Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 1, January 2024, pp. 591~602
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i1.pp591-602

The hybrid of BERT and deep learning models for Indonesian sentiment analysis

Dwi Guna Mandhasiya, Hendri Murfi, Alhadi Bustamam
Department of Mathematics, Faculty of Mathematics and Science, Universitas Indonesia, Depok, Indonesia

Article Info

Article history:

Received Jun 28, 2023
Revised Sep 27, 2023
Accepted Oct 25, 2023

Keywords:

BERT
Deep learning
Machine learning
Neural network
Sentiment analysis

ABSTRACT

Artificial intelligence (AI) is one example of how data science innovation has advanced quickly in recent years and has greatly improved human existence. Neural networks, which are a type of machine learning model, are a fundamental component of deep learning in the field of AI. Deep learning models can carry out feature extraction and classification tasks in a single design because of their numerous neural network layers. Modern machine learning algorithms have been shown to perform worse than this model on tasks including text classification, audio recognition, imaginary, and pattern recognition. Deep learning models have outperformed AI-based methods in sentiment analysis and other text categorization tasks. Text data can originate from a number of places, including social media. Sentiment analysis is the computational examination of textual expressions of ideas and feelings. This study employs the convolutional neural network (CNN), long-short term memory (LSTM), CNN-LSTM, and LSTM-CNN models in a deep learning framework using bidirectional encoder representations from transformers (BERT) data representation to assess the performance of machine learning. The implementation of the model utilises YouTube discussion data pertaining to political films associated with the Indonesian presidential election of 2024. Confusion metrics, including as accuracy, precision, and recall, are then used to analyse the model's performance.

Corresponding Author:

Dwi Guna Mandhasiya
Department of Mathematics, Faculty of Mathematics and Science, Universitas Indonesia
Depok-10430, Jawa Barat, Indonesia
Email: dwi.guna11@sci.ui.ac.id

1. INTRODUCTION

Approximately 78% of Indonesians have access to the internet, 88.99% of people use the internet for social networking, 66.13% for news or information, and 63.08% for entertainment [1]. With 53.81% of Indonesia's 270.20 million people being members of generation Z (who was born form 1997 until 2012) and millenial generation (who was born from 1981 until 1996), these two age groups are the most active users of social media [2]. Public opinion on social media can represent political communication between patterns of social media users, political motivation, and political participation. The basic assumption of democracy is that people should keep up with the latest news, become interested in politics, and participate in the political process [3]. On social media, we can find a lot of information or big data related to politics and it can be a source of information, especially information related to the 2024 Indonesia presidential election. In this study, we analyze the sentiments of social media users regarding the 2024 Indonesia presidential election.

The study of sentiment analysis examines views, feelings, assessments, conclusions, attitudes, and emotions towards various things, including goods, services, groups, people, occasions, subjects, and their characteristics [4]. Sentiment analysis is an analysis of text data to determine the sentiment or opinion on the text data that reflects the sentiment or opinion of the author about certain entities. Manual sentiment analysis necessitates increasing amounts of time and effort as the volume of data increases. One method that can be used to solve these problems is machine learning [5]. From the perspective of machine learning, sentiment analysis is a problem that requires supervised learning, i.e. classification. Thus, training a machine learning model requires data with sentiment labels [6].

One branch of artificial intelligence (AI) science is machine learning. A computational approach that uses experience to enhance performance or produce precise predictions is known as machine learning [7]. Machine learning uses task-related datasets to teach machines. From the provided dataset, the machine extracts patterns and creates it is own set of rules [8]. There are three distinct categories of machine learning: reinforcement learning, unsupervised learning, and supervised learning [9]. This work employs supervised learning to categorise sentiment into positive and negative categories.

Deep learning has gone beyond AI-based approaches in a variety of text classification tasks, including sentiment analysis. AI in analyzing big data with deep learning in comments can produce sentiment analysis from neural network by natural language processing (NLP) [10]. Altmetric research has also succeeded in measuring attention through the digital world from research results. Allows us to measure the social impact of a research article directly [11]. In developing deep learning, various neural network models have been implemented in NLP tasks such as in [12] NLP task on recurrent attention network memory for aspect sentiment analysis, NLP task on adaptive recursive neural network for target-dependent Twitter sentiment classification [13], NLP task on parameterized convolutional neural network (CNN) for aspect sentiment classification [14], NLP task on interactive attention networks for aspect-level sentiment classification [15], NLP task on long-short term memory (LSTM) for target-dependent sentiment classification [16], NLP task on aspect level sentiment classification with deep memory network [17], NLP task on attention-based LSTM for aspect-level sentiment classification [18], and NLP task on LSTM [19].

The initial stage of conducting sentiment analysis involves converting the textual input into a numerical representation vector. The process of text representation involves addressing the fundamental issue of converting text into numerical form, enabling mathematical calculations to be performed on the text. The initial method for representing text is bag of words, in which text is represented by a vector with dimensions in the form of words. Furthermore, the element value of the vector states the importance of the word in the document, for example, term frequency (TF), which is the frequency of the word appearing in the document, inverse document frequency (IDF), which is the opposite of the frequency of the document containing the word, and term frequency-inverse document frequency (TFIDF). One of the weaknesses of the bag of words method is that it does not store the order in which the word appears in the document and also the semantics of the word. This weakness is overcome by the sequence of word method, in which documents are represented in vector form with the dimensions expressing the order in which words appear in the document. Furthermore, each word is expressed in vector form (embedding), for example fine-tuning the weights in the training process.

Transformer is a language model which is claimed by its inventor as “all you need” for natural language. In general, transformers aim to predict pairs of sentences from a given sentence. There are two popular models developed based on transformer, namely bidirectional encoder representation from transformer (BERT) which are developed from encoders and aim to produce contextual representations and generative pre-trained transformer (GPT) which are based on decoders that aim to predict the next word by given part of a sentence [20].

