

Optimization Chatbot Services Based on DNN-Bert for Mental Health of University Students

Azmi Abiyyu Dzaky ^{1*}, Junta Zeniarja ^{2**}, Catur Supriyanto ^{3**}, Guruh Fajar Shidik ^{4**}, Cinantya Paramita ^{5**},
Egia Rosi Subhiyanto ^{6**}, Sindhu Rakasiwi ^{7**},

* Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia

** Research Center for Intelligent Distributed Surveillance and Security, Universitas Dian Nuswantoro, Semarang, 50131, Indonesia

azmiabiyyudzaky456@gmail.com ¹, juntazeniara@research.dinus.ac.id ², catur.supriyanto@dsn.dinus.ac.id ³,

guruh.fajar@research.dinus.ac.id ⁴, cinantya.paramita@dsn.dinus.ac.id ⁵, egia@dinus.ac.id ⁶, sindhu.rakasiwi@dsn.dinus.ac.id ⁷

Article Info

Article history:

Received 2024-03-22

Revised 2024-05-08

Accepted 2024-05-15

Keyword:

BERT,

Chatbot,

College Student,

DNN,

NLP.

ABSTRACT

Attention to mental health is increasing in Indonesia, especially with the recent increase in the number of cases of stress and suicide among students. Therefore, this research aims to provide a solution to overcome mental health problems by introducing a chatbot system based on Deep Neural Networks (DNN) and BiDirectional Encoder Representation Transformers (BERT). The primary objective is to enhance accessibility and offer a more effective solution concerning the mental health of students. This chatbot utilizes Natural Language Processing (NLP) and Deep Learning to provide appropriate responses to mild mental health issues. The dataset, comprising objectives, tags, patterns, and responses, underwent processing using Indonesian language rules within NLP. Subsequently, the system was trained and tested using the DNN model for classification, integrated with the TokenSimilarity model to identify word similarities. Experimental results indicate that the DNN model yielded the best outcomes, with a training accuracy of 98.97%, validation accuracy of 71.74%, and testing accuracy of 71.73%. Integration with the TokenSimilarity model enhanced the responses provided by the chatbot. TokenSimilarity searches for input similarities from users within the knowledge generated from the training data. If the similarity is high, the input is then processed by the DNN model to provide the chatbot response. This integration of both models has proven to enhance the responsiveness of the chatbot in providing various responses even when the user inputs remain the same. The chatbot also demonstrates the capability to recognize various inputs more effectively with similar intentions or purposes. Additionally, the chatbot exhibits the ability to comprehend user inputs although there are many writing errors.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Student mental health has now become a major concern in Indonesia, with a significant increase in cases of stress and suicide. Thus, it needs to receive great attention. According to data from the National Criminal Information Center (PUSIKNAS) of the Indonesian National Police (POLRI), there were 971 cases of suicide in Indonesia from January to October 2023. This indicates that Indonesia is facing a mental health crisis. Unaddressed mental illnesses contribute significantly to the escalating suicide rates [1]. Research notes

that 25% of Indonesian students experience depression, 51.1% suffer from excessive anxiety, and 38.9% undergo stress. This evidence demonstrates that a significant proportion of students are grappling with mental health issues. Moreover, students in their final semesters are more likely to experience depression and stress [2].

The causes behind students contemplating suicide involve mental health disorders and lifestyle influences. The stress experienced by students is commonly associated with familial issues, environmental factors, and educational stress triggers

[3]. Meanwhile, senior semester students in college face greater pressure, such as academic-related stress [4]. Educational problems and feelings of disappointment with parents [5]. In this situation, more serious attention is required to address the mental health of students. This is addressed by identifying existing risks and providing necessary support to prevent further deterioration of their condition.

University students actively accessing social media, raising problems of inaccurate self-diagnosis which can significantly elevate stress levels. Consequently, Access to mental health services for students is crucial, and there is a need for improvement, such as reducing wait times, facilitating easy access to alternative treatments, and fostering a supportive environment. These enhancements are essential to support students experiencing mental health challenges [6]. However, awareness regarding mental health in Indonesia remains low, primarily due to negative societal stigma. This issue is further increased by low literacy levels, limited access to services, financial constraints, and challenges in seeking professional assistance [7]. This situation blocks access to seeking assistance and leads to delays in addressing mental health concerns.

The utilization of technology in healthcare services, such as chatbots for patient-expert interactions, and machine learning algorithms, can enhance service efficiency. This facilitates user access to receive better healthcare services [8]. The integration of AI into healthcare services has transitioned into a digital version, aiming to improve experiences and optimize mental healthcare services [9]. Currently, a majority of mental health patients seek information online, turning to online sources when experiencing symptoms of mental health issues [10]. Question-and-answer systems have proven effective in providing interactive services due to their easy accessibility and continuous availability [11]. Additionally, electronically integrated expressive writing-based therapies have shown positive impacts on individuals indicating depression or health-related stress [12].

The developed online mental health chatbot system effectively addresses early-stage mental health issues with improved flexibility. This chatbot provides chat sessions like counseling, supported by natural language processing (NLP), Deep Neural Network (DNN), and Bidirectional Encoder Representations from Transformers (BERT). The advantage of DNN lies in its ability to automatically learn relevant features from data without requiring manual feature extraction, DNN has proven to deliver excellent results in various text processing tasks [13]. The utilization of the Natural Language ToolKit (NLTK) within the Python library proves beneficial in comprehending human language. It encompasses various modules designed for tasks such as tokenization, stemming, tagging, and parsing in NLP-related assignments [14].

