Measuring the accuracy of LSTM and BiLSTM models in the application of artificial intelligence by applying chatbot programme

Prasnurzaki Anki[1], Alhadi Bustamam[2]
[1,2]Department of Mathematics, Universitas Indonesia, Indonesia
[2]Data Science Centre, Universitas Indonesia, Indonesia

Article Info

ABSTRACT

Python programme contains a question and answer system that derived from data sets that have used and implemented the chatbot in this modern era. where the data collected is in the form of corpuses containing extensive metadata-rich fictional conversations derived from extracted film scripts, commonly called cornell movie dialogue corpus. The various models have been used chatbots in python programmes, and LSTM and BiLSTM models were specifically used in this study. Where the form of accuracy will be reported as a result of the implementation of LSTM and BiLSTM models in the chatbot programme. The programme performance will be influenced by the data from the model selection, because the level of accuracy is determined by the target programme being taken. So this is the main factor that determines which model to choose. Based on considerations required for choosing the programme model, in the end the LSTM and the BiLSTM models are chosen and will be applied to the programme. Based on the LSTM and BiLSTM chatbot programmes that have been tested, it can be concluded that the best parameters come from a pair of BiLSTM chatbots using the BiLTSM model with an average accuracy value of 0.995217.

Corresponding Author:

Alhadi Bustamam
Gedung D, Kampus Baru FMIFA, Universitas Indonesia
Depok, Jawa Barat 16424, Indonesia
Email: alhadi@sci.ui.ac.id

1. INTRODUCTION

Chatbots are automated systems created to help users by answering their questions. For businesses, chatbots can provide a better way to connect with their customers and increase customer satisfaction levels. Customers get a better, more convenient way to get answers to their questions without waiting on the phone or sending frequent emails [1]. Artificial intelligence (AI) has made an impact in everyday activities by designing and providing evaluation of sophisticated applications and devices, which can perform various functions. Chatbot is an artificial intelligence programme, which is based on the development of AI, it is hoped that the chatbot's ability to imitate human agents in conversation. Chatbots have become so common in their presence that they can reduce service costs and can handle multiple customers simultaneously [2]. Hopefully, future chatbots can improve business sector performance by increasing customer satisfaction levels by saving time. They will also save customer service employees time; customers can use chatbots to get information that previously required humans to answer questions manually.

A fetch model contains several forms based on matches derived from user input and the chatbot can generate answers based on the forms that the user has filled in. Here knowledge used in chatbots is a form of

human hand code. Chatbot knowledge construction is time-consuming and difficult. Therefore, it is very important to have an automated knowledge extraction mechanism to build various forms of chatbots [3]. The use of models that can improve chatbot performance in answering questions automatically can be considered to compare which two models will influence chatbot performance and determine which model has a better fit.

A system that can receive feedback and respond from users and can keep the conversation going, is called a chatbot. The encoder-decoder architecture is used in building parts of the chatbot [4]. A chatbot is a simple robot that contains a programme to answer questions from users. After that, the answer data will be generated from the questions asked by the user. The semantic question-answering system has developed in which words that are uncertain are the form of the question [5]. Application of question and answer system in the form of a chatbot is expected to answer these challenges.

Artificial intelligence (AI) which is the latest technological advancement is very helpful in the development of new virtual assistants to be efficient (online chatbots). Meanwhile, the study also analyzes how existing technological advances made on new chatbots have an impact on future customer support. Going forward, technological innovations in AI allow chatbots to perform increasingly complex tasks [6]. NLP (natural language processing) is a mechanism that can be used to support computer machines by simulating human abilities that function to understand language [7]. Natural language processing is another area where the stance of deep learning can have a huge impact on experimentation that could occur over the next few years [8]. In NLP models, LSTM considers the order dependence between word sequences that the test will perform on the programme to capture dependencies in both the long and short-range forms. BiLSTM can perform both directional scans, allowing simultaneous access to both contexts in forward and backward directions. Therefore, BiLSTM can solve sequence model tasks with better performance than LSTM [9]. Based on the study of these references, this journal will determine whether the BiLSTM model will perform better than the LSTM model in use in NLP.

