A Computational Approach in Analyzing The Empathy to Online Donations during COVID-19

Akmal Setiawan Wijaya , Ahmad R. Pratama , Dhomas Hatta Fudholi
[1] Universitas Islam Indonesia, Yogyakarta, Indonesia

ABSTRACT

The COVID-19 pandemic has had a negative impact on many aspects of life. The global economic downturn is one of these negative consequences. Nonetheless, even though everyone feels the threat of this pandemic for themselves, some people still have the empathy to help others. An empirical analysis of this empathy attitude is expected to be a catalyst in realizing a social force for the community to work together to combat this pandemic. This study will examine how people felt about donating during the COVID-19 pandemic on Twitter. This study aims to (1) compare differences in donor desire before and during the COVID-19 pandemic using the developed model and (2) determine whether there is a significant difference in empathy for donating before and during the pandemic. This study employs computational social science (CSS) techniques to achieve this goal. The data was obtained from Twitter using the keyword "donation" in the 24 months preceding the pandemic and in the 24 months following the pandemic's arrival in Indonesia. Data analysis includes hypothesis testing using Mann-Whitney and Cohen's D statistical tests, showing a significant increase in online donation support among Indonesian Twitter users since the COVID-19 pandemic hit. From the results of data processing data obtained 159.995 data in accordance with the criteria to be analyzed. From the results of the Mann-Whitney test, all variables showed significant results before and during the Covid-19 pandemic, and in the results of the Cohen's d test, all variables got a large effect size. From the results of the two tests, it can open Twitter social media users who have increased empathy to donate during the Covid-19 pandemic in Indonesia.

Corresponding Author:

Ahmad R. Pratama,
Department of Informatics,
Universitas Islam Indonesia, Yogyakarta, Indonesia,
Email: ahmad.rafie@uii.ac.id

1. INTRODUCTION

The COVID-19 pandemic began in late 2019 in Wuhan, China, and by early 2020 had spread to almost every country on the planet. The pandemic has had a devastating effect on human life. According to the SMERU survey results from October to November 2020, nearly 75% of the interviewed households experienced a decreased income. Food insecurity in households increased by 11.7% [1]. Furthermore, COVID-19 had an impact on MSMEs; according to the Ministry of National Development Planning/Bappenas, 80% of MSMEs experienced a significant decrease in income [2]. Because the COVID-19 pandemic has had so many negative consequences, President Joko Widodo declared it a national disaster with far-reaching consequences [3]. The magnitude of the pandemic's impact has encouraged people to help one another to alleviate the pandemic's burden.

Meanwhile, online donation campaigns on social media have been active since 2013 and have been shown to contribute to an increase in interest in giving alms [4]. Twitter, founded in 2006, has since grown to become one of the most popular social media platforms in the world, with millions of users across the globe. According to [5], Indonesia is one of the top five Twitter users in the world. Posts on Twitter, also known as Tweets, allow users to publish their activities or express their opinions on any topics of their interest. Twitter content has evolved into a realistic representation of people's lives that must be investigated [6, 7]. In relation to this pandemic condition, Twitter has become a means for people to encourage one another and help alleviate the burden on society. Twitter has also proven to be a tool capable of assisting other users [8]. This phenomenon makes Twitter an intriguing social media platform to investigate. As a result, many researchers have used Twitter data to investigate a wide range of topics, including technology [9, 10], the impact of the COVID-19 pandemic [11, 12], and mental health [13].

Sentiment analysis is one of the research topics that can be investigated using Twitter. [14] successfully completed a study of Twitter users' sentiment regarding COVID-19 in six countries. Of all countries considered for the sentiment analysis, the UK has the highest negativity, at 23.03%, followed by France at 22.71% and the USA at 22.01%. India had a negativity of 18.39% using a simple Lexicon-based approach, while it was 35.92% in the case of France, 35.68% in the case of the UK, and 35.38% in the case of the USA, and with the least negativity of 31.03% was India. [15] used the same lexicon method to classify data into positive words and phrases, which account for 24% (1079 entries) of the lexicon size, whereas the negative words and phrases account for 76% (3351 entries) of the lexicon size. Besides that, Lexicon can also add accuracy to other methods [16]. [17] also successfully conducted research using this Lexicon-based approach with an accuracy of 86.61%.

