000-mark for three days in a row https://t.co/rt6pxYvklH
4999	_musangPutih	ARMY Nation	RT @5hdisaster: Malaysia's government are always on about LGBT than dealing with the pandemic and it really shows their inability to focus…

5000 rows × 3 columns

Figure 7 Original fetched tweets

	user	location	tweets
0	M_Wayne_Brown	Bangkok, Thailand	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government has…
1	JoshLee75167117	Damansara, Selangor	_covid I really wish for Sabah gain another thousands cluster forthcoming. What a dumb decision
2	TerryDieHeiden	Hell 🔥	In relation to covid ke.. kalau covid tu alot of other countries are handling it better than Malaysia…
3	COVID_19Disease	NaN	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government…
4	bsindia	India	Malaysia will extend its current restrictions until the Covid-19 pandemic is further contained, Prime Minister…
...	...	...	...
4995	awan_biru89	Kuala Lumpur	_Cincau Malaysia will allow COVID-19 vaccines to be sold via the private market soon. Khairy said, "If you want to go fast, you pa…
4996	syedalaqib	NaN	JKJAV Covid-19 vaccine administration above 250,000-mark for three days in a row
4997	broz6969	NaN	MP SPEAKS | Malaysia will never win the war against Covid-19 if…
4998	jasze36	Malaysia	JKJAV Covid-19 vaccine administration above 250,000-mark for three days in a row
4999	_musangPutih	ARMY Nation	Malaysia's government are always on about LGBT than dealing with the pandemic and it really shows their inability to focus…

5000 rows × 3 columns

Figure 8 Cleaned tweets

	user	location	tweets
0	M_Wayne_Brown	Bangkok, Thailand	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government has…
1	JoshLee75167117	Damansara, Selangor	_covid I really wish for Sabah gain another thousands cluster forthcoming. What a dumb decision
2	TerryDieHeiden	Hell 🔥	In relation to covid ke.. kalau covid tu alot of other countries are handling it better than Malaysia…
3	COVID_19Disease	NaN	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government…
4	bsindia	India	Malaysia will extend its current restrictions until the Covid-19 pandemic is further contained, Prime Minister…
...	...	...	...
4980	malaymail	Malaysia	Nearly 110,000 receive first Covid jab as vaccination gathers pace in Sibu
4987	LowElsielow	NaN	PN will ruin Malaysia before Covid does.KerajaanGagal
4988	codebluenews	Kuala Lumpur	Malaysia achieved a daily high record of 268,604 Covid-19 doses administered nationwide on 24 June.Malaysia reach…
4993	NewsBFM	Kuala Lumpur	Federal Territories Minister Annuar Musa says people are seemingly unafraid of Covid-19, which he believes is an in…
4996	syedalaqib	NaN	JKJAV Covid-19 vaccine administration above 250,000-mark for three days in a row

1255 rows × 3 columns

Figure 9 Drop duplicate tweets

	user	location	tweets	Subjectivity	Polarity
0	M_Wayne_Brown	Bangkok, Thailand	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government has...	0.000000	0.0000
1	JoshLee75167117	Damansara, Selangor	_covid I really wish for Sabah gain another thousands cluster forthcoming. What a dumb decision	0.350000	-0.0875
2	TerryDieHeiden	Hell 🔥	In relation to covid ke.. kalau covid tu alot of other countries are handling it better than Malaysia...	0.437500	0.1875
3	COVID_19Disease	NaN	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government...	0.000000	0.0000
4	bsindia	India	Malaysia will extend its current restrictions until the Covid-19 pandemic is further contained, Prime Minister...	0.450000	0.0000
...	...	...	...	...	...
4980	malaymail	Malaysia	Nearly 110,000 receive first Covid jab as vaccination gathers pace in Sibu	0.366667	0.1750
4987	LowElsielow	NaN	PN will ruin Malaysia before Covid does.KerajaanGagal	0.000000	0.0000
4988	codebluenews	Kuala Lumpur	Malaysia achieved a daily high record of 268,604 Covid-19 doses administered nationwide on 24 June.Malaysia reach...	0.270000	0.0800
4993	NewsBFM	Kuala Lumpur	Federal Territories Minister Annuar Musa says people are seemingly unafraid of Covid-19, which he believes is an in...	0.000000	0.0000
4996	syedalaqib	NaN	JKJAV Covid-19 vaccine administration above 250,000-mark for three days in a row	0.100000	0.0000

