JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.7 No.1 January 2025: 45-50
e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12618

Effectiveness Analysis of RoBERTa and DistilBERT in Emotion Classification Task on Social Media Text Data

Ghinaa Zain Nabiilah

Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
ghinaa.nabiilah@binus.ac.id

Correspondence: ghinaa.nabiilah@binus.ac.id

Abstract – *The development of social media provides various benefits in various ways, especially in the dissemination of information and communication. Through social media, users can express their opinions, or even their feelings. In this regard, sometimes users also convey information or opinions according to the user's feelings or emotions. This triggers the impact of aggressive online behavior, including cyberbullying, which triggers unhealthy debates on social media. The development of deep learning models has also been developed in several ways, especially emotion classification. In addition to using deep learning models, the development of classification tasks has also been carried out using transformer architectures, such as BERT. The development of the BERT model continues to be carried out, so this study will analyze and explore the application of BERT model development, such as RoBERTa and DistilBERT. The optimal result of this study is with an accuracy value of 92.69% using the RoBERTa model.*

Keywords: DistilBERT, Emotion Classification, RoBERTa

I. INTRODUCTION

Social media has become a major platform for sharing information, opinions and emotions (Bhimani et al., 2019). Despite offering many benefits, such as connectivity and the spread of information, the expression of emotions on social media also has the potential for negative impacts. Posts filled with negative emotions, such as anger or hatred, tend to spread faster than neutral content. This can trigger divisions, increase polarization in society, and affect social relationships between individuals and groups. Hate that is continuously posted can also lead to real-world conflicts, such as violence or riots (Bayer et al., 2020).

In addition, negative emotions such as anger or frustration are often the cause of aggressive online behavior, including cyberbullying. Victims of cyberbullying can experience emotional distress, low self-esteem, anxiety, or even depression. In extreme cases, this can lead to fatal actions, such as suicide. Emotional content, especially those containing fear or anger, is more likely to influence users to believe and spread false information. Misinformation driven by these emotions can harm individuals, organizations, and even threaten social stability (Nisar et al., 2019).

Continuous interaction with negative emotional content can affect the mental health of social media users. Fear, envy, or anxiety triggered by certain posts can worsen an individual's psychological condition, especially if users often compare themselves to others (Naslund et al., 2020).

Based on this, it is necessary to classify the content shared by users on social media to help reduce the impact of the spread of emotional content on social media.

The development of deep learning models has brought significant progress in text classification tasks, with the presence of architectures such as RNN (Recurrent Neural Networks), LSTM (Long Short-Term Memory), and transformer models such as BERT (Bidirectional Encoder Representations from Transformers) (Cust et al., 2019). These models are able to understand the context and relationships between words in text in more depth, thereby improving accuracy and efficiency in a variety of applications, including sentiment analysis, emotion classification, and information extraction (Dong et al., 2021) Moreover, transformer-based models have become the new standard in natural language processing (NLP), enabling the handling of large-scale and multi-language data with optimal results (Wolf et al., 2020).

One application of the Transformer model is to use the BERT model for text classification using COVID-19 fake news data. Based on this study, the BERT model is able to provide an accuracy of 99.56 using the BERT – Base model type. Meanwhile, if using the COVID-19 English tweet dataset data type, it obtains an accuracy of 98.44 using the same BERT model type (Qasim et al., 2022).

In addition to BERT Base, the development of the BERT model is also ongoing, so that there are various other models ready to be used, such as Multilingual BERT, in emotion classification using Hindi text, the MBERT model is able to provide optimal accuracy of 93.88% (Kumar et al., 2023).

So based on this, this study aims to conduct Analysis and exploration of the development of other BERT models such as RoBERTa and DistillBERT to evaluate the most optimal performance in emotion classification based on social media data. Several previous studies have used the RoBERTa model for sentiment classification, this study combines RoBERTa with Recurrent Neural Networks (RNN) with optimal results of 95% (Cheruku et al., 2023).

Research conducted by Pingshan Liu et al, also used the RoBERTa model to classify emotions, the optimal results of this study were with an accuracy of 99% (P. Liu & Lv, 2023).

