ILKOM Jurnal Ilmiah Vol. 14, No. 3, December 2022, pp. 348-354
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN

Herlina Jayadianti [a,1,*]; Wilis Kaswidjanti [a,2]; Agung Tri Utomo [a,3]; Shoffan Saifullah [a,b,4]; Felix Andika Dwiyanto [b,c,5]; Rafal Drezewski [b,6]

[a] *Universitas Pembangunan Nasional Veteran Yogyakarta, Jl. Ring Road Utara No.104 Yogyakarta and 55283, Indonesia*
[b] *AGH University of Science and Technology, Cracow, Poland*
[c] *Universitas Negeri Malang, Jl. Semarang No.5, Sumbersari, Malang and 65145, Indonesia*
[1] *herlina.jayadianti@upnyk.ac.id;* [2] *wilisk@upnyk.ac.id;* [3] *agungtriu@gmail.com;* [4] *shoffans@upnyk.ac.id;* [5]*dwiyanto@agh.edu.pl;* [6]*drezew@agh.edu.pl;*
* *Corresponding author*

Article history: Received October 24, 2022; Revised November 09, 2022; Accepted November 29, 2022; Available online December 20, 2022

Abstract

Reviews are a form of user experience information on a product or service that can be used as a reference for potential consumers' preferences to buy, use, or consume a product. They can be also used by business entities to find out public opinion about their product or the performance of their business products. It will be very difficult to process the review data manually and it will take a long time. Therefore, sentiment analysis automation can be used to get polarity information from existing reviews. In this study, IndoBERT with Recurrent Convolutional Neural Network (RCNN) was used to automate sentiment analysis of Indonesian reviews. The data used was a sentiment analysis dataset obtained from IndoNLU with sentiment consisting of negative sentiment, neutral sentiment, and positive sentiment. The results of the test showed that IndoBERT with the Recurrent Convolutional Neural Network (RCNN) had better results than the IndoBERT base. IndoBERT with Recurrent Convolutional Neural Network (RCNN) obtained 95.16% accuracy, 94.05% precision, 92.74% recall and 93.27% f1 score.

Keywords: Sentiment Analysis; IndoBERT; Recurrent Convolutional Neural Network; Pretained Language Models

Introduction

Reviews are a form of user's sharing experience on a product or service [1]. Reviews can be used as a reference for potential consumers to determine their preferences to buy, use, consume a product or not. By business entities it can be used to find out public's opinion on the product or the performance of its business product. This can be utilized to improve the next business decision. In this digitalization era, consumers often submit reviews of a product or service using video or text. In text reviews, consumers often write them on online platforms, such as social media or comment fields [2]. However, manually processing a large amount of review data would be very difficult and time-consuming, therefore automated sentiment analysis can be employed [3] to derive polarity information from existing reviews.

Popular methods used to perform sentiment analysis are machine learning methods [4], [5], including naive bayes [6], [7], logistic regression [8] and support vector machines [9], [10]. Although the methods show promising performance, manual extraction features are required. Sometimes to perform an in-depth analysis more complex features are required. To overcome these weaknesses, deep learning can be used. It extracts features directly from the data during the training process so that manual extraction is not required.

The most widely used methods In deep learning for sentiment analysis are recurrent neural networks [11], [12], and convolutional neural networks [13], [14]. In the research conducted [15], proposed the recurrent convolutional neural network (RCNN) method. This method is the development of a deep learning by combining the recurrent neural network (RNN) and convolutional neural network (CNN) methods, so that the recurrent convolutional neural network (RCNN) method obtains advantages of the two models. In sentiment analysis, RCNN perform better results than naive Bayes, support vector machines, recurrent neural network, and convolutional neural network [16], [17]

http://dx.doi.org/10.33096/ilkom.v14i3.1505.348-354

and better than the rule-based method [18]. Despite getting promising performance, deep learning methods have a complex architecture so that to get optimal results a large dataset is required. Building a large dataset requires high cost. To overcome this problem, transfer learning techniques can be used, namely by using a pretrained model that has been previously trained with a very large dataset.

