IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 2, June 2024, pp. 1753~1761
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i2.pp1753-1761

Determining community happiness index with transformers and attention-based deep learning

Hilman Singgih Wicaksana[1], Retno Kusumaningrum[2], Rahmat Gernowo[3]
[1]Master Program of Information System, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia
[2]Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia
[3]Department of Pyshics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

Article Info

Article history:

Received Aug 9, 2023
Revised Sep 19, 2023
Accepted Sep 29, 2023

Keywords:

Aspect-based sentiment
Analysis
Bidirectional encoder
Representations from
Transformers
Happiness index
Long short-term memory
Twitter

ABSTRACT

In the current digital era, evaluating the quality of people's lives and their happiness index is closely related to their expressions and opinions on Twitter social media. Measuring population welfare goes beyond monetary aspects, focusing more on subjective well-being, and sentiment analysis helps evaluate people's perceptions of happiness aspects. Aspect-based sentiment analysis (ABSA) effectively identifies sentiments on predetermined aspects. The previous study has used word-to-vector (Word2Vec) and long short-term memory (LSTM) methods with or without attention mechanism (AM) to solve ABSA cases. However, the problem with the previous study is that Word2Vec has the disadvantage of being unable to handle the context of words in a sentence. Therefore, this study will address the problem with bidirectional encoder representations from transformers (BERT), which has the advantage of performing bidirectional training. Bayesian optimization as a hyperparameter tuning technique is used to find the best combination of parameters during the training process. Here we show that BERT-LSTM-AM outperforms the Word2Vec-LSTM-AM model in predicting aspect and sentiment. Furthermore, we found that BERT is the best state-of-the-art embedding technique for representing words in a sentence. Our results demonstrate how BERT as an embedding technique can significantly improve the model performance over Word2Vec.

Corresponding Author:

Hilman Singgih Wicaksana
Master Program of Information System, School of Postgraduate Studies, Diponegoro University
Street. Imam Bardjo SH No. 5, Semarang, Central Java, Indonesia
Email: singgih.hilman@gmail.com/hilmansinggihw@students.undip.ac.id

1. INTRODUCTION

In the current digital era, people are accustomed to using social media to express an opinion, expression, or response to news and information. This is the most critical foundation in knowing the situation, conditions, and circumstances experienced by the community so that the happiness index parameter can be measured. The happiness index has proven to be effective in evaluating social validity and providing the understanding needed to address problems and improve the overall quality of life of the community. Therefore, the use of social media and the happiness index parameter are two factors that are closely related to evaluating the quality of people's lives in the current digital era [1].

The index of community happiness that has been determined by the central statistics agency or badan pusat statistik (BPS) in Indonesian consists of 9 aspects: health, education, employment, income, security, social relations, availability of free time, family harmony, home conditions, and environmental conditions [2]. Nowadays, it is increasingly emphasised that measuring the population's welfare is important, not only through monetary aspects. The happiness indicators created are not only intended to describe the conditions of material prosperity but focus more on the subjective well-being of each individual. In this context, sentiment analysis is needed to evaluate how people respond to and perceive aspects of happiness. Sentiment analysis can help measure the extent of people's opinions, views, and feelings regarding these aspects. Therefore, sentiment analysis can help obtain more complete and in-depth information about individuals' subjective well-being and provide a more comprehensive representation of people's happiness.

Sentiment analysis is a technique that can be used to identify sentiments or feelings contained in a language, opinion, and others. This technique is widely applied to analyse, anticipate, and assess the view of text data [3]. Three types of sentiment analysis can be used: document-level sentiment analysis, sentence-level sentiment analysis and aspect-based sentiment analysis [4]. Document-level sentiment analysis can only determine the overall sentiment in a document. At the same time, sentence-level sentiment analysis can only decide on the idea in each sentence separately. Therefore, aspect-based sentiment analysis is more suitable to be used in the case of sentiment analysis on the happiness index because it can help in identifying sentiments on each aspect of happiness, thus providing more detailed information and assisting in evaluating the happiness index set by the central statistics agency (BPS).

Aspect-based sentiment analysis (ABSA) is a sentiment analysis approach that can generate sentiment ratings on predetermined aspects [5]. The system has been developed in a study to perform aspect-based sentiment analysis on hotel review data where the elements consist of food, room, service, location, and others. The approach uses the long short-term memory (LSTM) model with word-to-vector (Word2Vec) as the word embedding technique. It obtained an f1-score value of 75.28% for the best model based on the first hidden layer size of 1,200 neurons with tanh activation function and the second discreet layer size of 600 neurons with rectified linear unit (ReLU) activation function [6].

