SENTIMENT ANALYSIS OF COVID-19 VACCINATION POSTS ON FACEBOOK IN INDONESIA WITH CROWDTANGLE

Adita Rianto¹, Ahmad R. Pratama²)

¹Program Studi Informatika – Program Sarjana, Universitas Islam Indonesia
²Jurusan Informatika, Universitas Islam Indonesia

https://www.ui.ac.id/
adita.rianto@student.uii.ac.id ¹, ahmad.rafie@uiei.ac.id ²)

(*) Corresponding Author

Abstract

COVID-19 was declared a pandemic by the World Health Organization (WHO) in early 2020. The Indonesian government has taken steps to stop COVID-19 from spreading, one of which is vaccination. However, not everyone thinks vaccination is a good idea. Like in other countries, Indonesian people responded in different ways to COVID-19 vaccination posts on social media, be it from government officials or news portals. Public responses can be used to help the government make a better vaccination strategy to end the pandemic in Indonesia. Using the lexicon method in determining the sentiment in COVID-19 vaccination posts on Facebook, this research found that unlike news portals that tended to post a more balanced content (36% positive, 20% negative, and 44% neutral out of 23,623 posts, min score = -19, max score = 24, mean = 0.25, SD = 1.43), government accounts posted much more positive content, both in quality (min score = -15, max score = 40, mean score = 4.16, SD = 6.76) and quantity (69% positive) than they did the neutral (15%) and the negative content (16%) out of 723 posts. Subsequent analysis with Two-Way ANOVA tests discovered that COVID-19 vaccination posts by the news portals elicited more varied reactions from the public than government accounts that tended to elicit mostly positive reactions. Also, both the content sentiment of COVID-19 vaccination posts in Indonesia and the account types making those posts, as well as their interaction terms do have an impact on how the public responds to them.

Keywords: vaccine, COVID-19, sentiment analysis, government, news portal, Facebook, CrowdTangle
INTRODUCTION

In December 2019 in Wuhan, China, a disease caused by the novel coronavirus was found which we now often call COVID-19. This disease was declared as a pandemic by the World Health Organization (WHO) on March 11, 2020, due to the uncontrolled spread of the virus and not only attacking Wuhan, China but also spreading to all other parts of the world (WHO, 2020).

For society and countries, the COVID-19 pandemic has had a major adverse impact. Some of the impacts that occurred were trade and tourism business actors who experienced a decrease in income, a decline in imports of up to 3.7% in the first quarter, to the occurrence of mass layoffs carried out by many companies (Yamali & Pramuni, 2020). Therefore, many governments tried to rush the vaccination program to tackle the COVID-19 pandemic. In Indonesia, this vaccination program is regulated by an executive order straight from the president himself (Peraturan Presiden, 2021).

Vaccines have been long proven to be effective in dealing with viral outbreaks. Research shows that vaccines have many benefits for the world, especially the community. Some of the benefits of taking vaccines collectively include increasing life expectancy, minimizing parents' fear of the threat of diseases that can cause paralysis to the loss of their children's lives, overcoming epidemics in an area, and reducing high costs because they can avoid diseases and disabilities that occur due to a disease (Schuchat, 2011). Nevertheless, not everyone has a positive view of vaccines in general and even more so with the COVID-19 vaccine in particular.

Sentiment analysis is an activity used to extract and analyze opinions, sentiments, attitudes, and perceptions of a subject such as products, services, or things that are happening somewhere. (Birjali et al., 2021). Previously there was research conducted on sentiment analysis regarding the pros and cons of the COVID-19 vaccine, in this study, it was stated that the results of the majority of public sentiment on Twitter had neutral indications (Rachman & Pramana, 2020).

Another past study was also conducted to look at public sentiment on Twitter regarding the vaccinations. The research shows the results at that time the dominant community had a positive response with the percentage of tweets reaching 48% with a total of 471 tweets, then for the percentage of negative tweets by 23% with a total of 224 tweets and neutral by 29% with a total of 281 tweets (Rakhamawati et al., 2020).

This study aimed to find out how Indonesian people responded to the COVID-19 vaccination posts on social media, especially Facebook, and to identify the types of content posted by government officials and news portal accounts in Indonesia regarding COVID-19 vaccination program in this country with sentiment analyses. The findings of this research would be helpful for evaluation and review purposes, especially for the government in making better, more effective, and more equitable strategies and policies for the vaccination, especially in the next stage.