Meanwhile, hypothesis testing is done to answer the question of whether a hypothesis is true or false. One of the most widely used hypothesis tests is the Mann-Whitney U Test. [9][9][9] conducts research by testing to validate the hypothesis with Mann-Whitney with rejection results. [18] also conducted research with the Mann-Whitney test, in which they failed to reject H0. In addition, Cohen's D Test is another common method used to complement hypotheses tests, particularly when it is necessary to see the effect size rather than simply answering the question of statistical significance. For example, [19] has successfully conducted research with Cohen's D Test to see the effect size of the retinal layer.

Since the COVID-19 pandemic hit Indonesia, there have been changes in how people give their donations. Many abandoned traditional ways of donating in favor of online donations [20]. The main research question to answer was, Is there any significant difference in support for online donations before and during the COVID-19 pandemic, and if so, how much?. The answer to this question is important for additional knowledge about Indonesian society and to help policymakers at any level in their decision-making. Understanding the changes in how people donate and the motivations behind these changes can be important for policymakers in order to tailor their outreach and message to better engage with potential donors, promote and support effective and secure donation channels, and anticipate and plan for potential changes in funding streams for charities and other organizations.

The rest of the paper is organized as follows. Section 2. is the research method. In this section, we present the stages of the research. Then, the result and the analysis of the research are presented in Section 3. In this section, we carry out the method described in Section 2. Finally, Section 4. concludes the paper. In this section, we conclude from the analysis results obtained in Section 3.

2. RESEARCH METHOD

Techniques from the field of computational social science (CSS) were utilized throughout this study. Figure 1 depicts the progression of this study's four stages, which can be used to understand how this research was conducted. This study began with a literature review before delving into data collection, processing, and analysis, including visualization and hypothesis tests.
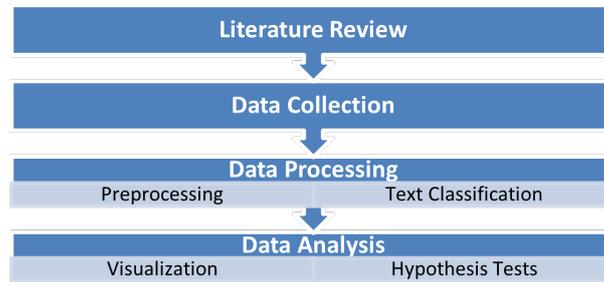
Figure 1. stages with CSS techniques

As shown in Figure 1, the first step of this research is a literature review as presented earlier (section 1), followed by data collection (section 2.1), data processing that involves preprocessing and text classification (section 2.3), and finally, data analysis that consists of visualization (section 2.4) and hypothesis tests (section 2.5).

2.1. Data Collection

This study used the social media platform Twitter as a data source. Python and the twepy library are used as crawling tools in Twitter, using the V2 media API. Data collection through the use of the keyword "donation." The data was collected 24 months before COVID-19 arrived in Indonesia and 24 months after. In other words, the data collected spanned from February 2018 until January 2022. Figure 2 shows one example of Twitter data collected in this study. From the example in Figure 2, the data will be taken in the form of Tweets, Date time, Retweets, Likes, and the number of replies.
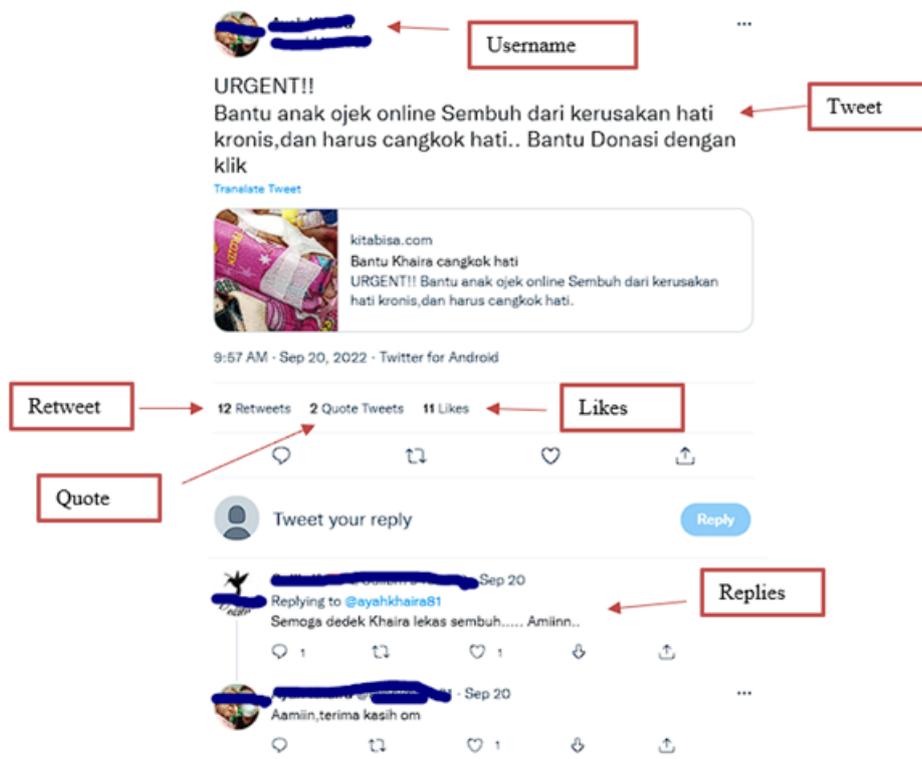

