

# Classification Performance Comparison of BERT and IndoBERT on Self-Report of COVID-19 Status on Social Media

## Porównanie wyników klasyfikacji BERT i IndoBERT w zakresie samodzielnego zgłaszania statusu COVID-19 w mediach społecznościowych

Irwan Budiman, Mohammad Reza Faisal\*, Astina Faridhah, Andi Farmadi, Muhammad Itqan Mazdadi, Triando Hamonangan Saragih, Friska Abadi

*Department of Computer Science, Lambung Mangkurat University, Banjarbaru 70714, Indonesia*

### Abstract

Messages shared on social media platforms like X are automatically categorized into two groups: those who self-report COVID-19 status and those who do not. However, it is essential to note that these messages cannot be a reliable monitoring tool for tracking the spread of the COVID-19 pandemic. The classification of social media messages can be achieved through the application of classification algorithms. Many deep learning-based algorithms, such as Convolutional Neural Networks (CNN) or Long Short-Term Memory (LSTM), have been used for text classification. However, CNN has limitations in understanding global context, while LSTM focuses more on understanding word-by-word sequences. Apart from that, both require a lot of data to learn. Currently, an algorithm is being developed for text classification that can cover the shortcomings of the previous algorithm, namely Bidirectional Encoder Representations from Transformers (BERT). Currently, there are many variants of BERT development. The primary objective of this study was to compare the effectiveness of two classification models, namely BERT and IndoBERT, in identifying self-report messages of COVID-19 status. Both BERT and IndoBERT models were evaluated using raw and preprocessed text data from X. The study's findings revealed that the IndoBERT model exhibited superior performance, achieving an accuracy rate of 94%, whereas the BERT model achieved a performance rate of 82%.

*Keywords:* Text classification; Covid-19 status; X; BERT; IndoBERT

### Streszczenie

Wiadomości udostępniane na platformach mediów społecznościowych, takich jak X, są automatycznie dzielone na dwie grupy: te, które samodzielnie zgłaszają swój status COVID-19, i te, które tego nie robią. Należy jednak pamiętać, że komunikaty te nie mogą stanowić wiarygodnego narzędzia monitorowania umożliwiającego śledzenie rozprzestrzeniania się pandemii Covid-19. Klasyfikację komunikatów w mediach społecznościowych można osiągnąć poprzez zastosowanie algorytmów klasyfikacyjnych. Do klasyfikacji tekstu wykorzystano wiele algorytmów opartych na głębokim uczeniu się, takich jak konwolucyjne sieci neuronowe (CNN) czy pamięć długoterminowa (LSTM). Jednak CNN ma ograniczenia w rozumieniu kontekstu globalnego, podczas gdy LSTM koncentruje się bardziej na zrozumieniu sekwencji słowo po słowie. Poza tym oba wymagają dużej ilości danych do nauki. Obecnie opracowywany jest algorytm klasyfikacji tekstu, który może pokryć wady poprzedniego algorytmu, a mianowicie dwukierunkowych reprezentacji enkoderów z transformatorów (BERT). Obecnie istnieje wiele wariantów rozwoju BERT. Podstawowym celem tego badania było porównanie skuteczności dwóch modeli klasyfikacji, a mianowicie BERT i IndoBERT, w identyfikowaniu komunikatów samoopisowych na temat statusu COVID-19. Zarówno modele BERT, jak i IndoBERT oceniano przy użyciu surowych i wstępnie przetworzonych danych tekstowych z X. Wyniki badania wykazały, że model IndoBERT wykazał się doskonałą wydajnością, osiągając współczynnik dokładności na poziomie 94%. Natomiast model BERT osiągnął wskaźnik wydajności na poziomie 82%.

