International Journal of Modern Trends in Social Sciences
EISSN: 2600-8777
IJMTSS
Volume 4 Issue 17 (September 2021) PP. 01-08
DOI 10.35631/IJMTSS.417001

INTERNATIONAL JOURNAL OF MODERN TRENDS IN SOCIAL SCIENCES (IJMTSS)
www.ijmtss.com

COVID-19 SENTIMENT ANALYSIS ON FACEBOOK COMMENTS

Sajidah Ibrahim[1*], Nor Zairah Ab Rahim[2], Fajar Ibnu Fatihan[3], Nur Azaliah Abu Bakar[4]

[1] Razak Faculty of Technology and Informatics (UTM), Malaysia
Email: sajidah@graduate.utm.my
[2] Razak Faculty of Technology and Informatics (UTM), Malaysia
Email: nzairah@utm.my
[3] Razak Faculty of Technology and Informatics (UTM), Malaysia
Email: ibnufatihan@graduate.utm.my
[4] Razak Faculty of Technology and Informatics (UTM), Malaysia
Email: azaliah@utm.my
[*] Corresponding Author

Article Info:

Article history:
Received date:14.01.2021
Revised date: 25.01.2021
Accepted date: 19.07.2021
Published date: 09.09.2021

To cite this document:

Ibrahim, S., Ab Rahim, N. Z., Fatihan, F. I., & Abu Bakar, N. A. (2021). Covid-19 Sentiment Analysis on Facebook Comments. *International Journal of Modern Trends in Social Sciences, 4* (17), 01-08.

DOI: 10.35631/IJMTSS.417001

Abstract:

Malaysia recorded its first COVID-19 case on 9th March 2020 and recorded a total of 59,817 by end of November 2020. Buzz in social media over COVID-19 and measures by the Government to curb infection spread among citizens. The study aims to understand Malaysian public awareness and perception of COVID-19 related issues on Facebook during the 1st and 2nd week of Movement Control Order (MCO). Data mining was conducted on DG Tan Sri Noor Hisham Abdullah's official Facebook account user comments and a total of 77,351 comments was collected between 18 March and 14 April 2020. The analyses included data pre-processing and sentiment analysis to identify and explore sentiments in discussion topics within the first two weeks of lockdown. The results yield majority of comments are in the Malay language and mix languages of English and Malay as a secondary type. Secondly, sentiment analysis showed that people have a positive reaction towards the frontliners and all efforts by the Ministry of Health towards fighting the pandemic. Many positive remarks are given in form of prayers, which is in line with the Islamic teaching of positive thinking and optimism, especially during crises. In conclusion, sentiment analysis is effective in producing useful insights about trends of COVID-19 discussion on social media, collecting public perception and feedback of COVID-19 efforts by the Government, and gives a different viewing angle of the current situation on the ground. These findings can be useful for health officials or the Government in developing communication mitigation plans or conduct extensive studies on pertaining issues within areas of concern.

International Journal of Modern Trends in Social Sciences
IJMTSS
EISSN: 2600-8777
Volume 4 Issue 17 (September 2021) PP. 01-08
DOI 10.35631/IJMTSS.417001

Keywords:

Sentiment Analysis, COVID-19, Machine Learning, Social Media

Introduction

Throughout history, mankind has experienced many episodes of infectious disease outbreaks, irrespective of ethnicity and geographical location (Rudra, 2020)**.** Most recently, a new disease outbreak from Wuhan, China has spread globally, with only 584 cases reported in 23 January 2020 has increased 13-fold to more than 118,000 cases in 114 countries and COVID-19 officially declared as a pandemic by the World Health Organization (WHO(b), 2020), (WHO(a), 2020). During this pandemic, a clear guideline and strong leadership is needed to steer public out of baseless fear and mass panic. Crisis management theories and models proposed on modern world focuses more on step-by-step procedures to follow, whereas from the perspective of Islam, crises are managed beginning with knowledge and skill, endurance and having faith of a higher power (Eid & Arnout, 2020).