RESEARCH METHOD

Population and Samples

In this study, the population is all public posts on Facebook discussing COVID-19 vaccination program in Indonesia. The samples used to represent that population were all public posts with the keyword of "covid vaccine" by selected verified (i.e., blue tick) Indonesian news portal (n = 10) and government accounts (n = 10) made between August 2020 and August 2021.

Data collection

Data collection in this study was done through CrowdTangle, which is a tool provided by Facebook that is used to make it easier to view, follow, analyze, and report on what is happening on public content on social media. On its website, CrowdTangle explains what data can be recorded such as the type of post, the account who made the post, the number of interactions or views that occur (CrowdTangle, 2021). The CrowdTangle data collection technique has been carried out several times by researchers worldwide, such as in a study conducted in Sweden to see a comparison of party accounts in attracting participants or followers on social media Facebook in the one month before the Swedish national election in 2018 (Larsson, 2020). Then data retrieval using CrowdTangle is also carried out in research on content that is attractive to the public and to see if incorrect health information gets high public attractiveness. This study uses CrowdTangle to retrieve data from one of the French public media accounts (Berriche & Altay, 2020).

Meanwhile, in Indonesia, there are at least two recent studies using CrowdTangle as data collection method. Both studies investigate Indonesian public sentiment toward educational policy amid COVID-19 pandemic on two different social media, namely Facebook (Widangsa & Pratama, 2021) and Instagram (Afif & Pratama, 2021).
Data collection with CrowdTangle is done by using the historical data feature on CrowdTangle web, entering keywords, filling in the type of posts, the type of accounts, and the desired time period to be sampled. After that, the data will be sent out to the requester in CSV format via email. The previous account types have been made into a list with the criteria previously mentioned on the CrowdTangle web, the list is divided into two, namely a list for government accounts and a list for news accounts as summarized in Table 1.

Table 1. The list for government accounts and news accounts.

<table>
<thead>
<tr>
<th>Government</th>
<th>News portal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anies Baswedan</td>
<td>BBC News Indonesia</td>
</tr>
<tr>
<td>Badan Nasional</td>
<td>Detik.com</td>
</tr>
<tr>
<td>Penanggulangan</td>
<td>KompasTv</td>
</tr>
<tr>
<td>Bencana</td>
<td>Kompas.com</td>
</tr>
<tr>
<td>DPR RI</td>
<td>IDN Times</td>
</tr>
<tr>
<td>Ganjar Pranowo</td>
<td>Kompas.com</td>
</tr>
<tr>
<td>Humas Pemda DIY</td>
<td>Kompas.com</td>
</tr>
<tr>
<td>Kementerian Kesehatan</td>
<td>Kumparan RI</td>
</tr>
<tr>
<td>Pemprov DKI Jakarta</td>
<td>Liputan6</td>
</tr>
<tr>
<td>Presiden Joko Widodo</td>
<td>Merdeka.com</td>
</tr>
<tr>
<td>Ridwan Kamil</td>
<td>Tempo Media</td>
</tr>
<tr>
<td>UNICEF Indonesia</td>
<td>tvOneNews</td>
</tr>
</tbody>
</table>

Procedure

![Figure 1. Research Procedure](image)

Figure 1 shows all steps taken in conducting this study. After the data collection was done, the next steps were conducting some preprocessing and data analyses, all of which were done in R 3.6.0 with Rstudio 1.2.1335. The full R code used for data analysis in this study, as well as the complete datasets generated from CrowdTangle are made available to the public on our GitHub repository.

Text Preprocessing

Pre-processing is the first step taken before data processing with the aim of making it easier to label the content. This process includes 1) converting all letters to lowercase, 2) removing URLs, 3) removing hashtags, 4) removing numbers, removing punctuation marks, 5) removing stopwords that consist of words that are not important or used far too common in any types of sentences, thus making it more difficult to distinguishes between different types of sentiment, removing excess spaces, and 6) stemming or converting entire words to basic words.

Sentiment Analysis on Posts Content

The first analysis carried out in this study is sentiment analysis on post content. For each post by the government accounts and news portals, a lexicon-based scoring method is used to determine if the content can be classified into positive, negative, or neutral posts. The lexicon method determines the tendency of sentiment by using a collection of words or phrases and statistical and semantic methods (Apriliant, 2019). This method has also been used to detect sarcasm on Indonesian social media (Lunando & Purwarianti, 2013).

