Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2024, 2024:71
https://doi.org/10.28919/cmbn/7858
ISSN: 2052-2541

EMOTION RECOGNITION INDONESIAN LANGUAGE FROM TWITTER USING INDOBERT AND BI-LSTM

STEPHEN WILLIAM[1,2], KENNY[1,2], ANDRY CHOWANDA[2,*]

[1]Computer Science Department, BINUS Graduate Program, Master of Computer Science, Indonesia

[2]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

Copyright © 2024 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Emotions recognition has been a very challenging task and topic in Natural Language Processing area. Nevertheless, there have not been many research have been done in the local language, such as Bahasa Indonesia, compared to the English language. This is due to the lack of resources available publicly. Therefore, we proposed a new labelled dataset scraped from social media platforms (i.e. Twitter), based on Ekman's six basic emotions (Happy, Fear, Anger, Disgust, Sadness, and Surprise) plus one neutral label. Furthermore, we implemented several fine-tuned models from the IndoBERT model to model the emotions recognition system. Moreover, Bidirectional Long Short-Term Memory (Bi-LSTM) architectures were also implemented to model the emotions recognition system. Finally, the models were trained to the collected dataset (7,629 tweets). The results show that the fine-tuned model resulted in an accuracy score of 92.3%, which outperforms both the baseline model and the Bi-LSTM with the accuracy of 90.7% and 84.0%, respectively.

Keywords: emotion recognition; natural language processing; deep learning; transfer learning; Indonesian tweet.

2020 AMS Subject Classification: 68T50, 68T60.

---

*Corresponding author
E-mail address: achowanda@binus.edu
Received December 18, 2022

1. INTRODUCTION

The concept of emotions is one of the most complex things to study. It is an essential aspect of daily human lives because we can infer many things from the emotions shown by our interlocutors. However, the knowledge that we can infer from emotions is limited to our daily lives, as it is helpful in many fields, from enhancing immersion in a video game playthrough to gaining knowledge from citizens' points of view during the presidential election [1, 2]. Recognizing emotions is also a paramount when we are building an affect aware system. There are several implementations in emotions recognitions in affect aware systems, such as: stress or depression detection [23], stress detection [22; 23] and chat bots or virtual humans [19, 21]. While there are several Facial Emotions Recognition (FER) and Speech Emotion recognition implementations, the field of emotions recognition done via text is still limited, even though people tend to express things through text in social media [3]. There are many ways to infer knowledge or informatio from text-based data. One of the ways that we can use is to implement what is known as Natural Language Processing (NLP). Natural Language Processing (NLP) is a subfield of linguistics, computer science, and artificial intelligence that focuses on the interactions between computers and human natural language The main focus of the Natural Language Processing (NLP) is mainly how to program computers to process and analyze large amounts of human natural language data. Researchers have been studying new ways to make Natural Language Processing more efficient and effective in text classification throughout the years. The General Language Understanding Evaluation (GLUE) benchmark, which is a collection of resources used to evaluate a model's ability in a Natural Language Processing task, usually has a different model tailored to a specific Natural Language Processing task so that it may perform exceptionally [4]. People have been dealing with these emotions recognition tasks through various approaches. Ma, et al. [5] has used the Bi-LSTM model network and achieved a micro averaged F1 score of 75.57%. Purnama [6] proposed a K-Nearest Neighbor (KNN) based approach, which performs with an accuracy of 71,26 % on a Bahasa Indonesia dataset gathered from three different news websites. While these models have proven themselves to still be relevant, in 2018, the team at Google Brain released a paper about a generic solution that outperforms these specifically tailored models in the The General Language Understanding Evaluation benchmark.