*Słowa kluczowe:* Klasyfikacja tekstu; Covid-19 status; X; BERT; IndoBERT

\*Corresponding author

Email address: [reza.faisal@ulm.ac.id](mailto:reza.faisal@ulm.ac.id) (M. R. Faisal)

Published under Creative Common License (CC BY-SA v4.0)

## 1. Introduction

Covid-19 is an ailment that arises from the Coronavirus, making its debut in Wuhan at the culmination of 2019. Subsequently, the first accounts of COVID-19 cases in Indonesia were disclosed in March 2020. Some prevalent indications of COVID-19 in individuals include heightened body temperature, coughing, difficulty in respiration, and diminished olfactory function. The

COVID-19 pandemic, as of the conclusion of 2019, has registered an excess of 3 million documented cases across the globe, along with roughly 208,516 fatalities as of April 2020 [1]. This elevated level of mortality is attributable to a delay in promptly recognizing those afflicted with Covid-19 [2]. Consequently, those affected continue to engage in their customary activities, thereby facilitating the transmission of the virus to their close associates.

The worldwide dissemination of COVID-19 has evolved into a widespread pandemic, significantly impacting individuals across the globe. Throughout this global health crisis, social media, particularly the popular platform X, has emerged as a predominant means of disseminating information pertaining to the Covid-19 virus.

X has emerged as a widely embraced platform, boasting a staggering number of over 3.7 million active users who diligently disseminate approximately 10 million posts per diem [3]. Amongst this vast array of posts, one intriguing facet worth mentioning is the propensity for individuals to utilize this platform as a conduit for sharing insightful information about the COVID-19 crisis, including personal anecdotes relating to symptoms experienced or even instances of infection. Furthermore, the online community of social media enthusiasts has aptly utilized X as a platform for recounting the adverse impact that the pandemic has had on their respective families. Given its real-time nature, X lends itself seamlessly to monitoring the progression of the Covid-19 pandemic.

Identification of status of COVID-19 can be automated through the classification of self-report messages employing classification algorithms extensively developed by scholars in the field. Scholars have employed various techniques, such as Long Short-Term Memory (LSTM), Bi-directional LSTM (BiLSTM), word2vec, Support Vector Machines (SVM), and Convolutional Neural Networks (CNN), for text classification.

The classification of emotions in the text was undertaken in the study [4] utilizing the LSTM method, which yielded an accuracy score of 73.15%. In [3], the investigation focused on analysing emotions and sentiments in posts using BERT, resulting in an accuracy rate of 92%. In a subsequent study conducted by [5], the examination of emotions and sentiments deployed the BERT model, achieving a performance level of 92%. Furthermore, the researchers devised the IndoBERT model to classify Indonesian texts [6]. The findings of this study demonstrate that the IndoBERT model exhibits improved efficacy when applied to Indonesian texts with limited training data.

The BERT model has gained popularity among researchers in recent times. In contrast to other language models, BERT was developed as a pre-trained deep bidirectionally trained model using unlabeled text data. This was achieved by integrating left- and right-sided context layers [7]. Consequently, BERT models can be tailored to various machine learning tasks, such as classification and question answering, by simply incorporating a single layer [8].

However, BERT continues to possess limitations, specifically about the restricted availability of words in the Indonesian language. Consequently, numerous scholars have endeavoured to design BERT architectures by their respective languages. Presently, the BERT model has been made accessible in a multitude of languages, including Indonesian, exemplified by the IndoBERT [6]. Another study on text classification em-

ployed the IndoBERT model as a basis and conducted an experiment to investigate the impact of text preprocessing on classification performance [9].

Current literature on COVID-19 on social media predominantly focuses on sentiment analysis. Conversely, there is a dearth of research on text classification that aims to identify self-reported messages regarding COVID-19 status, and the available datasets primarily consist of texts in the English language [1], [10], [11].

A study was undertaken to compare the performance of BERT and IndoBERT models in identifying self-reported COVID-19 status from social media messages in Indonesian. The study involved training and testing each model using both raw and preprocessed text. The objective was to determine which model combination offers the highest classification performance.