1255 rows × 5 columns

Figure 10 Polarity and Subjectivity

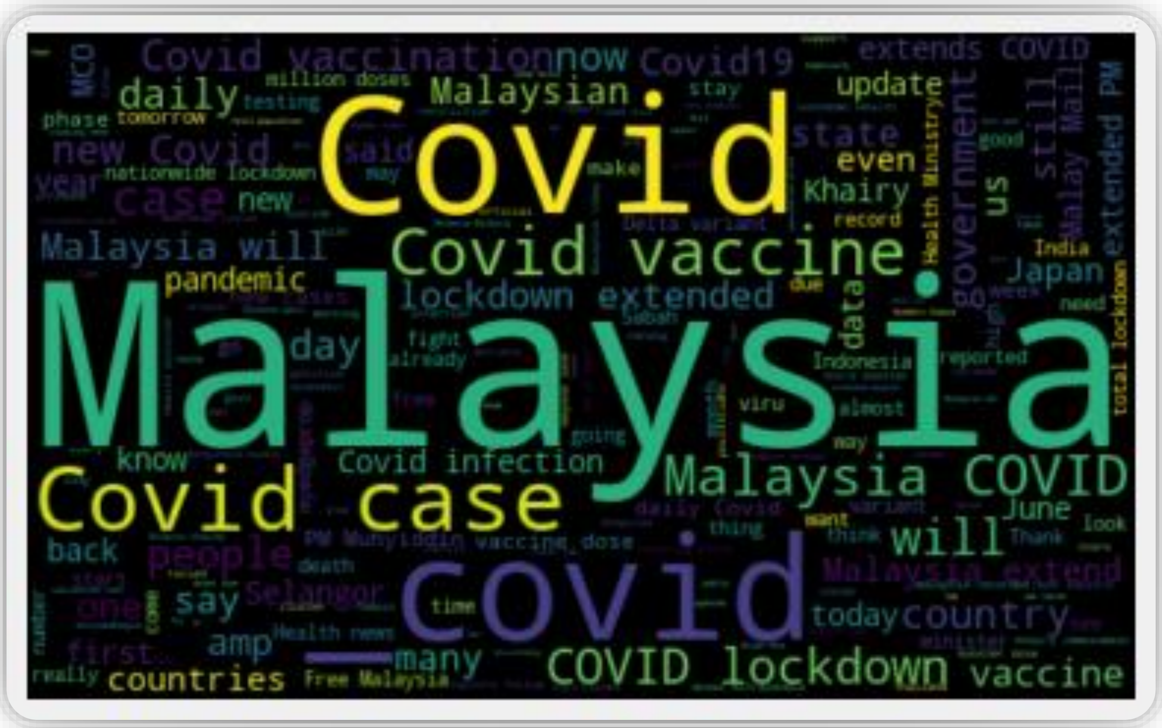


Figure 11 Wordcloud

	user	location	tweets	Subjectivity	Polarity	Sentiments
0	M_Wayne_Brown	Bangkok, Thailand	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government has...	0.000000	0.0000	Neutral
1	JoshLee75167117	Damansara, Selangor	_covid I really wish for Sabah gain another thousands cluster forthcoming. What a dumb decision	0.350000	-0.0875	Negative
2	TerryDieHeiden	Hell 🔥	In relation to covid ke.. kalau covid tu alot of other countries are handling it better than Malaysia...	0.437500	0.1875	Positive
3	COVID_19Disease	NaN	Malaysia's nationwide lockdown that was scheduled to expire Monday is being extended indefinitely, the government...	0.000000	0.0000	Neutral
4	bsindia	India	Malaysia will extend its current restrictions until the Covid-19 pandemic is further contained, Prime Minister...	0.450000	0.0000	Neutral
...	...	...	...	...	...	...
4980	malaymail	Malaysia	Nearly 110,000 receive first Covid jab as vaccination gathers pace in Sibu	0.366667	0.1750	Positive
4987	LowElsielow	NaN	PN will ruin Malaysia before Covid does.KerajaanGagal	0.000000	0.0000	Neutral
4988	codebluenews	Kuala Lumpur	Malaysia achieved a daily high record of 268,604 Covid-19 doses administered nationwide on 24 June.Malaysia reach...	0.270000	0.0800	Positive
4993	NewsBFM	Kuala Lumpur	Federal Territories Minister Annuar Musa says people are seemingly unafraid of Covid-19, which he believes is an in...	0.000000	0.0000	Neutral
4996	syedalaqib	NaN	JKJAV Covid-19 vaccine administration above 250,000-mark for three days in a row	0.100000	0.0000	Neutral