BERT, which stands for Bidirectional Encoder Representations from Transformers, has garnered much attention since its public release in 2018 by the Google Brain team because it has outperformed many tailored solutions in the The General Language Understanding Evaluation benchmark [7]. Unlike recent language representation models, BERT is designed to pre-train deep bidirectional representations from an unlabeled text by jointly conditioning both left and proper context in all layers. One of the reasons for using the BERT model is that the BERT model has already been pre-trained by the Google team, so the model only needed to be fine-tuned to specific tasks, which is emotion recognition. BERT also does not need too much training data because of the transfer learning concept it uses. Although BERT is better than most of its predecessors in most cases, it lacks flexibility in understanding other languages since BERT initial release was only for English and Chinese [7]. Recently, a new BERT model specifically trained to handle Bahasa NLP tasks called IndoBERT was released [8] and it performs better than the previous general-purpose BERT and the multi-lingual BERT (mBERT). There are many English language datasets for emotion recognition tasks, but the available corpus for Bahasa emotion recognition tasks is still small. The available publicly available only contain around 4,300 data divided into four classes of emotion [9]. We propose a more extensive dataset for the Bahasa emotion recognition task, divided into Ekman's six basic emotions [10]. This research aims to provide novel contributions by: 1) Creating a larger Bahasa emotion recognition dataset; 2) Comparing the IndoBERT model that we are using with other popular deep learning algorithms, such as Bi-LSTM. This work is organized as follows: Section 1 will cover the introduction of the study; Section 2 lists some literature study and the works that are related to the study; Section 3 will elaborate on the design, development, and evaluation methodology used in the experiment; Section 4 discusses the result of the experiment; and Section 5 will give a conclusion based on our finding and provides future research direction

2. RECENT WORK

Emotions classification from text has been a challenging Natural Language Processing problem for a long time. Various research has aimed for more accurate emotions detection from many languages. There have been various ways to tackle emotions classification problems; most use machine learning and deep learning to predict emotions. Machine Learning is a technique that uses advanced algorithms to read data and discover informative patterns based on the data. Several machine learning models solve various Natural Language Processing tasks such as emotion recognition and predicting missing words. Many machine learning algorithms can solve classification problems, such as Support Vector Machine (SVM), K Nearest Neighbour (K-NN), and Naive Bayes algorithm. Support Vector Machine (SVM) is one of the most implemented machine learning models for linearly non-separable problems, such as classification. In 2012, the SVM model was used to handle an emotion classification task. It classifies the emotions into six emotion classes: Anger, Joy, Disgust, Fear, Sadness, and Surprise. Their proposed method achieved an average score of 89.95 % across the six emotions, outperforming the state-of-the-art models [11]. Another machine learning algorithm called KNN and Naïve Bayes was used to classifying emotion from a Bahasa dataset. The classification algorithm used by this research is supervised classification K-NN and Naive Bayes algorithm. The research concludes that the F measure of K-NN is higher than Naïve Bayes in the text document classification [6]. Another group of researchers, Chowanda et. al. [20] explored several machine learning and deep learning architectures to model the emotions recognition system in English language. The researcher used AffectiveTweets datasets to model the emotions recognition system. The best model was achieved by the one trained by Generelized Linear Model (GLM) where it gets 92 %, 90.1 %, 90.2 %, 90. 2 % for accuracy score, F1-Score, Recall Score and Precision Score.

As machine learning grows, new techniques are introduced, such as Deep Learning [12]. Deep learning is a subfield of machine learning that imitates human brain work. Deep learning consists of many neuron layers: one input layer, one or more hidden layers, and one output layer. Deep learning algorithms work by processing values in every layer and adjusting the input weight to do better at the next iteration. Many deep learning classes solve classification problems, such as Convolution Neural Network (CNN), which are usually used when handling two-dimensional data such as image and Recurrent Neural Network (RNN), which are used when handling and aiming for a pattern/relation in sequence data. However, the Recurrent Neural Network technique still has its drawbacks. While it can see the previous input, it cannot see far enough back from the input, and thus the true meaning of the input may be lost. Then comes Long Short-Term Memory (LSTM) to solve the problem. Long Short-Term Memory is a modified model that uses Recurrent Neural Network as the base. Long Short-Term Memory has a "memory" cell that retains previous calculations made by the blocks [13]. This memory cell helps Long Short-Term Memory cover the drawbacks that Recurrent Neural Network used to have. Long Short-Term Memory blocks can also modify these memory cells by using "gates". These gates differentiate between "forget gate", "update gate", and "output gate". The forget gate will determine how much of the old memory will affect the new calculations, while the update gate will determine how much the new calculations will affect the old memory. Lastly, the output gate will determine the current block's output, which will also work as the input for the next block. In research comparing Long Short-Term Memory and Convolution Neural Network in the emotion recognition field, Long Short-Term Memory beat Convolution Neural Network‘s accuracy with a 5.33% margin [14]. The proposed method in this paper combines the emotional word vector and semantic word vector. The research uses the Five-fold Cross Validation method to evaluate text recognition models. The paper concludes that Long Short-Term Memory is better to model the sequential data than Convolution Neural Network. Even though Long Short-Term Memory has managed to cover the base Recurrent Neural Network models' drawbacks, Long-Short Term Memory still has drawbacks. It is slow to train, but it can only see the context from left to right.