Social media platforms are medium where people congregate and freely expresses opinions, thus these platforms can offer meaningful insights on public's perspectives towards social change and during crisis situation. These social media platforms collect and store vital information concerning emotions and reactions of mass public, which can be useful in producing insightful feedback to interested parties. Facebook, Instagram, Twitter and Reddit are among popular free social media sites, and globally, Facebook has the highest percentage of social media users. In Malaysia alone, Facebook records 84.47% of social media users for the last one year (StatCounter, 2020). Hence, Facebook is chosen as the best platform to conduct this study among Malaysian population.

This study examines emerging thematic issues and sentiment analysis of the general public towards COVID-19 pandemic through comments on Facebook posts. Malaysian's emotions, beliefs, and thoughts are gathered to allow policy makers make better informed decisions. The focus of this paper is on Malaysian discussion on Malaysia's Director General (DG) of Health's official Facebook account who posted about COVID-19 crisis. Posts collected during first two weeks of Movement Control Order (MCO) dated 18 March to 30 March 2020 (Week 1) and 1 April to 14 April 2020 (Week 2). The objectives of the paper are as follows: (1) to deploy sentiment analysis model to analyze comments collected and (2) to identify sentiment of Malaysians towards Malaysian government's effort in handling of COVID-19 and MCO.

Literature Review

Human behaviors are influenced by their opinions, beliefs and perceptions of reality. Making decisions and choices, at some point, depends on opinion of others. Opinions contains influential elements like emotions, sentiments, information, attitutes and evaluations. These factors are central subjects in sentiment analysis and opinion mining study. Sentiment analysis (also known as opinion mining) is a study of people's attitudes, feelings, assessments, appraisals, and emotions about things, entities, situations, events, and themes, as well as their attributes (Liu, 2012).

With regards to COVID-19 and sentiment analysis, there are lack of social-media based data research looking into the spread of information and general public's emotions on the disease. For the year 2020, a handfull of studies conducted relating to COVID-19 and sentiment analysis

International Journal of Modern Trends in Social Sciences
EISSN: 2600-8777
IJMTSS
Volume 4 Issue 17 (September 2021) PP. 01-08
DOI 10.35631/IJMTSS.417001

on social media. In China, Shen et al. studied relationship between symptoms and diagnosis posts made on Sina Weibo, a social media platform specific to China, with prediction of COVID-19 cases (Shen, et al., 2020) wheras Huang et al. found location distribution and epidimiological characteristics of patients based on comments related to COVID-19 made to Sina Weibo site (Huang, et al., 2020). These next two studies collected datasets in their native language and font. On the context of Italian social communication, a study on capturing sentiment of Italians (in Italian language) during lockdown decree produced disharmony result and found a need for more reliable classification on text analysis, possibly require human intervention to conduct classification process (Sciandra, 2020). In Myanmar, Facebook is one of the main social media platform for expressing opinions and feelings. Since comments collected are in Myanmar native language and uses Zawgi font type, Zawgyi fonts are converted to Unicode before conducting any machine learning processes (Aung & Pa, 2020). Sciandra dataset are collected using Twitter API whilst Aung & Pa collects via Facepager (data crawling tool).

Common methods of sentiment analysis and classification can be found in several studies are Naïve Bayes and SVM, and bi-gram or tri-gram. Sethi et al. (2020) built a model to conduct sentiment prediction based on gathered tweets with hashtag #COVID19 and #coronoavirus found SVM classification with unigram feature set works the best in any settings. Aung and Pa (2020) on the other hand, utilizes Word2Vec to train word associations from large high-dimensional corpus of semantics and word vectors, which then later found Logistic Regression performed the best for sentiment analysis.

Methodology

We collected comments from Malaysia's DG of Health Facebook account to monitor the Malaysian social media discussion about COVID-19. His Facebook account was chosen for the study due to his credibility as the DG of Health and being a health practitioner, he embodies persona of a trustworthy person (Ipsos MORI, 2005). It is also important to consider DG's leadership in handling the crisis in Malaysia, as one of the criteria of selecting this Facebook account for the purpose of this study. The first week of MCO is the first time Malaysian government impose movement and economic restriction onto Malaysian as preventive measures against the spread of COVID-19. Public reactions on the restrictive movements order and were collected at its initial phase and other COVID-19 progress are continuously observed and collected for the purpose of this study.

