IJoICT
ISSN 2356-5462
https://journals.telkomuniversity.ac.id/ijoict

Intl. Journal on ICT
Vol. 12, No. 1, Jun 2026. pp. 92-99
doi.org/10.21108/ijoict.v12i1.11389

Classification of Extroverted and Introverted Personality based on the "*Peringatan Darurat*" response observed on Social Media X using IndoBERT

Aura Sabina Pranika[1], Yuliant Sibaroni[2]*

[1,2,3] *School of Computing, Telkom University*
Jl. Telekomunikasi No. 1 Terusan Buah Batu, Bandung, Jawa Barat, Indonesia,40257

*yuliant@telkomuniversity.ac.id

Abstract

The Trending topics such as " *Peringatan Darurat* " on X (formerly Twitter) generated widespread public discussion, reflecting diverse personality traits, such as extroversion and introversion. These responses generated linguistic patterns that could be used to identify the responder's personality traits through natural language processing. This study proposes a personality classification model to classify user responses into extroversion and introversion categories using a pre-tuned IndoBERT model. The main challenge in this task was the extreme class imbalance in the dataset, which consisted of 2,070 extrovert samples and only 160 introvert samples. To mitigate the model's bias toward the majority class, a class weighting strategy was incorporated into the CrossEntropy loss function during the fine-tuning process. Experimental evaluation on the test dataset yielded an overall accuracy of 82%, a weighted F1 score of 85%, a macro F1 score of 55%, and an ROC-AUC of 69%. The proposed model achieved strong performance in identifying the majority extrovert class, with a precision of 94%, a recall of 86%, and an F1 score of 90%, but it still performed poorly on the minority class (introvert). These results indicate that the class weighting used in the IndoBERT model is very effective but still falls short on the minority class. These findings suggest that more sophisticated imbalance handling techniques are still needed to achieve more balanced personality classification performance.

Keywords: Personality classification, extrovert, introvert, IndoBERT, social media, communication styles

I. INTRODUCTION

The Indonesian public is currently faced with various political issues that have sparked intense debate, particularly regarding the Constitutional Court's (MK) rulings on the 2024 regional head elections in cases No. 60/PUU-XXII/2024 and No. 70/PUU-XXII/2024 [1]. This decree addresses the requirements for political parties to nominate regional leadership candidates and the minimum age limit for such candidates. These decisions came under intense scrutiny following attempts by several parties to overturn the Constitutional Court's ruling, sparking public concern about the integrity of democracy and the fairness of elections in Indonesia.
In response, the public widely shared the "Emergency Warning" symbol on social media platform X (the Garuda symbol on a blue background), generating over 118,000 posts featuring the phrase [2][3]. Furthermore, the hashtag #KawalPususanMK was widely used to mobilize public participation in safeguarding the democratic process. Public responses to this emergency warning reflected linguistic patterns correlated with behavioral expressions of personality traits, including extroversion and introversion. Extroverted individuals tend to actively participate in public discussions and voice their opinions openly, while introverts are more selective and reflective in their communication.

Received on 03 Jun 2026. Revised on 23 Jun 2026. Accepted and Published on 25 Aug 2026.

In computer science, personality types can be identified through text analysis using Natural Language Processing (NLP). Various techniques have been used for this classification, including Support Vector Machines (SVM), Naïve Bayes, and deep learning-based approaches [4][5]. The use of traditional machine learning methods such as SVM and Naïve Bayes in personality response classification has the limitation of relying on static bag-of-words, thus failing to capture contextual, sarcastic, and emotional nuances. The IndoBERT model can overcome this weakness by utilizing a bidirectional self-attention mechanism specifically trained on Indonesian [6][7]. Due to its ability to capture contextual word meaning bidirectionally, these models provide superior classification performance compared to traditional methods [8].