The potential pretrained models for Indonesian are multilingual BERT [19], [20], and IndoBERT [21]. For sentiment analysis using multilingual BERT indicated promising results on the English dataset [22], but for sentiment analysis on the Indonesian dataset the performance was not better than Naive Bayes [23]. In multilingual BERT, the BERT model is trained using a dataset from the Wikipedia corpus where the writing is more formal and less noisy. To overcome this problem, a pretrained IndoBERT model can be used which has previously been trained with a larger and more diverse Indonesian language dataset (Indo4B), including online news, social media, Wikipedia, online articles, subtitles from video recordings, and parallel datasets. Sentiment analysis using IndoBERT showed better results compared to multilingual BERT and other pretrained models [21]. Then it obtained better performance compared to kNN, SVM, naive bayes, decision tree, and random forest [24]. Even so, the development of a pretrained model based on the BERT model is still at an early stage, there are still many opportunities to develop the model [25].

The developments of the BERT-based model has been carried out by [12] by conducting experiments on the pretrained multilingual BERT model for sentiment analysis on the Vietnamese language dataset, namely doing fine tuning by using all the outputs of the multilingual BERT pretrained model as input for the classification model. It resulted in multilingual BERT with the recurrent convolutional neural network (RCNN) model as the best performance model compared to multilingual BERT base and multilingual BERT with other classification models.

Based on this background, the current research is an adaptation of research from [12] for sentiment analysis on Indonesian reviews, namely fine-tuning the pretrained IndoBERT model with the recurrent convolutional neural network (RCNN) model to obtain the accuracy value of sentiment analysis on better review on the Indonesian language.

Method

In this study, a transfer learning process was carried out for sentiment analysis by using the pretrained IndoBERT model and fine tuning. All output tokens from the pretrained IndoBERT model would be used as input for the RCNN classification model. An illustration of the architecture of the model can be seen in **Figure 1**.

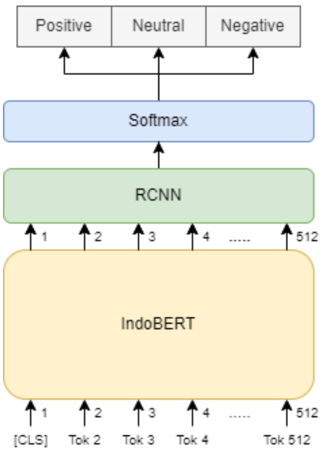

Figure 1. Architecture of IndoBERT-RCNN

This research started with collecting data taken from IndoNLU. The data was the smsa dataset. There were three sentiments in the dataset, namely positive, neutral, and negative. The dataset was a dataset obtained from several social media, including Twitter, Zomato, TripAdvisor, Facebook, Instagram, Qraved. The dataset consisted of 11000 training data, 1260 validation data, and 500 test data. However, in this study, only training data and validation data were used, because the test data on the SMSA dataset from IndoNLU had been masked. The validation data of the SMSA dataset was employed as test data, and the training data of the SMSA dataset would be divided into training data and validation data with a proportion of 9:1.

A. IndoBERT

IndoBERT is a pretrained model introduced by [7]. It is a pretrained model based on the BERT model [19] which was trained with 4 billion words of Indonesian text data derived from online news, social media, Wikipedia, online articles, subtitles from video recordings, and a parallel dataset which was later called Indo4B [26]. BERT or Bidirectional Encoder Representations from Transformers itself is a model with several structural layers of bidirectional transformer encoder based on transformer architecture [27].

In the BERT training framework there were two stages, pretraining and fine tuning [28]. In the pretraining stage, BERT was trained with datasets that did not have labels. At this stage, BERT was trained to use two self-supervised tasks, namely Masked LM (Masked Language Model) and Next Sentence Prediction. BERT would learn to understand a language and its context. For IndoBERT, at the pretraining stage, it was trained using the Indo4B dataset. Then in the fine tuning stage, the trained model in the pre-training stage was adjusted and then trained with a specific dataset that had a label.

For single sentence classification tasks such as sentiment analysis [29], basically the BERT model would only use the output token [C] to predict the label of the data. The illustration for the single sentence classification task can be seen in **Figure 2**.