This research focuses on text processing in the Indonesian language, where issues such as overstemming and understemming commonly happen during data preprocessing. The Sastrawi library utilized in Indonesian text processing has

proven to reduce instances of overstemming and understemming, while also enhancing processing speed [15]. The BERT algorithm is used in NLP development to train language models on the order of words in sentences. The transformer-based IndoBERT model derived from BERT is designed to process Indonesian language data, used because it can provide advanced performance when handling Indonesian language datasets [16].

In this study, the utilization of the chatbot is specifically delimited to the domain of college or university students' interactions, with the primary role being to provide solutions and information about mild mental health issues experienced by college students. However, it does not completely replace the role of a professional. In case of complex questions or issues from users, the chatbot will direct them to consult with a professional. The research reveals the chatbot's high accuracy of 98.97%, validation of 71.74%, and the 71.73% test results showcasing its effective and varied responses even though there are writing errors from users. This method supports users with mental health queries. Additionally, the chatbot aims to offer counseling and early help to students, purposed to reduce counseling demands, improve access to help and enhance mental health services efficiency.

II. METHODOLOGY

A. Datasets

This dataset was gathered using various collection methods, including personal counseling sessions involving 4 individuals from students' college with a psychologist. Throughout the counseling sessions, conversations were recorded, summarized, and transformed into textual data. Besides that, this dataset was collected through a Google Form list question which was filled out by around 193 respondents, most of whom were general students from various backgrounds. The questionnaire consists of 8 questions that highlight various aspects that influence mental health, in line with the focus of existing research, including personal life, family, individual capabilities, friendships, relationships, anxiety, and mental health. From the respondents who filled it out, a total of 1544 questions and statements were gathered.

The collected questions and statements through several stages of processing, including the removal of duplicates, elimination of irrelevant data, and filtration of frequently occurring, considered important, and challenging questions or statements. These questions and statements were then presented to psychologists to obtain answers, and during counseling sessions, the conversations were recorded, and subsequently summarized into textual data. In the entire dataset, there are a total of 1803 distinct tags comprising patterns and different responses. Patterns represent the formulation of questions or statements, while responses represent the answers provided by psychologists during counseling sessions.

```

C:\Users\ACER> OneDrive > Documents > 000INDSS > BOT_PRIVATE > DEPLOY > {} Dataset: copyjian > {} ]> ]> ]> {} 1493
7463 "patterns": ["orang jujur memangnya bisa sukses?"],
7464 "responses": ["jika anda berfikir seperti itu maka lingkungan anda penuh dengan orang yang tidak jujur. Orang jujur...
7465
7466
7467 "tag": "mental health apa",
7468 "patterns": ["apa itu mental health?"],
7469 "responses": ["mental health kesehatan mental tidak selalu berarti kita harus selalu merasa bahagia. Sebaliknya, ke...
7470
7471
7472 "tag": "mental health apa",
7473 "patterns": ["jelaskan apa itu mental health?"],
7474 "responses": ["mental health kesehatan mental tidak selalu berarti kita harus selalu merasa bahagia. Sebaliknya, ke...
7475
7476
7477 "tag": "mental health apa singkat",
7478 "patterns": ["jelaskan secara singkat apa itu kesehatan mental?"],
7479 "responses": ["mental health itu tidak kita tidak harus selalu bahagia. Sebaliknya, kesehatan mental berkaitan deng...
7480
7481
7482 "tag": "mental health apa singkat",
7483 "patterns": ["jelaskan secara singkat apa itu kesehatan mental?"],
7484 "responses": ["mental health itu tidak kita tidak harus selalu bahagia. Sebaliknya, kesehatan mental berkaitan deng...
7485
7486
7487 "tag": "mental health penting",
7488 "patterns": ["apakah mental health penting?"],
7489 "responses": ["kesehatan mental adalah hal yang sangat penting karena berdampak langsung pada potensi maksimal kita...
7490
7491
7492 "tag": "mental health penting",
7493 "patterns": ["seberapa penting mental health?"],
7494 "responses": ["kesehatan mental adalah hal yang sangat penting karena berdampak langsung pada potensi maksimal kita...
7495
7496
7497 "tag": "mental health penting",
7498 "patterns": ["apakah mental kesehatan mental?"],
7499

```

Figure 1. Dataset Sampels

TABLE 1
DATASET DESCRIPTION

Atribut Name	Description
Tag	The tags contain keywords or terms that reflect the key points of the patterns and responses within the dataset. Example: "parents," "stress," "sad," etc. <i>(translated from Indonesian).</i>
Patterns	Patterns contain various question sentence structures that correspond to the responses. Example: "I'm sad," "why are my parents always angry?" "I want to suicide," etc. <i>(translated from Indonesian).</i>
Responses	Responses contain answers to patterns and serve as the basic responses for the chatbot. Example: "What makes you sad?" "Your parents love you," etc. <i>(translated from Indonesian).</i>

B. Purposed Method

This research implements Deep Learning models on text data consisting of various patterns and responses. In the data preprocessing stage, the separation of words in a sentence into token forms or Tokenization is conducted. The BERT model "Indolem/indobert-base-uncased" is used for tokenization, disregarding punctuation marks such as "!", ",", "?", "" because in processing this dataset, these ignored marks have a small impact on the dataset's word data. The BERT model is employed for tokenization due to IndoBERT's superiority compared to other tokenization models used in the application of the dataset at hand. StemmerFactory from the Sastrawi library is used in the process of converting tokens into base words or stemming due to Sastrawi's capability to reduce occurrences of overstemming and understemming in Indonesian language text data.