BERT is a pre-trained language representation model that utilises deep learning techniques to learn the contextual link between words in a text [20]. BERT representatives have shown great advantages over text representation in many NLP tasks, including the classification of sentiments [21]. In BERT sentiment analysis, one fully connected layer is added for sentiment classification called BERT-NN. Neural network in the form of fully-connected is used as a classification layer. Currently, BERT stands as the most advanced transformer model for text representation [22].

In previous studies, the CNN and LSTM methods were popular feature selection methods used for sentiment analysis [22]. To improve the performance of the CNN and LSTM models, several studies have tried to use a hybrid model of the model architecture in sentiment analysis problems. Research from [18] used the CNN-LSTM hybrid model for sentiment analysis in English. The result showed that the CNN-LSTM hybrid model has better accuracy than the CNN model and the LSTM model. Then [6] used the CNN-LSTM and LSTM-CNN hybrid models employing two text representation techniques, BERT and embedding, for the problem of sentiment analysis in Indonesian. The findings indicate that the BERT representation surpasses the embedding representation in sentiment analysis. In this paper, we evaluated if CNN, LSTM, CNN-LSTM, and LSTM-CNN models can increase the performance of BERT for Indonesian sentiment analysis with a case study of YouTube comment data related to the 2024 presidential election.

2. THE PROPOSED MODEL ARCHITECTURE

2.1. BERT-CNN model

Figure 1 illustrates the sequential process that each input text goes through in the BERT-CNN deep learning model. Sentences with a total of p words are fed into the model, and BERT excites the input text into a collection of n tokens. Each token will be subjected to a conversion process that yields a numeric vector of 768 dimensions. The output produced by BERT is a 768-dimensional vector, which functions as a contextualized representation for every character. The vectors are arranged in a matrix of dimensions n×768. The convolutional layer applies a filter to the n×768 matrix derived from BERT, partitioning it and generating a vector output with dimensions $n-s+1$. Therefore, the resulting output of this layer is a vector with a size equal to $n-s+1$ times a. The pooling layer employs max pooling to extract the most salient features from each output vector of the convolution layer. Max pooling selects the highest value among the existing features, considering a pool size of z. The operation in this layer generates a vector of size $n-s+1z$, which can be organised as a matrix of dimensions $a \times n-s+1z$. The vectors obtained from the pooling layer are subsequently transformed into k distinct categorization categories. The study employed a value of k=2, encompassing both positive and negative emotion categories. The softmax activation function is utilized in this layer to generate output values ranging from 0 to 1. These values indicate the probability of an input sentence being categorized into its respective sentiment category.

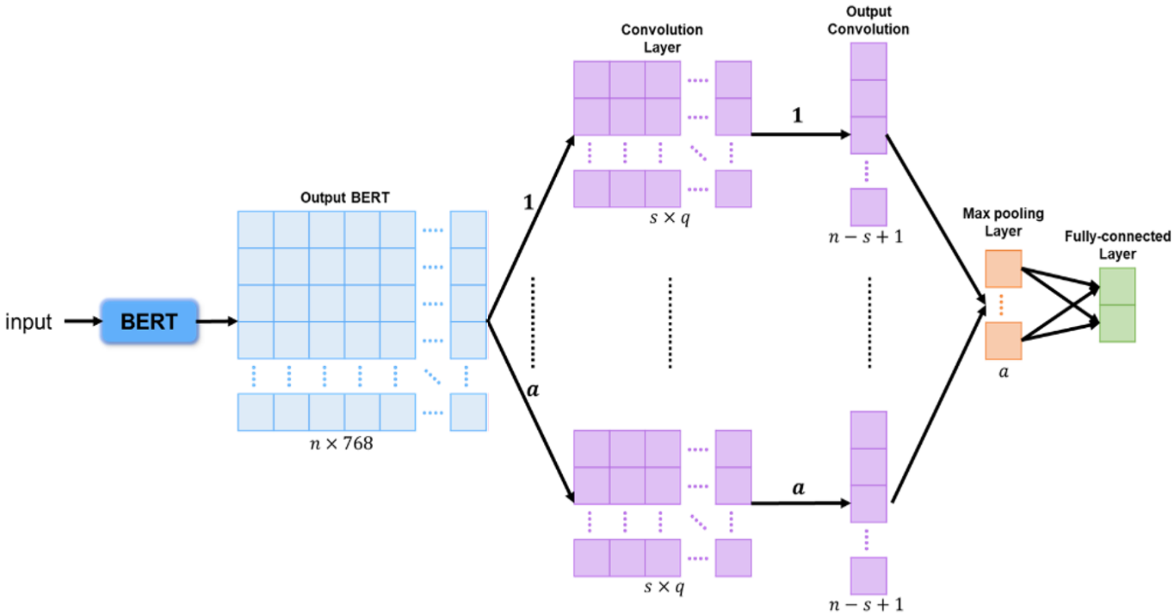


Figure 1. BERT-CNN model architecture

2.2. BERT-LSTM model

Figure 2 depicts the sequence of steps that each input sentence undergoes in the BERT-LSTM deep learning model. The model accepts input in the form of sentences composed of a total of p words, and then BERT converts the input sentence into a collection of n tokens. The LSTM sequentially analyzes the contextualized embedding vector of each token, beginning with the first token and continuing until the n token. The LSTM in this model produces a matrix with dimensions $n \times q$ as its output. This matrix is obtained by extracting the LSTM output at each discrete time step. The LSTM output, with dimensions $n \times q$, is then converted into k separate classification categories. Nevertheless, k was assigned a value of 2, which corresponds to the classification of positive and negative emotions. The softmax activation function is utilized in this layer, generating output values that span from 0 to 1. These numbers indicate the probability that an input sentence corresponds to a certain sentiment category.

2.3. Hybrid deep learning

Although a single deep learning model has relatively good performance in solving problems in a particular task, each deep learning model has its advantages and disadvantages. Therefore, an approach that combines two or more models is introduced as a means to combine the advantages of these models to fill some of the deficiencies of each model. This approach is called hybrid deep learning. This model retains all the advantages of each of its constituent models. While CNN can learn important features of words or phrases in the text, LSTM processes words in sentences sequentially and can learn long-term dependencies from the text. These two models are combined in a different order in the hope of getting better results.