1255 rows × 6 columns

Figure 12 Sentiment Analysis

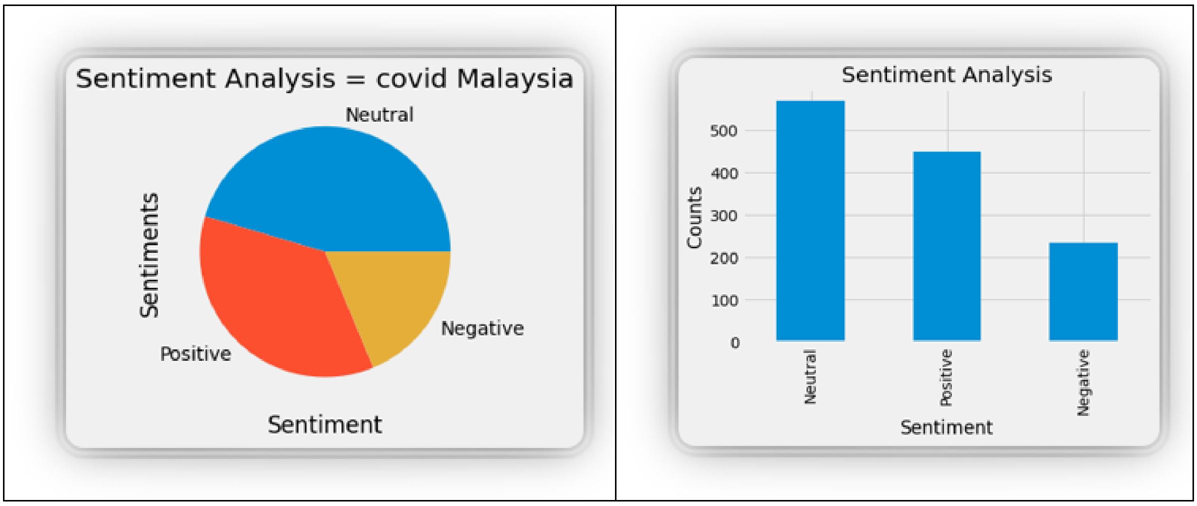


Figure 13 Pie and Bar chart for sentiment distribution

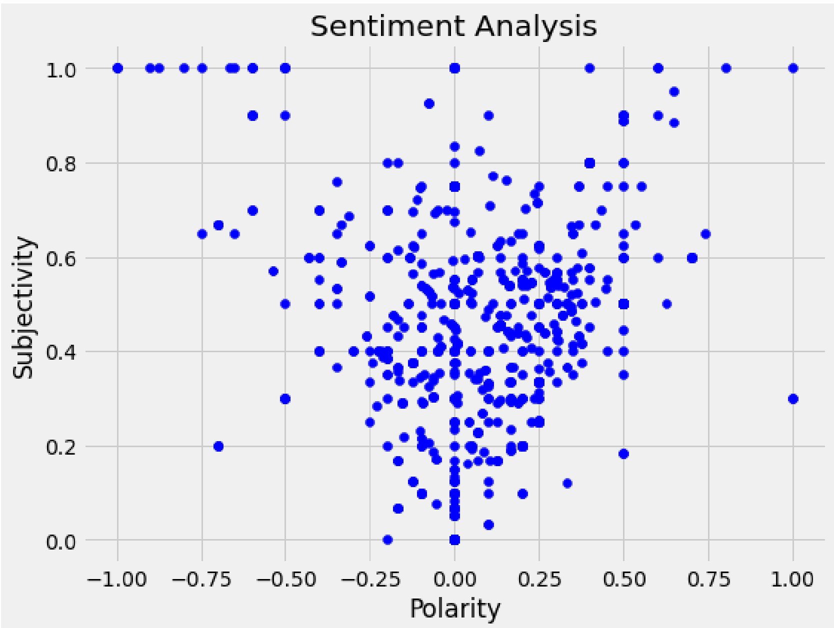


Figure 14 Plotted graph for sentiment distribution

Training accuracy Score   : 0.99800796812749
Validation accuracy Score : 0.852589641434263
number of Iteration : [60]

	precision	recall	f1-score	support
Negative	0.77	0.90	0.83	40
Neutral	0.92	0.81	0.86	130
Positive	0.81	0.90	0.85	81
accuracy			0.85	251
macro avg	0.83	0.87	0.85	251
weighted avg	0.86	0.85	0.85	251