The discussion conducted on several chatbot backgrounds indicates that consumers' problems, in general, can be presented through several recorded questions that relate to various constraints, such as data storage and limited customer service hours. In order to provide answers given by consumers, a programme is needed to optimize the results of these services. In connection with this, the modeling theory will be discussed in this journal. It is used as the basis when chatbots are deployed in question and answer systems that use Python programmes with LSTM and BiLSTM as models. Then, it is expected that from this research will be seen a comparison between the sentence response generated by the chatbot with the LSTM model and the BiLSTM model with the sentence response in the data set. The solution methodology that will be used in this research is measuringthe accuracy of LSTM and BiLSTM models by applying chatbot programme. In more detail, we will start from understanding the background of the importance of the role of the chatbot in the question and answer system, determining the steps for making a chatbot, applying various models, methods, applying data into the programme, write results and discussions, to make conclusions, that have been described in more detail from the session 1 to 6. The major contribution of this paper is to determine the most effective model that can be applied to the chatbot programme based on the comparison of the accuracy results of the two models.

2.   RESEARCH METHOD

As the various forms of chatbots increasingly integrate the design of AI mechanisms (such as game theory, data mining and optimisation techniques), they comply with these networks' rules and dynamics. This form can be characterised by real multi-actor-based conversations that require technical resources, specialisedknowledge and communication skills to maintain online interactions [10]. Summarized by the acronym AI, this is a science that focuses on handling the production of human knowledge, and can offer to the machine the ability to imitate human reasoning and intelligence [11]. AI technology can provide improvements to conversations and collaborate between humans and machines. This technology can be used to create better interactions between humans and machines [12].

The LSTM model, the BiLSTM model and several pairs of parameters, it is also the greedy method can be used in building a chatbot programme by using a set of sentences derived from the data set. The chatbot programme is run based on input in the form of commands from the user, where the results of the programme are a collection of sentences containing information that matches user input based on the relevance of questions and answers.

2.1. Steps in making a chatbot

Identifying the identity of the data, inputting the data about answer and question, using the programme of chatbot and then evaluating the output are the 4 steps that must be done when creating a chatbot. First, the identity of the data, cornell movie dialogue corpus is used as data, it contains a collection

of data in the form of a corpus in which it includes a vast collection of fictional conversations rich in meta data extracted from the film script [13]. The data used is from 2018 in the form of text data. Second, in the question and answer system data input contains sentences that come from the film dialogue, then the data functions as user input that is entered into the programme, then after that it is executed into input from the programme user. The third is the chatbot programme development. Fourthly, the following considerations need to be taken to prepare thechatbot programme. A sequence-to-sequence translation will come in several optionss. Selecting the LSTM model will generate the most accurate chatbot response in the end. The final step in the output evaluationis to determine whether the model can provide accurate results or not.

2.2. LSTM model

The LSTM model network is known as a model that has had influence in the past and shows the ability to learn from sequential data [14]. The implementation of encoder in this model will be contained in the last hidden statement of the LSTM [4].

$$f_{m+1} = \sigma\big(\theta^{(h\to f)}h_m + \theta^{(x\to f)}x_{m+1} + b_f\big) \, forget \, gate \tag{1}$$

$$i_{m+1} = \sigma(\theta^{(h\to i)}h_m + \theta^{(x\to i)}x_{m+1} + b_i) \, Input \, gate \tag{2}$$

$$\tilde{c}_{m+1} = \tanh(\theta^{(h\to c)}h_m + \theta^{(w\to c)}x_{m+1}) \, update \, candidate \tag{3}$$

$$c_{m+1} = f_{m+1}\odot c_m + i_{m+1}\odot\tilde{c}_{m+1} \, memory \, cell \, update \tag{4}$$

$$o_{m+1} = \sigma(\theta^{(h\to o)}h_m + \theta^{(x\to o)}x_{m+1} + b_o \, Output \, gate \tag{5}$$

$$h_{m+1} = o_{m+1}\odot\tanh(c_{m+1}) \, Output \tag{6}$$

The operator Θ is included in the elemental product. The result of adding hidden states that can be implemented in h_m with c_m. as a representation of memory cells is the definition of the LSTM model. The gate form is the value of the memory cells present at each mth time which consists of the sum of the two qualities: c_{m-1} is the value of the previous memory cell, while c_m is the value of the memory cell after it has changed, after being calculated from the previous input in the current x_m form and the previous hidden state in h_{m-1} form. Furthermore, h_m is calculated from the cell memory, where the non-linear function path during the update is not passed by the memory cell, so information can be worked on over a remote networ [4]. The preceding hidden state will be determined by each gate controlled by a weighted vector (e.g. $\theta^{(h\to f)}$) and input current (e.g. $\theta^{(x\to f)}$), plus a vector offset (e.g. b_f). The LSTM status after reading tokens is represented (h_m, c_m) in an operation that can informally be summarized as $(h_m, c_m) = \text{LSTM } \llbracket(x_m, (h_{m-1}, c_{m-1}))$. And the results obtained that LSTM can outperform standard artificial neural networks in a variety of problems displayed in square-shaped gates with dotted edges. the next word w_{m+1} is stripped using h_m existing in the LSTM language model. LSTM outperforms standard recurrent neural networks in various problems, such as language modelling problems [4].