IndoBERT is a variant of the BERT model trained specifically for the Indonesian language [9]. Numerous studies have demonstrated that IndoBERT achieves high performance across various NLP tasks, including sentiment analysis, text classification, and emotion detection [10][11][12]. Consequently, IndoBERT is highly suitable for personality classification based on X (formerly Twitter) user responses to trending issues like this, given that the dataset comprises short Indonesian-language texts rich in diverse user expressions and opinions [13].

Investigating this phenomenon is crucial; without this research, automated systems will face a severe challenge in detecting public anxiety during political crises due to the heavily nuanced and non-standard nature of Indonesian social media texts. Furthermore, it is important to clarify that due to the situational nature of tweets, this study primarily classifies online communication styles and behavioral expressions that align with extrovert/introvert tendencies rather than long-term, stable psychological traits. To be able to identify the personality of user X based on the response to the hashtag, in this paper, the classification of social media responses X related to introvert and extrovert personalities is carried out using IndoBERT.

II. LITERATURE REVIEW

Table I summarizes several recent studies that use machine learning and deep learning techniques for various text classification tasks. In the majority of previous studies in text classification (sentiment analysis, emotion detection, and personality classification), the use of transformer-based models consistently demonstrated superior performance compared to conventional approaches. This finding confirms the effectiveness of transformer architectures in capturing contextual semantic information from Indonesian text. However, only a small number of studies have specifically examined personality classification, particularly the classification of extrovert and introvert personalities from social media data.

TABLE 1.
RESEARCH IN PERSONALITY CLASSIFICATION AND BERT-BASED CLASSIFICATION

Author	Dataset	Method	Task	Result
Fikry & Yusra [6]	Twitter	SVM	Personality Classification	Successfully classified introvert and extrovert users
Khairani et al. [11]	Instagram Comments	IndoBERT, IndoBERTweet	Emotion Detection	IndoBERT: 88.81%; IndoBERTweet: 92.54%
Pradhisa & Fadillah [9]	Google Play Reviews	IndoBERT	Sentiment Analysis	Accuracy above 88%
Widiansyah et al. [10]	M-Paspor Reviews	IndoBERT	Aspect-Based Sentiment Analysis	High classification performance
Baharuddin & Naufal [12]	Indonesian Exam Questions	IndoBERT	Educational Text Classification	Good classification performance
Maulino et al. [13]	Twitter	BERT	Personality Classification	Demonstrated transformer effectiveness

Of these studies, most have neglected the problem of class imbalance, which is common in personality classification datasets. As shown in Table I, existing research primarily emphasizes classification performance without addressing the challenges associated with imbalanced class distributions. This limitation motivates this study, which uses a pre-trained IndoBERT model combined with a class-weighted Cross-Entropy loss function to mitigate bias toward the majority class. By focusing on the classification of extrovert and introvert personalities in a highly imbalanced dataset, this study aims to provide a more robust personality classification framework under such imbalanced conditions.

III. RESEARCH METHOD

A series of processes in this research were carried out including dataset collection, data labeling, data preprocessing, IndoBERT fine tuning and model evaluation. Below is the explanation of each process.

A. *Dataset Collection*

The research data was obtained through a scraping process on the social media platform X using the keywords "*Pilkada 2024*" and "*Peringatan Darurat*". The resulting dataset consists of 2.230 Indonesian-language tweets collected in 2024, which were subsequently categorized into two target classes: introvert and extrovert.

B. *Data Labelling*

Data labeling is the process of assigning specific categories or classes to raw data to facilitate machine learning classification tasks [6]. In this study, manual labeling was performed on tweets collected from social media platform X to differentiate between two personality introvert and extrovert.The introvert label was assigned to tweets reflecting individuals who tend to be more reflective, cautious in expressing opinions, and less inclined to emphasize open social interaction. In contrast, the extrovert label was designated for tweets exhibiting characteristics of individuals who are more open in expressing their opinions, highly communicative, and actively engaged with their social environment [4].