## 2. Dataset

The datasets utilized in this investigation are derived from preceding investigations [12]. The dataset comprises 1000 messages sourced from X encompassing the keyword covid, as demonstrated in. The dataset is segregated into two classifications, namely, positive and negative. A positive message is defined as a message that, alongside the keyword COVID-19, exhibits one or more symptoms of COVID-19, and its connotation indicates that the individual expressing the complaint is presumed positive for COVID-19. The quantity of positive messages totals 500 messages. Examples of messages within the positive category can be observed in Table 1.

Table 1: Positive messages

Post	Translation	Class Label
doakan aku cepat sembuh dari covid	Pray for my speedy recovery from Covid	positive
Hari ini saya meriang, dan ketika periksa ke klinik ternyata saya positif covid	Today I had a fever, and when I went to the clinic it turned out I was positive for Covid	positive

Negative category messages are messages that do not pertain to positive COVID-19 cases. The total count of negative messages amounts to 500. Instances of messages belonging to negative categories are shown in Table 2.

Table 2: Negative message

Post	Translation	Class Label
Hari saya test swab namun hasilnya negatif covid	Today I had a swab test but the results were negative for Covid	negative
Nafas agak sesak, aku kira gejala covid ternyata tidak	Breathing a bit short, I thought it was a symptom of Covid but it wasn't	negative

### 3. Research Implementation

The various phases executed throughout this investigation can be visually observed and analyzed in a detailed manner by referencing the comprehensive illustration labelled as Figure 1.

The initial phase entails gathering datasets, as expounded upon in the preceding section. Subsequently, the text is subjected to preprocessing in order to achieve normalization through the utilization of tokenization, stemming, and stopword elimination techniques [13], [14], [15]. Tokenization is a procedure executed to fragment sentences into segments of lexemes, punctuation, and other significant articulations in agreement with the stipulations of the utilized linguistic system. Stemming is the procedure of transforming a term that possesses an inflection into its foundational term (radical structure) by eliminating affixes such as prefixes, suffixes, and confixes. Stopwords Elimination is the procedure of excluding lexemes that lack any import.

Additional preprocessing measures were undertaken by incorporating specific tokens, namely [CLS] at the commencement of the sentence, [SEP] at its conclusion, and [PAD] in instances of shorter sentences [16].

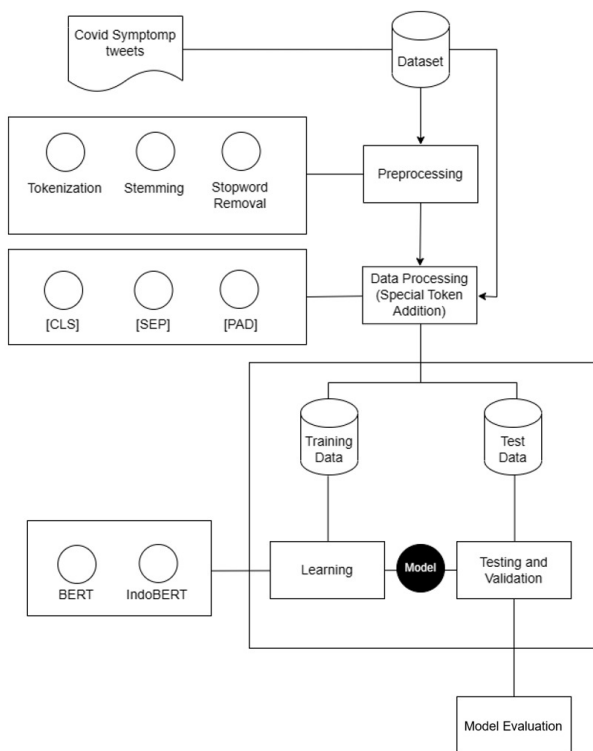


Figure 1: Research Flow.

The subsequent phase entails the division of the data utilizing the hold-out technique. This technique effectively separates the data into two distinct categories: training data and test data. The training data accounts for 80% of the composition, while the test data accounts for the remaining 20% [17], [18]. Following this, the learning phase commences, with the objective of constructing a classification model through the utilization of two methods, namely BERT and IndoBERT. As a

result of this stage, four models are created, incorporating a combination of methods and techniques, as illustrated in Table 3. It is important to note that the hyperparameter values employed in both the BERT and IndoBERT methods remain consistent. Furthermore, the models that are constructed utilizing the column values of the preprocessing are of no significance. These models do not undergo any text data processing techniques such as stemming or stopword removal.