Figure 2. Example of Twitter data collected in this study

2.2. Data Preprocessing

The preprocessing stage comes after the data collection stage. The following tasks are completed during the preprocessing stage: 1) Removing duplicate data, 2) Removing URLs, 3) Removing Non-ASCII characters, 4) Removing special Twitter characters such as hashtags, usernames, and RT, 5) Removing numbers, 6) Removing symbols and punctuation marks, 7) Removing stopwords, 8) Removing single-letter words, and 9) Converting to lowercase form.

2.3. Text Classification

Following completing all preprocessing steps, the next step is text classification using Python and the orange data mining application. The classification of text is accomplished in two stages. Using the InSet Lexicon dictionary, [21] the first step consists of classifying positive or negative sentiments. Tweets containing negative sentiments are deleted at this stage. The second step is identifying tweets containing a donation request or suggestion. All tweets that do not contain the target word are excluded from the data analysis phase.

2.4. Visualization

The next step after text classification is visualization, which is carried out to understand the data. For visualization, bar charts and time series charts will be used. For example, the frequency of words that appear when Bigrams and Trigrams are used is displayed in bar charts. Bigrams and trigrams are used to identify a few words that frequently appear in chunks between words on a bar chart. A time series chart is used to determine how many tweets, retweets, likes, and replies are sent each month.

2.5. Hypothesis Tests

The last step after visualization is hypothesis testing. The Mann-Whitney and Cohen's d tests were used in this study to determine whether there was a significant difference in the values of each variable before and during the COVID-19 pandemic and, if so, how significant the effect was. The Mann-Whitney test was conducted to find out from the hypothesis "is there any significant difference between before the Covid-19 pandemic entered Indonesia and when the Covid-19 pandemic was in Indonesia" based on decision making as follows:

1. If the p-value <0.05, then the hypothesis is accepted
2. If the p-value >0.05, then the hypothesis is rejected

Cohen's d test in this research intends to see the effect size on the data before the pandemic, and when the Covid-19 pandemic entered Indonesia, the effect size rules are as follows:

1. If the Effect size <0.2, then the effect size is ignored
2. If the Effect size is 0.20.5, then the effect size is small
3. If the effect size is 0.5-0.8, then the effect size is medium
4. If the Effect size >0.8, then the effect size is large

3. RESULT AND ANALYSIS

Table 1 summarizes the dataset in this study in three stages. The raw column contains all tweets after the data collection step. The post-preprocessing contains all remaining tweets after the data preprocessing step. At the preprocessing step, the tweet data was reduced by 3,175,650 because, at this stage, duplicate tweets were deleted. This duplicate tweet appears a lot because there are retweets considered tweets. The post-classification column is the final dataset for data analysis purposes. The reduced data is data that does not meet the criteria for analysis. For the criteria, use you from dictionary [21] with positive sentiment and the word in Table 2. The post-classification column contains all remaining tweets after the text classification step, at the data classification step, reduced by 561,773.