Figure 15 Classification report for Multiclass classification

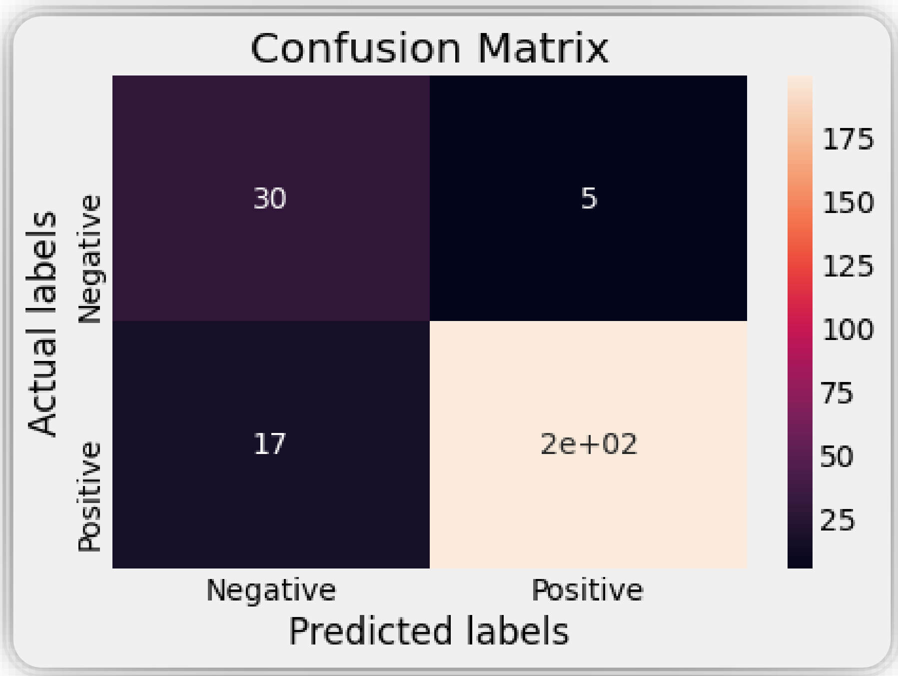


Figure 16 Confusion Matrix for Binary classification

Training accuracy Score   : 0.9950199203187251
Validation accuracy Score : 0.9123505976095617
number of Iteration : [29]

	precision	recall	f1-score	support
0	0.64	0.86	0.73	35
1	0.98	0.92	0.95	216
accuracy			0.91	251
macro avg	0.81	0.89	0.84	251
weighted avg	0.93	0.91	0.92	251

Figure 17 Classification report for Binary classification

Table 1 Multiclass Classification accuracy analysis

Training Accuracy	Validation Accuracy	No. of Iteration
99.80%	85.25%	60

Table 2 Binary classification accuracy analysis

Training Accuracy	Validation Accuracy	No. of Iteration	Precision Accuracy	Recall Accuracy
99.50%	91.23%	23	98%	92%

Figure 7 shows the original fetched tweets using tweepy. There was a total of 5000 tweets fetched with the keyword covid and Malaysia. After cleaning and dropping duplicated tweets, we are left with only 1255 tweets as in Figure 9.

Figure 10 displays the polarity and subjectivity of each individual tweets. Subjectivity refers to how much the tweet is related to the specified keywords with the range from 0 to 1. The higher the value, the more relatable the tweet is with the specified keywords. Polarity ranges from -1 to 1. Polarity indicates whether the text carries a positive or negative sentiment. Negative value means negative sentiment and positive value indicates positive sentiment as in Figure 12.

Figure 13 illustrates the pie chart and bar chart of the sentiment analysis results. As we can see, neutral sentiments dominate the chart. It can be caused by low subjectivity value of the tweets which lands them at neutral polarity value as shown in the plotted graph in Figure 14.

Figure 16 illustrates the confusion matrix for binary classification which is useful for quickly calculating precision and recall given the predicted labels from a model. A confusion matrix for binary classification shows the four different outcomes: true positive, false positive, true negative, and false negative. The actual values form the columns, and the predicted values (labels) form the rows. The intersection of the rows and columns show one of the four outcomes. For example, if we predict a data point is positive, but it actually is negative, this is a false positive.

Going from the confusion matrix to the recall and precision requires finding the respective values in the matrix and applying the equations:

$$recall = \frac{true\ positives}{true\ positives\ +false\ negatives} \qquad precision = \frac{true\ positives}{true\ positives + false\ positives}$$

Recall is the ability of a classification model to identify all relevant instances. From the confusion matrix in Figure 16, we can calculate the recall using the formula:

Recall = 199 / (199 + 5) = 0.9212

Precision is the ability of a classification model to return only relevant instances. From the confusion matrix in Figure 16, we can calculate the precision using the formula:

Precision = 199/ (199 + 5) = 0.9755

In some situations, we might know that we want to maximize either recall or precision at the expense of the other metric. For example, in preliminary disease screening of patients for follow-up examinations, we would probably want a recall near 1.0 — we want to find all patients who actually have the disease — and we can accept a low precision if the cost of the follow-up examination is not significant. However, in cases where we want to find an optimal blend of precision and recall we can combine the two metrics using what is called the F1 score.