Table 3: List of Classification Models

No	Preprocessing	Classification Method	Hyper Parameter
1	No	BERT	Batch size=16 Learning rate=2e-5 Epoch=10
2	Yes		
3	No	IndoBert	
4	Yes		

The testing and validation phase carries out the anticipation of the class label for the test data through four model classifications that were subsequently established in the preceding stage. Additionally, the performance outcomes for the four models are contrasted and scrutinized at the model evaluation stage. The assessment of the performance of the classification model at this stage employs the employment of the confusion matrix to determine accuracy, specificity, and sensitivity [14]. The depiction of the confusion matrix can be observed in Table 4.

Table 4: Confusion matrix

Actual Class	Predicted Class	
	Class	Class
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

True Positive (TP) is an outcome where the model correctly predicts the positive class. True Negative (TN) is an outcome where the model correctly predicts the negative class. False Positive (FP) is an outcome where the model incorrectly predicts the positive class. False Negative (FN) is an outcome where the model incorrectly predicts the negative class. Those values can be used to calculate accuracy, sensitivity, and specificity classification performance values. The formula of the three classification performances can be seen below.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \times 100\% \quad (1)$$

$$Sensitivity = \frac{(TP)}{(TP+FN)} \times 100\% \quad (2)$$

$$Specificity = \frac{(TN)}{(TN+FP)} \times 100\% \quad (3)$$

### 4. Results and Discussion

The results obtained from processing textual data in the dedicated token supplementary phase are presented in Table 5.

Table 5: Result of tokenization

	BERT	IndoBERT
Text	"Qadarullah", "pak", "suami", "negatif", "NS-1", "positif", "covid", ",", "plus", "ku", "jg", "sama", "positif", "covid", ".", "Suami", "cm", "bergejala", "3", "hari", "di", "awal", ",", "demam", "40", "derajat", ",", "skrg" "pilek", "ma", "batuk2", ".", "Ku", "baru", "gejala", "kemarin", ",", "demam", "38", "drjt", ",", "pusing", "ma", "linu2", "badan"	'qadar', '##ullah', 'pak', 'suami', 'negatif', 'ns', '-', '1', 'positif', 'cov', '##id', ',', 'plus', 'ku', 'jg', 'sama', 'positif', 'cov', '##id', ',', 'suami', 'cm', 'berg', '##eja', '##la', '3', 'hari', 'di', 'awal', '(', 'demam', '40', 'derajat', ')', 'sk', '##r', '##g', 'pil', '##ek', 'ma', 'batuk', '##2', ',', 'ku', 'baru', 'ada', 'gejala', 'kemarin', ',', 'demam', '38', 'dr', '##j', '##t', ',', 'pusing', 'ma', 'lin', '##u', '##2', 'badan'
Tokenization results using each pre-train Vocabulary	'q', '##ada', '##rul', '##lah', 'pak', 'sua', '##mi', 'negatif', 'ns', '-', '1', 'positif', 'co', '##vid', ',', 'plus', 'ku', 'jg', 'sama', 'positif', 'co', '##vid', ',', 'sua', '##mi', 'cm', 'berge', '##jala', '3', 'hari', 'di', 'awal', '(', 'dem', '##am', '40', 'dera', '##jat', ')', 'sk', '##rg', 'pile', '##k', 'ma', 'batu', '##k', '##2', ',', 'ku', 'baru', 'ada', 'ge', '##jala', 'kem', '##arin', ',', 'dem', '##am', '38', 'dr', '##jt', ',', 'pus', '##ing', 'ma', 'lin', '##u', '##2', 'badan'	'qadar', '##ullah', 'pak', 'suami', 'negatif', 'ns', '-', '1', 'positif', 'cov', '##id', ',', 'plus', 'ku', 'jg', 'sama', 'positif', 'cov', '##id', ',', 'suami', 'cm', 'berg', '##eja', '##la', '3', 'hari', 'di', 'awal', '(', 'demam', '40', 'derajat', ')', 'sk', '##r', '##g', 'pil', '##ek', 'ma', 'batuk', '##2', ',', 'ku', 'baru', 'ada', 'gejala', 'kemarin', ',', 'demam', '38', 'dr', '##j', '##t', ',', 'pusing', 'ma', 'lin', '##u', '##2', 'badan'