The manual labeling process by four evaluators followed a strict annotation guideline: tweets containing high-frequency punctuation (e.g., multiple exclamation marks), direct confrontational syntax, and highly expressive assertions were labeled as *Extrovert*. Tweets characterized by analytical, reserved, passive-reactive, or low-profile commentary were labeled as *Introvert*. To measure inter-annotator reliability, **Fleiss' Kappa** was calculated, yielding a score of $\kappa = 0.82$, indicating substantial agreement and high consensus consistency.

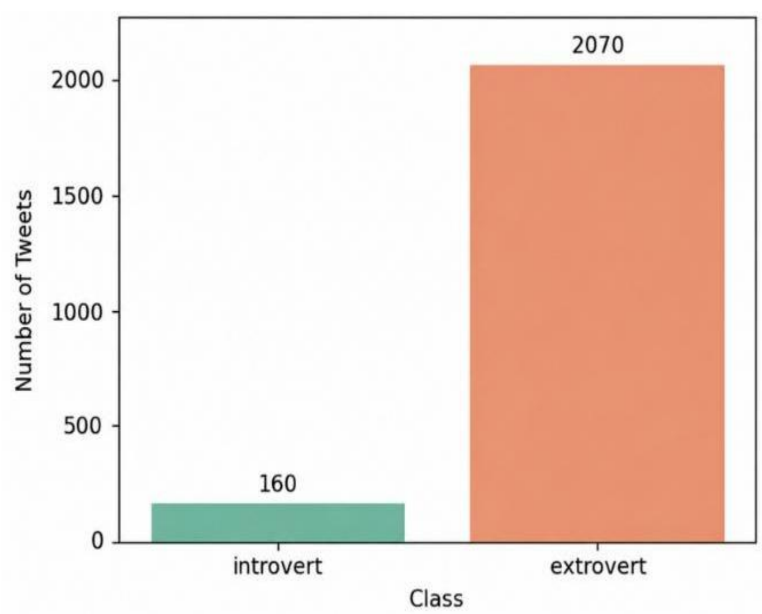

Fig 1. Distribution of personality class

The labeling process was carried out by four evaluators consisting of the researcher and three assistants. The final labeling results produced 160 tweets categorized as introvert and 2.070 tweets categorized as extrovert, resulting in a total of 2.230 labeled tweets. Based on the label distribution, the extrovert class contains a larger volume of data than the introvert class. This disparity indicates a class imbalance, which may cause the model to perform better at identifying extroverted traits. This condition is a primary factor driving the performance discrepancy between the two classes, which will be further discussed in the model evaluation section.

The dataset used exhibits a very significant level of class imbalance, with 2,070 extrovert data points (the majority class) compared to 160 introvert data points (the minority class), or approximately a 13:1 ratio. By default, loss functions such as cross-entropy loss treat prediction errors for each sample with the same weight ($\omega = 1.0$). As a result, during the backpropagation process, model weight updates are dominated by errors from the majority class, causing the model to become “lazy” and predict the majority of the data as the Extrovert class to maximize accuracy. To address this without having to manipulate the number of physical data samples (such as through oversampling or under-sampling), the Class-Weighting technique is applied to the Cross-Entropy Loss function.

C. *Data Preprocessing*

Data preprocessing was performed to enhance the quality of the collected text data prior to model training. This stage consisted of four sequential steps: data cleaning, stop-word removal, lemmatization, and emoticon removal.

1) Data Cleaning

   This step involved removing irrelevant elements from the tweets, such as URLs, mentions, hashtags, punctuation marks, numbers, and special characters. The primary objective of data cleaning is to produce a refined text structure and minimize noise within the dataset.

2) Stop – Word Removal

   Stopwords refer to frequently occurring words that do not reflect specific personality traits. Removing these words allows the model to prioritize more meaningful words that reflect specific personality traits.