To solve the Long-Short Term Memory problem, a modified version of Long Short-Term Memory called Bidirectional Long Short-Term Memory (Bi-LSTM) was introduced. The Bidirectional Long Short-Term Memory (Bi-LSTM) can sees the context from both directions and improves the Long Short-Term Memory model that can only see from left to right perspective. Moreover, in research that compares the performance of the Bi-LSTM model with the Long Short-Term Memory model, the experiment results show that the Bi-LSTM outperformed the Long Short-Term Memory model in overall performance and most emotion detection performance [15, 16]. Although Bi-LSTM can see the context from both directions, this is not a proper bidirectional context since it just looks at the input from left to right, right to the left, and concatenates the results. Bidirectional Encoder Representations from Transformers (BERT) was introduced in 2018 [7] to improve this solution. BERT solves the many issues that Long Short-Term Memory has. It is faster to train and perceive context in a bidirectional manner simultaneously, rather than perceiving it one by one then concatenating the result. The BERT model is built upon the transformer model's attention mechanism and stacks 12 pieces of the encoder upon itself rather than using transformers' decoder. IndoBERT, a BERT based model explicitly created to tackle Bahasa natural language tasks, was released in 2020. It was trained with a vocabulary utilizing the Sentence-Piece with a Byte Pair Encoding (BPE) tokenizer as its generation method. It was tested on a dataset available on the IndoNLU web page and is compared against other multilingual models such as XML-R, mBERT, and Scratch. The result shows that IndoBERT outperforms other multilingual models on 8 out of the 12 tasks available on the IndoNLU benchmark.

3. METHODOLOGY

Four main methods were introduced in this research, they are: Data Gathering, Data Labelling, Data Pre-processing, Modeling and Evaluation. The first three phases are focusing in the data collection and preparation. The data were collected from a social media platform (i.e. Twitter) and were annotated. The data collected then were used to model emotions recognition from text in the Modeling and Evaluation. Several deep learning architectures were applied to model the emotions recognition. Moreover, the emotions recognition models were also be evaluated in Modeling and Evaluation. The details of each method in this research are comprehensively demonstrated in the following sub-sections (Data Gathering, Data Labelling, Data Pre-processing, Modeling and Evaluation).

Data Gathering

There are several sources of data that can be collected to model the emotions recognition. In this research, we use social media platform to collect the data. The social media platform generally consists of tweets from people who have different backgrounds, emotions and moods when they were texting their opinions, comments, thought and ideas into the social media platform. Hence, social media platforms will be the best source to collect the data for emotions recognition modeling. The data were collected through crawling technique on social media (i.e. Twitter). We use the Tweepy API to get 2,000 tweets for each emotion class totalling 12,000 tweets for all classes. The filters that we implemented are the language filter set to Bahasa Indonesia or "id", and we also implemented an emotion filter by using one English word, two Bahasa Indonesia words, and one Bahasa Indonesia Slang word. The following process removes the mentioned username and replaces it with the tag [USERNAME] and the URL [URL]. All the tweets collected and processed were stored in a Coma Separated Value (CSV) file format to be processed in the next step.

Data Labelling

Ekman's theory argues that people express six basic emotions: Joy (Happy), Anger, Sadness, Fear, Surprise, and Disgust. We used this theory as the classifier when annotating the dataset. Most of the researchers in this area are using similar model of emotions (i.e. six basic emotions plus neutral), with some of the variations of the emotions models. We have two annotators who annotated the tweet collected from a social media platform (i.e. Twitter). First step, we annotated it manually by reading every sentence and giving it a class label. To avoid bias and to improve the reliability of the annotated data, we then rate the inter-rater reliability using the Kappa coefficient (κ) and manage to achieve a score of 87.7 %. The result indicates that both annotators were in almost perfect agreement in labelling the collected tweets from a social media platform (i.e. Twitter). Moreover, we also removed the data where the two annotators disagree in the label, resulting in only 7,629 data (from 12,000 tweets originally) on the six basic emotions plus one neutral emotion class. Table 1 shows the data distribution of the collected dataset from a social media platform (i.e. Twitter).