After the process of tokenization, there is an inclusion of specific tokens. These tokens are the [CLS] token, which is positioned at the beginning of the sentence, the [SEP] token, which is placed at the end of the sentence, and the [PAD] token, which is assigned to sentences that have a word count lower than the maximum number of words in a predefined sentence. Moreover, each token undergoes encoding based on the vocabulary index, and the outcomes of this encoding procedure are presented in Table 6.

Table 6: Result of special token addition

Special token addition	BERT	IndoBERT
	'[CLS]', 'q', '##ada', '##rul', '##lah', 'sua', '##mi', 'negatif', 'ns', '-', '1', 'positif', 'co', '##vid', ',', 'plus', 'ku', 'jg', 'positif', 'co', '##vid', ',', 'sua', '##mi', 'ge', '##jala', '3', 'hari', 'dem', '##am', '40', 'dera', '##jat', ',', 'pile', '##k', 'ma', 'batu', '##k', '##2', ',', 'ku', 'ge', '##jala', 'kem', '##arin', ',', 'dem', '##am', '38', 'dr', '##jt', ',', 'pus', '##ing', 'ma', 'lin', '##u', '##2', 'badan', '[SEP]', '[PAD]', ..., '[PAD]'	'[CLS]', 'qadar', '##ullah', 'sua- mi', 'negatif', 'ns', '-', '1', 'positif', 'cov', '##vid', ',', 'plus', 'ku', 'jg', 'positif', 'cov', '##id', ',', 'sua- mi', 'gejala', '##la', '3', 'hari', 'demam', '40', 'derajat', ',', 'pil', '##ek', 'ma', 'batuk', '##2', ',', 'ku', 'gejala', 'kemarin', ',', 'demam', '38', 'dr', '##j', '##t', ',', 'pusing', 'ma', 'lin', '##u', '##2', 'badan', '[SEP]', '[PAD]', ..., '[PAD]'

The subsequent stage involves executing an encoding procedure that transforms each token into an index. The outcomes are observable in Table 7.

Table 7: Result of encoding step

Encoding	BERT	IndoBERT
	101, 159, 11466, 31387, 16910, 10559, 10555, 81523, 28235, 118, 122, 66150, 10348, 41194, 117, 10608, 14886, 52817, 66150, 10348, 41194, 119, 10559, 10555, 50346, 29378, 124, 117, 10188, 11064, 10606, 75323, 15714, 78191, 10167, 10507, 19787, 10167, 10835, 119, 14123, 25463, 29378, 47304, 48499, 117, 10188, 11064, 11330,	2, 28216, 2421, 4425, 3778, 23803, 17, 21, 4590, 28912, 1528, 16, 8211, 1984, 23599, 4590, 28912, 1528, 18, 4425, 6004, 23, 1843, 16, 7322, 3828, 8033, 16, 4086, 1533, 2262, 14575, 952, 18, 1984, 6004, 4454, 16, 7322, 5885, 2455, 954, 930, 16, 12693, 2262, 4560, 943, 952, 2835, 3, 0, ..., 0

	11407, 30373, 117, 43955, 10285, 10507, 29715, 10136, 10835, 19605, 102, 0, ..., 0	
--	---	--

The first step involves generating an attention mask, which serves to differentiate between the word token's significance and the padding's insignificance. Examination of the outcomes is presented in Table 8.

