Sentiment Analysis of Covid Vaccination Policy In Indonesia Using Random Forest

Nurirwan Saputra[1,*] Ahmad Riyadi[1], Meilany Nonsi Tentua[1]

[1]Faculty of Science and Technology, Universitas PGRI Yogyakarta
Jl. PGRI No. 117, Sonosewu, 55182, Yogyakarta, Indonesia
[*]Corresponding author. Email:nurirwan@upy.ac.id

ABSTRACT

The World Health Organization (WHO) on March 11, 2020, has declared the novel coronavirus (COVID-19) outbreak a global pandemic and Presidential Decree of the Republic of Indonesia Number 12 of 2020 concerning the Determination of Non-Natural Disasters for the Spread of CORONA VIRUS DISEASE 2019 (COVID-19) as a National Disaster and vaccinating its citizens starting on January 13, 2021, but community is divided into two, there are those who agree with the government's policy and some who do not agree with the government's policy, besides that there are also those who are forced to accept it because the covid vaccine has become a requirement in state life, for travel, for work, entertainment and so on. With the difference of opinion among the public, a sentiment analysis study was conducted to see how high the level of accuracy was based on the level of data ranging from 500 data, 1000 data, 2000 data, 3000 data and 4000 using the Random Forest method. Data collection was carried out from January to September 2021 from Republic Indonesia's Ministry of Health Fanpage and retrieve 100-500 random data every month. There are 4 stages Preprocessing : cleaning, case folding, filtering, and tokenizing, and labelling with positive, netral and negative. The results of the model test show the highest F1-Score level in the data with a total of 2000 with a combination of uni-tri Gram tokenization. Meanwhile, the lowest F1-Score level is when the model uses 500 data with unigram tokenization. The use of more than 2000 data in the formation of the model using the random forest method showed a significant decrease in the F1-Score level. The experimental results show that when the 4000 data used with uni-tri gram tokenization concatenation has a lower F1-Score level, even almost the same as the F1-score level in 1000 data without using a tokenization combination. In general, the results of model testing using the Random Forest method show an increase in the F1-Score level when using a combination of uni-tri gram tokenization with less than 2000 data used.

Keywords: Sentiment Analysis, Covid, Random Forest, Data Mining

1. INTRODUCTION

The World Health Organization (WHO) on March 11, 2020, has declared the novel coronavirus (COVID-19) outbreak a global pandemic [1], the President of the Republic of Indonesia period 2019-2024 [2], Ir. H. Joko Widodo officially declared COVID-19 a national disaster. This stipulation was stated through Presidential Decree of the Republic of Indonesia Number 12 of 2020 concerning the Determination of Non-Natural Disasters for the Spread of CORONA VIRUS DISEASE 2019 (COVID-19) as a National Disaster [3].

During its journey, COVID-19 cases are highly elevated compared to other diseases and very complete information is presented every day by the Republic of Indonesia's COVID-19 Handling Task Force [4]. As of today, May 23, 2022, the COVID-19 death rate in Indonesia is 2.6%, while the cure rate is 97.4% [4].

The steps taken by the government of the Republic of Indonesia to overcome this pandemic followed the WHO directive, namely by vaccinating its citizens starting on January 13, 2021 [5], even though at that time it had not passed the third phase of clinical trials (Decree of the Minister of Health Number HK.01.07/Menkes/9860/2020), and Sinovac vaccine is the first choice of the Indonesian

government [6]. With this step, the community is divided into two, there are those who agree with the government's policy and some who do not agree with the government's policy, besides that there are also those who are forced to accept it because the covid vaccine has become a requirement in state life, for travel, for work, entertainment and so on. One of the reasons people do not agree and do not want to receive the covid vaccine is :

1. Health Law of Republic Indonesia Number 36 Year 2009 Chapter III Part One Article 5 Paragraph 3 which states that everyone has the right to independently and responsibly determine the health services needed for themselves.
2. Nuremberg Code : the voluntary consent of the human subject is absolutely essential [7].
3. Vaccines using mRNA technology have the potential to cause microvascular injury (inflammation and blood clots called microthrombi) in the brain, heart, liver, and kidneys [8]. In addition, RNA-based covid vaccines have the potential to cause more diseases than the COVID-19 outbreak itself [9].
4. There is no guarantee of not contracting the virus after the covid vaccination [10]
5. There were cases of death after being vaccinated against covid [11]–[19] and there were also autopsies that did not have comorbidities [20] but the Chairman of the National Commission for Post-Immunization Adverse Events (Komnas KIPI) Prof. Hindra Irawan Satari confirmed that until now there was no who died from COVID-19 vaccination [21]. In contrast to COVID-19, whose death was quickly determined to be due to COVID-19, although there were also cases of one of the 196 COVID tombs being dismantled again because the bodies were proven negative for COVID-19 [22].
6. No one can guarantee the absence of long-term side effects of the covid vaccine.