Table 1. *The Collected Dataset Distribution*

No	Emotions	Numbers	Percentage
1.	Happy	1,120	15%
2.	Sad	1,008	13%
3.	Anger	854	11%
4.	Surprise	1,210	16%
5.	Disgusted	1,301	17%
6.	Fear	1,017	13%
7.	Neutral	1,119	15%
TOTAL		**7,629**	**100%**

Data Pre-processing

The labelled dataset is shown in the form of sentences (the tweets) and labels (annotated labels). Before training the models using the annotated dataset collected from a social media platform (i.e. Twitter), several text pre-processing were applied to the data to improve the quality of the data. First, we convert all the sentences into lower case (case folding), next the sentences were split based on the punctuation token. The sentences then were tokenized into tokens (words). The tokenized words then are converted into numbers in a vector space (vectorization). Each model has different techniques of words vectorization (words embedding). The Bi-LSTM model implements a word embedding model, and the data is used to train the embedding model. Meanwhile, the BERT model already has the pre-trained embedding provided the community. In this research we implemented the IndoBERT pre-trained embedding models. The IndoBERT pre-trained embedding models were trained specifically to solve text processing in Bahasa Indonesia language. Finally, we then shuffled and split the dataset into the train set and test set by a 9 to 1 ratio, respectively. Finally, the setting was applied to all the proposed models.

Modeling and Evaluation

This research proposes using three models to evaluate: the original IndoBERT, our proposed fine-tuned IndoBERT, and a Bi-LSTM network. In the IndoBERT architecture, this research implements a pre-trained BERT based IndoBERT model from the Huggingface library, a BERT model specifically created to tackle Bahasa based NLP problems. We used the BERT paper's recommended parameters for the baseline model [7] and fine-tuned the base model parameters, combining learning rate, batch size, and epochs. The proposed Bi-LSTM model is implemented using the Keras library. It has three layers: Word Embedding Layer, Bidirectional LSTM Layer, and the dense layer. The embedding layer uses the FastText technique from the Gensim library to pre-trained the weight of every word based on its meaning, not randomly assigned at the training step [17]. The embedding layers' job is to assign a value representing every word the machine can understand. The Bidirectional Layer uses an activation function of tanh with the dropout value of 0.2 and a recurrent dropout of 0.2. Dropout values are used to prevent neural networks from overfitting [18]. The dense layer is a layer that specifies the dimension of the output vector. The activation function decides the content of the vector. The default activation function for dense in Keras is Linear without being specified. Since we are doing multiple classifications, we use Softmax for our activation function. Finally to evaluate the proposed emotions recognition models, performance metrics such as Precision, Recall, F1-Score, and Accuracy were used in this research.

4. RESULTS & DISCUSSIONS

The experiments were conducted with IndoBERT and Bi-LSTM architectures with three different models' settings. The models were then evaluated on their performance by classifying the six basic emotions plus Neutral emotions. The hyperparameters explored in the BERT models are the batch size, number of epochs, and learning rate. The hyperparameters were chosen based on the recommendation made by the creator of the BERT architecture on fine-tuning the BERT pre-trained model. In comparison, the parameters changed in the Bi-LSTM models are the word embedding layer's parameters, such as window range and embedding size. For the Bi-LSTM models, we have tried ten configurations with different parameters and achieved the top 3 overall accuracy results: 84% (Bi-LSTM 1 model or M4 model), 82.2% (Bi-LSTM 2 model or M5 model), and 81.8% (Bi-LSTM 3 model or M6 model) (see Figure 1). Figure 1 illustrates the overall results of the experiments. M0 denotes the baseline, while M1 – M3 denotes the fine-tuned IndoBERT model: IndoBERT1, IndoBERT2 and IndoBERT3, respectively, and M4 – M6 denotes the fine-tuned Bi-LSTM models: Bi-LSTM1, Bi-LSTM2 and Bi-LSTM3, respectively. The M0 or the BERT baseline model achieved the best accuracy of 90.7%, where the M2 model or IndoBERT2 model achieved the best accuracy of 92.3%. The M3 model or IndoBERT3 model achieved the best accuracy of 92.1%. Finally, M1 has the best overall accuracy of 92.3%, barely pulling ahead of M2's best accuracy of 92.2%. M1 was trained with a hyperparameter setting of batch size 16, the epoch of 2, and the learning rate of 1e-5. In general, the models trained with IndoBERT architectures performed better compared to the models trained with Bi-LSTM architectures. Moreover, there are no significant differences between models in the same group (i.e. group of models trained by IndoBERT architectures and group of models trained by Bi-LSTM architectures). The models trained by the Bi-LSTM architectures still cannot compete with the baseline model trained by IndoBERT (M0 model). The best model trained using Bi-LSTM architectures achieved the best accuracy score of 84% (Bi-LSTM 1 model or M4 model). However, all the models trained with the BERT architectures with IndoBERT embedding models outperformed the baseline model trained by IndoBERT (M0 model). The best model trained using Bi-LSTM architectures achieved the best accuracy score of 92.3% (IndoBERT1 model or M1 model).