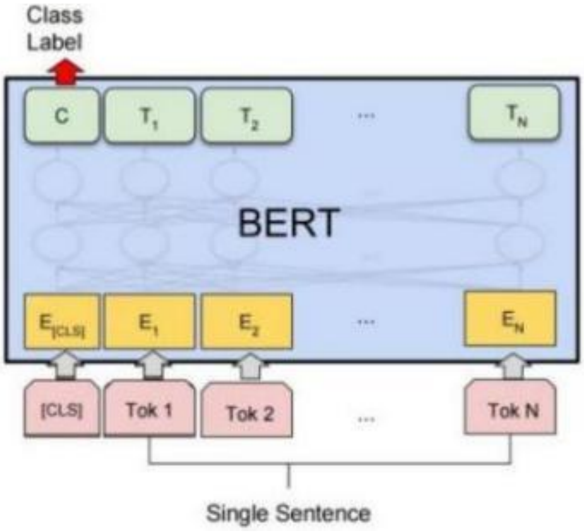

Figure 2. Single Sentence BERT

B. Recurrent Convolutional Neural Network (RCNN)

RCNN or recurrent convolutional neural network is a model that combines recurrent and convolutional structures [15]. By combining both, this model benefits from the recurrent neural network model and the convolutional neural network model. The structure of RCNN can be seen in **Figure 3**.

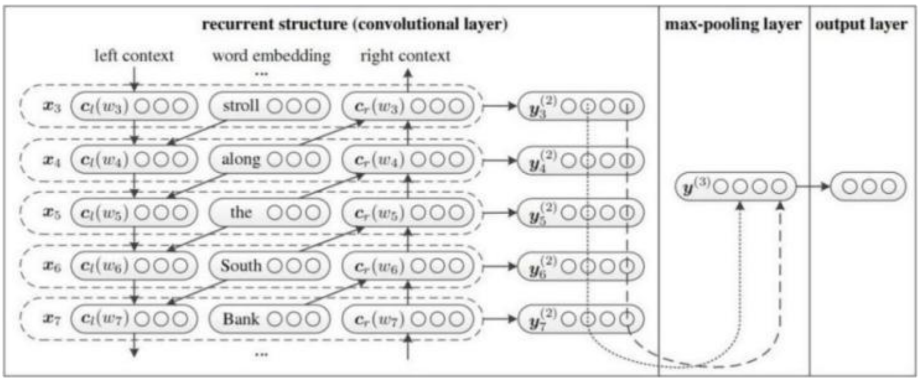

Figure 3. Architecture of RCNN

The recurrent structure used a bi-directional recurrent neural network used to capture contextual information [30] as much as possible when studying word representation. The main idea was to create a representation consisting of the left context obtained from the forward RNN with **Equation 1**, whereas the right context were obtained from the backward RNN with **Equation 2**.

$$c_l(w_i) = f\left(c_l(w_{i-1})W^{(l)} + e(w_{i-1})W^{(sl)}\right) \quad (1)$$

$$c_r(w_i) = f\left(c_r(w_{i+1})W^{(r)} + e(w_{i+1})W^{(sr)}\right) \quad (2)$$

The left-side context for the first embedding word used the same shared parameters as $c_l(w_1)$. The right-side context of the last embedding word shared the parameter with $c_r(w_n)$, and for the non-linear activation function by default the activation tanh was used. From **Equation 1**, and **Equation 2**, it showed that all left context and right context can be captured with context vector. Then the representations in the form of left-side context vector $c_l(w_i)$, embedding word $e(w_i)$, and the obtained right-side context vector $c_r(w_i)$ were combined with **Equation 3**.

$$x_i = [c_l(w_i): e(w_i): c_r(w_i)] \tag{3}$$

After getting the word representation, the next step was to find the value of $y_i^{(2)}$as can be seen in **Figure 3** with **Equation 4**. $y_i^{(2)}$containing a semantic vector used to find the most useful semantics in embedding word.

$$y_i^{(2)} = tanh\left(x_i W^{(2)} + b^{(2)}\right) \tag{4}$$

As can be seen in Figure 3, the next step was to apply a max-pooling layer for feature extraction from each representation by **Equation 5**. The max function in **Equation 5** took the maximum value of all representation elements from the embedding word.

$$y^{(3)} = max_{i=1}^{n}\, y_i^{(2)} \tag{5}$$

Finally, to get the output layer, the calculation was carried out with **Equation 6**. Because the output layer produced logits, the activation function Softmax was used for classification of more than two classes. Logits itself was a rough probability output of the classified sentence. The activation function Softmax was used to convert the rough probability into a probability, when added together, they would get the final result 1 with **Equation 7**.