One of parallel computerized models is the LSTM model. Parallel computing is a type of computation in which various process calculations can be carried out simultaneously, while the application of parallel computing can run algorithm more quickly in the appearance of the model used in this study [15], [16]. Based on that study, we choose parallel computing models to researched more deeply.

2.3. BiLSTM model

The BiLSTM model is a model that combines the advantages of the BiRNN model and the LSTM model [17]. The BiLSTM model is used to propagate the use of forward and reverse directions. The BiLSTM model is a two-way network used to store future data and past data, which is more effective in the LSTM model [18]. In the feature-based model, traits related to shape knowledge are processed by feature suffixes in the neural network. Embeddings are a technique used to handle the sparse matrix of the bag of words. One application of feature suffixes in neural networks is that they can be inserted by constructing the invisible embeddings of words from their spelling or morphology. One way to do this is to incorporate additional two-way RNN layers, one of which is for each word in the vocabulary. The BiLSTM model is one of many parallel computations. The first step is to encode $w^{(p)}$ dan the $w^{(q)}$ query using two LSTMs. This process is known as Bidirectional LSTM (BiLSTM)

$$h^{(q)} = BiLSTM(w^{(q)}; \theta^{(q)}) \tag{7}$$

$$h^{(p)} = BiLSTM(w^{(p)}; \theta^{(p)}) \tag{8}$$

The questions are represented by the vector *u*, vertically combining final states from left to right, and are represented by matching the ending state vertically from left to right.

$$u = \left[\overrightarrow{h^{(q)}}_{M_{(q)}}; \overleftarrow{h^{(q)}}_0\right] \tag{9}$$

Vector $\left(u^{(q)}\right)^T$ is the result of applying the vector *u* with equation $h_m = g(x_m, h_{m-1}), m = 1, 2, \dots M$, (based on [4]) which has been transposed, W_a is a weight matrix with index α, $h_m^{(p)}$ is the result of implementing the hidden state with (8), and vector $\widetilde{(\alpha_m)}$ is a representation of what is expected and is calculated by,

$$\widetilde{\alpha_m} = \left(u^{(q)}\right)^T W_a h_m^{(p)} \tag{10}$$

$$\alpha = SoftMax(\widetilde{\alpha}) \tag{11}$$

$$o = \sum_{m=1}^{M} \alpha_m h_m^{(p)} \tag{12}$$

In (11), the vector α is the result of the SoftMax function of $\widetilde{\alpha}$. In (12), these vectors can be arranged equal to the corresponding element in $h^{(p)}$, assuming that the candidate's answer (vector *o*) is the span of the original text. The score of each candidate for answer *a* is calculated by the product in,

$$\hat{c} = \underset{c}{arg}\max o.x_c \tag{13}$$

2.4. Greedy method

The next step after choosing a model is to determine the programme method. The greedy method was chosen as the programme method that is used because it is the implementation of the LSTM model, when it is run, the programme can process data in a faster time, so that it can increase the accuracy of the selected model [19]. While the Greedy algorithm is well understood to be able to produce reasonable estimates for a wide class of functions, it can be seen that it performs much better than one might expect from a greedy algorithm [20]. Based on [18], relatively increasing candidate solutions by trying to approach the optimal solution is Greedy 's algorithm. The greedy method is an implementation of the LSTM and BiLSTM models, which is at run time, the time used in processing data is faster and it can increase the accuracy of the selected model.