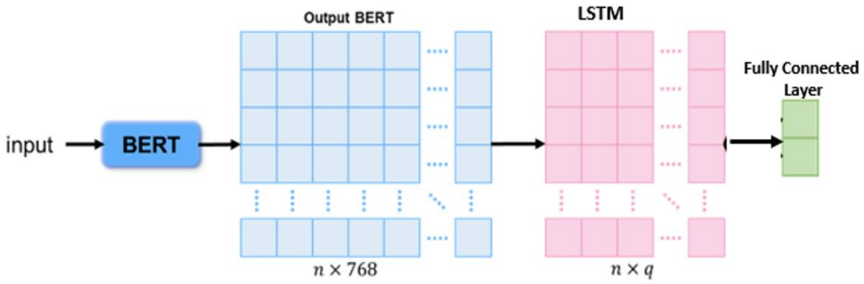


Figure 2. BERT-LSTM model architecture

2.3.1. BERT-CNN-LSTM model

The CNN-LSTM deep learning model depicts the sequential steps that each input sentence undergoes, as seen in Figure 3. The output of the BERT model, which represents the text, is first processed using CNN to discover the important features in the data. The output of the CNN is then inputted into the LSTM model, which develops a unique representation by carefully evaluating the sequence of the data. Furthermore, a fully connected layer is employed to convert the output of the LSTM into two separate classification categories.

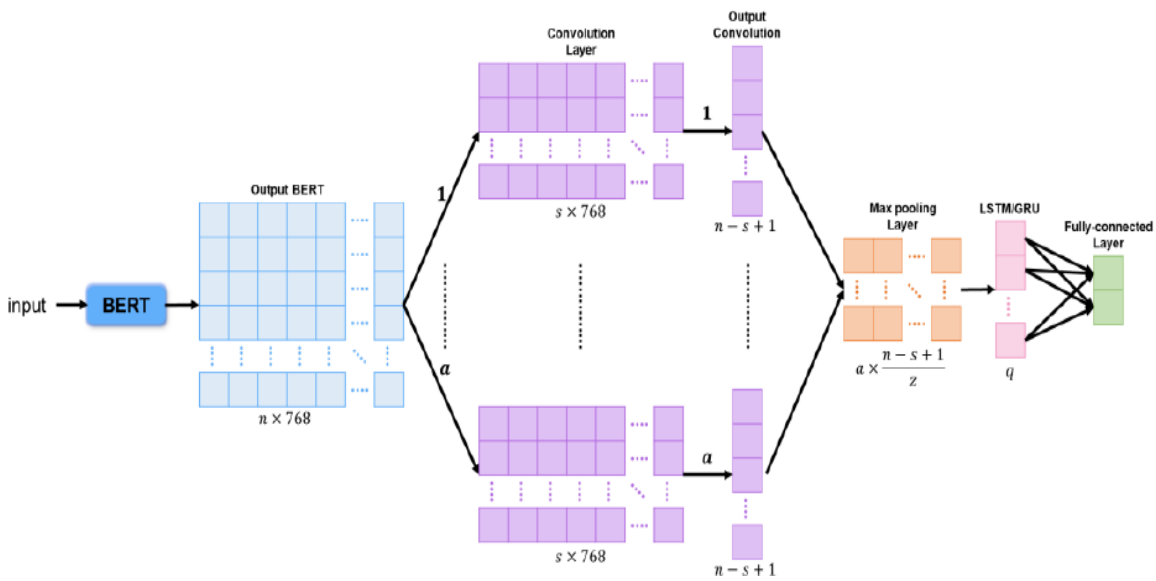


Figure 3. BERT-CNN-LSTM model architecture

2.3.2. BERT-LSTM-CNN

In the LSTM-CNN deep learning model, the process that each input sentence goes through is illustrated in Figure 4. Each input sentence is subjected to a technique that roughly mimics the CNN-LSTM model. The difference resides in the fact that LSTM first takes the BERT output to obtain the feature representation, while also taking into account the sequential arrangement of features in the data. The LSTM output is then inputted into a CNN, which aims to detect essential features within the data.

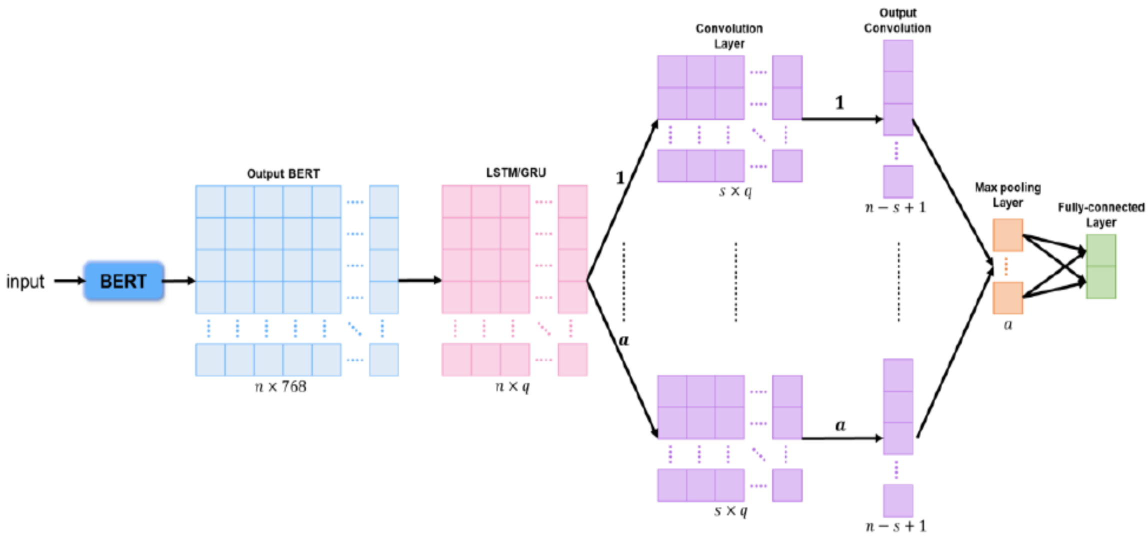


Figure 4. BERT-LSTM-CNN model architecture

3. METHOD

This chapter describes the text representation method used and its application in deep learning models for sentiment analysis problems in Indonesian. This study evaluated the performance of the deep learning models with BERT as a text representation method. Prior to doing sentiment analysis using deep learning models, the textual data input is transformed into a numerical representation using BERT. An assessment of the deep learning models, specifically CNN, LSTM, CNN-LSTM, and LSTM-CNN, is conducted after analyzing the textual representation. With a case study of social media user sentiments regarding the 2024 Indonesia presidential election.