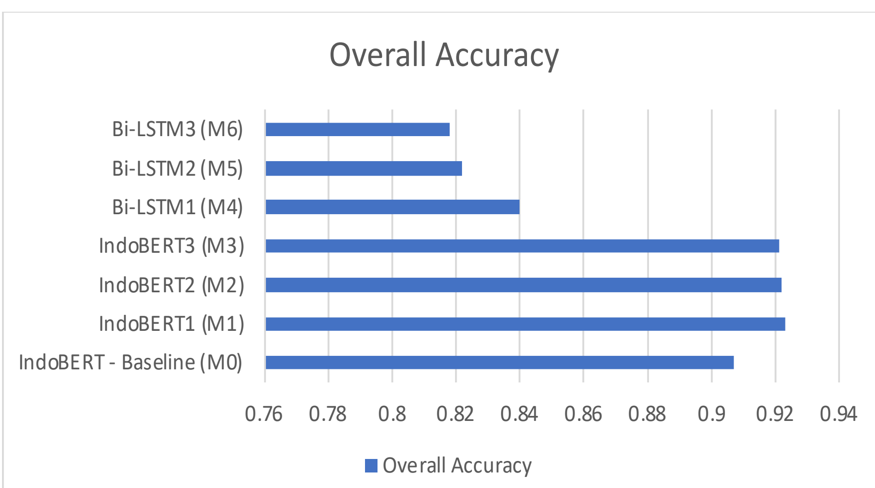

Figure 1. *Results – Overall Accuracy*

Table 2 demonstrates the results of the experiments, which contains the six emotions plus one neutral class and its respective Precision, Recall, and F1 Score denoted by P, R and F1, respectively. Even though the baseline model has a pretty good result on classifying the six basic emotions, the baseline model ran into trouble when classifying the neutral class. Out of the 238 neutral data (tweets) on the test set, it only managed to accurately classify 154 of them, which is just a 75% from the total. Nevertheless, we can see that the fine-tuned model has achieved a better precision score on classifying the neutral emotion than the baseline model. We also evaluated the fine-tuned IndoBERT model against the older Bi-LSTM model, and IndoBERT outperformed it on all aspects of the classification, be it the overall accuracy or each class's accuracy. Even though the BERT M1's Happy emotion precision score is lower than Bi-LSTM M4's, we can see that the BERT M1's has a perfect recall score of 1, showing the model predicted successfully all the actual positive as a positive. This is essential in emotion recognition as emotion recognition is the basis for many other medical fields, such as stress detection and depression detection, where we care about our wrongly classified patients. With the M1's precision being lower than M4's and its recall higher, we turn to the F1 score, which shows the harmonic mean between our precision and recall score to help us decide which model is better. In this case, the BERT M1's model has a higher F1 score rather than its counterpart. So, we can safely assume here that the BERT M1's model is our best performing model.

Table 2. *Detailed Overall Training Results*

	Anger			**Sadness**			**Happy**			**Disgust**			**Fear**			**Surprised**			**Neutral**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
M0	.87	.92	.89	.95	.94	.94	.98	.93	.96	.92	.96	.94	.89	.97	.93	.96	.99	.98	.75	.65	.70
M1	.92	.94	.93	.90	.98	.94	.91	1	.97	.94	1	.97	.94	.99	.96	.94	.97	.96	.88	.56	.69
M2	.93	.94	.93	.90	.99	.94	.97	.95	.96	.94	1	.97	.93	.99	.96	.93	.99	.96	.83	.58	.68
M3	.91	.93	.92	.90	.98	.94	.97	.95	.96	.94	1	.97	.95	.99	.97	.92	.99	.95	.82	.58	.68
M4	.89	.82	.86	.86	.84	.85	.93	.97	.95	.83	.94	.88	.87	.86	.88	.91	.85	.88	.56	.55	.55
M5	.84	.84	.89	.85	.87	.86	.92	.92	.92	.85	.84	.85	.84	.88	.86	.85	.93	.89	.60	.58	.59
M6	.80	.89	.84	.81	.85	.83	.84	.85	.85	.96	.93	.95	.80	.81	.81	.86	.93	.89	.62	.50	.55