The Lemmatization process utilizes the WordNet_lemmatizer from NLTK to return words to their

base form, considering the meaning of words within the sentence. The final step in data preprocessing involves partitioning the dataset into a training set, a testing set, and a validation set which is taken from 50% of the test data. This distribution purpose is to facilitate the model's learning process and subsequent evaluation. The adjustment of data comparison quantities will align with the optimal performance achievable by the model based on the available dataset. Various data divisions will be tested, and the best outcome will be selected for implementation.

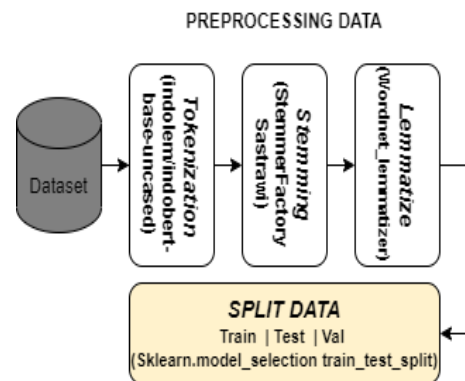


Figure 2. Preprocessing Data

In classification tasks, the utilization of DNN involves multiple Dense layers, activation functions, dropout, and batch normalization to enhance the model's training stability. To prevent overfitting, regularization is applied to the dense layers. This DNN model is designed to handle complex input data, enabling effective adaptation throughout the training process. The categorical cross-entropy loss function is also employed in designing the model, optimized with learning rate, momentum, and decay during the training process using a specified batch size. Training results are displayed with verbosity for clarity.

Additionally, the TensorBoard callback is utilized to log and record histograms throughout the training process. The TokenSimilarity model based on BERT, specifically IndoBERT, is utilized to measure the similarity level between two tokens. The process begins by loading the pre-trained IndoBERT language model and preprocessing the data. Each token undergoes processing by the model to generate embedding representations, employing the mean pooling technique from the model's output. The next process involves predicting the similarity level between two tokens using cosine similarity. If the similarity value exceeds the threshold of 80%, the pattern corresponding to the user input will be classified by the DNN model. The flow of the Token Similarity Model can be seen in Figure 3.

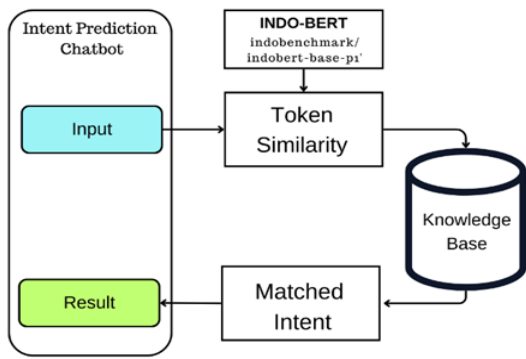


Figure 3. Token Similarity Model

The DNN model then performs classification using the Bag of Words (BoW) method on the received patterns. The BoW process cleanses and preprocesses the submitted patterns, and its output serves as input to the DNN model. Following the training process, the DNN model predicts the classification response from the input, resulting in generating random responses for the bot. The integration of TokenSimilarity and DNN enables contextual and diverse responses. The DNN classification process utilizes token similarity information from TokenSimilarity, thereby enhancing intent recognition. The final response stems from the DNN classification based on the pattern closest to the user input. For an outline of how this chatbot works, see Figure 4.

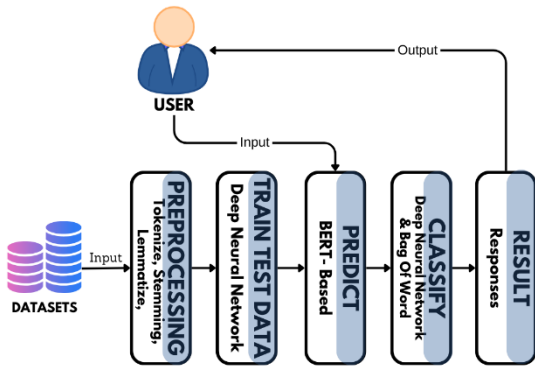


Figure 4. Chatbot Flowchart

C. Deep Neural Network (DNN)

DNN utilizes a neural network architecture with hidden layers positioned between the input and output layers by segmenting data hierarchically. Each hidden layer functions to make the feature learning process more complex and hierarchical. Generally, DNN is employed in Artificial Intelligence applications to achieve higher classification accuracy compared to other classification models [17].

$$z^n = a^{n-1}w^n + b^n \tag{1}$$

$$a^n = g(Z^n) \tag{2}$$

In Equation (1) and (2) n represents the index of a neural network layer within a specific range, with z^n denoting the activation vector at each n layer. $a^{n-1}/text$ represents the output of the preceding layer and serves as the input for the subsequent layer. a^n will provide the output of the final layer, representing the neuron activations in that layer. Its interpretation depends on the specific task of the DNN itself, such as classification or regression. g is a non-linear activation function such as ReLU, Sigmoid, Tanh, Softmax, etc. How the DNN model works can be seen in Figure 5.