3.1. Bidirectional encoder representation from transformer

The BERT model architecture is a multi-layer bidirectional transformer encoder that uses transformers, but only up to the encoder process. The transformer proposed by [23] is a device that analyzes the contextual relationships between words in text using an encoder-decoder architecture, as illustrated on the left and right sides in Figure 5. The transformer architecture utilizes multi-head attention at each encoder and decoder layer. Multi-head attention applies a specific mechanism called self-attention. This mechanism connects each word to all other words in the sentence which will help the model to learn the contextual representation of the text. The multi-head attention mechanism applies some self-attention to obtain a more accurate representation.

Task-specific BERT designs offer the capability to represent a pair of sentences as either a sequential sequence of tokens or a single phrase. To generate a representation for a particular token, the appropriate tokens, segments, and embeddings are combined. For the classification task, a unique token is assigned to the initial word in the sequence, and the completely connected layer is connected to the last encoder layer. The classification of sentences or pairs of sentences is achieved by employing a softmax layer [24]. BERT was developed by utilizing deep learning techniques to enhance the capabilities of previous methods such as embeddings from language model (ELMo) and OpenAI GPT. BERT utilizes a language paradigm that diverges from the traditional left-to-right or right-to-left methodology. BERT aims to analyze the contextual information before and after each layer and provide a representation that can effectively collect information from both ways. The comparison between the BERT architecture and Open AI GPT and ELMo is depicted in Figure 6.

There are two approaches to implementing the BERT model, namely the feature-based approach and the fine-tuning approach [21]. In this study, the BERT model will be used with a feature-based approach. This approach studies and represents text with pre-trained models, namely models that have been pre-trained on large datasets. Typically, the BERT model comes in two sizes: $\text{BERT}_{\text{BASE}}$ and $\text{BERT}_{\text{LARGE}}$ [20]. The $\text{BERT}_{\text{BASE}}$ model consists of 12 encoder layers (transformer blocks), with a hidden size of 768, 12 self-attention heads, and a total of 110 million parameters. On the other hand, the $\text{BERT}_{\text{LARGE}}$ model has 24 encoder layers, a hidden size of 1,024, 16 self-attention heads, and a total of 340 million parameters. The study employs the $\text{BERT}_{\text{BASE}}$ model, as depicted in Figure 7.