For these reasons, there are some people who doubt that the government's policy of requiring vaccinations to the public is the best way to get through this pandemic.

With the difference of opinion among the public, a sentiment analysis study was conducted to see how high the level of accuracy was based on the level of data ranging from 500 data, 1000 data, 2000 data, 3000 data and 4000 using the Random Forest method.

2. MATERIALS AND METHODS

Data collection was carried out from January to September 2021 from Republic Indonesia's Ministry of Health Fanpage (https://www.facebook.com/KementerianKesehatanRI) and retrieve 100-500 random data every month.

FIGURE 1. Display of the Indonesian Ministry of Health's Facebook fanpage

Every comment on the covid status posted on the fanpage will be taken as data in conducting sentiment analysis, enter the data input stage, at this stage the data will be stored in csv form. After the data obtained in the csv form, then do the preprocessing. There are 4 stages Preprocessing : cleaning, case folding, filtering, and tokenizing. After the data is processed in the preprocessing stage, the next stage is labeling positive and negative sentences using positive, netral and negative. Table 1 is an example of a comment that has been labeled.

TABLE 2. example of a comment that has been labeled

No	Komentar	Label
1.	Biar dagangannya laku bos, dah mepet masa ekspayer	Negatif
2.	Pertanyaan yang sama. Tp ga divaksin ga bisa kemana mana	Negatif
3.	iya, licik ya! Mengaitkan vaksin dgn sgla macam aktivitas khidupan. Menghalaljan sgla cara..	Negatif
4.	Pemerintah sudah berusaha semaksimal mungkin memutus mata rantai virus bahkan sampai geram menangani pandemi ini......	Positif
5.	tinggal vaksin dosis 2, insyaallah besok mudah mudahan ttp sehat semangat	Positif
6.	Itu surat vaksin harus bayar apa nggk??	Netral

The next stage is data sharing to start the classification process. The distribution of data is divided into two, namely the distribution of training data and test data, with the amount of training data as much as 90% and test data as much as 10%.

The classification of sentiment analysis is carried out using the Random Forest approach. Random forests, or random decision forests, are ensemble learning methods of classification, regression, and other tasks that work by building various decision trees during training. For classification tasks, the Random Forest output is the class selected by most trees. Random Forest prediction modeling simulation can be seen in Figure 3.
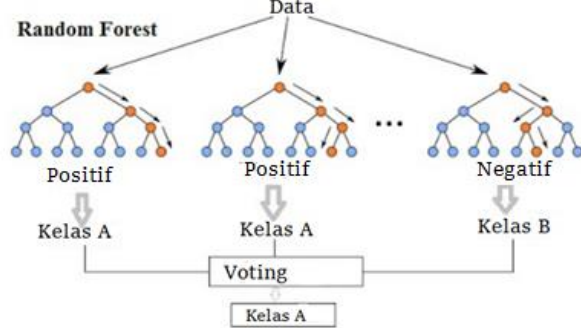
FIGURE 2. Modeling with Random Forest

1. Model testing

The performance of the model to make predictions can be measured by counting the number of predicted classes. Table 2 is a table of confusion matrix for calculating the accuracy value.

TABLE 2. Confusion matrix

		Positive	Negative
Predicted Value	Positive	true positive (tp)	false negative (fp)
	Negative	false positive (fn)	true negative (tn)

The metric value can be calculated from the confusion matrix with the following equation:

$$Accuracy = \frac{tp+tn}{tp+fn+fp+tn}$$

(1)

$$Precision = \frac{tp+tn}{tp+fp}$$

(2)

$$Recall = \frac{tp}{tp+fn}$$

(3)

$$fscore = \frac{(\beta^2+1)tp}{(\beta^2+1)tp+\beta^2 fn+fp}$$

(4)

Metrics precision, recall, and f score are used to measure the model's performance in making predictions in the positive class.