This method uses a score calculation on words that have positive and negative meanings contained in the dictionary that has been made. For each word in the positive dictionary, it will be given a score of +1 while for each word in the negative dictionary, it will be given a score of -1. After all words are scored, the total score will be calculated. A post gets a positive label if the total score is greater than 0, a negative label if it is less than 0, and a neutral label if equal to 0 (Setyobudi et al., 2018). Some examples of positive, negative, and neutral content are shown in Table 2, Table 3, and Table 4 respectively.

Table 2. Examples of posts with positive sentiment.

<table>
<thead>
<tr>
<th>Government</th>
<th>News portal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indonesia Galang</td>
<td>Majelis Ulama Indonesia (MUI) menerbitkan fatwa kehalalan CoronaVac yaitu vaksin Covid-19.</td>
</tr>
<tr>
<td>Solidaritas Anggota APEC untuk Membangun Ketahanan Kesehatan di Kawasan Asia Pasifik dalam the 10TH APEC Virtual High-Level Meeting on Health and The Economy (VHLMHE).</td>
<td></td>
</tr>
<tr>
<td>Majelis Ulama Indonesia (MUI) issued a fatwa on the halalness of CoronaVac, namely the Covid-19 vaccine.</td>
<td></td>
</tr>
</tbody>
</table>

1 [https://github.com/aditariant/c19vaccinefbposts](https://github.com/aditariant/c19vaccinefbposts)
WHO and the European Union announce a new collaboration to support the COVID-19 response and strengthen the health system in Indonesia. PT Pos Indonesia (Persero) is ready to optimize the distribution of Cash Social Assistance (BST) or cash assistance from the government to the public.

Finance Minister Sri Mulyani Indrawati said that the Indonesian economy in 2021 is projected to start recovering at around 5 percent.

President Joko Widodo said that 2020 was a year full of challenges and uncertainties. The global death toll from COVID-19 has reached four million, as growing gaps in vaccine access expose poorer countries to outbreaks of more infectious strains.
Sentiment Analysis on Public Reactions

The second sentiment analysis was carried out on the public reactions on each post. In this study, the focus of the public reactions used for sentiment analysis is "Love" and "Care" as positive sentiments and "Angry" and "Sad" as negative sentiments. The grouping of sentiments based on these reactions comes from a previous study in the literature (Lee et al., 2020), which classifies "Love" and "Like" as positive sentiments and reactions "Angry" and "Sad" as negative sentiments. However, there are modifications made to the "Like" reactions because according to another research (Spottswood & Wohn, 2019) the use of the "Like" button for the public is mostly treated only as a sign that they have seen the post. So, it is less representative to be judged as a direct emotional reaction. Therefore, this study substitutes the "Like" reaction to "Care" as a new reaction made by Facebook which shows the meaning of support in a positive way (Facebook, 2020). Then, the sentiment will be calculated to see whether a post gets a positive or negative sentiment by the public as its audience.

Descriptive Analysis

In looking at the general picture of the data, a descriptive analysis was conducted. This stage will show the amount of posted data obtained, the classification of content, and the number of reactions that the public has given to the post. These data can be used to find out how the proportion of posted content is distributed, and the reactions obtained from government accounts and news portals.

Data visualization

Data visualization is used to represent data that has been processed previously so that the information obtained is easier to understand by the reader. In this case, word clouds will be used to see what words appear most frequently when government accounts and news accounts post something about COVID-19 vaccine on Facebook, both when the content is deemed as positive and negative.

Inferential statistical analysis

Inferential statistical analysis in this study is used to see whether there are similarities in the results in the form of public reactions obtained against the type of account and on the content posted from both the government and news portals. The tests used to perform inferential statistics in this study were ANOVA test and Mann-Whitney U test.

ANOVA

Testing several populations to see the similarity of the mean can use a method commonly called ANOVA. This method is comprehensive enough to detect group differences in a single dependent variable (Pritasari et al., 2013). In this study, ANOVA was used to see if there were differences in public responses between posts from the government and posts from news portals. This ANOVA is also used in this study to see if a post in one type of account affects the reactions that arise from the community. This ANOVA test uses the independent variable in the form of content from the post, while the dependent variable uses the public reaction that appears.