In addition, another study has also been developed by [7] using a combination model of Word2Vec and LSTM with an attention mechanism on hotel review data with the same aspects as previously determined. However, in his study, using a double fully-connected layer to improve the performance of the LSTM model. Thus, the best model performance produces an f1-score value of 76.28% based on the parameters of the hidden layer unit of 128 neurons, a dropout parameter of 0.3, and a recurrent dropout of 0.3. Thus, the model's performance with an attention mechanism is superior to that without an attention mechanism, which only obtained an f1-score value of 75.28%.

In terms of previous studies, the study conducted by Jayanto *et al*. [6] and Cendani *et al*. [7] both use the Word2Vec model. The survey conducted by Cendani *et al*. [7] added applying the attention mechanism layer after the LSTM layer to improve the model's performance. In contrast, the study conducted by Jayanto *et al*. [6] did not use the attention mechanism layer. However, using Word2Vec as a word embedding technique has problems overcoming the context of words in a sentence. This can be overcome by applying bidirectional encoder representations from transformers (BERT), where BERT can overcome these problems by training in two directions, as has been done by Ingkafi [8]. Therefore, this study will propose a combination of BERT and LSTM models with attention mechanisms to improve the model's performance in predicting aspects and sentiments, which can then be used to identify the community happiness index.

2. METHOD

The study was conducted in three stages: dataset preparation, word embedding technique, and model building. The model building comprises six stages: data splitting, hyperparameter tuning, model training, classification model, testing, and evaluation. The entire process of this study is shown in Figure 1.

2.1. Dataset preparation

This section collects a dataset of 5,400 Indonesian tweets from a previous study [8]. Furthermore, the dataset is subjected to data pre-processing, which includes data cleaning, case folding, tokenization, word normalization, and data variation. Data cleaning is done to clean unnecessary characters, hyperlinks, Unicode, and so on [9]. Case folding is done to change capital letters to lowercase letters in a sentence as a whole [10]. Tokenization is done to separate words per word from a sentence using BERT Tokenizer [11]. Word normalization is done by converting informal words into formal ones, according to the *Kamus Besar Bahasa Indonesia* (KBBI) or The Big Indonesian Dictionary, if in English. Data variation is done manually by inserting, deleting, or rearranging existing data. Table 1 explains an example of data pre-processing carried out in this study with an example sentence in Indonesian, namely "*Kayaknya aku memang harus banyak banyak belajar sejarah lagi deh*" which if in English is "I think I really have to learn a lot of history again".

In addition, the one-hot encoding stage is also carried out to convert aspect and sentiment classes into numerical form. The aspect task uses 9 classes, including social relations (*hubungan sosial*), security (*keamanan*), family harmony (*keharmonisan keluarga*), health (*kesehatan*), leisure availability (*ketersediaan* *waktu luang*), living environment (*lingkungan hidup*), employment (*pekerjaan*), income (*pendapatan*), and education (*pendidikan*). Meanwhile, the sentiment class uses three classes which include negative (*negatif*), neutral (*netral*), and positive (*positif*). Table 2 gives an implementation of its one-hot encoding representation.

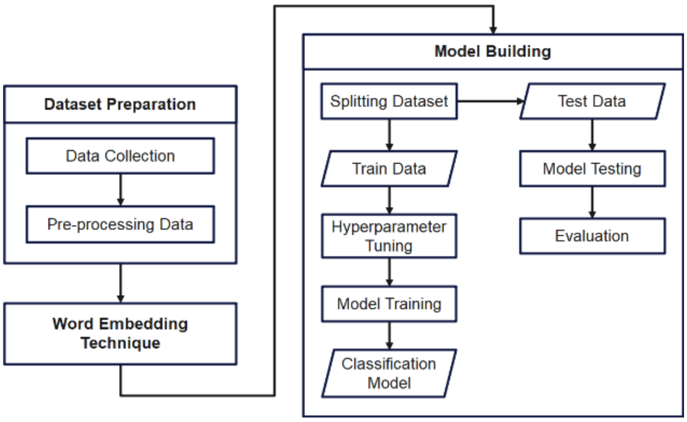

Figure 1. Research overview

Table 1. Example of pre-processing data

The Phases	Before Implementation	After Implementation
Data Cleaning	*Kayaknya aku memang harus banyak banyak belajar sejarah lagi deh (((:*	*Kayaknya aku memang harus banyak banyak belajar sejarah lagi deh*
Case Folding	*Kayaknya aku memang harus banyak banyak belajar sejarah lagi deh*	*kayaknya aku memang harus banyak banyak belajar sejarah lagi deh*
Tokenization	*kayaknya aku memang harus banyak banyak belajar sejarah lagi deh*	["*kayaknya*", "*aku*", "*memang*", "*harus*", "*banyak*", "*banyak*", "*belajar*", "*sejarah*", "*lagi*", "*deh*"]
Word Normalization	*kayaknya aku memang harus banyak banyak belajar sejarah lagi deh*	*sepertinya aku memang harus banyak belajar sejarah lagi deh*
Data Variation	*sepertinya aku memang harus banyak belajar sejarah lagi deh*	*sepertinya aku memang harus banyak belajar sejarah lagi*