3) Lemmatization

   Lemmatization is the process of converting words to their basic form. This process is carried out to minimize the variation in word features that actually have identical core meanings, thereby increasing the robustness of these features.

4) Emoticon Removal

   Emoticon removal is carried out to focus the feature extraction and processing process on text features only.

Although transformer models generally require minimal preprocessing, stopword removal and lemmatization were still incorporated into this study considering that Indonesian tweets about politics contain many unique features, highly irregular and non-standard slang, and recurring local structural inconsistencies.

D. *IndoBERT Fine – Tuning*

IndoBERT fine-tuning is the process of fine-tuning the IndoBERT model using the dataset in this study. Built on the Transformer architecture, IndoBERT uses a bidirectional self-attention mechanism to capture deep contextual representations in Indonesian text [15].

The IndoBERT model was specifically fine-tuned for personality classification using an emergency alert response text dataset. The input text was first tokenized and processed through an IndoBERT layer, which was then passed to a fully connected classification layer equipped with a softmax activation function to generate the probability of an introvert or extrovert class. Next, hyperparameter tuning, learning rate, batch size, and number of epochs were performed to optimize model performance.

E. *Model Evaluation*

Model evaluation was performed using the 10-fold cross-validation method to ensure more robust and stable results. 10-fold cross-validation is a machine learning model evaluation method that divides the dataset into 10 parts (folds). All of these parts (folds) are then used to train and test the model in turn. The evaluation results are then averaged to obtain a more stable and fairer performance estimate. This method is expected to increase the credibility of the final performance claims of the evaluated model. Model performance was comprehensively assessed using a confusion matrix, with metrics including accuracy, precision, recall, and F1-score (Macro-Averaged F1-Score and Weighted-Average F1-Score). The formula for accuracy, recall, precision, and F1 score can be seen in equations (1) to (6).

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

$$Precison = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Where:
TP (True Positive) : positive instances correctly classified as positive.
TN (True Negative) : negative instances correctly classified as negative.
FP (False Positive) : negative instances incorrectly classified as positive.
FN (False Negative) ; positive instances incorrectly classified as negative.

The Macro-Averaged F1-Score calculates the arithmetic mean of the F1-Scores for each class (Extrovert and Introvert) separately, with each class given equal weight (1:1). The formula is:

$$F1_{Macro} = \frac{1}{K}\sum_{c=1}^{K} F1_c = \frac{(F1_{Ekstrovert} + F1_{Introvert})}{2} \quad (5)$$

Variable descriptions:
$F1_{Macro}$: Average macro F1-Score
K : Total number of classified classes (K = 2, namely Extrovert and Introvert)
$F1_c$: F1-Score value for class c
$F1_{Extrovert}$: F1-Score value obtained for the *Extrovert* class
$F1_{Introvert}$: F1-Score value obtained for the *Introvert* class

The Macro-Averaged F1-Score approach is particularly appropriate for use on datasets with imbalanced class proportions, as the Macro-Average provides an objective measure of how well the model can predict the minority class (Introverts) without being affected by high performance in the majority class (Extroverts).

The weighted-average F1-score calculates the average F1-score by taking into account the proportion of the sample size (support) in each class relative to the total dataset. The formula is:

$$F1_{Weighted} = \left(\frac{N_{Ekstrovert}}{N_{total}} \times F1_{Ekstrovert}\right) + \left(\frac{N_{Introvert}}{N_{total}} \times F1_{Introvert}\right) \quad (6)$$

Variable descriptions:
$F1_{Weighted}$: Weighted average F1-Score
K : Total number of classes (K = 2, namely Extrovert and Introvert)
Nc : Number of samples (support) in class c ($N_{Extrovert}$ = 2,070 data points, $N_{Introvert}$ = 160 data points)
N_{total} : Total number of data points (2,230 data points)
$F1_c$: F1-Score obtained for class c (Introvert /extrovert)