2.5. Seq2seq model

In Figure 1, the important things that must be understood about the seq2seq model implementation are presented. The seq2seq model functions to generate various responses to user input, so it can implement a question and answer system, where the Python-based Jupyter Notebook Software is chosen as a programme that can view programme input and output [19].
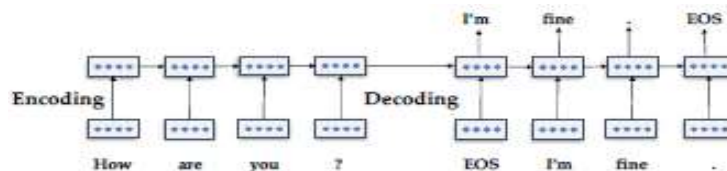

Figure 1. Generation of neural responses in dialogue at the encoder decoder model [19]

3. DATA IMPLEMENTATION IN THE PYTHON PROGRAMME

Systematically, the programme planning is compiled in implementing data from the Jupyter Notebook software in the Python programme as follows: Firstly, choose the appropriate software, they are choosing a programme that can process data well, having performance in software data processing and having the availability of supporting attributes that needed in creating the programme. Secondly, the programme performance is influenced by the selection of model that follows the characteristics of the data, so when choosing a model and supporting attributes, it is necessary to pay attention to it. The outcome of a programme

about its ability to determine a high or low level of accuracy are the major factors in choosing which programme to used. So, seeing that the LSTM Model was chosen as the model applied to this programme, it is because it is in accordance with the model selection requirements. In addition to the main factors, supporting factors are also the determining factors whether the processing of data is in line with the criteria in question. It will make the data verifiable. The input sentence that has been processed by the seq2seq model with other programme structures and models, then the output sentence will be issued, it is as a chatbot programme result. Finally, establish the method of programme evaluation. Several choices of evaluation methods that can be used include: loss, accuracy, val_loss and val_accuracy, so that from these cases [19] text classification evaluation methods such as simple binary detection tasks can be considered its use.

Spam detection for example, it assigns a positive or negative label to spam according to the category an email document, it should be able to define, and also it must be able to identify an item as spam or not. As presented in (14):

$$accuracy = \frac{tp+tn}{tp+fp+tn+fn} \tag{14}$$

Based on the theory [20], there are several things that need to be prepared for the value of losses, they include x for observation and stated is for the function of losses that is expressed by (15):

$$L(\hat{y}, y) = The\ amount\ of\ difference\ from\ the\ true\ y\ value \tag{15}$$

Here is a calculation to determine the closeness of the output classification. Here is a calculation to determine the closeness of the output classification $\hat{y} = \sigma(w.x + b))$ to the actual output (y, which is 0 or 1). The training process is used to calculate the error rate that is derived in calculated model, it is used to observe the validation set loss function. It is as defined in [21] about the meaning of loss value validation. The model that has been selected and trained, it is then evaluated for its effectiveness as a classification task. This can be done by calculating the percentage of samples that have been classified, as follows:

$$classification\ rate = \frac{number\ of\ samples\ correctly\ classified}{total\ number\ of\ samples} \tag{16}$$

The equation (16) shows the misclassification can be calculated and equipped with a classification rate, where is the calculation of the model error rate as follows:

$$error\ rate = \frac{number\ of\ loss\ function\ over\ validations\ set}{total\ number\ of\ validations\ set} \tag{17}$$

The loss function calculated on the predefined model validation set will result in the percentage, as was the case of [22], through the graph presented in Figure 2, we can compare the losses the dataset of training and the validation set of val_loss. From the graph, obtained that the result of a validation loss may be higher or lower than the loss value of the training data set, so this condition is called underfitted or overfitted.

Accuracy is then chosen for the evaluation method of chatbot programme, after comparing each programme's evalution method that accordingly. After getting the suitability value between the form of two sentences, it will be used as the accuracy value that will be used as a chatbot programme evaluation method. The selected model will then be applied to find the differences of words located in sentences during the training period [23].
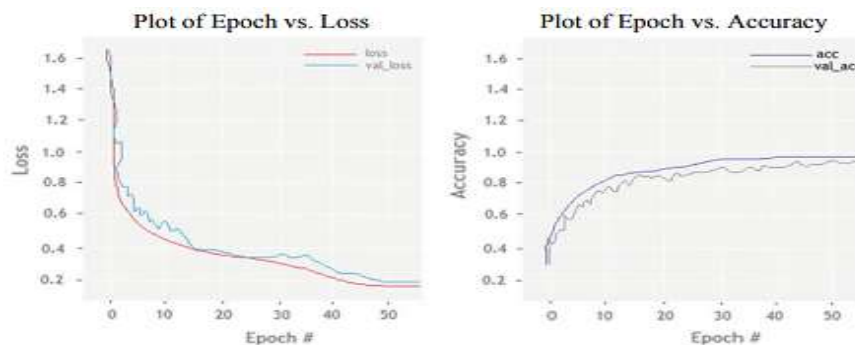

