Detecting emotions using a combination of bidirectional encoder representations from transformers embedding and bidirectional long short-term memory

Aji Prasetya Wibawa[1], Denis Eka Cahyani[2], Didik Dwi Prasetya[1], Langlang Gumilar[1], Andrew Nafalski[3]

[1]Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia
[2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang, Malang, Indonesia
[3]Department of Electrical Engineering, University of South Australia, Adelaide, Australia

Article Info

ABSTRACT

One of the most difficult topics in natural language understanding (NLU) is emotion detection in text because human emotions are difficult to understand without knowing facial expressions. Because the structure of Indonesian differs from other languages, this study focuses on emotion detection in Indonesian text. The nine experimental scenarios of this study incorporate word embedding (bidirectional encoder representations from transformers (BERT), Word2Vec, and GloVe) and emotion detection models (bidirectional long short-term memory (BiLSTM), LSTM, and convolutional neural network (CNN)). With values of 88.28%, 88.42%, and 89.20% for Commuter Line, Transjakarta, and Commuter Line+Transjakarta, respectively, BERT-BiLSTM generates the highest accuracy on the data. In general, BiLSTM produces the highest accuracy, followed by LSTM, and finally CNN. When it came to word embedding, BERT embedding outperformed Word2Vec and GloVe. In addition, the BERT-BiLSTM model generates the highest precision, recall, and F1-measure values in each data scenario when compared to other models. According to the results of this study, BERT-BiLSTM can enhance the performance of the classification model when compared to previous studies that only used BERT or BiLSTM for emotion detection in Indonesian texts.

Corresponding Author:

Denis Eka Cahyani
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Malang
Semarang street No. 5, Sumbersari, Lowokwaru, Malang, East Java 65145, Indonesia
Email: denis.eka.cahyani.fmipa@um.ac.id

1. INTRODUCTION

Recently, emotion detection in text is one of the challenging topics in automatic language understanding (natural language understanding (NLU)) [1]. Understanding human emotions using text without facial expressions is a complex job, so research in this area is important to develop [2]. Meanwhile, social media is growing rapidly and becoming a place to convey the emotions of its users in the form of text comments [3]. Therefore, a system that can understand the context of the sentence is needed so that the system can detect emotions in social media comment texts automatically. In addition, the detection of emotions in the text develops in various languages [4]–[6]. Text emotion detection in one language cannot be automatically applied in another language. This is because a language has a different structure from other languages [7].

Previous studies have developed emotion detection in texts. Several methods have been used to detect emotion in text, including multinomial naive Bayes (MNB) [8], hybrid support vector machine (SVM) and k-nearest neighbors (KNN) [9], and logistic regression (LR) [10]. These investigations implement machine learning techniques. Methods of machine learning have limitations that necessitate the use of labeled data for the development of classification models, and the ensuing accuracy performance must be enhanced. Meanwhile, research related to deep learning is also developing. Several deep learning techniques are used for text emotion detection, namely bidirectional encoder representations from transformers (BERT) [11], [12], bidirectional long short-term memory (BiLSTM) [13], [14], long short-term memory (LSTM) [15], convolutional neural network (CNN) [16]. This research has limitations because they used only one deep learning method for the emotion detection process in the text. These deep learning methods can be combined in order to create a more precise deep learning model by leveraging the benefits of each technique.

Combinatorial research on deep learning has been established previously such as the BiLSTM-CNN combination [17] for English text classification. The combination of BERT and BiLSTM can be used for text emotion detection and sentiment analysis in English [18], [19]. These studies have limitations that can only be applied in English. However, previous research has not combined BERT and BiLSTM to detect emotions in Indonesian texts. Indonesian has a different language structure from other languages such as English, so this research cannot be applied to different languages. The development of emotion detection in Indonesian text requires a dataset and applies pre-processed data in Indonesian. Currently, there is no research that combines the BERT-BiLSTM method for Indonesian text. The development of emotion detection in Indonesian text using a combination of deep learning methods, namely BERT-BiLSTM, is needed.

Text emotion detection utilizes text classification techniques to classify text into distinct emotions. Word embedding is one of the phases that comprise the classification of text. Word embedding is the process of converting text into a vector representation of a specific point in space [20]. Word embedding techniques include BERT [21], Word2Vec [22], and GloVe [23]. Word embedding is crucial to text classification because it can affect the performance of the classification model [24]. Inappropriate use of word embeddings will lead to increased computation times and suboptimal classification performance.

The main contribution of this research is using the BERT-BiLSTM model to identify emotions in Indonesian text. In addition, this research contributes to the comparison of word embeddings, namely BERT, Word2Vec, and GloVe, which are employed in text classification. This study also compares deep learning classification models, namely BiLSTM, LSTM, and CNN. Word embedding is combined with a classification model to obtain a model for emotion detection in text. The purpose of this study was to obtain a combination of word embedding and classification models with the best performance for detecting emotions in Indonesian texts. The combination of the best models will improve the classification performance so that it is more optimal. This study employs a social media comment dataset in Indonesian, namely Twitter for Commuter Line and Transjakarta reviews [11].