IV. RESULTS AND DISCUSSION

Two analyses evaluated the performance of the classification model. The first used four pairs of metrics: accuracy, precision, recall, and F1-score, and the second used ROC-AUC evaluation. Furthermore, an analysis of the model's performance in cases of data imbalance was also conducted.

a. The Performance of IndoBERT in Personality Text Classification

To obtain a more robust evaluation, 10–fold cross-validation was employed by partitioning the dataset into ten equal subsets. In each iteration, nine subsets were used for training and one subset for testing. The procedure was repeated ten times so that each subset served as the test set once. The results of performance measurements using 4 evaluation metrics: accuracy, precision, recall and F1-score can be seen in Table II.

TABLE II
THE PERFORMANCE OF INDOBERT PER FOLD.

Fold	Accuracy	Precision	Recall	F1 - Score
1.	76.68%	92.82%	81.16%	86.60%
2.	87.00%	94.06%	91.79%	92.91%
3.	82.96%	92.89%	88.41%	90.59%
4.	84.30%	96.24%	86.47%	91.09%
5.	84.30%	95.26%	87.44%	91.18%
6.	74.44%	93.10%	78.26%	85.04%
7.	80.27%	94.54%	83.57%	88.72%
8.	86.55%	94.03%	91.30%	92.65%
9.	81.61%	95.11%	84.54%	89.51%
10.	78.92%	94.44%	82.13%	87.86%
Average	81.70%	94.25%	85.51%	89.62%

Model performance was evaluated based on 10-fold cross-validation tests to ensure the stability of the results. Overall, the IndoBERT model demonstrated consistent performance with an average accuracy of 81.70%, a precision of 94.25%, a recall of 85.06%, and an F1-score of 89.62%. The high precision and F1-score values indicate that the IndoBERT model is appropriate for classifying responses related to introvert and extrovert personality traits.

Table III presents the classification performance of the proposed model for 2,230 test instances, consisting of 160 introvert samples and 2,070 extrovert samples. The model achieved an overall accuracy of 82%, demonstrating good performance when evaluated on the entire dataset. However, these results should be interpreted with caution, as the highly imbalanced dataset requires more specific analysis of each class.

TABLE III
CLASSIFICATION PERFORMANCE OF EACH CLASS

Class	Precision	Recall	F1-Score	Support
Introvert	0.15	0.33	0.20	160
Extrovert	0.94	0.86	0.90	2070
Macro Average	0.55	0.59	0.55	2230
Weighted Average	0.89	0.82	0.85	2230
Accuracy			0.82	

The per-class evaluation revealed a significant difference in performance between the two classes. The model effectively classified the extroverted majority class, achieving an F1 score of 90%, but only achieved an F-1 score of around 20% for the introverted class, indicating that the model still failed to recognize a significant portion of the minority class instances. The macro F1 score of 55%, which gives equal weight to both classes, is significantly lower than the weighted F1 score of 85%, which is dominated by the majority class. This suggests that the high overall accuracy is primarily driven by the model's strong performance on the majority (extroverted) instances rather than balanced classification across both classes.

These findings confirm that class imbalance affects overall classification performance. Although the model achieved high accuracy, its ability to detect the minority (introverted) class remains limited. Consequently, additional imbalance-adjustment techniques, such as oversampling, cost-sensitive learning, or hybrid deep learning approaches, are needed to improve minority class recognition while maintaining overall performance.

b. ROC – AUC and Discussion

The ROC-AUC evaluation results show that the proposed IndoBERT model achieved an average AUC of 69%. Based on these results, it can be concluded that the proposed model has a fairly good ability to distinguish between introvert and extrovert personality classes, although it still has significant potential for further improvement.