Table 2. One-hot encoding representation

Tasks	Classes	Number of Classes	Representation Results
Aspect	Social Relations (*Hubungan Sosial*)		[1, 0, 0, 0, 0, 0, 0, 0, 0]
	Security (*Keamanan*)		[0, 1, 0, 0, 0, 0, 0, 0, 0]
	Family Harmony (*Keharmonisan Keluarga*)		[0, 0, 1, 0, 0, 0, 0, 0, 0]
	Health (*Kesehatan*)		[0, 0, 0, 1, 0, 0, 0, 0, 0]
	Leisure Availability (*Ketersediaan Waktu Luang*)	9	[0, 0, 0, 0, 1, 0, 0, 0, 0]
	Living Environment (*Lingkungan Hidup*)		[0, 0, 0, 0, 0, 1, 0, 0, 0]
	Employment (*Pekerjaan*)		[0, 0, 0, 0, 0, 0, 1, 0, 0]
	Income (*Pendapatan*)		[0, 0, 0, 0, 0, 0, 0, 1, 0]
	Education (*Pendidikan*)		[0, 0, 0, 0, 0, 0, 0, 0, 1]
Sentiment	Negative (*Negatif*)		[1, 0, 0]
	Neutral (*Netral*)	3	[0, 1, 0]
	Positive (*Positif*)		[0, 0, 1]

2.2. Word embedding technique

The word embedding technique is a form of word representation that connects human understanding of knowledge meaningfully with machine understanding. The representation can be a set of real numbers (vector). The technique is divided into 3 types, namely traditional word embedding, static word embedding, and contextualized word embedding [12]. Based on the previously mentioned types of word embedding, BERT belongs to the contextualized word embedding type, while Word2Vec belongs to the static word embedding type.

The word embedding technique performed in this study is BERT as the primary technique and Word2Vec as the benchmark technique. The BERT embedding technique is performed with a previously trained model to be retrained with the dataset in this study or what is referred to as the fine-tuning process. Meanwhile, the Word2Vec embedding technique is first trained on the existing dataset using the skip-gram architecture with its default settings. This embedding technique will process input as a sentence with a specified sentence length of 64 as its max length.

2.3. Model building

In building the model, the primary step is to divide the dataset into training, validation, and testing data with a percentage of 80%, 10%, and 10%, respectively. Training and validation data are used during the model training process, while test data can be performed when the model has been trained. This study uses hyperparameter tuning using bayesian optimization [13] to find the best parameters from each experiment conducted. This technique will produce the best parameter combination with high validation accuracy to be used in the testing process on test data. Table 3 describes the parameters and values of the model training hyperparameter.

According to Table 3, hyperparameter tuning was performed by conducting training using three different parameter combinations. These parameters included dropout, learning rate, and hidden unit in LSTM. The best parameters with optimal validation accuracy can be determined using bayesian optimization. During the training process, the validation accuracy value of each scenario is generated, then the best validation accuracy is selected. After that, the model can be used to make predictions on new data that has never been seen. Table 4 describes the summary of several scenarios when performing hyperparameter tuning with bayesian optimization.

Table 3. Model training hyperparameter

Parameters	Values
Dropout	0.1, 0.3, 0.5
Learning Rate	0.00001, 0.0001, 0.001, 0.01
Hidden units of LSTM	128, 256, 512

Table 4. Bayesian optimization scenario

Scenario	Dropout	Learning Rate	Hidden Units of LSTM	Validation Accuracy
Scenario 1	0.1	0.00001	128	val_acc_1
Scenario 2	0.3	0.00001	128	val_acc_2
Scenario 3	0.5	0.00001	128	val_acc_3
Scenario 34	0.1	0.01	512	val_acc_{34}
Scenario 35	0.3	0.01	512	val_acc_{35}
Scenario 36	0.5	0.01	512	val_acc_{36}

According to Table 4, there are 36 possible scenarios for hyperparameter tuning using bayesian optimization. Each scenario involves a unique combination of parameters such as dropout, learning rate, and hidden units of the LSTM, as outlined in Table 3. After each scenario is run, validate accuracy values from scenario 1 symbolized as val_acc_1 to validate accuracy values from scenario 36 symbolized as val_acc_{36}, and then select the best validation accuracy value determined based on the highest value. Therefore, the scenario can be run properly during testing.