2. RESEARCH METHOD

This section describes the research methodology employed in this study. This study does not describe the data collection phase because it utilized datasets from prior studies [11]. Using the previous dataset enables comparisons between the findings of this research and those of earlier investigations. The research methodology includes five steps: text pre-processing, word embedding, modeling, hyperparameter tuning, and performance evaluation. Figure 1 depicts the study methodology for detecting emotions in Indonesian texts.

2.1. Text pre-processing

Text pre-processing is the process of selecting text data to be more structured so that the data is ready for the next stage of processing [25]. The stages of pre-processing text on Indonesian tweets include case folding, tokenizing, filtering, stop word removal, and stemming. Case folding is used to generalize the use of capital letters to lowercase letters. In the case folding stage, characters other than letters and numbers will be deleted. Furthermore, tokenizing is the process of breaking sentences into words or tokens. The natural language toolkit (NLTK) library is used for the tokenizing process. Then, filtering is used to retrieve important words from the tokenizing results. Filtering will remove tweet attributes such as URLs, links, and mentions. The stop word removal stage removes common words that often appear but have no meaning. The list of stop words for Indonesian was built on this research. Finally, the stemming process involves mapping and parsing the form of a word into it is standard form. This study's stemming procedure uses the Sastrawi python library. This library implements the Nazief and Adriani algorithm [26] which is suitable for Indonesian.
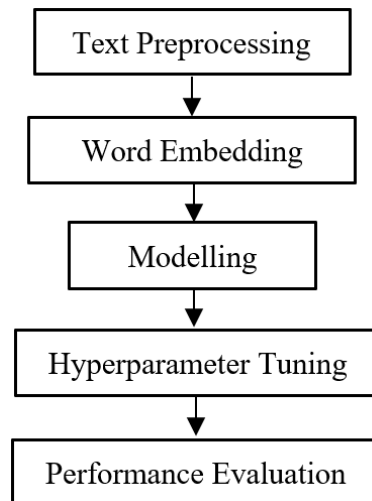
Figure 1. The research methodology for detecting emotions in Indonesian texts

2.2. Word embedding

The procedure of converting text to vector format is known as word embedding [27]. Word embeddings permit words with similar meanings to have equivalent visual representations. In addition to using BERT embedding, this study also uses other word embedding, namely Word2Vec and GloVe. BERT embedding is an efficient bidirectional transformer-based contextualized technique because it can be pre-trained and fine-tuned. In this study, Bert embedding with the type of BERT base multilingual cased model (multi_cased_L-12_H-768_A-12) was used. Meanwhile, Word2Vec is a statistical technique for efficiently learning a single word embedding from a text corpus. This study uses the Gensim library [28] for word2vec word embedding. Finally, global vectors for word representation (GloVe) are an expansion of the word2vec technique for efficiently learning word vectors. This research applies GloVe embedding to the type of Stanford's GloVe 6B 100d word embeddings [29].

2.3. Modelling

The emotion detection model in Indonesian text is built at this stage. This research uses Jupyter Notebook, Python 3.10, a Core i7 processor, 16 GB of RAM with a speed of 3,600 MHz, and a GPU Nvidia Cuda for model development. This study used the high-level Keras application programming interface (API) model class and Tensorflow 2.0. This study employed deep learning models for emotion detection are BiLSTM, LSTM, and CNN. Each model will be combined with the word embedding that has been described in the previous stage. This study uses nine scenarios combined with word embedding and emotion detection models, namely BERT-BiLSTM, BERT-LSTM, BERT-CNN, Word2Vec-BiLSTM, Word2Vec-LSTM, Word2Vec-CNN, GloVe-BiLSTM, GloVe-LSTM, and GloVe-CNN. The architectural design of one of the scenarios for the emotion detection model, the combination of BERT-BiLSTM, demonstrates in Figure 2.

BERT-BiLSTM combines BERT as the upstream part, namely the word embedding stage, and BiLSTM as the downstream part, namely the classification model development stage. Figure 2 shows the BERT-BiLSTM architectural design where the Indonesian text input is processed in the BERT embedding. Then the process is continued with model development using BiLSTM, which has two layers: the first layer is forward LSTM which captures the next sequential data linkage, and the second layer captures the previous sequential data linkage. Finally, the output will be in the form of text that has been classified based on the type of emotion, namely happiness, sadness, anger, fear, and surprise. The architectural design in other scenarios also applies in the same way by combining word embedding (BERT, Word2Vec, and GloVe) with deep learning classification models, namely BiLSTM, LSTM, and CNN.

In this investigation, the BiLSTM architecture employs a 256-unit bidirectional layer. Then, proceed with the 64-unit fully connected layer with the rectified linear unit (ReLU) activation function. The last layer is the output layer, with a single unit and an activation function called sigmoid to produce emotion output in the text. The LSTM architecture is also similar to the BiLSTM architecture, only differing in the LSTM layer.