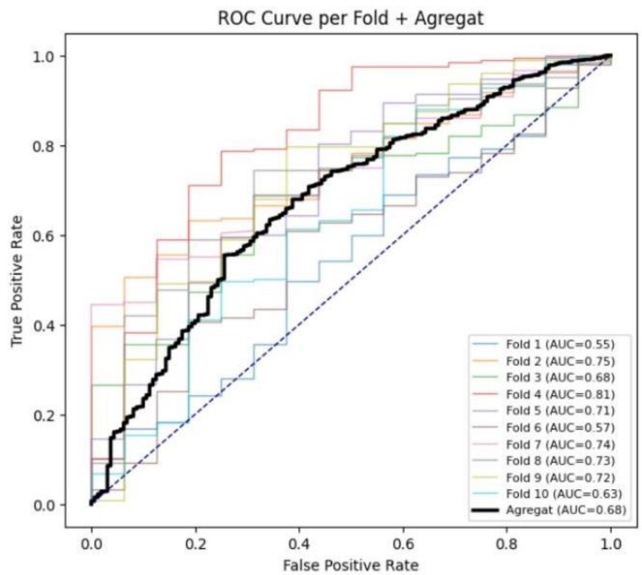

Fig 3. ROC Curve

c. Threats to Validity & Imbalance Discussion

Based on the test results, the IndoBERT model demonstrated very strong performance in the Extrovert class, with an F1 score of 90%. This superiority is supported by a sufficient training sample size of 2,070 data points, which allows the model to better extract variations in linguistic features, word context, and sentence structure that characterize extroverts. In contrast, in the Introvert class, the model only achieved a low F1 score of 20%. Although class weighting techniques were applied to give more attention to the minority class, performance in this class remained relatively low. This limitation is due to the small sample size of only 160 data points, which is not enough for a Transformer-based architecture like IndoBERT to fully learn language patterns. Additionally, the similarity in vocabulary (contextual overlap) resulting from the uniformity of the "Emergency Alert" topic causes the keywords and political terms used by both personality types to often be the same. The differences between the two lie only in nuances and emotional subtleties, which require a larger dataset for the model to distinguish them consistently. Based on the aggregate confusion matrix analysis, the model successfully classified 1,822 samples correctly out of a total of 2,230 data points. However, there were classification errors in which 300 Extrovert data points were predicted as Introvert, and 108 Introvert data points were predicted as Extrovert. The high error rate in the majority class being predicted as the minority class is a direct result of the application of fairly aggressive class-weighting, causing the model to be more inclined to predict the minority class in order to minimize high loss values. Meanwhile, the ROC-AUC evaluation yielded an average AUC value of 69%, indicating that the model exhibits moderate class separability. This value is significantly above that of a random guess (0.50), which proves and validates that the IndoBERT model successfully learns personality signals and patterns from text, even though its overall performance is still limited by the dataset imbalance.

I. CONCLUSION

Based on the results of the testing and analysis conducted, it can be concluded that the use of the IndoBERT model successfully classified the personality types of social media users on X based on their responses to the "Emergency Alert" issue, achieving an aggregate accuracy of 82%, a weighted average F1-score of 85%, and an ROC-AUC value of 0.6896. The evaluation results show that the "Emergency Alert" issue effectively provides a contrasting spectrum of language styles, where the extrovert group tends to express criticism and appeals demonstratively and openly, while the introvert group conveys opinions in a more reflective and analytical manner. However, the imbalance in the number of data samples between the Extrovert class (2,070 samples) and the Introvert class (160 samples) is the primary factor limiting the model's optimal performance. Although the application of class-weighting techniques to the loss function has helped give greater attention to the minority class, the classification performance for the Introvert class remains relatively low due to limited data variation and high overlap in political vocabulary within the homogeneous research topic.