Figure 2 illustrates the confusion matrix for the best model in this research (i.e. the M1 or IndoBERT1 model). The model achieved the highest accuracy of 0.923. However, we can see here that the model has reasonably predicted the six basic emotions but fell short on predicting the Neutral emotion. This may be due to the model's capabilities in recognizing the sentence's context are still low. For example, if a person says "Sudah, ngapain marah mulu", which in English translates into "Get over it, there is no need to be angry", the model may classify this as an Angry emotion because of the angry keyword in the sentence.

Figure 2. *The Best Model (M1 / IndoBERT1) Confusion Matrix*

This paper presents the research that compares the default IndoBERT model, Fine-tuned IndoBERT, and Bi-LSTM model. This experiment used the dataset obtained from scraping the Twitter API endpoint. Then, the obtained dataset is annotated by two different annotators. We then rate the inter-rater reliability using the Kappa coefficient (κ) and manage to achieve a score of 0.877. We also removed the data where the two annotators disagree, resulting in 7,629 data on the six basic emotions plus one neutral emotion class. Hence, this research also contributes a dataset in emotions recognition from text in a local language (i.e. Bahasa Indonesia). The dataset will be soon published publically, so the community can benefit from the dataset collected and annotated. The best performing model can also classify the Disgusted emotion as the best out of all seven classes. This may be because there are not many words representing the "disgusted" emotion in the Bahasa language. Meanwhile, the model finds it hard to classify the neutral emotion in a sentence because the models could not understand the text's context. The model classified and evaluated the precision, recall, and F1 score of 6 emotion classes of 'Anger', 'Disgusted', 'Fear', 'Happy', 'Sad', 'Surprised' and one neutral class. The evaluation result showed that two IndoBERT models (M0 and M1) perform well, with the baseline IndoBERT model (M0) achieved the best accuracy of 90.7%, and the fine-tuned IndoBERT model (M1) achieved the best accuracy of 93.8%. On the other hand, the best Bi-LSTM model (M4) achieved an accuracy of 84.0%.

There are two future directions to continue this research. The first one is the dataset collection. In the future, we will collect a larger dataset to increase the amount of data that we can train the model on and apply better pre-processing to the raw data acquired from social media (e.g., Twitter). The other social media platform also can be explored to increase the variation of the data. In addition, the possibility of applying new modalities such as speech or image may boost a model's accuracy when detecting emotion. The second future research direction is the deep learning architectures and embedding models. In the future, we will explore the possibility to automatically select and combine the best features from the embedding models. Moreover, several deep learning architectures (e.g. Transformer, Bi-LSTM, LSTM, Attention Convolutional Neural Network, and Graph Convolutional Neural Network) also can be combined to provide better results. Self-Supervised Contrastive Learning architectures can be explored to solve the number of the data (i.e. augmenting the data). Finally, the emotions recognition models trained in this research also can be implemented to an affective system such as Chat Bots or Virtual Humans [19, 21].

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] Andrew, A. N. Tjokrosetio, A. Chowanda, Dynamic difficulty adjustment with facial expression recognition for improving player satisfaction in a survival horror game, ICIC Express Lett. 14 (2020), 1097–1104. https://doi.org/10.24507/icicel.14.11.1097.

[2] W. Budiharto, M. Meiliana, Prediction and analysis of Indonesia presidential election from Twitter using

sentiment analysis, J. Big Data. 5 (2018), 51. https://doi.org/10.1186/s40537-018-0164-1.

[3] F.A. Acheampong, C. Wenyu, H. Nunoo-Mensah, Text-based emotion detection: Advances, challenges, and opportunities, Eng. Rep. 2 (2020), e12189. https://doi.org/10.1002/eng2.12189.

[4] A. Wang, A. Singh, J. Michael, et al. Glue: A multi-task benchmark and analysis platform for natural language understanding, in: 7th Int. Conf. Learn. Represent. ICLR 2019, 2019.