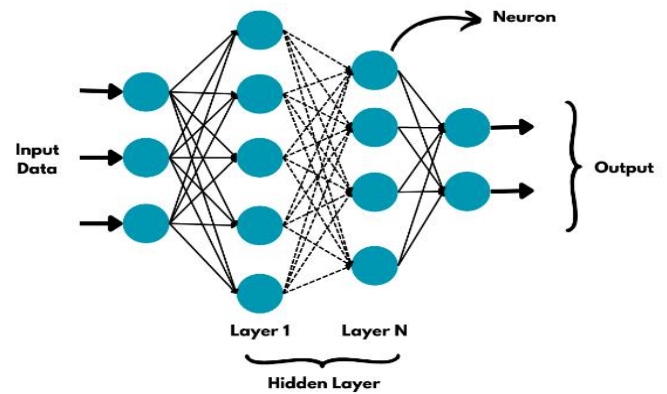


Figure 5. DNN Model System

D. Stochastic Gradient Descent with Momentum (SGDM)

SGDM commonly applies a dynamic momentum acquisition step. Parameter strategies like learning rate generate better results than static parameters throughout the training process [18]. This provides SGD with dynamic momentum approach more effective in optimizing chatbot models due to its adaptability in adjusting parameters during the training process.

$$w_t = \beta v_t + (1 - \beta)\Delta J(\theta_t) \tag{3}$$

$$\theta_{t+1} = \theta_t - \alpha v_t \tag{4}$$

In equation (3) the parameter w_t is updated based on the previous momentum value at each iteration t with θ_t . The momentum coefficient, β where $0 \leq \beta < 1$, optimizes the process by retaining a memory of gradient changes during training and memorizing gradient changes during training to stabilize the model's convergence, and the $\Delta J(\theta_t)$ represents the gradient of the cost function concerning θ_t .

In equation (4) θ_{t+1} is updated by subtracting the product of the learning rate α and the momentum value v_t . The value α determines the magnitude of parameter changes at each iteration, where a higher learning rate can accelerate convergence. However, if the learning rate is too large, the

model might overshoot the minimum value. Conversely, if the learning rate is too small, the convergence process toward the minimum value might become slower or even halt at a local minimum.

E. Bidirectional Encoder Representation from Transformers (BERT)

In general, BERT-based transformer architectures are trained on complex tasks at a large scale. BERT's advantage lies in its ability to compute similarity between words or phrases. The BERT model in the context of Conversational Question Answering (ConvQA) is considered effective in processing lengthy sequences of data, particularly in handling multiple interactions within conversations. The utilization of BERT to contextualize textual context has proven to enhance the model's performance [19].

IndoBERT “indobenchmark/indobert-base-pl” employs 124.5 million parameters for predicting the next sentence. IndoBERT stands out by achieving the highest average scores in text classification tasks [20]. The workflow of BERT comprises two stages, which are pre-training and fine-tuning. Pre-training involves self-supervised tasks like guessing hidden words and predicting if two adjacent sentences are related. Input representation is computed through word tokenization, merging embedding layers, and utilizing specific tokens for classification prediction [21]. Figure 6 shows how BERT works.

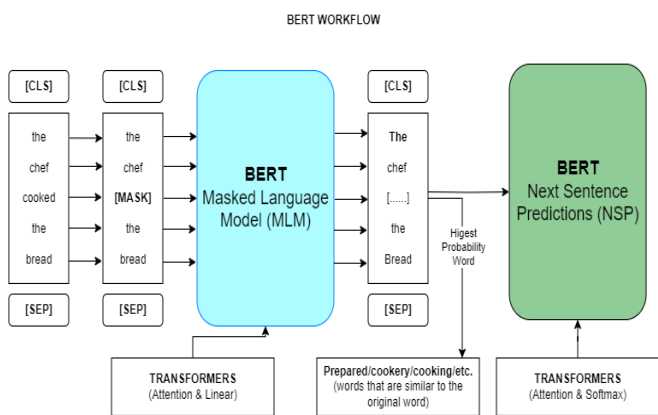


Figure 6. BERT Workflow

F. Bag of Word (BoW)

In this study, the Bag of Words (BoW) method is employed to evaluate the presence of words in a sentence by classifying text and computing weights for each word. BoW accepts a list of recognized vocabulary words, and then constructs a vector comprising values of 0 and 1.

Example of how BoW works:

- a) Tokenize
 - Sentence1: ["Saya", "suka", "pemrosesan", "bahasa", "alami"].
 - Sentence2: ["Pemrosesan", "bahasa", "alami", "sangat", "menarik"].
- b) Library
 - Includes unique words from all words that appear in both sentences.
 - Dictionary : ["Saya", "suka", "pemrosesan", "bahasa", "alami", "sangat", "menarik"].
- c) BoW Vector
 - Count the number of occurrences of each word in each sentence.
 - Sentence1 : [1, 1, 1, 1, 1, 0, 0] (1 "Saya", 1 "suka", 1 "pemrosesan", 1 "bahasa", 1 "alami", 0 "sangat", 0 "menarik")
 - Sentence2 : [0, 0, 1, 1, 1, 1, 1] (0 "Saya", 0 "suka", 1 "pemrosesan", 1 "bahasa", 1 "alami", 1 "sangat", 1 "menarik").