Figure 2. Comparison chart of epoch to loss and epoch to accuracy [22]

4.   APPLYING DATA IMPLEMENTATION IN THE PYTHON PROGRAMME

Accuracy was chosen as the method of programme evaluation in accordance with the selected chatbot programme after comparing the evaluation methods of each programme. In order to study the need to implement a chatbot in the question and answer system, this section will explain. The solution methodology that will be used in this research starts from understanding the background of the importance of the role of the chatbot in the question and answer system, determining the steps for making a chatbot, applying various models, methods, applying data into the programme, selecting evaluation methods that have been described in more detail from the session 1 to 3.

4.1.   Description of the problem

Many questions will certainly be asked by consumers based on certain data in the face of dynamics in this modern era. In order to provide easy access and convenient operational time, chatbots need to be created. So, the question and answer service between humans and chatbot programmes (machines) can run. This is as an implementation of the quastion & answer system data.

4.2.   Programme making

In Figure 3, we show how to steps by steps in programme making to build the chatbot.
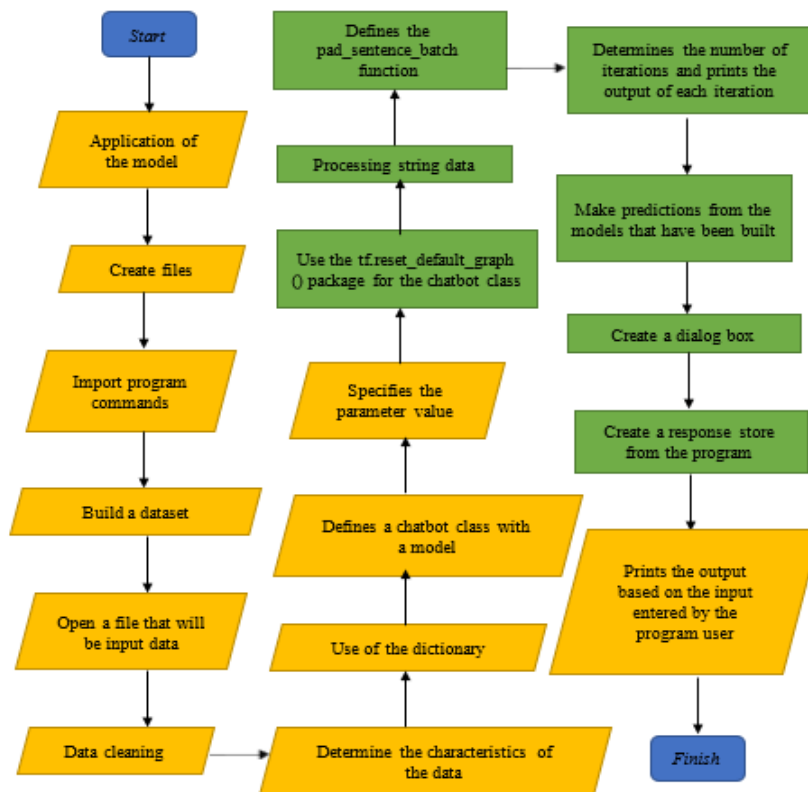

Figure 3. Programme making flowchart

5.   RESULTS AND DISCUSSIONS

From the chatbot programme that has been created, 6 files are generated, 3 files from the implementation of the LSTM model into the chatbot programme and 3 files from the BiLSTM model into the chatbot programme. In Table 1, the LSTM model is used to test the parameter pairs. Secondly, epoch (4 different amounts, 20, 30, 40, 50) will be tested with different numbers of epochs, adjusting for other parameter pairs, to determine which parameter pair is the most accurate in the two selected models [24]. In Table 1, the LSTM model is used to test the parameter pairs (File 1, File 2, and File 3), and BiLSTM model (File 4, File 5, and File 6). The purpose of having differentvalues in the parameter is to produce better accuracy and compare the differences between the number of differences in the same parameter, whether it can have better output results than the test results on the parameter pairs in the experiment.