This study has produced two models, namely, a model for predicting aspects and a model for predicting sentiment using different word embedding techniques, namely BERT and Word2Vec. The model architecture proposed in this study consists of an input layer which can be symbolized as x_0 to x_n where n is the word length of a sentence based on the specified max length. The embedding layer with BERT [14] and Word2Vec [15] word embedding techniques are performed separately. Furthermore, the output of BERT and Word2Vec is forwarded to the LSTM model as its input.

The LSTM model is a development of the recurrent neural network (RNN) model to overcome vanishing gradient or exploding gradient problems [16]. With the existence of 3 gates, which include the input gate, the forget gate and the output gate, it can function to control the flow of information in and out of the memory cell [17]. After using LSTM, it is continued with the addition of an attention mechanism layer to improve the quality of predictions or outputs produced by focusing on the most influential parts of the final result [18]. The attention mechanism was first proposed by Bahdanau *et al*. [19] using additive attention techniques and Luong *et al*. [20] using multiplicative attention techniques. After that, the use of attention mechanisms began to be applied to text classification proposed by Raffel and Ellis [21]. The dropout after the attention mechanism layer aims to reduce overfitting by randomly removing some neurons during training [22]. The output layer acts as a classifier layer that simultaneously provides prediction results in the form of aspects and sentiments. The output layer can also be referred to as a dense layer or fully-connected layer, a type of layer in a neural network that connects each neuron in the coating and each neuron in the previous layer [23].

This classification task is the primary purpose of using fully-connected layers in neural networks [24]. The activation function used is sigmoid because the activation is able to understand binary data consisting of 0 and 1 [25]. In addition, loss functions are also required to evaluate candidate prediction solutions and prediction errors [26]. The loss function used is a cross-entropy loss function with an output in the range of values between 0 and 1. There are two types of cross-entropy loss functions: categorical cross-entropy and binary cross-entropy [26], [27]. This study will use categorical cross-entropy to process categorical data on aspect and sentiment classes. The architecture of the proposed model is shown in Figure 2, and it is consistent with the statements mentioned earlier.

Evaluation of the model must include consideration of a variety of performance metrics. Adequate evaluation of the model's performance requires a thorough examination of parameters such as accuracy, precision, recall, and f1-score derived from the confusion matrix table. It is essential that these factors be taken into account in order to achieve a comprehensive evaluation [28]. Figure 3 shows the confusion matrix consisting of true positive (TP), false positive (FP), false negative (FN), and true negative (TN).

Based on the confusion matrix shown in Figure 3, we can use model evaluation metrics like precision, recall, f1-score, and accuracy. Precision is the ratio of positive correct predictions compared to all positive predicted results, recall is the ratio of positive correct predictions compared to all positive correct data, f1-score is a weighted average comparison of precision and recall, and accuracy is the ratio of correct predictions (positive and negative) to all data [29]. These metrics are described in (1), (2), (3), and (4).

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

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

$$f1 - score = \frac{2\times Recall \times Precision}{Recall + Precision} \tag{3}$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{4}$$