$$y^{(4)} = W^{(4)} y^{(3)} \quad b^{(4)} \tag{6}$$

$$y^{(5)} = softmax(y^{(4)} \tag{7}$$

Results and Discussion

In this study, fine tuning of IndoBERT with RCNN was carried out. Then to determine the effect on accuracy, a baseline was needed to be compared. It used the IndoBERT base model. Then hyperparameters for fine tuning were used based on recommendations from [19], including batch sizes 16 and 32, using the Adam optimizer with learning rates of 2e-5, 3e-5, and 5e-5, as well as epochs 2, 3, and 4. The results of the tests in this study can be seen in **Table 1**.

Table 1. Results of testing

Architecture	Batch size	Learning rate	Epoch	Accuracy	F1
IndoBERT-RCNN	16	$2e^{-5}$	2	94.37	92.44
			3	94.6	92.25
			4	94.92	93.04
		$\mathbf{3e^{-5}}$	2	94.92	92.96
			3	**95.16**	**93.27**
			4	93.97	92.26
		$5e^{-5}$	2	94.21	92.14
			3	91.27	89.21
			4	94.52	92.04
	32	$2e^{-5}$	2	94.6	92.12
			3	94.37	92.22
			4	94.68	92.56
		$3e^{-5}$	2	93.73	91.18
			3	92.06	88.66
			4	93.97	91.66
		$5e^{-5}$	2	92.78	91.11
			3	94.05	92.28
			4	93.02	89.68
IndoBERT *base*	16	$2e^{-5}$	2	93.57	90.48
			3	94.21	92.27
			4	94.6	92.5

Architecture	Batch size	Learning rate	Epoch	Accuracy	F1
		$3e^{-5}$	2	93.57	91.29
			3	94.21	92.39
			4	93.25	89.69
		$5e^{-5}$	2	93.02	90.69
			3	93.17	90.97
			4	93.49	91.33
	32	$2e^{-5}$	2	93.97	91.8
			3	94.6	92.69
			4	94.37	92.07
		$3e^{-5}$	2	94.44	91.73
			3	94.76	92.6
			4	94.21	91.84
		$5e-^{5}$	2	94.52	92.48
			3	92.62	88.68
			4	93.81	91.18

Based on the results of the tests conducted by IndoBERT with the Recurrent Convolutional Neural Network (RCNN) the best results were obtained with hyperparameter 16 batch sizes, learning rate 3e-5 and 3 epochs obtained accuracy values of 95.16%, and f1 score was 93.27%. IndoBERT base got the best results with an accuracy value of 95.16%, and f1 score was 93.27% with batch size 32, learning rate 2e-5 and epoch 3.

Based on these results, IndoBERT with RCNN obtained better results than IndoBERT base with 95.16% accuracy, and f1 score was 93.27% with batch size 16, learning rate 3e-5 and epoch 3.

Conclusion

The results showed that fine tuning IndoBERT with RCNN performed better than IndoBERT base on sentiment analysis of Indonesian reviews. IndoBERT fine tuning with RCNN on sentiment analysis of Indonesian reviews obtained the best results with an accuracy value of 95.16 and an f1 score was 93.27% with a batch size of 16, learning rate 3e-5 and epoch 3. This shows that fine tuning IndoBERT by using all outputs as input for the RCNN results in accuracy performance improvement for sentiment analysis of Indonesian reviews.

In the future research, IndoBERT or IndoBERT with RCNN can be utilized to carry out various other tasks in the field of natural language processing, such as: hoax detection, sarcasm detection, news categorization, topic analysis, emotion classification, aspect-based sentiment analysis, post tagging, named entities. recognition, question answering, and other natural language processing tasks [31].

References

[1] B. Yang, Y. Liu, Y. Liang, and M. Tang, "Exploiting user experience from online customer reviews for product design," *Int. J. Inf. Manage.*, vol. 46, pp. 173–186, Jun. 2019, doi: 10.1016/j.ijinfomgt.2018.12.006.