Convolutional+ReLU is followed by a Maxpool layer that repeats up to three times for feature detection in this study's fundamental CNN architecture. Then, the flattening procedure is performed, which consists of transforming the matrix into a vector as a completely connected input layer. Additionally,

a dropout layer with a dropout of 0.5 was introduced to prevent overfitting. In addition, the fully connected layer (FCL) with the activation function ReLU is added to the total number of neurons, which is 256. The final CNN layer employs sigmoid activation to generate emotion detection output for Indonesian text.
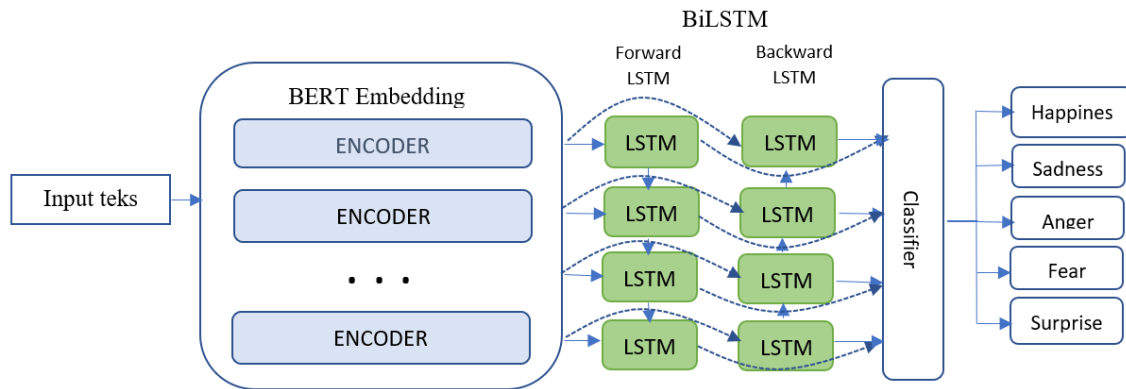

Figure 2. The architectural design of the BERT-BiLSTM model

2.4. Hyperparameter tuning

The hyperparameter optimization phase determines the parameters of the optimal experimental scenario. This study employs the parameters loss function, optimizer, batch_size, model activation, and epoch. In this study, the loss function employed is binary_crossentropy. Meanwhile, Adam is the optimizer used. Used Batch_Size is 1,024. The used Epoch is 12 with a validation split of 0.1. In BERT embedding, the length of the Max_sequence is 128. In this investigation, the ratio between training data and test data is 9:1.

2.5. Performance evaluation

In the model's evaluation phase, a confusion matrix is used to figure out how well the system can find emotions in Indonesian text. The confusion matrix has four values, namely true positive (TP), false positive (FP), false negative (FN), and true negative (TN) [30]. As shown in (1)-(4), the accuracy, precision, recall, and F1-measure calculations are performed during the evaluation phase.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$F1 - Measure = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \tag{4}$$

3. RESULTS AND DISCUSSION

The dataset based on the tweets of Transjakarta and Commuter Line users using Indonesian. This study employs the dataset utilized in previous research [11]. Table 1 illustrates the dataset that was used in this research. This study categorizes emotion text into five types of emotions: namely happiness, anger, sadness, fear, and surprise. The data scenarios used in this study are Commuter Line, Transjakarta, and Commuter Line+Transjakarta. In the Commuter Line+Transjakarta data scenario, we combine the two types of data. The total data on the Commuter Line is 20,395 tweets, and on the Transjakarta data, it is 10,649 tweets.

This research combines word embedding (BERT, Word2Vec, and GloVe) with models of classification (BiLSTM, LSTM, and CNN). The scenarios in this study consisted of nine types, namely BERT-BiLSTM, BERT-LSTM, BERT-CNN, Word2Vec-BiLSTM, Word2Vec-LSTM, Word2Vec-CNN, GloVe-BiLSTM, GloVe-LSTM, and GloVe-CNN. Figure 3 shows a comparison of the accuracy of emotion detection in Indonesian text.

Table 1. The dataset in this study

Type of emotion	Commuter Line	Transjakarta	Commuter Line+Transjakarta
Anger	15,365	7,619	22,984
Happiness	4,289	2,633	6,922
Sadness	507	265	772
Fear	190	116	306
Surprise	44	16	16

Figure 3(a) to 3(c) shows that the BERT-BiLSTM model generates the highest accuracy on the Commuter Line, Transjakarta, and Commuter Line+Transjakarta data with values of 88.28%, 88.42%, and 89.20%, respectively. Then followed by BERT-LSTM, and BERT-CNN models. In Word2Vec embedding, the model that generates the best accuracy is Word2Vec-BiLSTM, Word2Vec-LSTM, and Word2Vec-CNN in all data scenarios. Meanwhile, in GloVe embedding, the highest accuracy is obtained by models, namely GloVe-BiLSTM, GloVe-LSTM, and GloVe-CNN in all data scenarios.