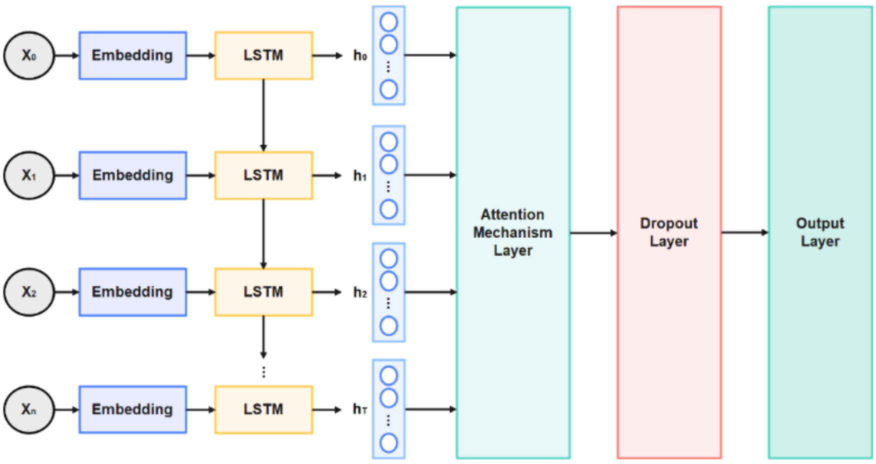

Figure 2. Proposed model architecture

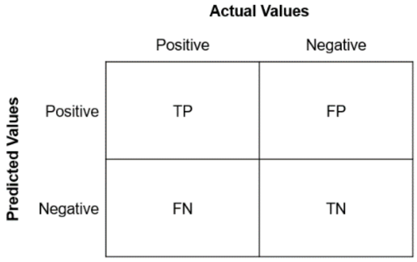

Figure 3. Confusion matrix

3. RESULTS AND DISCUSSION

This section will be explained in two parts: the selection and assessment models. Model selection is the first step in determining the best model on training data and validation data, and model assessment is a performance test of the best model selected previously on test data. The following paragraphs will discuss the model selection and model assessment, respectively.

3.1. Model selection

In the phase of selecting the model using bayesian optimization, several training sessions take place with different combinations of parameters. To obtain the best combination of parameters for aspect and sentiment tasks using the word embedding technique, the training has produced multiple scenarios based on the parameters outlined in Table 3.

During the training process, a total of 36 scenarios were selected for the best model with the highest validation accuracy value, as described in Table 4. The parameter combination includes a dropout of 0.3, a learning rate of 0.01, and hidden units of LSTM of 256 used for the aspect prediction task and a dropout of 0.1, a learning rate of 0.001, and hidden units of LSTM of 512 used for the sentiment prediction task. The scenario has the highest validation accuracy for aspect and sentiment prediction tasks, respectively, from the other scenarios, with a value of 0.97037 or approximately 97.04% for aspect prediction tasks and a value of 0.768519 or approximately 76.85% for sentiment prediction tasks. The scenarios were run if the BERT embedding technique was used.

However, if the Word2Vec embedding technique is used, the scenario will be different, so the parameter combination used is also different. A parameter combination that includes a dropout of 0.3, a learning rate of 0.01, and 512 hidden units of LSTM is used for the aspect prediction task, while a value of 0.1 dropout, 0.01 learning rate, and 256 hidden units of LSTM is used for the sentiment prediction task. The scenario has the highest validation accuracy for aspect and sentiment prediction tasks, respectively from the other scenarios with a value of 0.955556 or approximately 95.56% for aspect prediction task, while a value of 0.687037 or approximately 68.70% for sentiment prediction task. Figure 4 compares the model validation accuracy for solving aspect and sentiment prediction tasks.

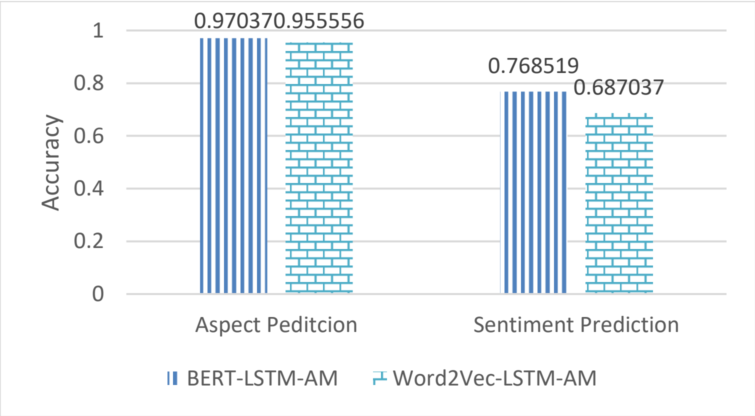

Figure 4. Model validation accuracy comparison

According to Figure 4, the BERT and LSTM model with attention mechanism (BERT-LSTM-AM) has the highest validation accuracy compared to the Word2Vec and LSTM model with attention mechanism (Word2Vec-LSTM-AM) in solving aspect and sentiment prediction tasks. After analyzing the results, the BERT-LSTM-AM model is better equipped to learn from the training data and accurately validate it during training. The BERT-LSTM-AM model produces a validation accuracy value on the aspect prediction task of 0.97037 or approximately 97.04%, and the validation accuracy value on the sentiment prediction task is 0.768519 or approximately 76.85%.

3.2. Model assessment

The effectiveness of the BERT-LSTM-AM and Word2Vec-LSTM-AM models in predicting aspects and sentiments are evaluated through a model assessment on relevant test data. This evaluation aims to determine the performance of each model based on the optimal parameter combination discussed in section 3.1. This evaluation will focus on deciding which model is the most effective at accurately representing a sentence as input and predicting its aspect and sentiment. It is crucial in selecting the best model for future data predictions. Figure 5 compares the accuracy of BERT-LSTM-AM and Word2Vec-LSTM-AM in testing models.