[2] A. Karunakaran, W. J. Orlikowski, and S. V. Scott, "Crowd-Based Accountability: Examining how social media commentary reconfigures organizational accountability," *Organ. Sci.*, vol. 33, no. 1, pp. 170–193, Jan. 2022, doi: 10.1287/orsc.2021.1546.

[3] N. H. Cahyana, S. Saifullah, Y. Fauziah, and A. S. Aribowo, "Text annotation automation for hate speech detection using SVM-classifier based on Feature Extraction," 2022.

[4] Y. Fauziah, S. Saifullah, and A. S. Aribowo, "Design text mining for anxiety detection using machine learning based-on social media data during COVID-19 pandemic," in *Proceeding of LPPM UPN "Veteran" Yogyakarta Conference Series 2020–Engineering and Science Series*, 2020, vol. 1, no. 1, pp. 253–261, doi: 10.31098/ess.v1i1.117.

[5] N. H. Cahyana, S. Saifullah, Y. Fauziah, A. S. Aribowo, and R. Drezewski, "Semi-supervised Text Annotation for Hate Speech Detection using K-Nearest Neighbors and Term Frequency-Inverse Document Frequency," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 10, 2022, doi: 10.14569/IJACSA.2022.0131020.

[6] Y. Nurdiansyah, S. Bukhori, and R. Hidayat, "Sentiment analysis system for movie review in Bahasa Indonesia using naive bayes classifier method," *J. Phys. Conf. Ser.*, vol. 1008, no. 1, pp. 1–7, Apr. 2018, doi: 10.1088/1742-6596/1008/1/012011.

[7] M. Wongkar and A. Angdresey, "Sentiment analysis using Naive Bayes Algorithm Of The Data Crawler:

Twitter," *2019 Fourth Int. Conf. Informatics Comput.*, pp. 1–5, Oct. 2019, doi: 10.1109/ICIC47613.2019.8985884.

[8] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, "Analisis klasifikasi sentimen terhadap sekolah daring pada twitter menggunakan Supervised Machine Learning," *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, Apr. 2021, doi: 10.28932/jutisi.v7i1.3216.

[9] M. Ahmad, S. Aftab, M. Salman, N. Hameed, I. Ali, and Z. Nawaz, "SVM Optimization for Sentiment Analysis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 4, pp. 393–398, 2018, doi: 10.14569/IJACSA.2018.090455.

[10] H. S. Utama, D. Rosiyadi, B. S. Prakoso, and D. Ariadarma, "Analisis Sentimen sistem ganjil genap di tol bekasi menggunakan Algoritma Support Vector Machine," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 243–250, Aug. 2019, doi: 10.29207/resti.v3i2.1050.

[11] D. R. Pant, P. Neupane, A. Poudel, A. K. Pokhrel, and B. K. Lama, "Recurrent Neural Network Based bitcoin price prediction by twitter sentiment analysis," *2018 IEEE 3rd Int. Conf. Comput. Commun. Secur.*, pp. 128–132, Oct. 2018, doi: 10.1109/CCCS.2018.8586824.

[12] Merinda Lestandy, Abdurrahim Abdurrahim, and Lailis Syafa'ah, "Analisis sentimen tweet vaksin COVID-19 menggunakan Recurrent Neural Network dan Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, pp. 802–808, Aug. 2021, doi: 10.29207/resti.v5i4.3308.

[13] X. Ouyang, P. Zhou, C. H. Li, and L. Liu, "Sentiment analysis using Convolutional Neural Network," *2015 IEEE Int. Conf. Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Auton. Secur. Comput. Pervasive Intell. Comput.*, pp. 2359–2364, Oct. 2015, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.349.

[14] Y. Yuliska, D. H. Qudsi, J. H. Lubis, K. U. Syaliman, and N. F. Najwa, "Analisis sentimen pada data saran mahasiswa terhadap kinerja departemen di perguruan tinggi menggunakan Convolutional Neural Network," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 5, p. 1067, Oct. 2021, doi: 10.25126/jtiik.2021854842.

[15] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent Convolutional Neural Networks for Text Classification," *Proc. Twenty-Ninth AAAI Conf. Artif. Intell.*, pp. 2267–2273, 2015, [Online]. Available: https://dl.acm.org/doi/10.5555/2886521.2886636.