Table 1. Data tested on LSTM and BiLSTM chatbot

Parameter Pair	File 1, File 4	File 2, File 5	File 3, File 6
size_layer	128	256	512
num_layers	2	2	2
embedded_size	64	128	256
learning_rate	0.001, 0.0015	0.001, 0.0015	0.001, 0.0015
batch_size	8	16	32
epoch	20,30,40,50	20,30,40,50	20,30,40,50

According to the number of parameters in LSTM Chatbot, in 3 different files (with each value of the size_layer parameter different in each file, namely: 128, 256, 512) will be tested for a total of 8 trials each, with 4 parameters with the number of values in the same parameters in each file.

− First, the size_layer (in the form of the number of layer sizes that will be applied to the programme, in general, multiples of 2 are used such as 16, 32 and 64, based on [25]) in the programme to be discussed is an applied size_layerwith the values 128, 256 and 512.
− Second, Num_layers (in the form of n number of layers that will be applied to the programme [26]) in the programme to be discussed is an appliedsize_layer with a value of 2.
− Third, the embedded_size (in the form of the number of sizes of the embedded vector that will be applied to the programme); in general, multiples of 2 are used such as 8, 16, 32 and 64 [26] in the programme to be discussed is an applied size_layer with values 128, 256 and 512.
− Finally, the batch_size (in the form of a batch size that will be applied) in the programme to be discussed will affect programme performance [24]. Therefore, in this programme, the batch_size will be tested with a value of 8, 16 and 32.

Besides having the same parameters, it will also be tested for various parameters with a number of values for different parameters. Based on the programme test conducted in [27], one of the parameters tested in a different number of values is epoch. In this programme, manyparameters will be tested with a number of different values. The Figure 4 presents the parameter pair data in each file, which is the best parameter pair of the 6 files tested and will not be rated if the value is> 1.0, this condition is called an overfit (where if the training level is very good in accuracy, but when testing the results are not good). In Figure 4 will be selected which has the best average accuracy, which is a scale of 0,0 to 1,0. So, from all the tests that have been done, it is obtained Parameter Pair 1 from file 6 is the best parameter pair of BiLSTM chatbot, which is with an average accuracy value of 0,995217. Based on the reference, on results of applying the BiLSTM model in domain-specific Chinese word segmentation, the accuracy rate is 95.7415% or 0,957415 [17]. Based on comparison result in this study and the results of the application of the BiLSTM model carried out by [17] is better the research reference.
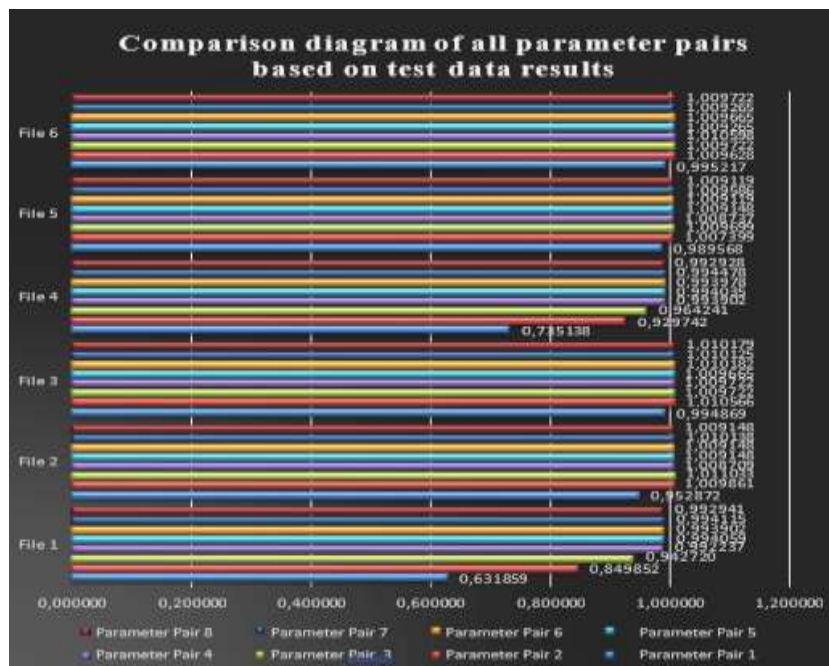

Figure 4. Parameter pair comparison chart based on file data results

Hope in order to achieve more effective result, to get better model performance can combine the models. Other than that, we hope to improve the BiLSTM Chatbot with make modifications to the model architecturethat will result in better accuracy [28], [29]. We will also increase the amount of data analysed, aiming to encourage researchers to propose methods that produce better, more efficient results [30].