The experiment has produced a BERT-LSTM-AM model with the highest testing accuracy against test data of 0.950092 or approximately 95.01% on the aspect prediction task, while on the sentiment prediction task, it has produced a testing accuracy value of 0.746765 or approximately 74.68%. The model has outperformed again against the Word2Vec-LSTM-AM model on both aspect and sentiment prediction tasks. This indicates that the BERT-LSTM-AM model better understands the test data. In addition, the word embedding technique used is very influential in providing a good word representation. The better the word embedding technique provides word representation, the better the model performs aspect and sentiment prediction tasks. Therefore, the BERT-LSTM-AM model can be excellent for testing new data due to the effect of using BERT as a word embedding technique in the model. Table 5 displays the precision, recall, f1-score, and accuracy values for the BERT-LSTM-AM and Word2Vec-LSTM-AM models.

According to the metrics used to assess its performance, the BERT-LSTM-AM model effectively predicts both aspects and sentiments. Regarding aspect prediction, the model displays remarkably high levels of precision, recall, f1-score, and accuracy, nearly reaching a perfect score of 1, meaning it can identify aspects accurately. Regarding sentiment prediction, while the precision, recall, f1-score, and accuracy values are not as high as those for aspect prediction, they are still strong enough to make reliable predictions. Therefore, the model can be trusted to make accurate sentiment predictions.

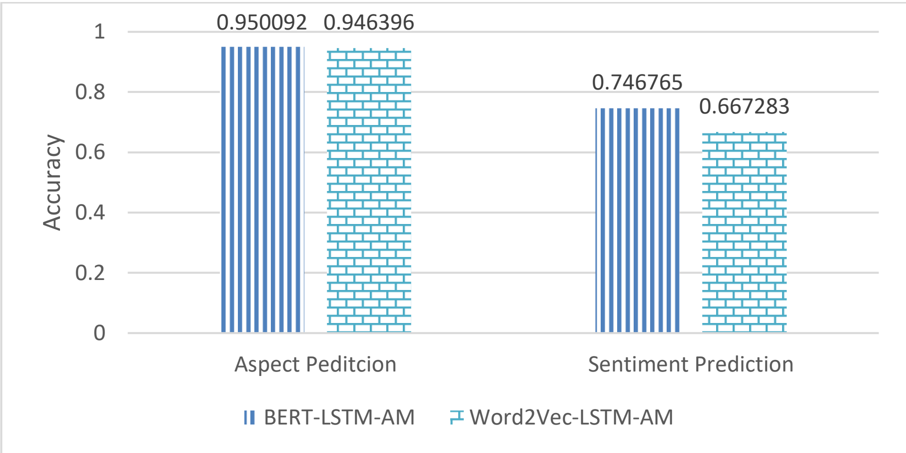

Figure 5. Model testing accuracy comparison

Table 5. Model performance evaluation

Models	Tasks	Evaluation Metrics			
		Precision	Recall	F1-score	Accuracy
BERT-LSTM-AM	Aspect Prediction	0.9491	0.9524	0.9500	0,9501
	Sentiment Prediction	0.7433	0.7422	0.7392	0,7468
Word2Vec-LSTM-AM	Aspect Prediction	0.9483	0.9487	0.9483	0.9464
	Sentiment Prediction	0.6671	0.6642	0.6646	0.6673

Although the results produced have very good values, some classification errors still occur. For example, there are some tweets with positive sentiment classes that are predicted as negative sentiment classes. This happens because some sentences have two words contained in the sentence that contain sentiment polarity. However, these sentences were well predicted in classifying the "*pekerjaan*" aspect of the word "*kerja*". For example, in the sentence "*mending lelah kerja dari pada lelah cari kerja semangat promo besok*", there are two words that can represent two different sentiments, namely the word "*mending*" for positive sentiment and the word "*lelah*" for negative sentiment. The word "*mending*" at the beginning of the sentence explains the comparative meaning, while the word "*lelah*" has a negative sentiment polarity. Therefore, the sentence should give a positive sentiment, but the model gives a wrong classification because there is a word containing a comparative meaning in a sentence and a word containing a negative sentiment polarity.

This study aligns with the previous study conducted by Ingkafi [8], which conducted aspect and sentiment prediction separately. However, what distinguishes this study from previous study is that it combines the BERT and LSTM models with attention mechanism (BERT-LSTM-AM), where BERT is the embedding technique used in the study. In addition, this study also develops from study conducted by Jayanto *et al*. [6] and Cendani *et al*. [7]. The study conducted by Jayanto *et al*. [6] used Word2Vec and LSTM models without Attention Mechanism, while Cendani *et al*. [7] used Word2Vec and LSTM models with attention mechanism. The study conducted by Jayanto *et al*. [6] and Cendani *et al*. [7] both use Word2Vec as the embedding technique used in their study.