In general, Figure 3 shows the classification model for emotion detection in Indonesian that obtains the best accuracy is BiLSTM, followed by LSTM, and finally CNN. While for word embedding, BERT embedding obtained the highest accuracy, followed by Word2Vec, and finally GloVe. In comparison to other data scenarios, the accuracy of the Commuter Line+Transjakarta data was highest in the scenario used. This is because by combining the two existing data, there are more text features for emotion detection in Indonesian when compared to using only one of the data. The more text features that can detect emotions in data, the higher the accuracy performance.

Precision, recall, and F1-measure are also used to measure how well the emotion detection model works. Measurements of precision, recall, and F1-measure will determine how effectively the model recognizes emotions in Indonesian. Table 2 demonstrates the values of precision, recall, and F1-measure in each data scenario. This value is the highest when compared to other models, as indicated by the table's bold font.

Table 2 shows that the BERT-BiLSTM model generates the best precision, recall, and F1-measure values compared with competing models in the Commuter Line, Transjakarta, and Commuter Line+Transjakarta scenarios. In all data scenarios, BERT-BiLSTM, BERT-LSTM, and BERT-CNN are the models that produce the highest precision, recall, and F1-measure values. The order of the highest precision, recall, and F1-measure values in Word2Vec embedding is Word2Vec-BiLSTM, Word2Vec-LSTM, and Word2Vec-CNN. Whereas in GloVe embedding, the GloVe-BiLSTM model obtained the highest precision, recall, and F1-measure values, followed by the GloVe-LSTM model, and finally the GloVe-CNN model.

In general, Table 2 shows that the model with the best precision, recall, and F1-measure values is BiLSTM, followed by LSTM, and CNN. Meanwhile, the word embedding that produce the best precision, recall, and F1-measure performances are BERT, Word2Vec, and GloVe, respectively. The data scenario that obtains better precision, recall, and F1-measure values than the other data scenarios is the combined data of Commuter Line+Transjakarta. The findings from the comparison of precision, recall, and F1-measure values are in line with the previously mentioned accuracy performance.

The BERT-BiLSTM model produces superior evaluation results compared to other models. Because the BERT-BiLSTM model integrates word embedding BERT as an upstream part and BiLSTM as a downstream part as a classification model for emotion detection in text. BERT has the ability to understand the contextual relationships in each word and can read long texts better. Meanwhile, BiLSTM can build a model with input in the form of sequential data, such as text. BiLSTM has the advantage of being able to store the selected memory because it has two layers: the first layer captures the previous sequential data linkage, and the second layer captures the next sequential data linkage. The combination of BERT and BiLSTM models combines the benefits of BERT and BiLSTM in order that the model can detect emotions in Indonesian texts more accurately than other models.

This study uses three classification models, namely BiLSTM, LSTM, and CNN, to detect emotions in Indonesian texts. BiLSTM gets the highest achievement score in contrast to other classification models because it will execute input both from the past to the future and from the future to the past. Whereas LSTM only stores past information because it receives only past input. BiLSTM uses two hidden states combined and can store information from the past and the future at any time. BiLSTM shows excellent results because it can comprehend context more on sequential data. So, this causes BiLSTM will perform more effectively than LSTM. BiLSTM and LSTM models are types of recurrent neural network (RNN) models that usually process sequential data and pay attention to the input sequence so that they are suitable for text data. Meanwhile, the CNN model is a model that is usually applied to image data because CNN has a convolution and pooling layer that can perform feature extraction on the image well. However, the CNN model also shows good performance in text data classification, although its performance is still below that of BiLSTM and LSTM.

Accuracy of Commuter Line Data

(a)

Accuracy of Transjakarta Data

(b)