Subsequently, the customized input is subsequently introduced into the BERT and IndoBERT networks, both of which consist of a series of 12-layer encoder transformers. Each encoder layer is composed of two sub-layers: the first one being a multi-head self-attention mechanism and the second one being a fully connected feed-forward network. Following the passage through all encoders, a vector output is generated for each token. However, only the vector output of the token [CLS] will be utilized as the vector input classifier.

Table 8: Result of attention mask

Attention mask	BERT	IndoBERT
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 1, 1, 1, 1,	1, 1, 1, 1, 1, 1, 1, 1,
	1, 1, 1, 1, 0, ..., 0	1, 1, 1, 0, ..., 0

The datasets transformed into the resultant input vector configuration are subsequently separated into training and test data. Following that, a phase of learning, testing, and validation ensues. The outcomes about the performance of the classification model can be observed in Table 9.

Table 9: Performance of classification models

No	Preprocessing	Classification Method	Classification Performance
1	No	BERT	Accuracy=81.50% Specificity=91.09% Sensitivity=71.72%
2	Yes	BERT	Accuracy=82.00% Specificity=89.11% Sensitivity=74.75%
3	No	IndoBERT	Accuracy=89.50% Specificity=82.83% Sensitivity=96.04%
4	Yes	IndoBERT	Accuracy=94.00% Specificity=92.00% Sensitivity=96.00%

Comparisons regarding the performance of each classification model can be observed in Figure 2. Upon examining these results, it becomes apparent that model

No. 4 outperforms the other models. Model No. 4 is a text-based model that has undergone preprocessing by applying stemming and stopword removal techniques.

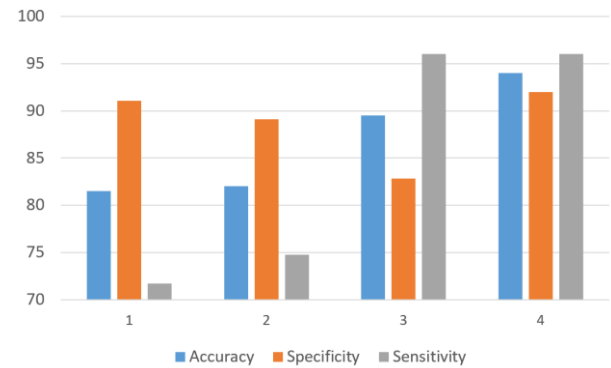


Figure 2: Comparison of classification models' performance.

Figure 3 shows a juxtaposition of the classification performance achieved by different classification methods. The outcomes manifest that models constructed utilizing the IndoBERT technique exhibit superior efficacy, as inferred from the accuracy and sensitivity metrics. Remarkably, the elevated sensitivity metric signifies the model's commendable aptitude in accurately discerning self-reported messages pertaining to COVID-19 status.

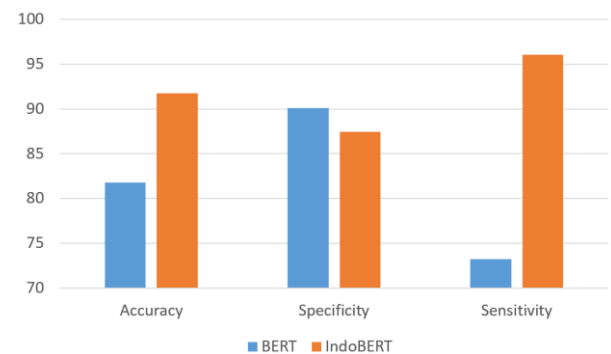


Figure 3. Comparison performance between BERT and IndoBERT models.

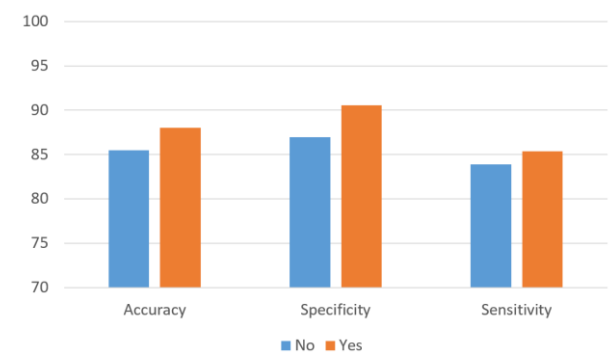


Figure 4. Effect of preprocessing on classification models' performance.