4. CONCLUSION

According to the findings of our study, the BERT-LSTM-AM model outperforms the Word2Vec-LSTM-AM model when it comes to forecasting sentiment and aspects. This is due to the difference in word embedding techniques used by the two models, with BERT and Word2Vec being the respective techniques used to represent words in a sentence. The BERT-LSTM-AM model has a 95.01% accuracy rate in predicting aspects and a 74.68% accuracy rate in predicting sentiment. In predicting aspects, the best parameter combination is a dropout of 0.3, a learning rate of 0.01, and hidden units in LSTM of 256. The best parameter combination in predicting sentiment includes a dropout of 0.1, a learning rate of 0.001, and hidden units in LSTM of 512. The parameters chosen for the model were determined through multiple scenarios run during the training process using Bayesian Optimization. This particular combination of parameters proved to be highly effective in achieving good model validity and test accuracy. In addition, this study is also influenced by the embedding technique used, namely BERT. The BERT embedding technique is the best state-of-the-art technique to produce a better word representation than Word2Vec.

ACKNOWLEDGEMENT

This research was conducted as a requirement to graduate in the Master Program of Information System, School of Postgraduate Studies, Diponegoro University. As the first author, I would like to thank Dr. Retno Kusumaningrum, S.Si., M.Kom. and Prof. Dr. Rahmat Gernowo, M.Si. as my supervisors, for their excellent guidance so that our manuscript has the best quality.

REFERENCES

[1] A. Carnett, L. Neely, M.-T. Chen, K. Cantrell, E. Santos, and S. Ala'i-Rosales, "How might indices of happiness inform early intervention research and decision making?," *Advances in Neurodevelopmental Disorders*, vol. 6, no. 4, pp. 567–576, Dec. 2022, doi: 10.1007/s41252-022-00288-0.
[2] U. Suchaini, W. P. S. Nugraha, I. K. D. Dwipayana, and S. A. Lestari, "The happiness index in indonesian: indeks kebahagiaan," *Badan Pusat Statistik RI*, pp. 1–185, 2021.
[3] A. Iqbal, R. Amin, J. Iqbal, R. Alroobaea, A. Binmahfoudh, and M. Hussain, "Sentiment analysis of consumer reviews using deep learning," *Sustainability*, vol. 14, no. 17, p. 10844, Aug. 2022, doi: 10.3390/su141710844.
[4] B. Liu, *Sentiment analysis and opinion mining*. Cham: Springer International Publishing, 2012. doi: 10.1007/978-3-031-02145-9.
[5] P. N. Andono, Sunardi, R. A. Nugroho, and B. Harjo, "Aspect-Based Sentiment Analysis for Hotel Review Using LDA, Semantic Similarity, and BERT," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 5, pp. 232–243, Oct. 2022, doi: 10.22266/ijies2022.1031.21.
[6] R. Jayanto, R. Kusumaningrum, and A. Wibowo, "Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory," *International Journal of Advances in Intelligent Informatics*, vol. 8, no. 3, p. 391, Nov. 2022, doi: 10.26555/ijain.v8i3.691.
[7] L. M. Cendani, R. Kusumaningrum, and S. N. Endah, "Aspect-based sentiment analysis of Indonesian-Language Hotel Reviews using long short-term memory with an attention mechanism," 2023, pp. 106–122. doi: 10.1007/978-3-031-15191-0_11.
[8] D. A. Ingkafi, "Aspect-based sentiment analysis in measuring the community happiness index of Semarang City on Twitter Social media using bidirectional encoder representations from transformers (BERT) in Indonesian: aspect-based sentiment analysis dalam pengukuran indeks," *Universitas Diponegoro Semarang*, pp. 1–60, 2022.
[9] A. Vohra and R. Garg, "Deep learning based sentiment analysis of public perception of working from home through tweets," *Journal of Intelligent Information Systems*, vol. 60, no. 1, pp. 255–274, Feb. 2023, doi: 10.1007/s10844-022-00736-2.
[10] N. A. M. Roslan, N. M. Diah, Z. Ibrahim, Y. Munarko, and A. E. Minarno, "Automatic plant recognition using convolutional neural network on malaysian medicinal herbs: the value of data augmentation," *International Journal of Advances in Intelligent Informatics*, vol. 9, no. 1, p. 136, Mar. 2023, doi: 10.26555/ijain.v9i1.1076.
[11] A. H. Oliaee, S. Das, J. Liu, and M. A. Rahman, "Using bidirectional encoder representations from transformers (BERT) to classify traffic crash severity types," *Natural Language Processing Journal*, vol. 3, p. 100007, Jun. 2023, doi: 10.1016/j.nlp.2023.100007.
[12] S. Selva Birunda and R. Kanniga Devi, "A review on word embedding techniques for text classification," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 59, pp. 267–281, 2021, doi: 10.1007/978-981-15-9651-3_23.
[13] M. Liu, Z. Wen, R. Zhou, and H. Su, "Bayesian optimization and ensemble learning algorithm combined method for deformation prediction of concrete dam," *Structures*, vol. 54, pp. 981–993, Aug. 2023, doi: 10.1016/j.istruc.2023.05.136.
[14] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 4171–4186, 2019.
[15] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 2013.