Systems based on the Bag of Words (BoW) approach have demonstrated the capability to attain elevated performance scores when employed in text classification algorithms, as evidenced by numerous studies [22].

III. RESULTS AND DISCUSSION

A. Results

This research uses the Google Colab platform, utilizing a 15 GB T4 GPU for computing tasks. Additionally, 12 GB RAM was used to facilitate data processing and training for both mental health chatbot models. Pyngrok is also used in developing user interfaces for chatbots.

The data preprocessing results involving tokenization using indoBERT, stemming using Sastrawi's stemmerFactory, and lemmatization using NLTK's WordNet_lemmatizer have produced good data processing. The preprocessed data was trained within the model. Throughout the design phase of the Deep Neural Network (DNN) classification model, experiments revealed optimal outcomes utilizing a 5-layer architecture, encompassing both input and output layers. This configuration comprised 3 Dense layers of size 256 with ReLU activation, 3 Dropout layers with a size of 0.4, and Batch Normalization to enhance model training stability. Overfitting prevention was implemented using L2 Regularization with a parameter of 0.01 specifically applied to the second Dense layer. The model build results can be seen in the Figure 7

TABLE 2
 EXPERIMENT TABLE WITH 256 DENSE AND 5 LAYERS

Train: Test: Val (%)	Learning Rate	Accuracy	Accuracy_Loss	Validation	Validation_Loss
90: 10: 5	0.001	0.9903	0.1531	0.5870	2.6615
	0.01	0.9897	0.0744	0.7174	2.6392
	0.1	0.2880	13.3454	0.1630	17.0445
80: 20: 10	0.001	0.9905	0.1605	0.6359	2.8650
	0.01	0.9932	0.0344	0.6413	3.8407
	0.1	0.2001	5.5967	0.1141	27.3859
70: 30: 15	0.001	0.9914	0.1781	0.5399	3.5361
	0.01	0.9852	0.2077	0.5399	5.7410
	0.1	0.0514	5.9707	0.0435	7.6468
60: 40: 20	0.001	0.9927	0.1835	0.4122	5.0395
	0.01	0.9900	0.0400	0.5399	5.7140
	0.1	0.0191	35.6367	0.0082	35.8925
50: 50: 25	0.001	0.9946	0.2040	0.4924	44.303
	0.01	0.9935	0.0320	0.4880	5.5722
	0.1	0.0153	62.437	0.0065	7.2020

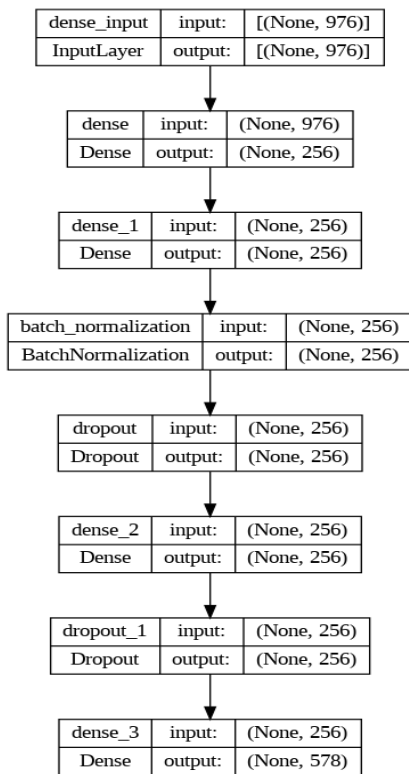


Figure 7. DNN Build Model

The Categorical Cross Entropy loss function is also utilized. Following several tests, the Stochastic Gradient Descent (SGD) optimizer with parameters: learning rate 0.01, momentum 0.9, and damage rate 1e-6 is executed across 1000 epochs of training with a batch size of 64. Training progress is displayed at verbosity level 1 for clarity, and histogram recording is logged using the TensorBoard callback.

Table 2 shows significant variations observed during experiments when splitting the data into 90: 10: 5 ratios with a dense layer of 256 and a learning rate of 0.01. Additionally, other outcomes indicate an accuracy of 98.97%, with Accuracy Loss of 0.0744, which, while not as promising as in other experiments, achieves the highest validation result of 71.74% with 2.6392 Validation Loss specifically with that split. The best test data results from the optimal experiment in the table produce the best accuracy of test accuracy 71.73% and test loss of 3.807. From the experimental findings, partitioning the data into segments and varying the dense layer's values did not significantly impact the accuracy or validation. Similarly, adding layers and modifying batch size showed minimal influence on the outcomes. However, the learning rate value proved to be notably impactful. A higher learning rate, such as 0.1, resulted in significantly poorer performance when applied to the model and dataset in this study. Conversely, a lower learning rate demonstrated considerably improved results.