Table 1. Summary of data collection

Variable	Raw	Post-Preprocessing	Post-Classification
Number of Month	48	48	48
Total Data	3,897,418	721,768	159,995
Number of Tweets	3,897,418	721,768	159,995
Number of Replies	1,046,696	628,500	125,426
Number of Likes	10,694,679	7,250,803	1,923,866
Number of Retweets	6,264,150,057	3,104,755	1,056,184

Table 2. A lexicon containing a list of words of encouragement to donate

Word	English translation	Frequency in dataset
Ayo	Lets go	17,380
Ayolah	Lets go	103
Segera	Soon	71
Lakukan	Do	1,935
Yuk	Lets go	35,562
Mari	Come on	12,997
Marilah	Come on	118
Klik	Click	18,711
Tolong	Help	8,751
Bantu	Help	77,373
Bantuan	Assistance	24,400
Mohon	Please	14,970
Link	Link	18,341
Via	Via	8,272
Dibuka	Open	5,089
Disalurkan	Distributed	8,429
Rekening	Account	14,124

Based on the results of preprocessing and data classification, Table 3 displays 159,995 data obtained following the data summary criteria. Table 3 displays the mean, median, and frequency of Tweets, Retweets, Likes, and Replies before and after the Covid-19 pandemic entered Indonesia. This table also indicates when the pandemic entered Indonesia. In the tweets and retweets variables, the mean, median, and frequency increased by more than 200%, while the likes variable increased by more than 600% on the mean and frequency and more than 900% on the median, and the replies variable increased by more than 400% on the mean and frequency and more than 500% on the median. In addition, during the Covid-19 pandemic, there has been an increase in empathetic responses to requests for donations.

Table 3. Summary statistics of all metrics before and during the COVID-19 pandemic

		Mean	Median	Frequency
Tweet	Before	1,627.20	1,391.00	39,053.00
	During	5,039.30	4,391.50	120,942.00
	Change	210%	216%	210%
Retweet	Before	10,824.50	7,565.00	259,789.00
	During	33,182.70	26,284.50	796,384.00
	Change	207%	247%	207%
Likes	Before	9,398.40	6,061.00	225,561.00
	During	70,762.30	64,552.00	1,698,294.00
	Change	653%	965%	653%
Replies	Before	836.5	614	20,075.00
	During	4,389.30	3,818.00	105,344.00
	Change	42 425%	522%	425%

Figure 3 depicts the visualization of monthly tweet data as a time series using line charts for all four metrics (i.e., tweets, retweets, likes, and replies). The X-axis represents the time period from February 2018 to January 2022. The Y-axis represents the relative frequency of each variable. The red line represents the dividing line between data prior to the pandemic, and the time the Covid-19 pandemic entered Indonesia. Before the pandemic (to the left of the red line), the frequency of all variables tended to be low, whereas, during the Covid-19 pandemic (to the right of the red line), there was an immediate and significant increase relative to before the pandemic. This is due to direct social restrictions that force individuals to interact via social media.
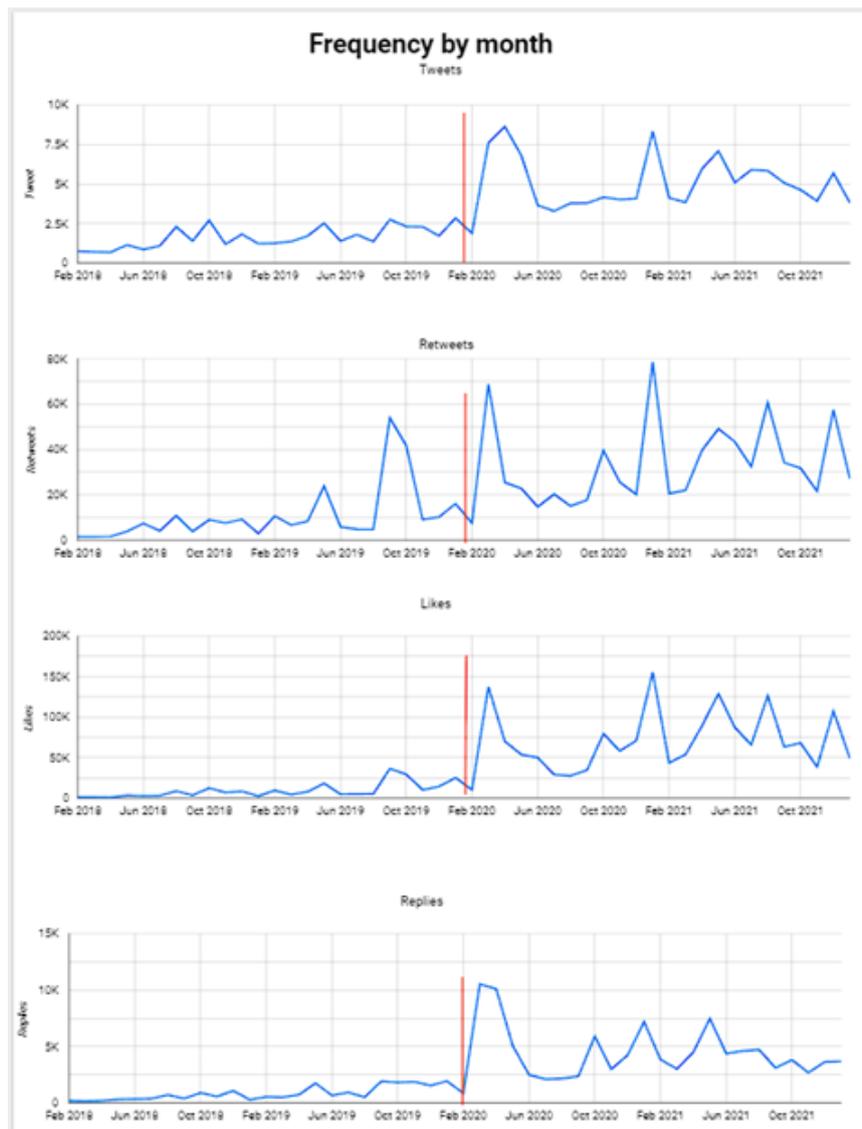