In addition to the BERT and RoBERTa models, DistilBERT is also known as one of the developments of the BERT model that is lightweight but has optimal results. One example is the research conducted by Mario Jojoa et al, who used the distilBERT model for sentiment classification. The optimal result is with an accuracy value of 0.823 (Jojoa et al., 2024).

Other studies also explored the application of DistilBERT for text classification indicating symptoms of Mental Health problems. The results of the study obtained an optimal accuracy of 96% (Diwakar & Raj, 2024).

II. METHODS

This study was conducted by comparing the application of the RoBERTa and DistilBERT models for emotion classification based on text data on social media. The data used in this study grouped texts into several groups such as happy, sadness, anger, fear, love, surprise. Figure 1 contains an overview of the stages of the research carried out.

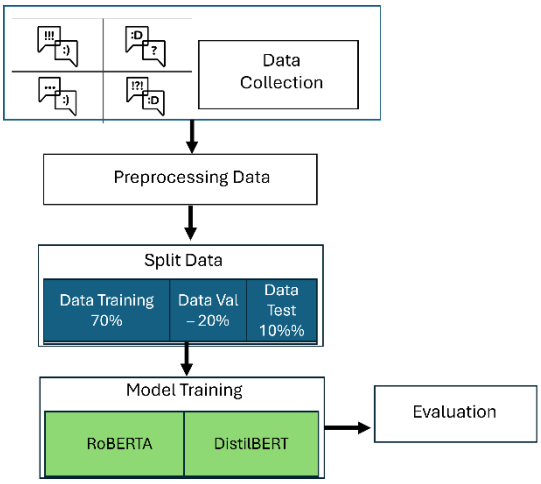

Figure 1. Research Flow

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12618

The research stages were carried out by collecting text data shared on social media. This data has been given several labels such as happy, sadness, anger, fear, love, surprise. After that, to get clean data, the data preprocessing process was carried out. In addition to cleaning the data, the preprocessing stage needs to be carried out to get consistent data, and reduce bias so that it is easier to process so that patterns in the data can be analyzed at the classification stage.

2.1 Dataset & Preparation

The dataset used in this study comes from social media texts that have been labeled and can be publicly accessed via Kaggle (Elgiriyewithana, 2024). The number of datasets is 21,459. Figure 2 contains the distribution of labels from the data.

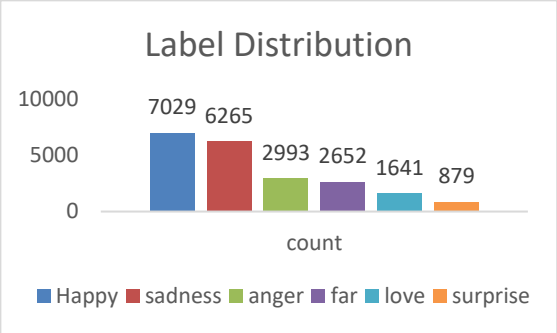

Figure 2. Label Distribution

The stages also carried out in this study are data preprocessing. In data preprocessing, the stages carried out are as follows

- Noise Removal, to remove punctuation, numbers, links or excess spaces.
- Case Folding, to change text to the same capitalization
- Tokenization, to break sentences into words
- Stemming, to remove affixes or change sentences into basic words.
- Stopword, to remove words that appear frequently but do not have a specific meaning.

After this stage, the data is then divided into three parts, namely train data, validation data and test data. The data division is carried out with a proportion of 70% train data, 20% validation data and 10% testing data.

2.2 RoBERTa

RoBERTa (Robustly Optimized BERT Approach) is a transformer-based model developed as an extension of BERT (Bidirectional Encoder Representations from Transformers). RoBERTa is designed to improve the performance of the BERT model by optimizing the training process and data usage. Unlike BERT, which uses the masked language modeling (MLM) and next sentence prediction (NSP) approaches, RoBERTa eliminates the NSP component and focuses entirely on MLM with larger data scales and longer training times (Y. Liu et al., 2019).

Additionally, RoBERTa improves efficiency by using larger batch sizes, more training steps, and tuning hyperparameters to maximize performance. RoBERTa shows significant advantages in a variety of natural language processing (NLP) tasks, including text classification, sentiment analysis, and information extraction. The model is able to capture deeper and more robust language representations, making it superior in understanding complex contexts and linguistic nuances (Malik et al., 2023).