[16] C. Du and L. Huang, "Sentiment classification via recurrent Convolutional Neural Networks," *DEStech Trans. Comput. Sci. Eng.*, no. cii, pp. 308–316, Dec. 2017, doi: 10.12783/dtcse/cii2017/17268.

[17] A. D. Arumsari and E. Winarko, "Analisis Sentimen pada Tweet Indonesia Menggunakan Recurrent Convolutional Neural Network," Universitas Gadjah Mada, 2017.

[18] Z. Mahmood *et al.*, "Deep sentiments in Roman Urdu text using Recurrent Convolutional Neural Network model," *Inf. Process. Manag.*, vol. 57, no. 4, pp. 1–14, Jul. 2020, doi: 10.1016/j.ipm.2020.102233.

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proc. 2019 Conf. North*, pp. 4171–4186, 2019, doi: 10.18653/v1/N19-1423.

[20] Q. T. Nguyen, T. L. Nguyen, N. H. Luong, and Q. H. Ngo, "Fine-Tuning BERT for Sentiment Analysis of Vietnamese Reviews," *2020 7th NAFOSTED Conf. Inf. Comput. Sci.*, pp. 302–307, Nov. 2020, doi: 10.1109/NICS51282.2020.9335899.

[21] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," *Proc. 1st Conf. Asia-Pacific Chapter Assoc. Comput. Linguist. 10th Int. Jt. Conf. Nat. Lang. Process.*, pp. 843–857, Sep. 2020, [Online]. Available: http://arxiv.org/abs/2009.05387.

[22] C. A. Putri, "Analisis Sentimen review film berbahasa inggris dengan pendekatan bidirectional encoder representations from transformers," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 2, pp. 181–193, Jan. 2020, doi: 10.35957/jatisi.v6i2.206.

[23] D. Fimoza, A. Amalia, and T. H. F. Harumy, "Sentiment Analysis for Movie Review in Bahasa Indonesia Using BERT," *2021 Int. Conf. Data Sci. Artif. Intell. Bus. Anal.*, pp. 27–34, Nov. 2021, doi: 10.1109/DATABIA53375.2021.9650096.

[24] K. S. Nugroho, A. Y. Sukmadewa, H. Wuswilahaken DW, F. A. Bachtiar, and N. Yudistira, "BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews," *6th Int. Conf. Sustain. Inf. Eng. Technol. 2021*, pp. 258–264, Sep. 2021, doi: 10.1145/3479645.3479679.

[25] H. Christian, D. Suhartono, A. Chowanda, and K. Z. Zamli, “Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging,” *J. Big Data*, vol. 8, no. 1, p. 68, Dec. 2021, doi: 10.1186/s40537-021-00459-1.

[26] S. M. Isa, G. Nico, and M. Permana, “Indobert for Indonesian Fake News Detection,” *ICIC Express Lett.*, vol. 16, no. 3, pp. 289–297, 2022.

[27] D. Fan, L. Wan, W. Xu, and S. Wang, “A bi-directional attention guided cross-modal network for music based dance generation,” *Comput. Electr. Eng.*, vol. 103, p. 108310, Oct. 2022, doi: 10.1016/j.compeleceng.2022.108310.

[28] M. Arevalillo-Herraez, P. Arnau-Gonzalez, and N. Ramzan, “On Adapting the DIET Architecture and the Rasa Conversational Toolkit for the Sentiment Analysis Task,” *IEEE Access*, vol. 10, pp. 107477–107487, 2022, doi: 10.1109/ACCESS.2022.3213061.

[29] S. Saifullah, Y. Fauziah, and A. S. Aribowo, “Comparison of Machine Learning for Sentiment Analysis in Detecting Anxiety Based on Social Media Data,” Jan. 2021, [Online]. Available: http://arxiv.org/abs/2101.06353.

[30] M. Peters *et al.*, “Deep Contextualized Word Representations,” *Proc. 2018 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Vol. 1 (Long Pap.*, pp. 2227–2237, 2018, doi: 10.18653/v1/N18-1202.

[31] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, “Pre-trained models for natural language processing: A survey,” *Sci. China Technol. Sci.*, vol. 63, no. 10, pp. 1872–1897, Oct. 2020, doi: 10.1007/s11431-020-1647-3.