3. RESULT AND DISCUSSION

The results of the sentiment analysis process used data in the amount of 500, 1000, 2000, 3000 and 4000 data using the 10-cross validation technique. The pre-processing combination used is a combination of Unigram and Uni-Tri Gram tokenization. Table 2 shows the results of testing the sentiment analysis model using various pre-processed combinations.

TABLE 3. Model Test Results

Jumlah Data	F1-Score	
	Unigram	Uni-Tri Gram
500	58,6	60,4
1000	62	65,9
2000	65,35	66,6
3000	62,8	62,6
4000	62,8	62,5

The results of the model test show the highest F1-Score level in the data with a total of 2000 with a combination of uni-tri Gram tokenization. Meanwhile, the lowest F1-Score level is when the model uses 500 data with unigram

tokenization. The use of more than 2000 data in the formation of the model using the random forest method showed a significant decrease in the F1-Score level. The experimental results show that when the 4000 data used with uni-tri gram tokenization concatenation has a lower F1-Score level, even almost the same as the F1-score level in 1000 data without using a tokenization combination.

In general, the results of model testing using the Random Forest method show an increase in the F1-Score level when using a combination of uni-tri gram tokenization with less than 2000 data used.

CONCLUSSION

Sentiment analysis about the covid-19 vaccine using the random Forest method has an F1-Score level of 66.6%. The data used to achieve the F1-Score level is 2000 data with a combination of uni-tri gram tokenization. In general, the F1-Score level with the Random Forest method will increase when using a tokenization combination. The data used to achieve the best F1-Score from the experiments that have been carried out is less than 2000 data.

REFERENCES

[1] D. Cucinotta and M. Vanelli, "WHO Declares COVID-19 a Pandemic," Acta Biomed., vol. 91, no. 1, pp. 157–160, 2020, doi: 10.23750/ABM.V91I1.9397.

[2] 2   "Presiden Joko Widodo • Presiden RI." https://www.presidenri.go.id/presiden-joko-widodo/ (accessed May 23, 2022).

[3] 3   A. (Pusat D. I. dan K. K. B. R. I. Wibowo and D. D. I. dan K. K. B. R. I. Arifin, "Presiden Tetapkan COVID-19 Sebagai Bencana Nasional - BNPB," Pusat Data Informasi dan Komunikasi Kebencanaan BNPB Republik Indonesia, 2020. https://bnpb.go.id/berita/presiden-tetapkan-covid19-sebagai-bencana-nasional (accessed May 23, 2022).

[4] 4   "Peta   Sebaran | Covid19.go.id." https://covid19.go.id/peta-sebaran (accessed May 23, 2022).

[5] 5   A. R. Rahmanti et al., "Social media sentiment analysis to monitor the performance of vaccination coverage during the early phase of the national COVID-19 vaccine rollout," Comput. Methods Programs Biomed., vol. 221, p. 106838, Jun. 2022, doi: 10.1016/J.CMPB.2022.106838.

[6] 6   "Uji Klinis Fase III Vaksin Sinovac Belum Usai, Kok BPOM Sudah Kasih Izin?" https://finance.detik.com/berita-ekonomi-bisnis/d-5336383/uji-klinis-fase-iii-vaksin-sinovac-belum-usai-kok-bpom-sudah-kasih-izin (accessed May 23, 2022).

[7] 7   "Book Review 4 Journal of Pharmacy & Law   1994-1995." https://heinonline.org/HOL/LandingPage?handle=hein.journals/jpharm4&div=16&id=&page= (accessed May 23, 2022).

[8] 8   J. P. Whelan, "Could Spike Protein in Moderna, Pfizer Vaccines Cause Blood Clots, Brain Inflammation and Heart Attacks? • Children's   Health   Defense." https://childrenshealthdefense.org/defender/moderna-pfizer-vaccines-blood-clots-inflammation-brain-heart/ (accessed May 23, 2022).

[9] 9   J. B. Classen, "Immunologist: Pfizer, Moderna Vaccines Could Cause Long-Term Chronic Illness • Children's Health Defense." https://childrenshealthdefense.org/defender/pfizer-moderna-vaccines-long-term-chronic-illness/ (accessed May 23, 2022).