Figure 4 displays a comparison of the impact of preprocessing on the efficacy of classification models. This comparison reveals that preprocessing steps, such as stemming and stopword elimination, have the potential

to enhance the classification performance of BERT and IndoBERT-based models.

## 5. Conclusion

The findings of this investigation demonstrated that the execution of the IndoBERT-founded approach exhibited the capability to generate the most outstanding model for discerning self-reported COVID-19 status messages when compared to the classification model founded on BERT. Additional evidence presented in this analysis indicates that the preprocessing impact of stemming and stopword elimination can enhance the performance of identifying self-reported messages about COVID-19 status.

The self-report COVID-19 status message identification model, based on the IndoBERT model, demonstrated the highest level of performance. Its accuracy was recorded at 94%, with a specificity of 92% and a sensitivity of 96%. Conversely, the BERT-based model achieved an accuracy of 82%, a specificity of 89.11%, and a sensitivity of 74.75%. Employing the unpaired t-test on the test results, a two-tailed P value of 0.0488 was obtained, indicating a statistically significant difference. Thus, it can be concluded that the IndoBERT-based model exhibits significantly superior performance. On the other hand, the impact of text preprocessing on the classification model's performance was found to be insignificant. Using the paired t test technique, a two-tailed P value of 0.0528 was observed.

In the subsequent investigation, an evaluation will be conducted regarding additional BERT-derived methodologies, including ALBERT, Roberta, and DistilBERT. The primary objective entails constructing a model effectively discerning self-reported COVID-19 status within messages.

## 6. Acknowledgements

The computation time for the computer system in this study was furnished by the Data Science Lab, which operates within the Computer Science Department of the Faculty of Mathematics and Natural Sciences at Lambung Mangkurat University. This investigation received financial backing from the Program Dosen Wajib Meneliti (PDWM) grant provided by PNPB Lambung Mangkurat University.