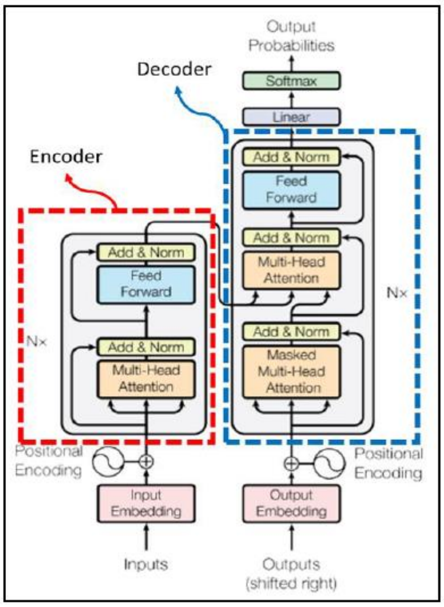


Figure 5. Transformer architecture

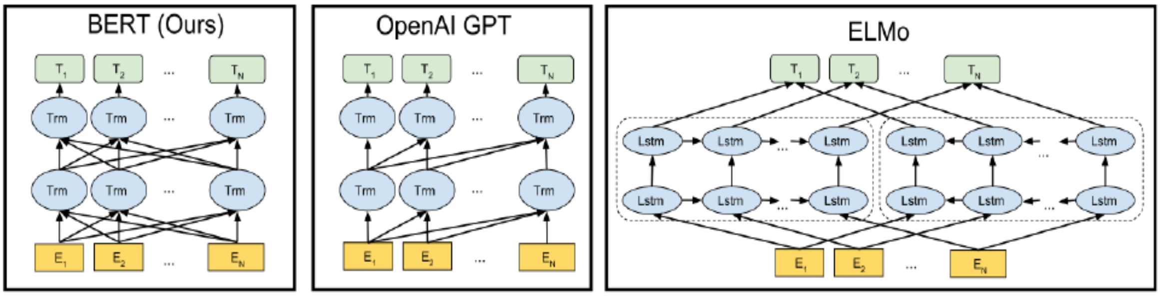


Figure 6. Comparison between BERT, OpenAI GPT, and ELMo architecture

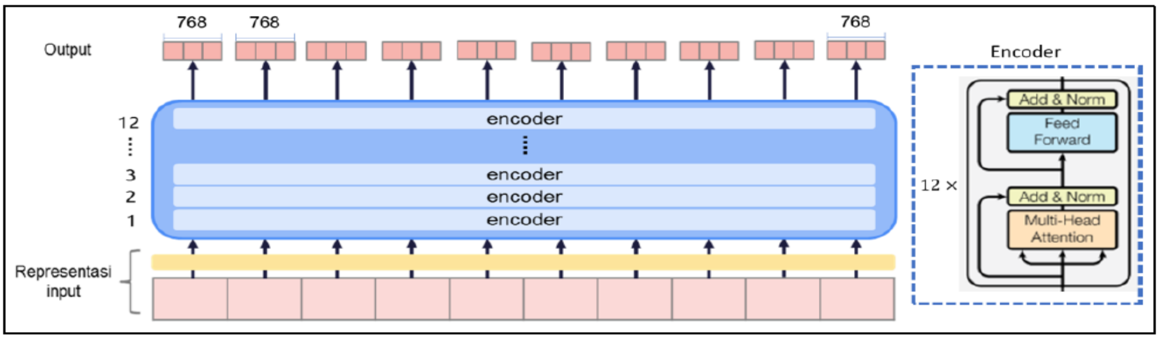


Figure 7. The BERTBASE architecture

3.2. Convolutional neural network

The CNN design has several layers, including the convolutional layer, pooling layer, and fully connected layer [25], as depicted in Figure 8. Suppose given a sentence consisting of p, first of all, the sentence will be added (padding) with some special words to set the length of the sentence to n words, where $n \geq p$. Furthermore, each word in the sentence is represented by a word representation vector of size k. Each vector representation of the words $x_1, x_2, \ldots, x_n$ is combined into a matrix $X_{1:n}$ using (1).

$$X_{1:n} = \boldsymbol{x_1} \oplus \boldsymbol{x_2} \oplus \dots \oplus \boldsymbol{x_n} \quad (1)$$

Therefore, the CNN input is a matrix $X_{1:n}$ with size $n \times k$.

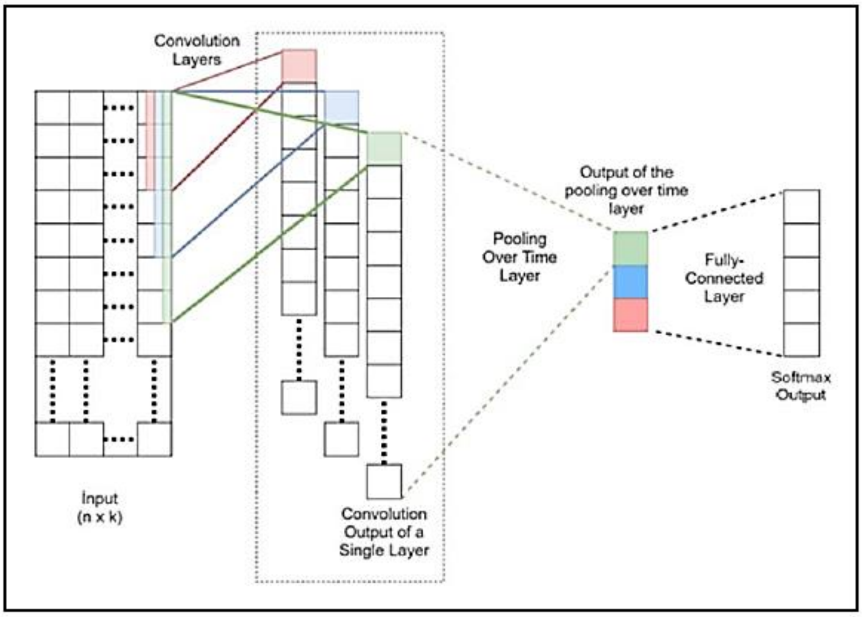


Figure 8. CNN architecture for text classification

3.2. Long-short term memory

The LSTM architecture, like the recurrent neural network (RNN), consists of a series of repeated modules. However, the LSTM has a more intricate structure [26]. The LSTM model receives an input vector xt with a certain dimension and the output from the previous hidden state ($ht - 1$) with a dimension of p. The LSTM possesses the capacity to determine the optimal moment for replacing, updating, or discarding the information stored in individual neurons inside the cell state Ct. The LSTM possesses the ability to effectively store and maintain information for long durations. The cell state in the LSTM experiences negligible alterations as it passes between cells via a sequence of simple linear operations. The LSTM architecture shows in Figure 9 and every step shows in Figures 9(a) to 9(d).

The LSTM's ability to alter the information stored in the cell state is regulated by a gate mechanism. The LSTM gate consists of three components: an input gate, a forget gate, and an output gate [27]. During the initial time step of the LSTM, the main objective is to determine the precise information that will be discarded from the previous cell state ($Ct - 1$) and the information that will be passed on to the cell state C_t. The result of this decision is determined by the sigmoid function of the forget gate. The forget gate functions at time t and receives an input vector xt with a dimension of d, as well as a prior vector $ht - 1$ with a dimension of p. The output is a scalar value that ranges from 0 to 1. If the forget gate is set to 0, the input from the previous time step, $Ct - 1$, is excluded from the calculation of Ct. On the other hand, if the forget gate is assigned a value of 1, all the information from the previous time step ($Ct - 1$) will be completely integrated into the calculations for the current time step (Ct). The calculation of the forget gate at time step (f_t) is determined by (2).

$$f_t = \sigma(w_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (2)$$

Where W_{fx} is dimension weight matrix $p{\times}d$, W_{fh} is a dimensional weight matrix $p{\times}p$, and b_f is a dimensional bias vector p. The LSTM input gate calculated using (3).

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (3)$$

Next, the value of the candidate vector at the time step-t is calculated using (4).