2.2 DistilBERT

DistilBERT is a lightweight version of the BERT (Bidirectional Encoder Representations from Transformers) model developed by Hugging Face with the aim of reducing computational complexity while maintaining performance close to the original

model. DistilBERT uses a knowledge distillation technique, where a smaller model is trained to replicate the behavior of a large model by extracting and condensing knowledge from BERT (Sanh et al., 2019).

This process allows DistilBERT to have about 40% fewer parameters and run 60% faster than BERT, while still maintaining about 97% of its original accuracy on a variety of NLP tasks. DistilBERT retains BERT's transformer architecture but reduces the number of encoder layers from 12 to 6. The model is optimized to capture language representations with high efficiency without sacrificing too much semantic detail (Li, 2024).

III. RESULTS AND DISCUSSION

This research was conducted using Google colab with 10 epochs, 16 batch sizes, adam optimizer, and 1e-5 learning rates. The type of RoBERTa model used is RoBERTa Base while the DistilBERT model used is distilbert-base-uncased. Figure 3 and Figure 4 contain the results of model evaluation using training data, where the classification results of both provide quite optimal results.

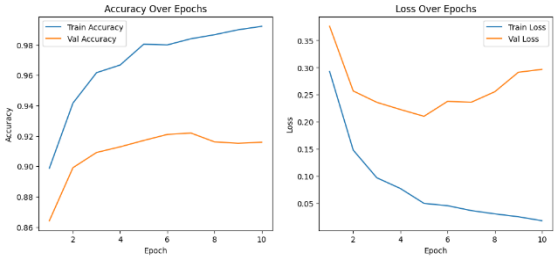

Figure 3. DistilBERT Training Result

Based on the results of the experiments conducted on the training data, the distilBERT model provides quite optimal accuracy. However, if viewed at a larger number of epochs, the model tends to experience an increasing difference between the training data and the validation data. Although the distance is not too far, it is still below 10%, so it can be concluded that there is an indication that the model is overfitting although not extreme. Meanwhile, the results of the model evaluation on the training data using RoBERTa are shown in Figure 4.

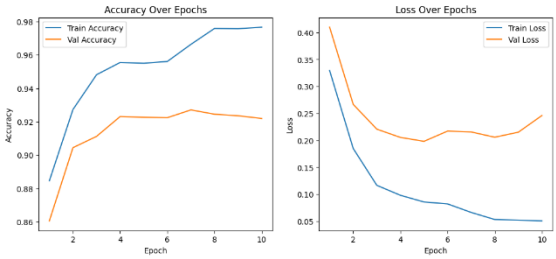

Figure 4. RoBERTa Training Result

Based on the experiments conducted on the RoBERTa model, it turns out that the results of the model evaluation on RoBERTa are also almost the same as the DistilBERT model. Meanwhile, the model evaluation on the testing data is in table 1.

Table 1 Testing Evaluation Result

Model	Precission	Recall	Acuracy	F1 - Score
DistilBERT	90%	90%	92.64%	90%
RoBERTa	**89%**	**91%**	**92.69%**	**90%**

Based on the results of the evaluation of the testing data, both models have the same accuracy and F1-Score results, namely F1-Score 90%. The resulting accuracy also has a very small difference. So it can be concluded that both models have optimal accuracy for classifying emotional data text. The development of the BERT model, namely RoBERTa and DistilBERT, is able to provide optimal values, although the distribution of labels in the data tends to be unbalanced. However, a fairly large amount of data can also affect this. In addition, the use of sufficient epoch values, selection of batch sizes, determination of the type of optimizer and learning rate can also affect the model in making it easier to find patterns in the text so that it can classify emotions based on the data.

IV. CONCLUSION

Based on the experiments conducted, the RoBERTa and DistilBERT models were able to provide optimal accuracy on emotional text data. Although it provides optimal results, the model tends to experience indications of overfitting, although not extreme. The results of this study prove that adjusting the number of parameters and sizes in the RoBERTa and

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12618

DistilBERT models does not affect model performance.