Figure 3. Line chart visualization of the frequency of each variable of interest by month

The ratio of the three variables (likes, replies, and retweets) to tweets is depicted in Figure 4. The likes/tweets and replies/tweets charts indicate an increase following the Covid-19 pandemic. Based on the retweet/tweet chart, it can be concluded that there was no increase before the Covid-19 pandemic and during it. Like Figure 3, there is an increase in the ratios of likes/tweets and replies/tweets, with an increase in the variables like and replies. Unlike the retweet variable, which increased during the Covid-19 pandemic, there is no increase in the ratio of retweets to tweets between before and during the pandemic. Possibility of not increasing the ratio of retweets to tweets because it permits the development of tweet and retweet variables.
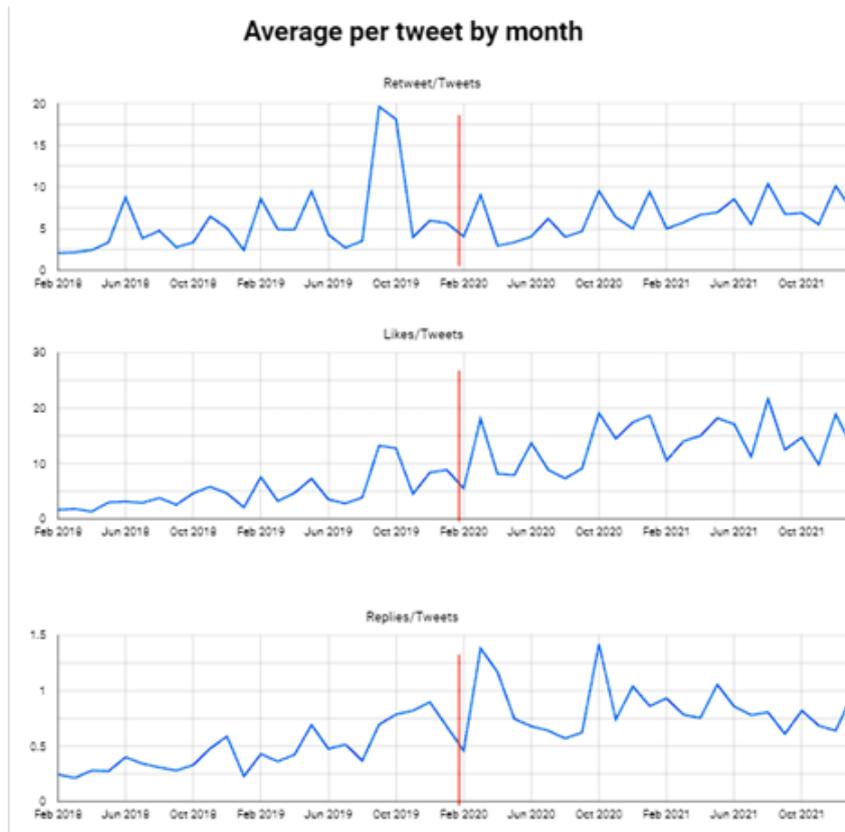
Figure 4. Line chart visualization of the average value of each variable of interest per tweet by month