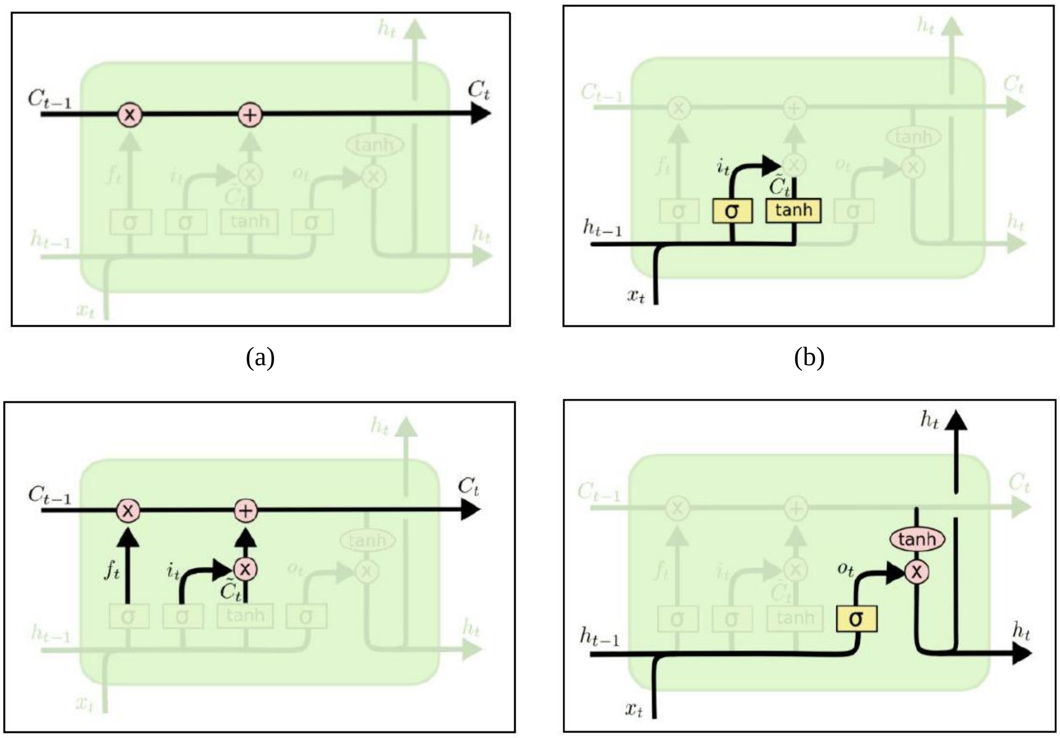
$$\tilde{C}_t = \tanh\,(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \tag{4}$$

The LSTM cell process update/forget gate calculated using (5).

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{5}$$

The LSTM Output Gate calculated using (6).

$$h_t = o_t * \tanh(C_t) \tag{6}$$



(a) (b) (c) (d)

Figure 9. LSTM architercture in of (a) cell state, (b) input gate, (c) cell process update/forget gate, and (d) output gate

4. RESULTS AND DISCUSSION

The simulation phase begins by collecting data. This study employed data extracted from video comments on YouTube on the 2024 presidential election. Furthermore, the data pre-processing phase was carried out, which involved performing data purification and applying one-hot encoding. Later on, BERT is used to represent the data. The resulting representation results are then used as input for doing sentiment analysis utilizing the CNN, LSTM, CNN-LSTM, and LSTM-CNN models. Afterwards, the model's performance is evaluated by utilizing a confusion matrix that includes measures of accuracy, precision, and recall. The performance assessment of the BERT-based deep learning model is carried out to ascertain the most effective model performance. Figure 10 depicts the consecutive stages of simulation.

4.1. Data gathering

This study collected data from video comments on YouTube pertaining to the 2024 presidential election. The acquired comment data comprises 4,844 comments that were mechanically extracted from several sources like *Dialog Kebangsaan Indonesia Bangkit TVOne*, *Catatan Demokrasi*, *Satu Meja the Forum, ILC, ROSI, AIMAN, and Asumsi*. The data is categorized based on its attributes, with a label of 0 representing negative sentiment and a label of 1 representing positive sentiment. The comment data consisted of 2,861 positive feelings and 1,623 negative sentiments, making it the most prominent categorization criteria among all the data. This data is utilized as training data to train the model.

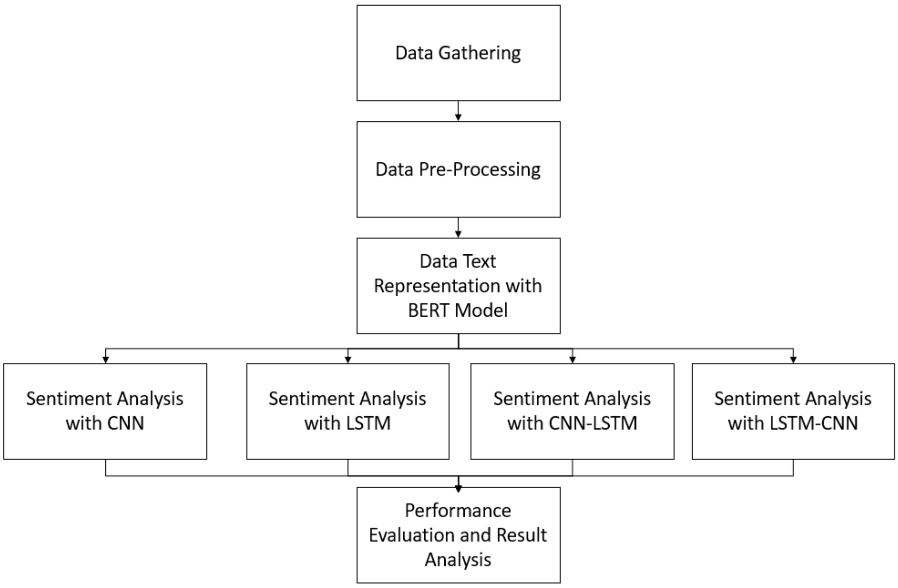


Figure 10. General stages of simulation

4.2. Data pre-processing

Once the data has been gathered, the subsequent step involves performing data pre-processing. The pre-processing procedure in this study comprises two steps, specifically data cleansing and one-hot encoding. The objective of the data cleaning process is to enhance the model's comprehension of data representation and optimize model performance. The data labels indicating negative and positive sentiments are categorical variables. Consequently, a one hot encoding procedure is conducted to transform the labels into numerical variables.

4.3. Data text representation with BERT model

Before text data is processed by BERT, the data is first adjusted to the input representation that can be accepted by the BERT model. Tokenization is the initial step in the input representation process, where the words in the input sentence are transformed into a set of tokens. BERT employs the WorldPiece model for performing the tokenization procedure. The input sentence is augmented with specific [CLS] and [SEP] tokens, which are placed at the start and end of the sentence.