6. CONCLUSION

After applying the BiLSTM model to the chatbot, we were able to deduce from all the test results of the programme that had been conducted with a variety of different parameter pairs, then it is obtained the result, if the Parameter Pair 1 (size_layer 512, num_layers 2, embedded_size 256, learning_rate 0.001, batch_size 32, epoch 20) from File 6 is the best parameter pair of BiLSTM Chatbot with an average accuracy value of 0.995217. For future work, the researcher should improve latest model, trying to increase the number of proportions in the data to be studied, so as to produce better research results.

REFERENCES

[1]   M. Nuruzzaman and O. K. Hussain, "IntelliBot A Dialogue-based chatbot for the insurance industry," *Knowledge-Based Systems*, vol. 196, p. 105810, 2020, doi: 10.1016/j.knosys.2020.105810.
[2]   E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Application*, vol. 2, p. 100006, 2020, doi: 10.1016/j.mlwa.2020.100006.
[3]   S. Arsovski, *et al.*, "Automatic knowledge extraction of any Chatbot from conversation," *Expert Systems With Applications*, vol. 137, pp. 343-348, 2019, doi: 10.1016/j.eswa.2019.07.014.
[4]   J. Eisenstein, "Natural Language Processing," *MIT press*, pp. 137-138, 2018.
[5]   A. C. O. Reddy and K. Madhavi, "Hierarchy based firefly optimized k-means clustering for complex question answering," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 17, no. 1, p. 265, 2020, doi: 10.11591/ijeecs.v17.i1.pp264-272.
[6]   E. Pantano and G. Pizzi, "Forecasting artificial intelligence on online customer assistance: Evidence from chatbot patents analysis," *Journal of Retailing and Consumer Services*, vol. 55, p.102096, 2020, doi: 10.1016/j.jretconser.2020.102096.
[7]   S. J. and S. Swamy, "A prior case study of natural language processing on different domain," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 5, p. 4928, 2020, doi: 10.11591/ijece.v10i5.pp4928-4936.
[8]   P. Patel and A. Thakkar, "The upsurge of deep learning for computer vision applications," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, p. 547, 2020, doi: 10.11591/ijece.v10i1.pp538-548.
[9]   W. Li, *et al.*, "Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification," *Neurocomputing*, vol. 381, p. 64, 2020, doi: 10.1016/j.neucom.2020.01.006.
[10] G. Murtarelli, A. Gregory, S. Romenti, "A conversation-based perspective for shaping ethical human–machine interactions: The particular challenge of chatbots," *Journal of Business Research*, vol. 129, pp. 927-935, 2020, doi: 10.1016/j.jbusres.2020.09.018.
[11] S. Berhil, H. Benlahmar, N. Labani, "A review paper on artificial intelligence at the service of human resources management," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 18, no. 1, p. 35, 2020, doi: 10.11591/ijeecs.v18.i1.pp32-40.
[12] S. R. Salkuti, "A survey of big data and machine learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 1, p. 2088-8708, 2020, doi: 10.11591/ijece.v10i1.pp575-580.
[13] R. Chidananda, "Cornell Movie-Dialogs Corpus," *Kaggle datasets*, 2018.
[14] M. Yaqub, *et al.*, "Modeling of a full-scale sewage treatment plant to predict the nutrientremoval efficiency using a long short-term memory (LSTM) neural network," *Journal of Water Process Engineering*, vol. 37, p. 3, 2020, doi: 10.1016/j.jwpe.2020.101388.
[15] H. Fang, *et al.*, "An efficient radial basis functions mesh deformation with greedy algorithm based on recurrence Choleskey decomposition and parallel computing Parallel computing and swarm intelligence based artificial intelligence model for multi-step-ahead hydrological time series prediction," *Journal of Computational Physics*, vol. 377, p. 186, 2019, doi: 10.1016/j.jcp.2018.10.029.
[16] J. M. Sadler, *et al.*, "Leveraging open source software and parallel computing for model predictive control of urban drainage systems using EPA-SWMM5," *Environmental Modelling & Software*, vol. 120, p. 11, 2019, doi: 10.1016/j.envsoft.2019.07.009.