Data Collection

Dataset collected on both weeks consists of 77,351 comments of English, Malay, Arabic and combination of three languages. The files produced by NCapture, an NVivo web extension (QSR International, 2020) in Excel format as unstructured data which enables us to conduct further analysis and review comments in simple format. The datasets are limited to this Facebook account and the specified duration is as measure to control the polarity and content of comments to suit the purpose of study within the context specified. Summary of dataset collected for Week 1 and Week 2 as per Table 1.

Table 1: Dataset Collected Week 1 And Week 2

Week	Number of Comments
1	43,733
2	33,618
Total	77,351

International Journal of Modern Trends in Social Sciences
EISSN: 2600-8777
IJMTSS
Volume 4 Issue 17 (September 2021) PP. 01-08
DOI 10.35631/IJMTSS.417001

Learning Phase

Both datasets underwent learning phase before being analyze for text filtering and sentiment analysis. Learning phase enables datasets to undergo pre-processing and language detection process thus the comments were cleaned from unnecessary characters, numbers, emoticons and then group the comments into languages.

Data Pre-Processing

Pre-processing was conducted using the Natural Language Toolkit (NLTK) package due to its easy-to-use interface to over more than 50 corpora and lexical resources. NLTK enables us to work with suites for classification, tokenization and tagging, which will be conducted later in the process. Steps conducted as follows:

- null handling of comments whereby comments that are null by unique user accounts are removed;
- text cleaning by removing URL, hashtags, numbers and non-Roman characters. This is to normalize text encoding and allow corpus to process comments better. URL, hashtags and non-Roman characters are removed as it does not contribute to text categorization of comments. Emoticons are translated to words as it will help in giving a better result in sentiment analysis; and
- removal of stop words to simplify and condense comments into related words only.

Language Detection

Three languages detected in pre-processed comments were English, Malay and Arabic and mix of languages were detected. This phase aims to separate the comments into English and Malay language groups by Week 1 and Week 2. Malay language and other languages will be translated into English for ease of analysis of sentiment. Furthermore, translating Malay language to English is also due to lack of comprehensive Malay text corpus available. Table 2 shows the number of comments for Week 1 and Week 2 in English, Malay and mix languages:

Table 2: Language Groups By Week

Week	Language	Number of Comments
1	English	1,266
	Malay	23,142
	Mix	19,325
2	English	6,187
	Malay	21,699
	Mix	5,732

Malay and mix languages comments are translated into English using the Translation API from Google Cloud Platform (GCP). Neural Machine Translation (NMT) model in Translation API is used to translate the Malay and mix languages into English. This model is the same model Google use for the Google Translate. It is to note that the Arabic language are Islamic prayers and well-wishes.

Text Filtering

Text filtering enables us to conduct sentiment analysis and text categorization onto a more precise and clean dataset. At this stage, on and off topic separation is performed on Week 1 and Week 2 datasets of both English and translated Malay language. On/off topic is not

International Journal of Modern Trends in Social Sciences
EISSN: 2600-8777
IJMTSS
Volume 4 Issue 17 (September 2021) PP. 01-08
DOI 10.35631/IJMTSS.417001

conducted on mix languages dataset due to its complexity to translate mix languages into English using Translation API GCP. Continuation from this is separated On/Off topic for English and translated Malay datasets will be processed for sentiment analysis.

On/off topic separation process must be conducted manually because many comments were written in jargon, short form or even mix of English and Malay languages in a sentence. Hence, machine learning process for on/off topic separation cannot be conducted due to this limitation. Criteria for on/off topics are:

- On topic: words, sentences, comments related to subject of discussion/post in Facebook; and
- Off topic: mentions (tag Username in comments), emoticons, Arabic words/wishes written Roman characters, provocateur comments, insensitive comments.

The initial corpus had total of 43,733 comments for Week 1 and 33,618 comments for Week 2. After conducting On/Off Topic separation, comparison of On/Off topic are as in Table 3.

Table 3: On/Off Topic Separation Ratio

Week	On	%	Off
1	4,145		20,263
2	27,203	:	683

Comparison between both weeks shows more comments related to discussion or post in the Facebook in Week 2 (97.56%) compared to Week 1 (16.98%), whereas Off topic comments are prevalent in Week 1 (83%) compared to Week 2 (2.45%). Next we attempt to identify sentiments among the comments collected.