However, in this study, both of them gave optimal results. Although it provides optimal results, the tendency for overfitting in the model cannot be ignored. So that for further research, it can explore the application of data preprocessing or techniques to handle the possibility of overfitting in this experiment, although the overfitting in this study was not extreme.

REFERENCES

Bayer, J. B., Triệu, P., & Ellison, N. B. (2020). Social Media Elements, Ecologies, and Effects. *Annual Review of Psychology*, *71*(1), 471–497. https://doi.org/10.1146/annurev-psych-010419-050944

Bhimani, H., Mention, A.-L., & Barlatier, P.-J. (2019). Social media and innovation: A systematic literature review and future research directions. *Technological Forecasting and Social Change*, *144*, 251–269. https://doi.org/10.1016/j.techfore.2018.10.007

Cheruku, R., Hussain, K., Kavati, I., Reddy, A. M., & Reddy, K. S. (2023). Sentiment classification with modified RoBERTa and recurrent neural networks. *Multimedia Tools and Applications*, *83*(10), 29399–29417. https://doi.org/10.1007/s11042-023-16833-5

Cust, E. E., Sweeting, A. J., Ball, K., & Robertson, S. (2019). Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance. *Journal of Sports Sciences*, *37*(5), 568–600. https://doi.org/10.1080/02640414.2018.1521769

Diwakar, & Raj, D. (2024). *DistilBERT-based Text Classification for Automated Diagnosis of Mental Health Conditions* (pp. 93–106). https://doi.org/10.1007/978-981-99-9621-6_6

Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review*, *40*, 100379. https://doi.org/10.1016/j.cosrev.2021.100379

Elgiriyewithana, N. (2024). Emotions. *Https://Www.Kaggle.Com/Datasets/Nelgiriyewithana/Emotions*.

Jojoa, M., Eftekhar, P., Nowrouzi-Kia, B., & Garcia-Zapirain, B. (2024). Natural language processing analysis applied to COVID-19 open-text opinions using a distilBERT model for sentiment categorization. *AI & SOCIETY*, *39*(3), 883–890. https://doi.org/10.1007/s00146-022-01594-w

Kumar, T., Mahrishi, M., & Sharma, G. (2023). Emotion recognition in Hindi text using multilingual BERT transformer. *Multimedia Tools and Applications*, *82*(27), 42373–42394. https://doi.org/10.1007/s11042-023-15150-1

Li, B. (2024). A Study of DistilBERT-Based Answer Extraction Machine Reading

Comprehension Algorithm. *Proceedings of the 2024 3rd International Conference on Cyber Security, Artificial Intelligence and Digital Economy*, 261–268. https://doi.org/10.1145/3672919.3672968

Liu, P., & Lv, S. (2023). Chinese RoBERTa Distillation For Emotion Classification. *The Computer Journal*, *66*(12), 3107–3118. https://doi.org/10.1093/comjnl/bxac153

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. http://arxiv.org/abs/1907.11692

Malik, M. S. I., Nazarova, A., Jamjoom, M. M., & Ignatov, D. I. (2023). Multilingual hope speech detection: A Robust framework using transfer learning of fine-tuning RoBERTa model. *Journal of King Saud University - Computer and Information Sciences*, *35*(8), 101736. https://doi.org/10.1016/j.jksuci.2023.101736

Naslund, J. A., Bondre, A., Torous, J., & Aschbrenner, K. A. (2020). Social Media and Mental Health: Benefits, Risks, and Opportunities for Research and Practice. *Journal of Technology in Behavioral Science*, *5*(3), 245–257. https://doi.org/10.1007/s41347-020-00134-x

Nisar, T. M., Prabhakar, G., & Strakova, L. (2019). Social media information benefits, knowledge management and smart organizations. *Journal of Business Research*, *94*, 264–272. https://doi.org/10.1016/j.jbusres.2018.05.005

Qasim, R., Bangyal, W. H., Alqarni, M. A., & Ali Almazroi, A. (2022). A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification. *Journal of Healthcare Engineering*, *2022*, 1–17. https://doi.org/10.1155/2022/3498123

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. http://arxiv.org/abs/1910.01108

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6