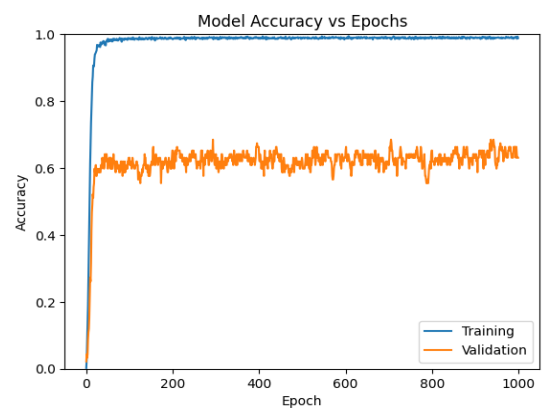


Figure 8. Accuracy Graphic

From Figure 8, illustrating the outcomes of 1000 iterations using the Deep Neural Network (DNN) model, it can be observed that the training accuracy demonstrates remarkably consistent performance, maintaining a stable value surpassing 98%. This signifies the DNN model's effective capability to learn and recognize patterns within the training dataset. However, on the validation side, it exhibits lower values, with metric values ranging between 0.5 to 0.71. This suggests that the model's proficiency, while adept at learning from the training data, may require further attention or strategies to improve performance on new, unseen data, indicating potential challenges in generalizing to new data points.

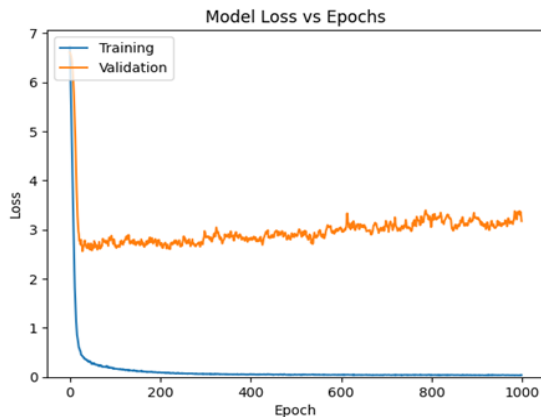


Figure 9. Model Loss Graphic

Figure 9 describes a noticeable trend where the training loss consistently diminishes, diverging from the validation loss that tends to ascend over time. The substantial difference between the training, test, and validation accuracy suggests a potential occurrence of overfitting within the developed model.

```
(Mulai Percakapan dengan Bot(quit/q/bye) untuk stop !)  
You: bagaimana caranya aku mengatasi stres  
1/1 [=====] - 0s 19ms/step  
1/1 [=====] - 0s 21ms/step  
bot: Oke saya memiliki beberapa cara untuk mengatasi stres  
1. Kelola waktu dengan baik. Sisihkan waktu untuk istirahat dan me  
2. Temukan relaksasi yang tepat untuk anda, mungkin anda bisa beri  
3. Jaga Pola makan dan pola hidup sehat anda bisa dengan makan mak  
Tapi dari semua itu akan lebih baik jika kita menghindari stres  
  
You: gimana sih cara ngatasin stres  
1/1 [=====] - 0s 17ms/step  
1/1 [=====] - 0s 19ms/step  
bot: Oke saya memiliki beberapa cara untuk mengatasi stres  
1. Kelola waktu dengan baik. Sisihkan waktu untuk istirahat dan me  
2. Temukan relaksasi yang tepat untuk anda, mungkin anda bisa beri  
3. Jaga Pola makan dan pola hidup sehat anda bisa dengan makan mak  
Tapi dari semua itu akan lebih baik jika kita menghindari stres  
  
You: kalua begitu bagaimana caranyi aku menjaag kesehatan mental  
1/1 [=====] - 0s 18ms/step  
1/1 [=====] - 0s 18ms/step  
aku tidak paham, coba jelaskan lagi...
```

Figure 10. DNN Classification Report

From Figure 10, it is evident that the Deep Neural Network (DNN) model is still capable of handling various inputs, recognizing and classifying diverse inputs into

patterns aligned with the knowledge acquired from training data, yielding good responses. However, when users provide input with numerous spelling errors, the DNN model fails to recognize the given sentence by the user.

```
Kamu : apa sih mental health itu  
Token input cocok dengan intent: mental health apa  
Pola terbaik: apa itu mental health?  
Kemiripan: 88.04%  
  
Kamu : kenapa isu mental health semakin parah dialami remaja ?  
Token input cocok dengan intent: mental health tanya 2  
Pola terbaik: apa yang membuat isu mental health semakin parah di  
Kemiripan: 88.17%
```

Figure 11. TokenSimilarity Model Result

In Figure 11, the implementation of the TokenSimilarity model based on BERT showcases exceptional performance by effectively establishing connections between input data and the knowledge embedded within the trained model.

Initial data processing using TokenSimilarity aims to identify similarities above 80% before directing the data to the DNN model for classification. However, if the similarity is less than 80%, the chatbot will respond by showing a lack of understanding and asking the user to provide a clearer explanation. This approach is designed to overcome the limitations of DNN models, which are demonstrated not satisfactory accuracy and had difficulty handling user typing errors. By using TokenSimilarity as an initial filter, data directed to the DNN model will undergo a selection process in the TokenSimilarity model, thereby increasing response accuracy. This minimizes the possibility that the chatbot response will not understand the user input and the classification response will be inaccurate.



Figure 12. Chatbot Response 1 (varied answers)



Figure 13. Chatbot Response 2 (varied answers)

Figures 12 and 13 provide a comprehensive visual representation of the obtained response outcomes by integrating the Deep Neural Network (DNN) model and the BERT TokenSimilarity model. Upon combined analysis, it is evident that the chatbot demonstrates remarkable capability in delivering varied responses even when presented with identical inputs.