Accuracy of Commuter Line+Transjakarta Data
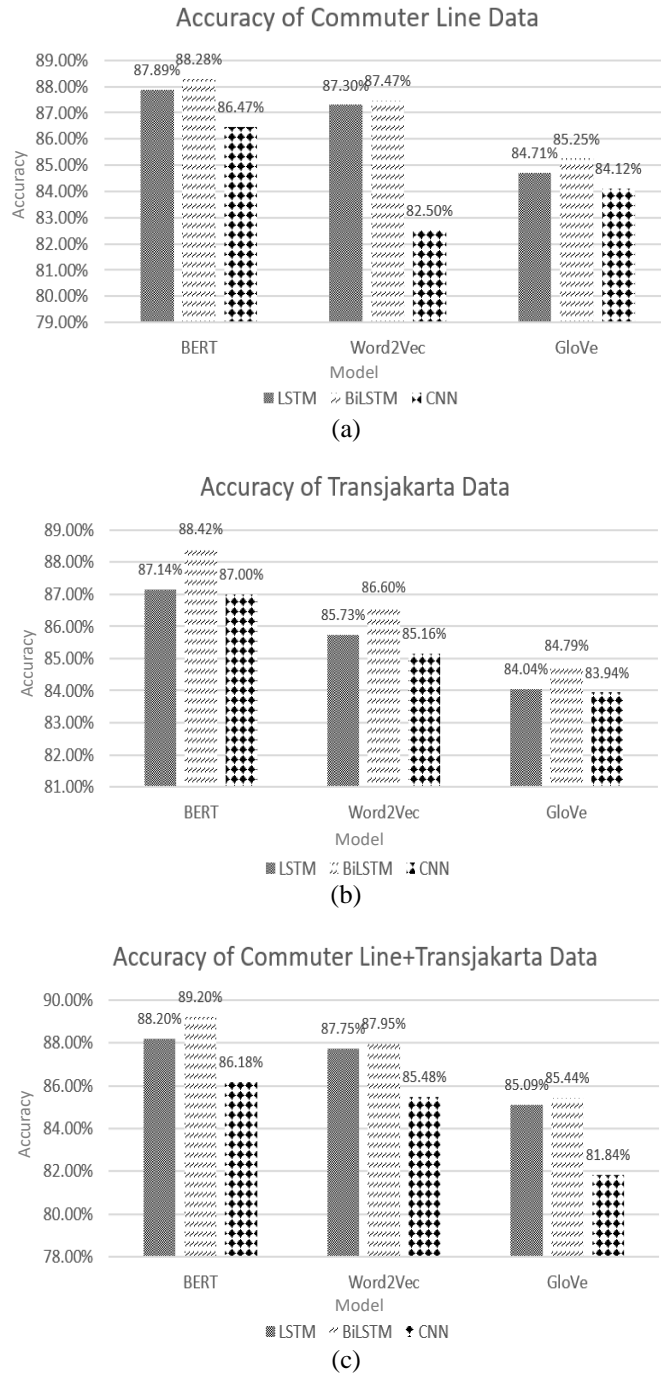

(c)

Figure 3. The comparison of accuracy (a) Commuter Line, (b) Transjakarta, and
(c) Commuter Line+Transjakarta

The word embedding used in this study shows BERT achieves superior results to Word2Vec and Glove. This is due to the fact that BERT produces word representations that are dynamically informed by surrounding words. BERT generates word representations that are affected by the words following and preceding them in a sentence. If the same word has multiple meanings (polysemy) in two separate sentences, BERT will produce two vectors for that word, which is used in two distinct contexts. In order for BERT to generate more precise features and enhance model performance, we must implement the suggested modifications. In contrast, each word in Word2Vec has a fixed representation irrespective of it is context. Word2Vec will generate the same single vector for polysemic words in both sentences; therefore, it cannot accurately represent the sentence context.

Word2Vec is superior to GloVe due to the fact that the embedding it generates indicates whether words are used in similar contexts. Word2Vec can determine the linear relationship between adjacent vector space words. This maximizes the probability of any term being close to the target, thereby enhancing accuracy. In the meantime, the Glove embeddings pertain to the probabilities that two words appear together in the corpus. GloVe is incapable of understanding the linear relationship between vector space concepts. This makes GloVe less precise than Word2Vec.

This study's findings were also contrasted to those of previous research [11]. Using BERT or BiLSTM, previous studies have developed a classification system for emotions. The values of accuracy, precision, recall, and F1-measure from previous studies are compared to those of the model proposed in this study in Table 3.

Table 3 shows that the BERT-BiLSTM model proposed in this study got better values for accuracy, precision, recall, and F1-measure in all data scenarios than other models. The next models that have the highest performance metric values are BERT and BiLSTM, which were proposed in previous studies. This study's findings have the potential to enhance the performance of emotion detection systems compared to previous research. Consequently, this study has the advantage of enhancing the accuracy, precision, recall, and F1-measure evaluation results for emotion detection in Indonesian texts.

Table 2. The comparison of precision, recall, and F1-measure

Data	Word embedding	Model	Precision	Recall	F1-measure
Commuter Line	BERT	LSTM	84.49%	87.89%	85.82%
		BiLSTM	**85.64%**	**88.28%**	**86.22%**
		CNN	82.55%	86.47%	84.42%
	Word2Vec	LSTM	83.94%	87.30%	84.57%
		BiLSTM	**85.36%**	**87.47%**	**84.89%**
		CNN	79.29%	82.50%	80.85%
	GloVe	LSTM	81.00%	84.71%	82.81%
		BiLSTM	**81.56%**	**85.25%**	**83.46%**
		CNN	80.79%	84.12%	82.40%
Transjakarta	BERT	LSTM	83.36%	87.14%	85.12%
		BiLSTM	**84.71%**	**87.56%**	**85.42%**
		CNN	83.45%	87.00%	85.27%
	Word2Vec	LSTM	82.67%	85.73%	83.50%
		BiLSTM	**82.89%**	**86.60%**	**84.63%**
		CNN	81.40%	85.16%	83.13%
	GloVe	LSTM	81.15%	84.04%	81.32%
		BiLSTM	**82.12%**	**84.79%**	**82.10%**
		CNN	80.30%	83.94%	82.08%
Commuter Line + Transjakarta	BERT	LSTM	84.97%	87.37%	84.97%
		BiLSTM	**85.84%**	**89.08%**	**87.07%**
		CNN	83.03%	86.18%	84.58%
	Word2Vec	LSTM	84.27%	87.75%	85.79%
		BiLSTM	**84.60%**	**87.95%**	**85.93%**
		CNN	82.23%	85.48%	83.81%
	GloVe	LSTM	81.71%	85.09%	83.29%
		BiLSTM	**82.03%**	**85.44%**	**83.53%**
		CNN	79.97%	81.84%	80.71%