Following the tokenization process, the subsequent step involves ensuring that the input sentence has equal length by applying padding. The process of padding involves the addition of a unique token, denoted as [PAD], until the sentence length reaches 128 tokens. The subsequent step involves numericization, which refers to the process of transforming tokens into integers in order to enable the BERT model to read the input. This level involves employing the vocabulary of the WordPiece model, which comprises 30,522 pairs of tokens and their corresponding unique integers. The numericalization stage produces a tokenID that BERT use to convert each input token into a numeric representation vector.

4.4. Sentiment analysis simulation

The data in this study is categorised into two distinct parts: training data and testing data. The dataset is partitioned into 80% training data and 20% testing data [6]. The training data is utilised to instruct the model, which is subsequently assessed through data testing. The data training and data testing, which have been subjected to text data representation using BERT, are subsequently utilised as input for the CNN, LSTM, CNN-LSTM, and LSTM-CNN models.

The model was constructed with 50 epochs, indicating that the training data was utilised 50 times during the learning process. The employed optimisation strategy is Adam, utilising a batch size of 32 and a learning rate of 1×10-3. Subsequently, the utilised loss function is categorical cross-entropy. The activation function utilised in CNN is rectified linear unit (ReLU), but in LSTM, the sigmoid activation function is employed in each gate and tanh in the recurrent output. Next, the fully-connected layer employs the softmax activation function. The hyperparameter values listed in Table 1 will be optimised in order to achieve the most effective combination of hyperparameters for each model. With a total of 432 hyperparameter combinations, testing each of these combinations on every model would be a time-consuming process.

Hence, Bayesian optimisation methods are employed to choose hyperparameter combinations that yield optimal performance on both the training and testing datasets. Bayesian optimisation use a probability model of the loss function to choose the most favourable hyperparameters, which are subsequently assessed using the real loss function.

4.5. Model performance evaluation

After building the model using training data, the next step is to evaluate the model using data testing. From the dataset and the five sentiment analysis models used, there are five metrics to consider. Table 2 shows the mean and standard deviation of the three trials evaluation of sentiment analysis using BERT as text representation. From the 4,844 YouTube comments dataset regarding the 2024 Indonesia presidential election, performance evaluation BERT-NN, BERT-CNN, BERT-LSTM, BERT-CNN-LSTM, and BERT-LSTM-CNN carried out using confusion matrix with accuracy, precision, and recall.

Table 1. List of hyperparameters and the candidate's values

Layer	Hyperparameter	Value
CNN	Number of filters	200; 250; 300
	Region size	5; 4; 5
	L2 CNN	0.001; 0.01
LSTM	Unit	100; 150; 200
	L2 Kernel	0.001; 0.01
	L2 recurrent	0.001; 0.01
Fully connected	L2 dense	0.001; 0.01

Table 2. The simulation results

Model	Accuracy	Precision	Recall
BERT-NN	0.8599±0.005233	0.8599±0.005233	0.8599±0.005233
BERT-CNN	**0.9896±0.004941**	**0.9881±0.00462**	**0.9896±0.004941**
BERT-LSTM	0.8696±0.000212	0.8693±0.007071	0.8695±0.007071
BERT-CNN-LSTM	0.8684±0.046103	0.8684±0.045891	0.8684±0.046103
BERT-LSTM-CNN	0.8757±0.000751	0.8738±0.000833	0.8742±0.000751

4.6. Discussion

In this study, evaluating the performance of the BERT-based deep learning model for sentiment analysis of social media users regarding the 2024 presidential election shows that the BERT-CNN model is the model with the best performance. BERT-CNN has an average accuracy of 0.9896; precision of 0.9881; recalls of 0.9896. Then the hybrid deep learning model BERT-LSTM-CNN with an average accuracy of 0.8757; precision of 0.8738; recall of 0.8684. BERT-LSTM model with an average accuracy of 0.8696; precision of 0.8693; recall of 0.8695. Then the BERT-CNN-LSTM model with an average accuracy, precision, and recall of 0.8684. Finally, the BERT-NN model has an average accuracy, precision, and recall of 0.8599.

The deep learning model for sentiment analysis CNN can improve BERT accuracy by 0.1297, precision by 0.1288, and recall by 0.1297. LSTM also increases the performance of BERT but not significantly with accuracy by 0.0097, precision by 0.0094, and recall by 0.0097. Hybrid deep learning model LSTM-CNN can improve BERT accuracy by 0.0158, precision by 0.0139, and recall by 0.0143. For the hybrid deep learning model CNN-LSTM can increase the performance of BERT but not significantly with accuracy, precision, and recall by 0.0085. The order of the performance evaluation results of the BERT-based deep learning model is CNN, LSTM-CNN, LSTM, and CNN-LSTM. Meanwhile, BERT-NN shows the lowest performance.

5. CONCLUSION

The evaluation performance of the BERT-based deep learning model was carried out for analyzing the sentiments of social media users regarding the 2024 presidential election. The text representation method uses BERT with CNN, LSTM, CNN-LSTM, and LSTM-CNN models. With the case study of the 2024 Indonesia presidential election, it can be concluded that the deep learning model improves the performance of the BERT model in conducting sentiment analysis. Especially on the CNN model with an average accuracy of 0.9896, precision of 0.9881, and recall of 0.9896 which has significantly improved BERT model performance.

ACKNOWLEDGEMENTS

Author would like to thank the Kantor Staf Presiden Republik Indonesia for supporting and assisting in providing data and information for the purposes of this research. In most cases, sponsor and financial support acknowledgments.

REFERENCES

[1] Intenet World Stats, "Internet users in to 20 country in June 2019," *Miniwatts Marketing Group*, 2019. http://www.internetworldstats.com/top20.htm (accessed Oct. 25, 2023).

[2] Badan Pusat Statistik, "The results of the population census (SP2020) in September 2020 recorded a population of 270.20 million people, in Indonesian: Hasil sensus penduduk (SP2020) pada September 2020 mencatat jumlah penduduk sebesar 270,20 juta jiwa," *Badan Pusat Statistik*, 2020. https://www.bps.go.id/pressrelease/2021/01/21/1854/hasil-sensus-penduduk-2020.htm (accessed Oct. 25, 2023).