[16] K. Kusum and S. P. Panda, "Sentiment analysis using global vector and long short-term memory," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, p. 414, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp414-422.

[17] F. Kurniawan, Y. Romadhoni, L. Zahrona, and J. Hammad, "Comparing LSTM and CNN Methods in Case Study on Public Discussion about Covid-19 in Twitter," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 402–409, 2022, doi: 10.14569/IJACSA.2022.0131048.

[18] A. Vaswani *et al.*, "Attention is all you need," Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[19] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.0473

[20] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," Aug. 2015, [Online]. Available: http://arxiv.org/abs/1508.04025

[21] C. Raffel and D. P. W. Ellis, "Feed-forward networks with attention can solve some long-term memory problems," Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.08756

[22] SrivastavaNitish, HintonGeoffrey, KrizhevskyAlex, SutskeverIlya, and SalakhutdinovRuslan, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[23] W. Ma and J. Lu, "An equivalence of fully connected layer and convolutional layer," Dec. 2017, [Online]. Available: http://arxiv.org/abs/1712.01252

[24] N. Singh and H. Sabrol, "Convolutional neural networks-an extensive arena of deep learning. a comprehensive study," *Archives of Computational Methods in Engineering*, vol. 28, no. 7, pp. 4755–4780, Dec. 2021, doi: 10.1007/s11831-021-09551-4.

[25] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: comparison of trends in practice and research for deep learning," Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.03378

[26] I. Goodfellow, Y. B. And, and A. Courville, "Deep learning (adaptive computation and machine learning)," *Massachusetts Institute of Technology*, vol. 8, no. 9, pp. 1–58, 2016.

[27]. Usha Ruby Dr.A, "Binary cross entropy with deep learning technique for Image classification," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 4, pp. 5393–5397, Aug. 2020, doi: 10.30534/ijatcse/2020/175942020.

[28] Y. H. Park, "Gradients in a deep neural network and their Python implementations," *Korean Journal of Mathematics*, vol. 30, no. 1, pp. 131–146, 2022, doi: 10.11568/kjm.2022.30.1.131.

[29] K. N. Alam *et al.*, "Deep learning-based sentiment analysis of COVID-19 vaccination responses from Twitter data," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–15, Dec. 2021, doi: 10.1155/2021/4321131.

BIOGRAPHIES OF AUTHORS

Hilman Singgih Wicaksana holds a Bachelor Degree of Computer Science (S.Kom.) from Telkom Institute of Technology Purwokerto in 2020. He is currently a student in the Master Program of Information System at Diponegoro University, Semarang, Central Java, Indonesia. He has research interests in Machine Learning, Deep Learning, Natural Language Processing, and Data Mining. He can be contacted at email: singgih.hilman@gmail.com or hilmansinggihw@students.undip.ac.id.

Retno Kusumaningrum holds a Bachelor Degree of Science (S.Si.) from Diponegoro University, a Master Degree in Computer Science (M.Kom.) from the University of Indonesia, and a Doctoral Degree in Computer Science (Dr.) from the University of Indonesia. She is a Lecturer in the Department of Informatics at Diponegoro University, Semarang, Central Java, Indonesia. Her research interests are Computer Vision, Pattern Recognition, Natural Language Processing, Topic Modelling, and Machine Learning. She can be contacted at email: retno@live.undip.ac.id.

Rahmat Gernowo holds a Bachelor Degree of Science (Drs.) from Bandung Institute of Technology, a Master Degree of Science (M. Si) from Bandung Institute of Technology, and a Doctoral Degree of Science (Dr.) from Gadjah Mada University. He is a Professor and Lecturer in the Department of Physics at Diponegoro University, Semarang, Central Java, Indonesia. Currently, he serves as a Head of the Doctoral Program of Information System at Diponegoro University. His research interests are in Geophysics & Atmospheric Science. He can be contacted at email: rahmatgernowo@lecturer.undip.ac.id.