Table 3. Metrics of the model BERT, BiLSTM, and BERT-BiLSTM

Data	Model	Accuracy	Precision	Recall	F1-Measure
Commuter Line	BERT-BiLSTM	**88.28%**	**85.64%**	**88.28%**	**86.22%**
	BERT	87.42%	83.47%	83.45%	83.41%
	BiLSTM	85.22%	82.07%	78.54%	80.01%
Transjakarta	BERT-BiLSTM	**88.42%**	**84.71%**	**87.56%**	**85.42%**
	BERT	86.27%	83.39%	83.54%	83.47%
	BiLSTM	82.83%	78.42%	77.37%	77.81%
Commuter Line + Transjakarta	BERT-BiLSTM	**89.20%**	**85.84%**	**89.08%**	**87.07%**
	BERT	86.99%	83.45%	83.79%	83.60%
	BiLSTM	86.41%	82.67%	81.67%	82.15%

4. CONCLUSION

This study successfully identified emotions in Indonesian tweets using the Commuter Line and Transjakarta review datasets. This study is a combination of nine different word embedding (BERT, Word2Vec, and GloVe) and emotion detection (BiLSTM, LSTM, and CNN) scenarios. The BERT-BiLSTM model generates the highest accuracy on Commuter Line, Transjakarta, and Commuter Line+Transjakarta

data with values of 88.28%, 88.42%, and 89.20%, respectively. In general, the classification model that obtains the best accuracy is BiLSTM, followed by LSTM, and finally CNN. While for word embedding, BERT embedding obtained the highest accuracy, followed by Word2Vec, and finally GloVe. Precision, recall, and F1-measure are used to evaluate the performance of the emotion detection model. The BERT-BiLSTM model also generates the highest precision, recall, and F1-measure values compared to other models in each data scenario. The results of this study show that BERT-BiLSTM can enhance the classification model's performance compared to previous studies that only used BERT or BiLSTM for emotion detection in Indonesian texts.

In future research, researchers suggest combining the BiLSTM model with other classification models, such as CNN, because combining the two deep learning models is expected to improve performance in emotion detection in text. The dataset in this study can also be expanded not only using social media reviews but also using formal text from Indonesian-language documents so as to expand the benefits of the emotion detection model built.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. Poria, A. Gelbukh, E. Cambria, A. Hussain, and G.-B. Huang, "EmoSenticSpace: A novel framework for affective common-sense reasoning," *Knowledge-Based Systems*, vol. 69, pp. 108–123, Oct. 2014, doi: 10.1016/j.knosys.2014.06.011.

[2] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, "Understanding emotions in text using deep learning and big data," *Computers in Human Behavior*, vol. 93, pp. 309–317, Apr. 2019, doi: 10.1016/j.chb.2018.12.029.

[3] J. Errasti, I. Amigo, and M. Villadangos, "Emotional uses of Facebook and Twitter: its relation with empathy, narcissism, and self-esteem in adolescence," *Psychological Reports*, vol. 120, no. 6, pp. 997–1018, Dec. 2017, doi: 10.1177/0033294117713496.

[4] K. Becker, V. P. Moreira, and A. G. L. dos Santos, "Multilingual emotion classification using supervised learning: Comparative experiments," *Information Processing and Management*, vol. 53, no. 3, pp. 684–704, May 2017, doi: 10.1016/j.ipm.2016.12.008.

[5] J. Charoensuk and O. Sornil, "A hierarchical emotion classification technique for Thai reviews," *Journal of ICT Research and Applications*, vol. 12, no. 3, Dec. 2018, doi: 10.5614/itbj.ict.res.appl.2018.12.3.6.

[6] A. Uçan, M. Dörterler, and E. Akçapınar Sezer, "A study of Turkish emotion classification with pretrained language models," *Journal of Information Science*, vol. 48, no. 6, pp. 857–865, Dec. 2022, doi: 10.1177/0165551520985507.

[7] H. L. Simanjuntak, "The translation of English noun phrase into Indonesian," *Scope: Journal of English Language Teaching*, vol. 4, no. 2, p. 81, Jun. 2020, doi: 10.30998/scope.v4i2.6507.