Sentiment Analysis

Facebook comments collected were analyzed for polarity using pre-trained model Natural Language API in GCP. Natural Language (NL) API inspects given text and identifies prevailing emotional opinions within the text, especially user's attitude as positive, negative or neutral. Firstly, the analysis is conducted using pre-trained model in NL API against training data. Then, the model is used on English and Malay translated comments, of both On and Off topic comments. We found that despite having the comments in English and translated to English from other languages, NL API could not identify sentiments of some comments. It is noted however that due to small sample size of Malay translated language, there are possibilities where machine learning will mismatch representativity of actual dataset with training set (Ceron, Curini, & Iacus, 2016). Sentiment distribution by week as in Table 4.

Table 4: Sentiment Distribution By Week

Week	Positive	Neutral	Negative
1	73.88%	14.5%	11.61%
2	82.12%	11.03%	6.84%

Due to reliance on Translation API GCP to translate Malay and other languages to English contributes towards possible mislabeling of sentiments. We believe that in order to obtain realistic and reliable results, use of supervised learning with human translation for each comment can yield better result (Ceron, Curini, & Iacus, 2016).

International Journal of Modern Trends in Social Sciences
EISSN: 2600-8777
IJMTSS
Volume 4 Issue 17 (September 2021) PP. 01-08
DOI 10.35631/IJMTSS.417001

Sentiment analysis of COVID-19 on social media study gained interest of several researchers. In comparison to this study, we found Afroz et. al, Tyagi et. al and Raheja & Asthana conducted similar studies, however, varies in social media platform or context. Afroz et. al and Tyagi et. al studied sentiment of public during lockdown phases in India on Twitter platform. Results from both studies surprisingly resonated similar outcome of this study, majority Positive sentiment for the early weeks of lockdown phases. On the other hand, Raheja & Asthana study focused on sentiment based on specific keywords, for example "COVID19" and "corona virus" produced primarily Neutral sentiment. We believe our findings are incomparable to Raheja & Asthana as the approach are different. (Tyagi, Goyal, & Gupta, 2021; Raheja & Asthana, 2021; Afroz, Boral, Sharma, & Gupta, 2021)

Conclusion

This Facebook comments analysis can give insights into public awareness and perception of the COVID-19 pandemic in the context of Malaysian public. Based on the data, Malaysian uses Malay language predominantly in social media communication. Due to lack of Malay language text corpus, translation of Malay language to English is necessary before sentiment analysis process can take place. Due to limitation of time to conduct this study, we conduct translation into English using Google Translation API, which possibly contribute towards some of the comments lost in translation. We found majority of Malaysians have positive sentiment towards COVID-19 and MCO efforts being done by the Government. Majority of the positive sentiment are well-wishes and congratulations towards the frontliners and leadership of DG of Health. Many of positive remarks are given in form of prayers, which is in line with Islamic teaching of positive thinking and optimism especially during crises, as in surah Al-Baqarah verse 216 "*Fighting has been enjoined upon you while it is hateful to you. But perhaps you hate a thing, and it is good for you; and perhaps you love a thing and it is bad for you. And Allah Knows, while you know not."* (The Noble Quran). It is with hope that outcome of this study can spark interests of public stakeholders into developing in-depth study towards producing comprehensive policy in handling health communication on social media.

References

Afroz, N., Boral, M., Sharma, V., & Gupta, M. (2021). Sentiment Analysis of COVID-19 Nationwide Lockdown Effect in India. *International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (pp. 561-567). Coimbatore: IEEE.

Agarwal, V., & K, S. B. (2020). COVID-19: Current Pandemic and Its Societal Impact. *International Journal of Advanced Science and Technology*, 432-439.

Aung, H. M., & Pa, W. P. (2020). Analysis of Word Vector Representation Representation Techniques with Machine-Learning Classifiers for Sentiment Analysis of Public Facebook Page's Comments in Myanmar Text. *2020 IEEE Conference on Computer Applications(ICCA)* (pp. 1-7). Yangon: IEEE.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python.* Sebastopol: O'Reilly Media Inc.