After the One-Way ANOVA test was carried out, this study also carried out a Two-Way ANOVA test to see and re-confirm whether there is an interaction between account types and content sentiment on public reactions. Account type and content will be the independent variable and reaction will be the dependent variable.

Mann-Whitney

The Mann-Whitney test is a non-parametric statistical analysis test that is used to compare the mean value of two populations (Putra & Mustika, 2014). This study uses the Mann-Whitney test to see the distribution of public responses to whether the account types are the same or different by comparing the two median values of the two types of accounts that have been obtained, namely government accounts and news portal accounts.

RESULTS AND DISCUSSION

The results from scoring and classification stage showed that government accounts tend to use a lot of positive narratives, resulting in a high positive score for each upload compared to news portal accounts. As seen in the boxplot depicted in Figure 2 where the highest positive score on government accounts reached +40, which is much higher than that of the news portal accounts where the highest positive score was merely +24. Meanwhile, for the negative content, there is no significant difference between government accounts and news portals. It can be seen that the score with the lowest negative sentiment posted by government accounts was -15 while on news accounts the lowest negative sentiment score was -19. The distribution of this data will be analyzed further to see whether the distribution of scores that look different will affect the public reactions to the two types of accounts as well as the content classifications or not.
It was further found that the posts by the government accounts consisted of 69% positive content, 16% negative content, and 15% neutral content from a total of 723 posts. Meanwhile, the news portal account had only 36% positive content, 20% negative content, and the rest 44% was neutral content, out of 23,623 posts in total as shown in Table 5. It further emphasizes the notion that unlike news portals that tend to post a more balanced content, government accounts posted much more positive content, both in terms of quality and quantity than they did the neutral and the negative content.

When reviewed further. It can be seen that in government-owned posts, whatever content is posted, the reactions given by the public is relatively the same, which is positive reactions at around 94-96% It is a different case for the posts by the news portal accounts where different content will produce different public reactions as well. In posts belonging to news portals, positive content received 73.6% positive reactions and 26.4% negative reactions by the public, whereas posts with negative sentiment content, received more negative reactions at 57% compared to the positive reactions at 43%. Lastly, the neutral content from news portal posts got 60.7% positive reactions and 39.3% negative reactions. Based on that result alone, it is quite an indication that public reactions vary based on the account type (i.e., government or news portal) and the content sentiment (i.e., positive, negative, or neutral). However, to confirm if that indication is statistically significant, further analysis required.

Before conducting any statistical analysis, some data visualizations were carried out with wordclouds to find out what words or topics appearing frequently in the dataset. As can be seen in Figure 3, in the positive content within government posts the three most frequent words are “pemerintah” (government), “pelayanan” (services), and “kementerian” (ministry). Meanwhile in Figure 4, the words “pemerintah” (government), “penyakit” (disease), and “imunisasi” (immunization) are the three most common words in the content with negative sentiment posted by the government accounts.

Table 5. Results of content classification on government accounts and news portals.

<table>
<thead>
<tr>
<th>Type</th>
<th>Positive</th>
<th>Negative</th>
<th>Neutral</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Governments</td>
<td>471,961 (66%)</td>
<td>27,939 (6%)</td>
<td>44,022 (6%)</td>
<td>723</td>
</tr>
<tr>
<td>News</td>
<td>24,000 (44%)</td>
<td>6,344 (26%)</td>
<td>17,661 (36%)</td>
<td>23,623</td>
</tr>
</tbody>
</table>

Table 5. Results of content classification on government accounts and news portals.

Figure 2. Boxplots of scoring and sentiment analysis results of all posts

Figure 3. Wordcloud from the content with positive sentiment by the government accounts
Figures 5 and 6 show word cloud data visualization regarding positive and negative content from news portals. The picture shows that in positive news content the words or topics that appear most often are "pemerintah" (government), "dosis" (dose), and "Sinovac". While the negative content of news portals, the words or topics that appear most often are "Sinovac", "pemerintah" (government), and "uji" (trial).

From the four wordclouds that have been presented previously, it can be seen that "government" is the most common topic for both positive and negative sentiment contents posted by either the government news portals accounts discussing the COVID-19 vaccine. It only makes sense since the government is after all the main actor behind this vaccination work, hence the spotlight. Meanwhile, "Sinovac" is commonly mentioned in news portal posts, both in positive and negative sentiment content, which is not the case for the government accounts. It seems like the government did not want to emphasize the vaccine brand unlike the news portal accounts.