[8] D. E. Cahyani, F. Meilani, and S. W. Sihwi, "Emotions comparison of Commuter Line and Transjakarta users based on twitter using multinomial naïve bayes classifier," in *2019 5th International Conference on Science and Technology (ICST)*, Jul. 2019, pp. 1–6, doi: 10.1109/ICST47872.2019.9166171.

[9] Q. Zhao, "Social emotion classification of Japanese text information based on SVM and KNN," *Journal of Ambient Intelligence and Humanized Computing*, Mar. 2021, doi: 10.1007/s12652-021-03034-x.

[10] F. M. Alotaibi, "Classifying text-based emotions using logistic regression," *VAWKUM Transactions on Computer Sciences*, pp. 31–37, Apr. 2019, doi: 10.21015/vtcs.v16i2.551.

[11] D. E. Cahyani, D. Satyananda, L. T. Oktoviana, A. S. Kaspambudi, A. S. Anwar, and W. A. Sekti, "Emotions classification using bidirectional encoder representations from transformers," in *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, Sep. 2021, pp. 165–169, doi: 10.1109/ICITACEE53184.2021.9617473.

[12] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, "Transformer models for text-based emotion detection: a review of BERT-based approaches," *Artificial Intelligence Review*, vol. 54, no. 8, pp. 5789–5829, Dec. 2021, doi: 10.1007/s10462-021-09958-2.

[13] R. Fei, Y. Zhu, Q. Yao, Q. Xu, and B. Hu, "A deep learning method based self-attention and Bi-directional LSTM in emotion classification," *Journal of Internet Technology*, vol. 21, no. 5, pp. 1447–1461, 2020.

[14] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-emotion neural network for emotion recognition from text," *IEEE Access*, vol. 7, pp. 111866–111878, 2019, doi: 10.1109/ACCESS.2019.2934529.

[15] V. Ganesh and M. Kamarasan, "Deep learning based long short term memory model for emotions with intensity level sentiment classification for twitter texts," *International Journal of Advanced Science and Technology*, vol. 29, no. 5, pp. 2916–2929, 2020.

[16] H. Jiang, D. Wu, R. Jiao, and Z. Wang, "Analytical comparison of two emotion classification models based on convolutional neural networks," *Complexity*, vol. 2021, pp. 1–9, Feb. 2021, doi: 10.1155/2021/6625141.

[17] Z. Rajabi, A. Shehu, and O. Uzuner, "A multi-channel BiLSTM-CNN model for multilabel emotion classification of informal text," in *2020 IEEE 14th International Conference on Semantic Computing (ICSC)*, Feb. 2020, pp. 303–306, doi: 10.1109/ICSC.2020.00060.

[18] R. Cai *et al.*, "Sentiment analysis about investors and consumers in energy market based on BERT-BiLSTM," *IEEE Access*, vol. 8, pp. 171408–171415, 2020, doi: 10.1109/ACCESS.2020.3024750.

[19] H. Al-Omari, M. A. Abdullah, and S. Shaikh, "EmoDet2: emotion detection in English textual dialogue using BERT and BiLSTM models," in *2020 11th International Conference on Information and Communication Systems (ICICS)*, Apr. 2020, pp. 226–232, doi: 10.1109/ICICS49469.2020.239539.

[20] Y. Wang *et al.*, "A comparison of word embeddings for the biomedical natural language processing," *Journal of Biomedical Informatics*, vol. 87, pp. 12–20, Nov. 2018, doi: 10.1016/j.jbi.2018.09.008.

[21] Q. Wang, P. Liu, Z. Zhu, H. Yin, Q. Zhang, and L. Zhang, "A text abstraction summary model based on BERT word embedding and reinforcement learning," *Applied Sciences*, vol. 9, no. 21, Nov. 2019, doi: 10.3390/app9214701.

[22] K. W. Church, "Word2Vec," *Natural Language Engineering*, vol. 23, no. 1, pp. 155–162, Jan. 2017, doi: 10.1017/S1351324916000334.

[23] R. Adipradana, B. P. Nayoga, R. Suryadi, and D. Suhartono, "Hoax analyzer for Indonesian news using RNNs with fasttext and glove embeddings," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 10, no. 4, pp. 2130–2136, Aug. 2021, doi: 10.11591/eei.v10i4.2956.

[24] O. Melamud, D. McClosky, S. Patwardhan, and M. Bansal, "The role of context types and dimensionality in learning word embeddings," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1030–1040, doi: 10.18653/v1/N16-1118.

[25] L. Hickman, S. Thapa, L. Tay, M. Cao, and P. Srinivasan, "Text preprocessing for text mining in organizational research: review and recommendations," *Organizational Research Methods*, vol. 25, no. 1, pp. 114–146, Jan. 2022, doi: 10.1177/1094428120971683.

[26] M. Adriani, J. Asian, B. Nazief, S. M. M. Tahaghoghi, and H. E. Williams, "Stemming Indonesian," *ACM Transactions on Asian Language Information Processing*, vol. 6, no. 4, pp. 1–33, Dec. 2007, doi: 10.1145/1316457.1316459.