## References

- [1] T. Mackey, V. Purushothaman, J. Li, N. Shah, M. Nali, C. Bardier, B. Liang, M. Cai, R. Cuomo, Machine learning to detect self-reporting of symptoms, testing access, and recovery associated with COVID-19 on Twitter: retrospective big data intelligence study, *JMIR public health and surveillance*, 6(2) (2020) 1-9, <https://doi.org/10.2196/19509>
- [2] A. Z. Klein, A. Magge, K. O'Connor, J. I. Flores Amaro, D. Weissenbacher, and G. Gonzalez Hernandez, Toward using Twitter for tracking COVID-19: a natural language processing pipeline and exploratory data set, *Journal of medical Internet research*, 23 (1) (2021) 1-6, <https://doi.org/10.2196/25314>
- [3] F. E. Ayo, O. Folorunso, F. T. Ibhralu, and I. A. Osinuga, Machine learning techniques for hate speech classification of Twitter data: State-of-The-Art, future challenges and research directions, *Computer Science Review*, 38 (2020) 1-34, <https://doi.org/10.1016/j.cosrev.2020.100311>
- [4] M. A. Riza, N. Charibaldi, U. Pembangunan, and N. Veteran, Emotion Detection in Twitter Social Media Using Long Short - Term Memory ( LSTM ) and Fast Text, 3 (1) (2021) 15–26, <https://doi.org/10.25139/jjair.v3i1.3827>
- [5] A. Chiorrini, C. Diamantini, A. Mircoli, and D. Potena, Emotion and sentiment analysis of posts using BERT, In *EDBT/ICDT Workshops*, 3 (2021) 1-7
- [6] B. Wilie, K. Vincentio, G.I. Winata, S. Cahyawijaya, X. Li, Z.Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, A. Purwarianti, IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding, *arXiv preprint arXiv:2009.05387*, (2020) 1-15
- [7] P. Ganesh, Y. Chen, X. Lou, M.A. Khan, Y. Yang, H. Sajjad, P. Nakov, D. Chen, M. Winslett, Compressing large-scale transformer-based models: A case study on BERT, *Transactions of the Association for Computational Linguistics*, 9 (2021) 1061–1080, [https://doi.org/10.1162/tacl\\_a\\_00413](https://doi.org/10.1162/tacl_a_00413)
- [8] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, In *Proceedings of naacL-HLT*, 1 (2019) 4171–4186
- [9] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP, *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference (2020) 757–770*, <http://dx.doi.org/10.18653/v1/2020.coling-main.66>
- [10] C. Menni, A.M. Valdes, M.B. Freidin, C.H. Sudre, L.H. Nguyen, D.A. Drew, S. Ganesh, T. Varsavsky, M.J. Cardoso, J.S. El-Sayed Moustafa, A. Visconti, Real-time tracking of self-reported symptoms to predict potential COVID-19, *Nature medicine*, 26 (7) (2020) 1037–1040, <https://doi.org/10.1038/s41591-020-0916-2>
- [11] M. A. Al-garadi, Y. Yang, S. Lakamana, A. Sarker, A Text Classification Approach for the Automatic Detection of Twitter Posts Containing Self-reported COVID-19 Symptoms, *Open Review*, (2020) 1–5
- [12] S. N. Sari, M. R. Faisal, D. Kartini, I. Budiman, Comparison of Feature Extraction with Supervised and Unsupervised Weighting in the Random Forest Algorithm for Monitoring Reports of COVID-19 Sufferers on Twitter, *Jurnal Komputasi*, 11 (1) (2023) 34–42, <http://dx.doi.org/10.23960%2Fkomputasi.v11i1.6650>
- [13] M. R. Faisal, I. Budiman, F. Abadi, M. Haekal, D. T. Nugrahadi, A comparison of word embedding-based extraction feature techniques and deep learning models of natural disaster messages classification, *Journal of Computer Sciences Institute*, 27 (2023) 145–153, <https://doi.org/10.35784/jcsi.3322>

- [14] M. Khairie, M. R. Faisal, R. Herteno, I. Budiman, F. Abadi, and M. I. Mazdadi, The Effect of Channel Size on Performance of 1D CNN Architecture for Automatic Detection of Self-Reported COVID-19 Symptoms on Twitter, in 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA) (2023) 621–625, <https://doi.org/10.1109/ISITIA59021.2023.10220444>
- [15] M. R. Faisal, I. Budiman, F. Abadi, D. T. Nugrahadi, M. Haekal, and I. Sutedja, Applying Features Based on Word Embedding Techniques to 1D CNN for Natural Disaster Messages Classification, 2022 5th International Conference on Computer and Informatics Engineering, IC2IE 2022, (2022) 192–197, <https://doi.org/10.1109/IC2IE56416.2022.9970188>
- [16] G. A. Pradnyana, W. Anggraeni, E. M. Yuniarno, and M. H. Purnomo, Fine-Tuning IndoBERT Model for Big Five Personality Prediction from Indonesian Social Media, in 2023 International Seminar on Intelligent Technology and Its Applications (ISITIA) (2023) 93–98, <https://doi.org/10.1109/ISITIA59021.2023.10221074>
- [17] M. F. Nafiz, D. Kartini, M. R. Faisal, F. Indriani, and T. Hamonangan, Automated Detection of COVID-19 Cough Sound using Mel-Spectrogram Images and Convolutional Neural Network, Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), 9 (3) (2023) 535–548, <http://dx.doi.org/10.26555/jiteki.v9i3.26374>
- [18] K. Y. Halim, D. T. Nugrahadi, M. R. Faisal, R. Herteno, and I. Budiman, Gender Classification Based on Electrocardiogram Signals Using Long Short Term Memory and Bidirectional Long Short Term Memory, Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI), 9 (3) (2023) 606–618, <http://dx.doi.org/10.26555/jiteki.v9i3.26354>