[3] K. Dommett and P. J. Verovšek, "Promoting democracy in the digital public sphere: applying theoretical ideals to online political communication," *Javnost*, vol. 28, no. 4, pp. 358–374, Sep. 2021, doi: 10.1080/13183222.2021.1955206.

[4] J. Liu, W. C. Chang, Y. Wu, and Y. Yang, "Deep learning for extreme multi-label text classification," in *SIGIR 2017-Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Aug. 2017, pp. 115–124, doi: 10.1145/3077136.3080834.

[5] A. Nandi and P. Sharma, "Comparative study of sentiment analysis techniques," in *Interdisciplinary Research in Technology and Management*, CRC Press, 2021, pp. 456–460.

[6] T. Gowandi, H. Murfi, and S. Nurrohmah, "Performance analysis of hybrid architectures of deep learning for Indonesian sentiment analysis," in *Communications in Computer and Information Science*, vol. 1489 CCIS, Springer Singapore, 2021, pp. 18–27.

[7] X. Liu, P. He, W. Chen, and J. Gao, "Multi-task deep neural networks for natural language understanding," in *ACL 2019-57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 2020, pp. 4487–4496, doi: 10.18653/v1/p19-1441.

[8] F. Chollet, *Deep learning with Python*. New York, 2018.

[9] K. P. Murphy, *Machine learning: A probabilistic perspective*. MIT Press, 2012.

[10] I. V. Tetko, P. Karpov, R. Van Deursen, and G. Godin, "State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis," *Nature Communications*, vol. 11, no. 1, Nov. 2020, doi: 10.1038/s41467-020-19266-y.

[11] C. García-Villar, "A critical review on altmetrics: can we measure the social impact factor?," *Insights into Imaging*, vol. 12, no. 1, Jul. 2021, doi: 10.1186/s13244-021-01033-2.

[12] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 452–461, doi: 10.18653/v1/d17-1047.

[13] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent Twitter sentiment classification," in *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014-Proceedings of the Conference*, 2014, vol. 2, pp. 49–54, doi: 10.3115/v1/p14-2009.

[14] B. Huang and K. Carley, "Parameterized convolutional neural networks for aspect level sentiment classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 1091–1096, doi: 10.18653/v1/d18-1136.

[15] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," in *IJCAI International Joint Conference on Artificial Intelligence*, Aug. 2017, pp. 4068–4074, doi: 10.24963/ijcai.2017/568.

[16] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," *arXiv preprint arXiv:1512.01100*, Dec. 2015, [Online]. Available: http://arxiv.org/abs/1512.01100.

[17] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 214–224, doi: 10.18653/v1/d16-1021.

[18] Y. Wang, M. Huang, L. Zhao, and X. Zhu, "Attention-based LSTM for aspect-level sentiment classification," in *EMNLP 2016-Conference on Empirical Methods in Natural Language Processing, Proceedings*, 2016, pp. 606–615, doi: 10.18653/v1/d16-1058.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[20] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019, vol. 1, pp. 4171–4186, doi: 10.18653/v1/N19-1423.

[21] Z. Gao, A. Feng, X. Song, and X. Wu, "Target-dependent sentiment classification with BERT," *IEEE Access*, vol. 7, pp. 154290–154299, 2019, doi: 10.1109/ACCESS.2019.2946594.

[22] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: a survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, 2018, doi: 10.1002/widm.1253.

[23] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, Jun. 2017.

[24] B. Wilie *et al.*, "IndoNLU: benchmark and resources for evaluating Indonesian natural language understanding," in *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 2020, pp. 843–857.

[25] C. C. Aggarwal, "Training deep neural networks," in *Neural Networks and Deep Learning*, Springer International Publishing, 2018, pp. 105–167.

[26] "Understanding LSTM networks," *colah's blog*, 2015. http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed Oct. 25, 2023).

[27] P. Anki, A. Bustamam, H. S. Al-Ash, and D. Sarwinda, "High accuracy conversational AI chatbot using deep recurrent neural networks based on BiLSTM model," in *2020 3rd International Conference on Information and Communications Technology, ICOIACT 2020*, Nov. 2020, pp. 382–387, doi: 10.1109/ICOIACT50329.2020.9332074.

BIOGRAPHIES OF AUTHORS

Dwi Guna Mandhasiya is Master Student at Faculty of Mathematics and Natural Science, Universitas Indonesia. He Holds a bachelor degree in Industrial Engineering, Faculty of Engineering, Universitas Indonesia. He does research in machine learning with applications in sentiment analysis. His research areas are management information and decision support. Currently, he is Staff of President at Republik Indonesia President Staff Office. He can be contacted at email: dwi.guna11@sci.ui.ac.id.

Hendri Murfi is an Associate Professor in the Department of Mathematics at Faculty of Mathematics and Natural Science, the Universitas Indonesia where he has been a faculty member since 1998. He has been the Head of Master Program in Mathematics since 2022. He completed his doctor degree at TU Berlin and his master and undergraduate studies at Universitas Indonesia. Hendri currently works at machine learning group and does research in machine learning with applications in topic detection, sentiment analysis, recommender system, and insurance. He can be connected at email: hendri@sci.ui.ac.id.

Alhadi Bustamam received the B.Sc. degree (Hons.) in computational mathematics, in 1996, the master's degree in computer science from Universitas Indonesia, in 2002, and the Ph.D. degree in bioinformatics from The Institute for Molecular Bioscience (IMB), University of Queensland, Australia, in 2011. Currently, he is a full Professor and the Head of the Bioinformatics and Advanced Computing Laboratory (BACL), Department of Mathematics. He is also the Head of Department of Mathematics, and the Chairperson of Data Science Centre (DSC) (https://dsc.ui.ac.id), Universitas Indonesia. His research interests include high-performance computing approaches to computational mathematics, computational biology, bioinformatics, computer science, data science, intelligent healthcare, and artificial intelligence (AI). He can be connected at email: alhadi@sci.ui.ac.id.