[27] S. Li, R. Pan, H. Luo, X. Liu, and G. Zhao, "Adaptive cross-contextual word embedding for word polysemy with unsupervised topic modeling," *Knowledge-Based Systems*, vol. 218, Apr. 2021, doi: 10.1016/j.knosys.2021.106827.

[28] R. Rehurek and P. Sojka, "Gensim--python framework for vector space modelling," *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, vol. 3, no. 2, 2011.

[29] J. Pennington, R. Socher, and C. Manning, "Glove: global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543, doi: 10.3115/v1/D14-1162.

[30] J. Xu, Y. Zhang, and D. Miao, "Three-way confusion matrix for classification: A measure driven view," *Information Sciences*, vol. 507, pp. 772–794, Jan. 2020, doi: 10.1016/j.ins.2019.06.064.

BIOGRAPHIES OF AUTHORS

Aji Prasetya Wibawa (ID) received the Bachelor of Engineering (S.T.) degree in electrical engineering from Universitas Brawijaya, Indonesia, in 2004 and the Master in Management of Technology (M.MT.) degree in information management technology from Institut Teknologi 10 November, Indonesia in 2007 and Ph.D. degrees in electrical and information engineering from University of South Australia, Australia in 2014. Currently, he is a Lecturer at State University of Malang, Department of Electrical Engineering, Indonesia. His interest research is natural language processing (NLP), machine translation, data science-analytics. He can be contacted at email: aji.prasetya.ft@um.ac.id.

Denis Eka Cahyani (ID) holds a Bachelor of Computer Science (S. Kom.) in computer science, Master of Computer Science (M.Kom.) in computer science, Universitas Indonesia in 2015 besides several professional certificates and skills. Holds a Bachelor of Informatics degree from Universitas Sebelas Maret, Indonesia in 2013. She is currently lecturing with the Department of Mathematics at Universitas Negeri Malang, Malang, Indonesia. She is a member of the Engineers and the Institute of Electrical and Electronics Engineers (IEEE) Indonesia Section. Her research areas of interest include data science, natural language processing, and artificial intelligent. She can be contacted at email: denis.eka.cahyani.fmipa@um.ac.id.

Didik Dwi Prasetya (ID) received the Bachelor of Engineering (S.T.) degree in informatics engineering from Universitas Ahmad Dahlan, Indonesia, in 2004 and the Master of Engineering (M.T.) degree in informatics engineering from Institut Teknologi Bandung in 2007 and D.Eng. degrees in Department of Information Engineering from Hiroshima University, Japan in 2021. Currently, he is a Lecturer at State University of Malang, Department of Electrical Engineering, Indonesia. His interest research is software engineering, web and mobile advanced learning technology, knowledge engineering, and data science. He can be contacted at email: didikdwi@um.ac.id.

Langlang Gumilar received the Bachelor of Applied Science (S.ST.) degree in electromechanics from STTN-BATAN, Indonesia, in 2013 and the Master of Engineering (M.T.) degree in electrical engineering from Universitas Indonesia, Indonesia in 2015. Currently, he is a Lecturer at State University of Malang, Department of Electrical Engineering, Indonesia. His interest research is electrical power system, power quality, renewable energy, and energy management. He can be contacted at email: langlang.gumilar.ft@um.ac.id.

Andrew Nafalski Since 2003 Andrew has been a Professor of Electrical Engineering, University of South Australia. He was a visiting professor at Kanazawa University, Kanazawa, Shinshu University, Nagano, Kyushu University, Fukuoka, Doshisha University, Kyoto in Japan, Physikalish-Technishe Bundesanstalt, Braunschweing in Germany, University of Wales, Cardiff in the UK, Toronto University, Ryerson University, University of Calgary in Canada, Ecole Supérieure d'Ingénieurs en Genie Electrique (ESIGELEC), Rouen in France, Massachusetts Institute of Technology, Cambridge and New York University, Brooklyn in the USA. Between February 25, 2000 and March 30, 2006, Andrew was Professor and Head of School of Electrical and Information Engineering, University of South Australia. Between July 1, 1996 and February 2000, he was a Professor and Head of School of Physics and Electronic Systems Engineering, University of South Australia. Between 1993 and 1996, he was a professor (since August 9, 1993) and Head of School of Electrical Engineering (January 1, 1994-June 30, 1996), University of South Australia. Since 1994, he has been Head of the Magnetics Technology Group, University of South Australia. Between 1992 and 1993 (18 months), he was a Senior Lecturer, School of Electrical and Computer Engineering, Curtin University of Technology, Perth, Western Australia. Between 1970 and 1991, he was with the Lublin University of Technology, Lublin, Poland where after being Junior Assistant, Assistant, Senior Assistant, Assistant Professor and Associate Professor he reached in 1990 the rank of a Professor and Head of Department of Information Technology. He was also Director of the University's Computing Centre for 13 months. He can be contacted at email: andrew.nafalski@unisa.edu.au.