[5] L. Ma, L. Zhang, W. Ye, et al. PKUSE at SemEval-2019 task 3: Emotion detection with emotion-oriented neural attention network, in: Proceedings of the 13th International Workshop on Semantic Evaluation, Association for Computational Linguistics, Minneapolis, Minnesota, USA, 2019: pp. 287–291. https://doi.org/10.18653/v1/S19-2049.

[6] K.E. Purnama, Classification of emotions in indonesian texts using K-NN method, Int. J. Inform. Electron. Eng. 2 (2012), 899-903.

[7] J. Devlin, M.W. Chang, K. Lee, et al. BERT: Pre-training of deep bidirectional transformers for language understanding, (2018). https://doi.org/10.48550/ARXIV.1810.04805.

[8] B. Wilie, K. Vincentio, G.I. Winata, et al. IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding, (2020). https://doi.org/10.48550/ARXIV.2009.05387.

[9] M.S. Saputri, R. Mahendra, M. Adriani, Emotion classification on Indonesian Twitter dataset, in: 2018 Int. Conf. Asian Lang. Process. (IALP). (2018). https://doi.org/10.1109/ialp.2018.8629262.

[10] P. Ekman, Handbook of cognition and emotion, John Wiley & Sons, 1999.

[11] D.K. Kirange, R.R. Deshmukh, Emotion classification of news headlines using SVM, Asian J. Comput. Sci. Inform. Technol. 2, (2012), 104-106.

[12] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature. 521 (2015), 436–444. https://doi.org/10.1038/nature14539.

[13] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[14] M.-H. Su, C.-H. Wu, K.-Y. Huang, Q.-B. Hong, LSTM-based text emotion recognition using semantic and emotional word vectors, in: 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), IEEE, Beijing, 2018: pp. 1–6. https://doi.org/10.1109/ACIIAsia.2018.8470378.

[15] H. Ren, J. Wan, Y. Ren, Emotion detection in cross-lingual text based on bidirectional LSTM, in: C.-N. Yang, S.-L. Peng, L.C. Jain (Eds.), Security with Intelligent Computing and Big-Data Services, Springer International Publishing, Cham, 2020: pp. 838–845. https://doi.org/10.1007/978-3-030-16946-6_68.

[16] S. Siami-Namini, N. Tavakoli, A.S. Namin, The performance of LSTM and BiLSTM in forecasting time series, in: 2019 IEEE International Conference on Big Data (Big Data), IEEE, Los Angeles, CA, USA, 2019: pp. 3285–3292. https://doi.org/10.1109/BigData47090.2019.9005997.

[17] P. Bojanowski, E. Grave, A. Joulin, et al. Enriching word vectors with subword information, Trans. Assoc. Comput. Linguistics 5 (2017), 135–146. https://doi.org/10.1162/tacl_a_00051.

[18] N. Srivastava, G. Hinton, A. Krizhevsky, et al. Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (2014), 1929-1958.

[19] R. Sutoyo, A. Chowanda, A. Kurniati, et al. Designing an emotionally realistic chatbot framework to enhance its believability with AIML and information states, Procedia Computer Sci. 157 (2019), 621–628. https://doi.org/10.1016/j.procs.2019.08.226.

[20] A. Chowanda, R. Sutoyo, Meiliana, et al. Exploring text-based emotions recognition machine learning techniques on social media conversation, Procedia Computer Sci. 179 (2021), 821–828. https://doi.org/10.1016/j.procs.2021.01.099.

[21] M. Volonte, Y.C. Hsu, K.Y. Liu, et al. Effects of interacting with a crowd of emotional virtual humans on users' affective and non-verbal behaviors, in: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE, Atlanta, GA, USA, 2020: pp. 293–302. https://doi.org/10.1109/VR46266.2020.00049.

[22] S. Gedam, S. Paul, A review on mental stress detection using wearable sensors and machine learning techniques, IEEE Access. 9 (2021), 84045–84066. https://doi.org/10.1109/access.2021.3085502.

[23] A. Fatima, Y. Li, T.T. Hills, M. Stella, DASentimental: Detecting depression, anxiety, and stress in texts via emotional recall, cognitive networks, and machine learning, Big Data Cognitive Comput. 5 (2021), 77. https://doi.org/10.3390/bdcc5040077.