Figure 5 depicts a bar chart with the bigram applied to the tweets data. With a frequency of over 17,500, "bantu donasi" (help donate) is at the top of the list. In the second position, the word "donasi untuk" (donation for) occurs more than 10,000 times. The word "terima kasih" (thank you) appears in third place, acquiring more than 7,500 words. There is a significant difference between "help donate" and "donate for"; the difference is almost doubled in the second order. Due to its high frequency, the term signifies that it is relied upon by tweet creators to pique donors' interest. Still from Figure 5, it can also be seen that the dominant words with bigram for the word donation are "bantu" (help), "untuk" (for), "open," "dengan" (with), "mau" (want), "buat" (make), "untuk" (for), "link," "ke" (to), "bisa" (can), and "yang" (which).
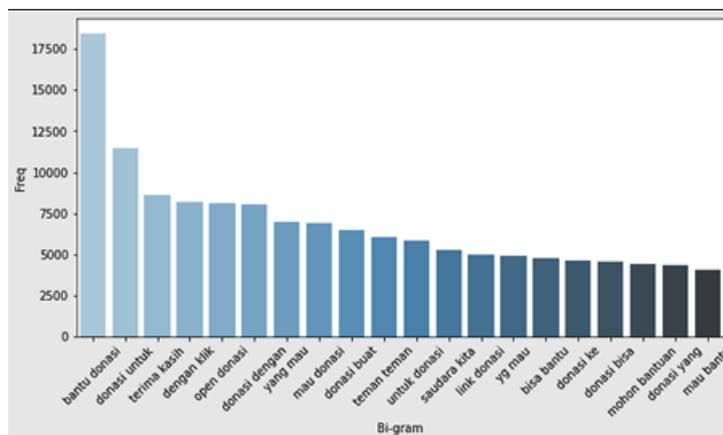

Figure 5. Bigram visualization of dataset

Figure 6 depicts a bar chart visualization with the trigram derived from the visualization of the words "help donate with" and "donate with clicks," having the same frequency with over 7,000 tweets. However, there is a difference of more Based on the trigram method. Nevertheless, it can be concluded that the phrases "help donate with" and "donate by clicking" have attracted the most Twitter users' empathy for online donations, as they have the highest volume (over 7,000) compared to other phrases (below 3,000).
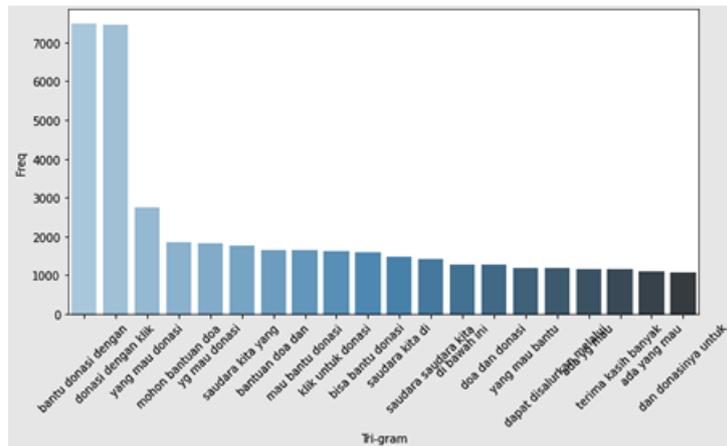

Figure 6. Trigram visualization of dataset

The final step of data analysis consisted of hypothesis tests using the Mann-Whitney U Test to determine whether the dataset before and during the COVID-19 pandemic differed significantly for each variable. Then, the Cohen's D Test was used to determine whether any statistical differences have a significant effect size. Table 4 summarizes the outcomes of both tests and each variable. The Mann-Whitney U Test indicates a significant difference between the number of tweets, retweets, likes, and replies in online donation content before and after the Covid-19 pandemic hit Indonesia. In addition, the subsequent Cohen's D Test results for all four variables are greater than 0.8, indicating that the significant difference in the number of tweets, retweets, likes, and replies in online donation content before and after the Covid-19 pandemic hit Indonesia has a large and meaningful effect size.