REFERENCES

[1] Mahkamah Konstitusi Republik Indonesia, "Putusan Nomor 60/PUU-XXII/2024 dan Nomor 70/PUU-XII/2024," 2024. [Online]. Available: https://www.mkri.id/putusan/60-70-2024

[2] Tim detikcom, "Peringatan Darurat Garuda Biru Itu 'Dinyalakan' Publik," *detikJabar*, Aug. 21, 2024. [Online]. Available: https://www.detik.com/jabar/berita/d-7501979/peringatan-darurat-garuda-biru-itu-dinyalakan-publik

[3] A. Milagsita, "Viral Peringatan Darurat Garuda Biru di Media Sosial, Apa Maksudnya?" *detikJateng*, Aug. 21, 2024. [Online]. Available: https://www.detik.com/jateng/berita/d-7501809/viral-peringatan-darurat-garuda-biru-di-media-sosial-apa-maksudnya

[4] S. Anggraini and M. Abidin, "Sudut pandang individu introvert dan ekstrovert dalam berinteraksi sosial," *JSHP: Jurnal Sosial Humaniora dan Pendidikan*, vol. 9, no. 1, pp. 105–118, 2025, doi: 10.32487/jshp.v9i1.2156.

[5] A. Subtinanda and N. Y., "Kepribadian ekstrovert dan introvert dalam konteks komunikasi antarpribadi mahasiswa Ilmu Komunikasi UNTIRTA," *JPN*, pp. 1–15, 2023. [Online]. Available: https://edu.pubmedia.id/index.php/jpn

[6] M. Fikry and Yusra, "Ekstrovert atau introvert: Klasifikasi kepribadian pengguna Twitter dengan metode support vector machine," *SITEKIN*, pp. 72–76, 2018. [Online]. Available:http://ejournal.uin-suska.ac.id/index.php/sitekin

[7] V. G. dos Santos and I. Paraboni, "Myers-Briggs Personality Classification from Social Media Text Using Pre-trained Language Models," *arXiv:2207.04476*, Jul. 2022. [Online]. Available: https://arxiv.org/abs/2207.04476

[8] B. J. and Yulianto, "Indonesian news classification using IndoBERT," *Intelligent Systems and Applications in Engineering*, pp. 454–460, 2022. [Online]. Available: http://www.ijisae.org

[9] K. C. Pradhisa and R. F., "Analisis sentimen ulasan pengguna e-commerce di Google Play Store menggunakan metode IndoBERT," *BITS*, vol. 6, no. 1, pp. 92–104, 2024, doi: 10.47065/bits.v6i1.5247.

[10] M. Widiansyah, F. F. Az-Zahra, and A. Pambudi, "Fine-tuning model IndoBERT untuk analisis sentimen berbasis aspek pada aplikasi M-Paspor," *Jurnal Informatika Unisla*, pp. 183–195, 2023. [Online]. Available: https://jurnalteknik.unisla.ac.id/index.php/informatika

[11] U. Khairani, V. Mutiawani, and H. Ahmadian, "Pengaruh tahapan preprocessing terhadap model IndoBERT dan IndoBERTweet untuk mendeteksi emosi pada komentar akun berita Instagram," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 11, no. 4, pp. 887–894, 2024, doi: 10.25126/jtiik.1148315.

[12] F. Baharuddin and M. F. Naufal, "Fine-tuning IndoBERT for Indonesian exam question classification based on Bloom's taxonomy," *JISEBI*, pp. 253–263, 2023. [Online]. Available:https://e-journal.unair.ac.id/index.php/JISEBI

[13] Y. A. Maulino, W. Maharani, and P. H. Gani, "Personality Classification on Twitter Social Media using BERT," *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, pp. 520–528, 2023, doi: 10.30865/mib.v7i1.559.

[14] H. Izzatilla and Z. Fatah, "Klasifikasi kepribadian introvert dan ekstrovert menggunakan k-nearest neighbors berdasarkan data perilaku sosial," *Jurnal Mahasiswa Teknik Informatika*, vol. 4, no. 2, 2025. [Online]. Available: https://jurnal.unw.ac.id/index.php/jamastika/article/view/4483

[15] 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 2019*, Minneapolis, Minnesota, USA, 2019, pp. 4171–4186.