Boon-Itt, S., & Skunkan, Y. (2020). Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modelling Study. *JMIR Public Health Surveillance*, 1-26.

Ceron, A., Curini, L., & Iacus, S. M. (2016). iSA: A fast, scalable and accurate algorithm for sentiment analysis of social media content. *Information Sciences, 105*(124), 105-124.

International Journal of Modern Trends in Social Sciences
EISSN: 2600-8777
IJMTSS
Volume 4 Issue 17 (September 2021) PP. 01-08
DOI 10.35631/IJMTSS.417001

Eid, N. A., & Arnout, B. A. (2020). Crisis and disaster management in the light of the Islamic approach: COVID-19 pandemic crisis as a model (a qualitative study using the grounded theory). *Journal of Public Affairs*, 1-14.

Huang, C., Xu, X., Cai, Y., Ge, Q., Zeng, G., Li, X., . . . Yang, L. (2020). Mining the Characteristics of COVID-19 Patients in China: Analysis of Social Media Posts. *Journal of Medical Internet Research, 22*(5).

Ipsos MORI. (2005, March 10). *Doctors Top Public Opinion Poll On Trustworthy Professions*. Retrieved from Ipsos MORI: https://www.ipsos.com/ipsos-mori/en-uk/doctors-top-public-opinion-poll-trustworthy-professions

Kekre, H. B., Sarode, T. K., & Save, J. K. (2014). Generation of an Effective Training Feature Vector using VQ for Classification of Image Database. *International Journal of Computer Applications, 108*(18), 15-23.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers.

Lwin, M. O., Lu, J., Sheldenkar, A., & Schulz, P. J. (2018). Strategic Uses of Facebook in Zika Outbreak Communication: Implications for the Crisis and Emergency Risk Communication Model. *International Journal of Environment Research and Public Health*, 1-19.

Nair, A. J., G, V., & Vinayak, A. (2021). Comparative study of Twitter Sentiment On COVID-19 Tweets. *5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1173-1178). Erode: IEEE.

Python. (2020, May 4). *Gensim 3.8.3*. Retrieved from Python Project: https://pypi.org/project/gensim/

QSR International. (2020). *About NCapture*. Retrieved from NVivo 11 for Windows Help: https://help-nv11.qsrinternational.com/desktop/concepts/about_ncapture.htm

Raheja, S., & Asthana, A. (2021). Sentimental Analysis of Twitter Comments on COVID-19. *11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 704-708). Noida: IEEE.

Rudra, T. (2020). A comparative analysis of global pandemic of two respective genre. *Journal of Otorinolanryngology, 12*(4), 140-143.

Sciandra, A. (2020). COVID-19 Outbreak through Tweeters' Words: Monitoring Italian Social Media Communication about COVID-19 with Text Mining and Word Embeddings. *IEEE Symposium on Computers and Communications* (pp. 1-6). Rennes: IEEE.

Sethi, M., Pandey, S., Trar, P., & Soni, P. (2020). Sentiment Identification in COVID-19 Specific Tweets. *International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 509-516). Coimbatore: IEEE.

Shen, C., Chen, A., Luo, C., Zhang, J., Feng, B., & Liao, W. (2020). Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillence Study. *Journal of Medical Internet Research, 22*(5).

StatCounter. (2020, November 26). *Social Media Stats Worldwide*. Retrieved from StatCounter GlobalStat: https://gs.statcounter.com/social-media-stats

Tyagi, P., Goyal, N., & Gupta, T. (2021). Analysis of COVID-19 Tweets During Lockdown Phases. *9th International Conference on Information and Education Technology (ICIET)* (pp. 471-475). Okayama: IEEE.

WHO(a). (2020, March 11). *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. Retrieved from World Health Organization: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020

International Journal of Modern Trends in Social Sciences
EISSN: 2600-8777
IJMTSS
Volume 4 Issue 17 (September 2021) PP. 01-08
DOI 10.35631/IJMTSS.417001

WHO(b). (2020, January 23). *WHO Director-General's statement of the advice of the IHR Emergency Committee on Novel Coronavirus*. Retrieved from World Health Organization: https://www.who.int/director-general/speeches/detail/who-director-general-s-statement-on-the-advice-of-the-ihr-emergency-committee-on-novel-coronavirus