Figure 14. Chatbot Response 3 (Handle Spelling Errors)

Conversely, the analysis in Figure 14 signifies the bot's consistency in delivering effective responses, even when encountering numerous spelling errors in the provided input. This indicates that the BERT model provides added capability to the DNN classification model in delivering consistent and effective chatbot responses in situations involving input variations or spelling errors.

B. Discussion

From the experimental results of the DNN model, the chatbot can provide satisfying responses, with a training accuracy of 98.97%, validation accuracy of 71.74%, and test accuracy of 71.73% indicating the presence of overfitting in the model. In his journal, G. H. G Krishna Vamsi utilized a DNN model to create a chatbot. He employed a limited dataset and suggested that employing a constrained dataset with complex algorithms might result in overfitting in

network models. therefore, careful precautions are needed to determine the optimal learning rate during the model development process [23].

The dataset limitations in this experiment guide to potential memorization of information by the model from the training data, preventing its ability to generalize to new data. The use of a complex Deep Neural Network (DNN) model increased the risk of overfitting due to the relatively small and limited dataset. Although DNN is effective for classification tasks, the model struggles to comprehend and interpret data effectively when faced with limited quantity and diversity of data. To address this, it is recommended to utilize larger and more diverse datasets. This suggestion stems from an experiment conducted using a public dataset from Kaggle with 81 tags, where the DNN achieved 99% accuracy on training data but only 31.87% on validation data, resulting in a training data loss of 0.2453 compared to a validation data loss of 4.1996.

The accuracy of DNN models is typically higher on replication datasets due to a principled evaluation protocol utilizing uncertainty-related information, contrary to conventional evaluation approaches [24]. In such conventional tests, possible variations or analyzes in the data may not be explicitly considered. That is, the model is tested with the assumption that the test data itself is a perfect representation of the data that would be encountered in the real world, without considering variations or appearances that may arise. Validation accuracy may not reflect a DNN's true quality, as it can still be vulnerable to adversarial attacks, misclassify data, lack human interpretability, or exhibit overconfidence on out-of-distribution data [25].

IV. CONCLUSION

This study highlights the effectiveness of preprocessing techniques, specifically, tokenization using IndoBERT, stemming from Sastrawi, and lemmatization via WordNet lemmatizer, in handling Indonesian language datasets. However, despite the success of these preprocessing methods, the Deep Neural Network (DNN) model utilized for text classification exhibited suboptimal performance in validation and testing accuracy. The analysis, underscores that this limitation is attributed to the scarcity and lack of diversity in the dataset.

The DNN model tends to overfit, memorizing existing data without discerning between different meanings within the dataset. This emphasizes the importance of addressing overfitting and developing models capable of better generalizing data. To mitigate these issues, this study proposes the integration of a TokenSimilarity model based on BERT as an additional approach. This model has shown promise in enhancing the performance of the DNN model by preprocessing initial user inputs before DNN classification.

This approach aims to enable the chatbot to provide better responses to various user inputs, including those containing typing errors, while also increasing response variability for

similar inputs. Thus, a combination of careful preprocessing techniques, appropriate overfitting management, a good dataset, and the application of additional models such as BERT-based TokenSimilarity can be effective steps in improving chatbot performance in handling Indonesian language datasets. This opens the potential for further development in the field of natural language processing and the development of text-based interactive systems.

ACKNOWLEDGEMENTS

The author wishes to extend sincere appreciation to the Research and Service Community Institute (LPPM) at Dian Nuswantoro University for their invaluable support. This study received support from the Basic Research Grant Scheme under contract no. 109/A.38-04/UDN-09/XI/2023.