The F1 score is the harmonic mean of precision and recall taking both metrics into account in the following equation:

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

By using the formula and the calculated precision and recall:

F_1 = 2*(0.9212*0.9755)/ (0.9212+0.9755) = 0.9476

Between precision and accuracy, Precision is how sure we are of our true positives whilst recall is how sure we are that we are not missing any positives.

Recall should be chosen if the idea of false positives is far better than false negatives, in other words, if the occurrence of false negatives is unaccepted/intolerable, that we would rather get some extra false positives(false alarms) over saving some false negatives.

Precision should be emphasized if we want to be more confident of our true positives. For example, spam emails. We would rather have some spam emails in our inbox rather than some regular emails in our spam box. So, the email company wants to be extra sure that email Y is spam before they put it in the spam box and you never get to see it.

In our use to detect negative sentiments, we should focus on precision reading.

8.0 LIMITATIONS

This project has several drawbacks. To begin, it is worth mentioning that this study examined the trends and frequency of terms associated with COVID-19. The list of keywords specified may have been insufficient. The keywords employed in this study can be expanded to include twitter searches by merging COVID-19 and its symptoms-related keywords. Additional research might be conducted to determine the most pertinent collection of keywords with a high degree of detail based on the number of tweets including symptoms and other keywords. Other than that, the tweets gathered for this study were in English, which may be a constraint for this project. As if want to focus on Malaysia only, there may be a lot of tweets in Bahasa Malaysia compared to English.

9.0 FUTURE IMPROVEMENT

Due to a lack of time and computing resources, several details were left for future works. It might be interesting to explore the following area:

- Can include another feature such as variable importance where it describes which features are relevant. Thus, it will help in better and lead to model improvements.
- Besides English, this project can consider on analyze other language such as Bahasa Malaysia or any multilingual analysis.
- This application is not only applicable to the Coronavirus health problem, but it may also be used as a model for determining sentiment emotion in future comparable instances.

10.0 CONCLUSIONS

This project aimed at analyzing the sentiments of people during Covid-19 pandemic in Malaysia being conducted. The two keywords are used in this analysis are 'covid' and 'Malaysia' for determining the subjectivity and polarity of each tweet. Based on NLP analysis, the sentiments such as positive, neutral and negative successfully been determined and classified using the polarity of each tweet. Moreover, this study offered an in-depth investigation of the sentiments and mindsets of individuals on Covid-19, enabling us to realize that people in Malaysia concern about the current situations.

11.0 REFERENCES

[1] N. T. Bhatt and A. J. Saket Swarndeep, (2020). "Sentiment Analysis using Machine Learning Technique: A Literature Survey," *Int. Res. J. Eng. Technol.*, 20

[2] P. Pandey, M. Santosh Mishra, and M. Devendra Rewarikar. (2020). "Real-Time Twitter Sentiment Analysis using Machine Learning using Different Classification Algorithm," *Int. Res. J. Eng. Technol.*, 2020, Accessed: Jun. 10, 2021. [Online]. Available: www.irjet.net.

[3] O. Y. Adwan, M. Al-Tawil, A. M. Huneiti, R. A. Shahin, A. A. Abu Zayed, and R. H. Al-Dibsi. (2020). "Twitter sentiment analysis approaches: A survey," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 15, pp. 79–93, 2020, doi: 10.3991/ijet.v15i15.14467.

[4] A. Alsaeedi and M. Z. Khan. (2021). "A Study on Sentiment Analysis Techniques of Twitter Data," 2019. Accessed: Jun. 10, 2021. [Online]. Available: www.ijacsa.thesai.org.

[5] AF. Hidayatullah. (2020). "Sentiment Analysis on Twitter using Neural Network: Indonesian Presidential Election 2019 Dataset," doi: 10.1088/1757-899X/1077/1/012001.

[6] "Step by Step: Twitter Sentiment Analysis in Python | by Yalin Yener | Towards Data Science." https://towardsdatascience.com/step-by-step-twitter-sentiment-analysis-in-python-d6f650ade58d (accessed Jun. 10, 2021).

[7] "Beyond Accuracy: Precision and Recall | by Will Koehrsen | Towards Data Science." https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c (accessed Jul. 2, 2021).

[8] "Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on? | by Salma Ghoneim | Towards Data Science." https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124 (accessed Jul. 2, 2021).