[17] D. Shao, *et al.*, "Domain-Specific Chinese Word Segmentation Based on Bi-Directional Long-Short Term Memory Model," *IEEE Access*, vol. 7, p. 12996, 2019, doi: 10.1109/ACCESS.2019.2892836.

[18] I. Attri and M. Dutta, "Bi-Lingual (English, Punjabi) Sarcastic Sentiment Analysis by using Classification Methods," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, p. 1386, 2019.

[19] D. Jurafsky and J. H. Martin, "Speech and Language Processing An Introduction to Natural Language Processing," *Prentince Hall*, 2020.

[20] L. Brown and S. Steinerberger, "Positive-Definite Functions, Exponential Sums and the Greedy Algorithm: a Curious Phenomenon," *Journal of Complexity*, vol. 61, p. 9, 2020, doi: 10.1016/j.jco.2020.101485.

[21] C. Lim, "An Evaluation of Machine Learning Approaches to Natural Language Processing for Legal Text Classification," *Imperial College London*, 2019.

[22] K. Kotecha, *et al.*, "Data Science and Intelligent Applications: Proceedings of ICDSIA 2020," *Springer Singapore*, vol. 52, p. 117, 2020, doi: 10.1007/978-981-15-4474-3.

[23] Y. Long, *et al.*, "Repeat Padding: Balancing words and sentence length for language comprehension in visual question answering," *Information Sciences*, vol. 529, pp. 170-171, 2020, doi: 10.1016/j.ins.2020.04.034.

[24] S. Kalya, M. Kulkarni, and K. S. Shivaprakasha, "Advances in Communication, Signal Processing, VLSI, and Embedded Systems: Select Proceedings of VSPICE 2019," *Springer Nature*, vol. 614, p. 307, 2019, doi: 10.1007/978-981-15-0626-0.

[25] H. Yin, *et al.*, "Intelligent Data Engineering and Automated Learning – IDEAL 2019, 20th International Conference, Manchester, UK, November 14–16, 2019, Proceedings, Part II," *Springer Nature*, vol. 11872, p. 182, 2019, doi: 10.1007/978-3-030-33617-2.

[26] L. Barolli, P. Hellinckx, T. Enokido, "Advances on Broad-Band Wireless Computing, Communication and Applications: Proceedings of the 14th International Conference on Broad-Band Wireless Computing, Communication and Applications (BWCCA-2019)," *Springer Nature*, vol. 97, 2019, doi: 10.1007/978-3-030-33506-9.

[27] K. Fereday, *et al.*, "A comparison of rolling averages versus discrete time epochs forassessing the worst-case scenario locomotor demands of professionalsoccer match-play," *Journal of Science and Medicine in Sport*, vol. 2, pp. 765-766, 2020, doi: 10.1016/j.jsams.2020.01.002.

[28] Y. Nie, P. Jiang, H. Zhang, "A novel hybrid model based on combined preprocessing method and advanced optimization algorithm for power load forecasting," *Applied Soft Computing Journal*, vol. 97, pp. 16-17, 2020, doi: 10.1016/j.asoc.2020.106809.

[29] P. Silitonga, *et al.*, "Comparison of Dengue Predictive Models Developed Using Artificial Neural Network and Discriminant Analysis with Small Dataset," *Appl. Sci.*, vol. 11, p. 15, 2021, doi: 10.3390/app11030943.

[30] G. Ardaneswari, A. Bustamam, D. Sarwinda, "Implementation of plaid model biclustering method on microarray of carcinoma and adenoma tumor gene expression data," *Journal of Physics: Conference Series*, vol. 893, p. 4, 2017, doi: 10.1088/1742-6596/893/1/012046.

BIOGRAPHIES OF AUTHORS

Prasnurzaki Anki B.Sc. received the BSc (honour) degree in mathematics from Universitas Indonesia in 2020. He is currently pursuing a master's degree in mathematics from Universitas Indonesia. His research interests are in the areas of computational mathematics, data science, and artifificial intelligence.

Assoc. Prof. Alhadi Bustamam, Ph.D. received the BSc (honour) degree 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 University of Queensland, Australia, in 2011. His research focuses on high-performance computing approaches to computational mathematics, computational biology, bioinformatics, computer science, data science, and artificial intelligence. Currently, he is working as an Associate Professor and the Head of Bioinformatics and Advanced Computing Laboratory (BACL) at the Department of Mathematics. He is also serving as the chairman of Data Science Centre (DSC) https://dsc.ui.ac.id at Universitas Indonesia.