REFERENCES

- [1] D. Tadmon and P. S. Bearman, "Differential spatial-social accessibility to mental health care and suicide," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 120, no. 19, pp. 1–8, 2023, doi: 10.1073/pnas.2301304120.
- [2] E. Astutik, S. K. Sebayang, S. I. Puspikawati, T. D. Tama, and D. M. S. K. Dewi, "Depression, anxiety, and stress among students in newly established remote university campus in Indonesia," *Malaysian J. Med. Heal. Sci.*, vol. 16, no. 1, pp. 270–277, 2020.
- [3] G. I. Rina Suprayanti, Fathra Annis Nauli, "GAMBARAN PERILAKU SELF INJURY DAN RISIKO BUNUH DIRI PADA MAHASISWA," *Heal. CARE J. Kesehat.*, no. December, pp. 31–37, 2021, doi: <https://doi.org/10.36763/healthcare.v10i2.133>.
- [4] S. D. Pratiwi and R. Djuwita, "Hubungan Gaya Hidup dengan Kesehatan Mental Remaja di Indonesia (Analisis Data Global School-Based Student Health Survey Indonesia 2015)," *J. Epidemiol. Kesehat. Komunitas*, vol. 7, no. 1, pp. 382–393, 2022, doi: 10.14710/jekk.v7i1.11905.
- [5] Y. Himayanti, R. Ratnawati, L. Yuliatun, and H. D. Windarwati, "Phenomenology Study of High School Students' Experiences Who Do Suicidal-Attempt in Malang City, Indonesia," *Int. J. Sci. Soc.*, vol. 2, no. 3, pp. 259–268, 2020, doi: 10.54783/ijssoc.v2i3.168.
- [6] M. Janota, V. Kovess-Masfety, C. Gobin-Bourdet, and M. M. Husky, "Use of mental health services and perceived barriers to access services among college students with suicidal ideation," *J. Behav. Cogn. Ther.*, vol. 32, no. 3, pp. 183–196, 2022, doi: 10.1016/j.jbct.2022.02.003.
- [7] L. Munira, P. Liamputtong, and P. Viwattanakulvanid, "Barriers and facilitators to access mental health services among people with mental disorders in Indonesia: A qualitative study," *Belitung Nurs. J.*, vol. 9, no. 2, pp. 110–117, 2023, doi: 10.33546/bnj.2521.
- [8] S. Prakasam, N. Balakrishnan, T. R. Kirthickram, B. Ajith Jerom, and S. Deepak, "Design and Development of AI-Powered Healthcare WhatsApp Chatbot," *ViTECoN 2023 - 2nd IEEE Int. Conf. Vis. Towar. Emerg. Trends Commun. Netw. Technol. Proc.*, 2023, doi: 10.1109/ViTECoN58111.2023.10157423.
- [9] S. D'Alfonso, "AI in mental health," *Curr. Opin. Psychol.*, vol. 36, pp. 112–117, 2020, doi: 10.1016/j.copsyc.2020.04.005.
- [10] A. Yakisir and Z. C. Duman, "Mental health-related internet use by people with a diagnosis of chronic psychiatric disorder, attending community mental health centres," *Heal. Inf. Libr. J.*, no. May, pp. 1–11, 2022, doi: 10.1111/hir.12449.
- [11] J. Madan and S. Madan, "Intelligent and Personalized Factoid Question and Answer System," *2022 10th Int. Conf. Reliab. Infocom Technol. Optim. (Trends Futur. Dir. ICRITO 2022)*, pp. 1–7, 2022, doi: 10.1109/ICRITO56286.2022.9964818.
- [12] T. Nabarian, Y. Wirani, M. P. Siddiq, Aseptianova, M. Z. Aziz, and A. D. Saputro, "E-counseling-based Expressive Writing Therapy Platform for Overcoming Student Mental Health Problems," *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 2022-October, no. October, pp. 250–255, 2022, doi: 10.23919/EECSI56542.2022.9946453.
- [13] M. Jaiwai, K. Shiangjen, S. Rawangyot, S. Dangmanee, T. Kunsuree, and A. Sa-Nguanthong, "Automatized Educational Chatbot using Deep Neural Network," *2021 Jt. 6th Int. Conf. Digit. Arts, Media Technol. with 4th ECTI North. Sect. Conf. Electr. Electron. Comput. Telecommun. Eng. ECTI DAMT NCON 2021*, pp. 85–89, 2021, doi: 10.1109/ECTIDAMTNCN51128.2021.9425716.
- [14] M. Wang and F. Hu, "The application of nltk library for python natural language processing in corpus research," *Theory Pract. Lang. Stud.*, vol. 11, no. 9, pp. 1041–1049, 2021, doi: 10.17507/tpls.1109.09.
- [15] M. A. Rosid, A. S. Fitriani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 874, no. 1, 2020, doi: 10.1088/1757-899X/874/1/012017.
- [16] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.
- [17] R. Anusuya and K. Sakkarapani, "Efficient Training Strategy for Classification Using Deep Neural Network With Hybridization of Quantum Enthusied Artificial Bee Colony Algorithm," *Int. J. Health Sci. (Qassim)*, vol. 6, no. March, pp. 6864–6884, 2022, doi: 10.53730/ijhs.v6ns4.9958.
- [18] Y. Liu, Y. Gao, and W. Yin, "An improved analysis of stochastic gradient descent with momentum," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. 2, pp. 1–11, 2020.
- [19] A. Al-Besher, K. Kumar, M. Sangeetha, and T. Butsa, "BERT for conversational question answering systems using semantic similarity estimation," *Comput. Mater. Contin.*, vol. 70, no. 3, pp. 4763–4780, 2022, doi: 10.32604/cmc.2022.021033.
- [20] B. Willie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," 2020, [Online]. Available: <http://arxiv.org/abs/2009.05387>
- [21] A. Rogers, O. Kovaleva, and A. Rumshisky, "A primer in bertology: What we know about how bert works," *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 842–866, 2020, doi: 10.1162/tacl_a_00349.
- [22] W. A. Qader, M. M. Ameen, and B. I. Ahmed, "An Overview of Bag of Words: Importance, Implementation, Applications, and Challenges," *Proc. 5th Int. Eng. Conf. IEC 2019*, pp. 200–204, 2019, doi: 10.1109/IEC47844.2019.8950616.
- [23] G. H. G. Krishna Vamsi, Akhtar Rasool, "Chatbot A Deep Neural Network Based Human to Machine Conversation Model," *Int. Conf. Comput. Netw. Technol.*, 2020, doi: <https://doi.org/10.1109/ICCCNT49239.2020.9225395>.
- [24] E. T. Anzaku, H. Wang, A. Van Messem, and W. De Neve, "A principled evaluation protocol for comparative investigation of the effectiveness of DNN classification models on similar-but-non-identical datasets," *arXiv Prepr.*, 2022, [Online]. Available: <http://arxiv.org/abs/2209.01848>
- [25] J. S. Hyatt and M. S. Lee, "Requirements for Developing Robust Neural Networks," pp. 2–5, 2019, [Online]. Available: <http://arxiv.org/abs/1910.02125>