Table 4. Summary of hypothesis tests

Variable	Mann-Whitney U Value	p-value	Cohens D	Effect Size
Tweets	9.133	< 0.001	2.636	Large
Retweets	4.939	< 0.001	1.425	Large
Likes	7.797	< 0.001	2.25	Large
Replies	7.053	< 0.001	2.036	Large

From the two tests above, it can be concluded that Twitter users in Indonesia showed greater empathy towards content requesting or encouraging online donations after the COVID-19 pandemic hit Indonesia than before. This finding is consistent with previous studies that concluded there was a link between Twitter usage prior to and during the COVID-19 pandemic that used a traditional survey method to Twitter users [22] as well as another one that collected data directly from Twitter in the form of tweet intensity before and after COVID-19 pandemic started [23].

4. CONCLUSION

The objectives of the study were to compare Twitter users empathy for online donations before and during the COVID-19 pandemic and to determine whether there was a significant difference in donor empathy before and during the pandemic. This study collects data from Twitter using the keyword "donation" in the 24 months preceding and following the pandemic's arrival in Indonesia using computational social science (CSS) techniques using data from 3,897,418 tweets. A series of data analyses, including summary statistics and time series visualizations, as well as hypothesis testing with Mann-Whitney and Cohen's D statistical tests, revealed a significant increase in support for online donations among Indonesian Twitter users since the COVID-19 pandemic struck. Based on several relevant metrics on Twitter, such as Tweets, Retweets, Likes, and Replies, the empathy to support online donations increased

by as little as twice and as much as ten times during the COVID-19 pandemic compared to before it started. The COVID-19 pandemic has had a negative impact on many aspects of life. One of these negative consequences is the global economic downturn. Nonetheless, even though everyone is concerned about the threat of the pandemic, some people have the compassion to help others. This study provides an empirical analysis of the empathy attitude expected to be a catalyst in realizing a social force for the community to work together to combat the COVID-19 pandemic by looking at how people felt about donating during the COVID-19 pandemic on Twitter.

5. ACKNOWLEDGEMENTS

6. DECLARATIONS

AUTHOR CONTIBUTION
The first author collected, processed, and analyzed the data as well as wrote the initial draft of the manuscript, and assisted in revising it throughout the peer-review stage. The second and corresponding author provided initial guidance, secured the funding for this research, critically edited the initial draft of the manuscript, and revised it throughout the peer-review stage. The third author provided supervision and critically edited the manuscript, and assisted in revising it throughout the peer-review stage.
FUNDING STATEMENT
COMPETING INTEREST All three authors have no competing interest to declare.

REFERENCES

[1] SMERU, *Ringkasan Eksekutif: Dampak Sosial Ekonomi Terhadap Rumah Tangga dan Rekomendasi Kebijakan Strategis untuk Indonesia*, Jakarta, 2021.

[2] Bappernas, *Studi Pembelajaran Penanganan Covid-19 Indonesia*, 2021.

[3] Peraturan Pemerintah RI, "Keputusan Presiden Republik Indonesia Nomor 12 Tahun 2020 tentang Penetapan Bencana Non-Alam Penyebaran CORONA VIRUS DISEASE 2019 (COVID-19) sebagai Bencana Nasional." 2020.

[4] T. A. van den Broek, A. Need, M. L. Ehrenhard, A. Priante, and D. Hiemstra, "The Influence of Network Structure and Prosocial Cultural Norms on Charitable Giving: A Multilevel Analysis of Movember's Fundraising Campaigns in 24 Countries," *Social Networks*, vol. 58, no. April, pp. 128–135, 2019.

[5] Kominfo, "Indonesia Peringkat Lima Pengguna Twitter," 2012.

[6] F. M. Firmansyah and J. J. Jones, *Did the Black Panther Movie Make Blacks Blacker? Examining Black Racial Identity on Twitter Before and After the Black Panther Movie Release.* Springer International Publishing, 2019, vol. 11864 LNCS.

[7] A. Kartino, M. Khairul Anam, Rahmaddeni, and Junadhi, "Analisis Akun Twitter Berpengaruh Terkait Covid-19 Menggunakan Social Network Analysis," *Rekayasa Sistem dan Teknologi Informasi*, vol. 5, no. 4, pp. 697–704, 2021.