[10] 10   "Komnas KIPI: Vaksinasi Bukan Jaminan Tidak Tertular Virus, Tetap Patuhi Prokes   Halaman   all   -   Kompas.com." https://nasional.kompas.com/read/2021/02/22/18410451/komnas-kipi-vaksinasi-bukan-jaminan-tidak-tertular-virus-tetap-patuhi-prokes?page=all (accessed May 23, 2022).

[11] 11   "Siswa SD di Tasikmlaya Meninggal Usai Divaksin, Dinkes Sebut karena DBD." https://mediaindonesia.com/nusantara/465021/siswa-sd-di-tasikmlaya-meninggal-usai-divaksin-dinkes-sebut-karena-dbd (accessed May 23, 2022).

[12] 12   "Siswa SD di Jombang Meninggal Usai Divaksin   Covid." https://mediaindonesia.com/nusantara/461243/siswa-sd-di-jombang-meninggal-usai-divaksin-covid (accessed May 23, 2022).

[13] 13   "Viral Bocah di Tanjungbalai Meninggal Usai Vaksin Covid-19, Ini Kata Dinkes | merdeka.com." https://www.merdeka.com/sumut/viral-bocah-di-tanjungbalai-meninggal-usai-vaksin-covid-19-ini-kata-dinkes.html (accessed May 23, 2022).

[14] 14   "Kasus Siswa SD Meninggal Usai vaksinasi Covid-19, Dinkes Jombang: Hasil Sementara Investigasi Bukan Disebabkan Vaksin -   metrotvnews.com." https://www.metrotvnews.com/play/kM6CQ5vO-kasus-siswa-sd-meninggal-usai-vaksinasi-covid-19-dinkes-jombang-hasil-sementara-investigasi-bukan-disebabkan-vaksin (accessed May 23, 2022).

[15] 15   "Siswa SD Meninggal Usai Terima Vaksin Covid-19, Komnas KIPI Sebut Tak Ada Kaitan   dengan   Vaksinasi." https://www.kompas.tv/article/247023/siswa-sd-meninggal-usai-terima-vaksin-covid-19-komnas-

kipi-sebut-tak-ada-kaitan-dengan-vaksinasi (accessed May 23, 2022).

[16] 16 "Anak di Magetan Meninggal Setelah Vaksinasi Covid-19, Komnas KIPI Langsung Lakukan Audit." https://www.suara.com/health/2022/01/05/17232 1/anak-di-magetan-meninggal-setelah-vaksinasi-covid-19-komnas-kipi-langsung-lakukan-audit (accessed May 23, 2022).

[17] 17 "Dinkes Garut Usut Kasus Anak Meninggal Usai Vaksinasi Covid." https://www.cnnindonesia.com/nasional/2022012 3213800-20-750208/dinkes-garut-usut-kasus-anak-meninggal-usai-vaksinasi-covid (accessed May 23, 2022).

[18] 18 "Dalam Sepekan 3 Pelajar SD di Jatim Meninggal Usai Divaksin COVID-19." https://news.detik.com/berita-jawa-timur/d-5883689/dalam-sepekan-3-pelajar-sd-di-jatim-meninggal-usai-divaksin-covid-19 (accessed May 23, 2022).

[19] 19 "Seorang Anak di Tasik Meninggal Usai 2 Hari Divaksin, Dinkes Beri Penjelasan." https://news.detik.com/berita/d-5902600/seorang-anak-di-tasik-meninggal-usai-2-hari-divaksin-dinkes-beri-penjelasan (accessed May 23, 2022).

[20] 20 "Keluarga Ungkap Hasil Autopsi Trio: Tak Ada Komorbid, Tapi Ada Flek Hitam di Paru | merdeka.com." https://www.merdeka.com/peristiwa/keluarga-ungkap-hasil-autopsi-trio-tak-ada-komorbid-tapi-ada-flek-hitam-di-paru.html (accessed May 23, 2022).

[21] 21 "Komnas KIPI: Tidak Ada Yang Meninggal Karena Vaksinasi COVID-19 – Sehat Negeriku." https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20210520/4537800/komnas-kipi-tidak-ada-yang-meninggal-karena-vaksinasi-covid-19/ (accessed May 23, 2022).

[22] 22 "196 Makam Pasien Covid-19 Dibongkar karena Jenazah Terbukti Negatif Corona Halaman all - Kompas.com." https://regional.kompas.com/read/2021/06/14/13 4312678/196-makam-pasien-covid-19-dibongkar-karena-jenazah-terbukti-negatif-corona?page=all (accessed May 23, 2022).