The results from One-Way ANOVA test as shown in Table 6 indicates no association between the content sentiment and public reactions on posts made by government accounts (p = 0.442). As indicated in the descriptive statistics previously, it seems like regardless of the content sentiment on posts made by the government accounts, most public reactions they received remained positive.

The same One-Way ANOVA test was also carried out on posts made by news portal accounts. As can be seen in Table 7, the test revealed a significant association between the content sentiment of a post made by news portal account and the public reactions it gets (p = < 0.001).
Content with positive sentiment posted by news portal received more positive public reactions. Likewise, the content with negative sentiment they posted received more negative public reactions.

Table 7. One-Way ANOVA test results on news portal accounts

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Sum sq</th>
<th>Mean sq</th>
<th>F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>2</td>
<td>10694</td>
<td>5347</td>
<td>9.518</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Residual</td>
<td>23620</td>
<td>13269814</td>
<td>562</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Furthermore, a Two-Way ANOVA test was conducted to see if both the types of account posting the content and the sentiment of the content posted affected public reaction on COVID-19 vaccine as well as to see if there is an interaction between the two. As it turned out, all three of them were statistically significant (p < 0.001) in affecting public reactions as shown in Table 8.

Table 8. Two-Way ANOVA test on content sentiment and account types

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Sum sq</th>
<th>Mean sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>2</td>
<td>7.68x10^6</td>
<td>384096</td>
<td>15.27</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Type</td>
<td>1</td>
<td>3.57x10^6</td>
<td>35706542</td>
<td>1420.23</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Content</td>
<td>2</td>
<td>1.52x10^6</td>
<td>7593461</td>
<td>30.20</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Type</td>
<td>Residual</td>
<td>24340</td>
<td>6.12x10^6</td>
<td>251410</td>
<td></td>
</tr>
</tbody>
</table>

This finding further indicated that government accounts tend to receive more positive reactions from the public in general compared to the news portal regardless of the content sentiment that they posted. However, it is also true that positive content tends to receive positive reactions and vice versa. Considering the fact that government accounts tend to post much more positive content, it does make sense that they receive more positive reactions from the public, too.

Table 9. Mann Whitney u-test results on people's responses to account types

<table>
<thead>
<tr>
<th></th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>394891</td>
</tr>
<tr>
<td>p-Value</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Lastly, to see whether the distribution of public reactions to different account types (i.e., government vs news portal) are the same or different, the Mann Whitney u-test was conducted to compare the two median values of the two types of accounts. In the test, it can be seen that the results show that there are differences in the distribution of public responses to the types of accounts who made the posts as shown in Table 9. It further proved that there are indeed differences between government accounts and news portal accounts in influencing the response of the Indonesian people.

CONCLUSION

In the case of the COVID-19 vaccination, the word “government” is the most often discussed word, both with positive and negative sentiment, on Facebook posts made by both the government and the news portal accounts. It emphasizes the central role of the government in COVID-19 vaccination, including in public sentiment toward it. Also, the word “Sinovac”, which is the most commonly found and administered Chinese-made COVID-19 vaccine in Indonesia, is the second most frequent word discussed by the news portal accounts, with slightly more occurrences in the negative sentiment than in the positive one. Based on the findings of this study, it can be concluded that both the difference in content sentiment of the COVID-19 vaccination posts on Facebook and the account types making those posts elicited different public reactions in Indonesia. News portal accounts, which posted a more balanced sentiment on their posts, received significantly varied public responses toward their posts. On the contrary, the government accounts, which posted far more positive sentiment content, received relatively homogeneous positive reactions to their Facebook posts about COVID-19 vaccination by the public. This research can hopefully become a reference for researchers to further analyze the public reactions to public posts on Facebook and other social media. Also, the findings can be used as an evaluation material and reference for the government in taking necessary actions regarding the COVID-19 vaccinations in Indonesia.

REFERENCES


Berriche, M., & Altay, S. (2020). Internet users engage more with phatic posts than with health misinformation on Facebook. Palgrave Communications, 6(1), 1–9. https://doi.org/10.1057/s41599-020-0452-