[8] R. A. Abbasi, O. Maqbool, M. Mushtaq, N. R. Aljohani, A. Daud, J. S. Alowibdi, and B. Shahzad, "Saving Lives Using Social Media: Analysis of the Role of Twitter for Personal Blood Donation Requests and Dissemination," *Telematics and Informatics*, vol. 35, no. 4, pp. 892–912, 2018.

[9] E. Mnif, K. Mouakhar, and A. Jarboui, "Blockchain Technology Awareness on Social Media: Insights from Twitter Analytics," *The Journal of High Technology Management Research*, vol. 32, no. 2, p. 100416, nov 2021.

[10] H. H. Hidayat, A. Ardiansyah, P. Arsil, and L. I. Rahmawati, "Pemetaan Kata Kunci dan Polaritas Sentimen Pengguna Twitter terhadap Kehalalan Produk," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, pp. 1–10, nov 2021.

[11] R. N. Aldekhyyel, S. Binkheder, S. N. Aldekhyyel, N. Alhumaid, M. Hassounah, A. AlMogbel, and A. A. Jamal, "The Saudi Ministries Twitter Communication Strategies During the COVID-19 Pandemic: A Qualitative Content Analysis Study," *Public Health in Practice*, vol. 3, no. June, pp. 1–9, jun 2022.

[12] M. Kfoury, D. Landy, and S. Fourmentin, "Combination of DES and Macrocyclic Host Molecules: Review and Perspectives," *Current Opinion in Green and Sustainable Chemistry*, vol. 36, no. August, p. 100630, aug 2022.

[13] S. Hinduja, M. Afrin, S. Mistry, and A. Krishna, "Machine Learning-Based Proactive Social-Sensor Service for Mental Health Monitoring Using Twitter Data," *International Journal of Information Management Data Insights*, vol. 2, no. 2, pp. 1–9, nov 2022.

[14] H. S. Hota, D. K. Sharma, and N. Verma, *Lexicon-Based Sentiment Analysis Using Twitter Data*.   Elsevier Inc., 2021.

[15] A. Al-Thubaity, Q. Alqahtani, and A. Aljandal, "Sentiment Lexicon for Sentiment Analysis of Saudi Dialect Tweets," in *Procedia Computer Science*, vol. 142.   Elsevier B.V., 2018, pp. 301–307.

[16] F. A. Effendi and Y. Sibaroni, "Sentiment Classification for Film Reviews by Reducing Additional Introduced Sentiment Bias," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 863 – 875, 2021.

[17] N. P. Dewi and Ubaidi, "Lexical Rule dan Pengaruh Penggunaan Lexicon pada Pos," *Matrix*, vol. 18, no. 1, pp. 65–72, 2018.

[18] T. Sriwidadi, "Penggunaan Uji Mann-Whitney pada Analisis Pengaruh Pelatihan Wiraniaga dalam Penjualan Produk Baru," *Binus Business Review*, vol. 2, no. 2, pp. 751–762, nov 2011.

[19] A. Bowring, F. J. Telschow, A. Schwartzman, and T. E. Nichols, "Confidence Sets for Cohen's d Effect Size Images," *NeuroImage*, vol. 226, no. February, pp. 1–28, feb 2021.

[20] F. M. Firmansyah and A. R. Pratama, "Anonymity in COVID-19 Online Donations: A Cross-Cultural Analysis on Fundraising Platforms," in *Advances in Intelligent Systems and Computing*, 2021, vol. 1364, no. April, pp. 34–47.

[21] F. Koto and G. Y. Rahmaningtyas, "Inset Lexicon: Evaluation of A Word List for Indonesian Sentiment Analysis in Microblogs," in *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017*, 2018, pp. 391–394.

[22] N. Rihhadatul'aisyi, S. Muthmainnah, H. P. Zahra, T. W. Putri, and F. T. Febrian, "Efek Twitter di Masa Pandemi COVID-19 pada Sikap dan Perilaku," *Jurnal Ilmu Komunikasi*, vol. 19, no. 2, pp. 205–220, oct 2021.

[23] D. Purwitasari, A. Apriantoni, and A. B. Raharjo, "Identifikasi Pengaruh Pandemi Covid-19 terhadap Perilaku Pengguna Twitter dengan Pendekatan Social Network Analysis," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 6, pp. 1309–1318, nov 2021.