TEKNOKOM : Jurnal Teknologi dan Rekayasa Sistem Komputer
Vol. 6, No. 2, September 2023, pp. 50 - 57
ISSN: 2686-3219, DOI: 10.31943/teknokom.v6i2.125

IMPLEMENTATION OF TWO LANGUAGE CHATBOT WEB TO FIND INFORMATION ON BATIK USING THE ARTIFICIAL NEURAL NETWORK MODEL

Taufik Maulana Tanjung[1*], Rahmat Hidayat[1]
[1]Department of Electrical Engineering, Universitas Singaperbangsa Karawang, Indonesia

ARTICLE INFO

History of the article:
Received May 8, 2023
Revised May 25, 2023
Accepted June 26, 2023
Published July 8, 2023

Keywords:
Artificial Intelligence
Artificial Neural Network
Chatbot
Database

ABSTRACT

Information technology plays an important role in the progress of a country in order to disseminate information massively and make it easier for the public to obtain information. Currently, the development of information technology continues to innovate, such as with the existence of artificial intelligence (AI) technology. This research will discuss AI technology, namely the BatikInd chatbot with the artificial neural network (ANN) method, which aims to make it easier for people to find information about the types of batik that, exist in all regions of Indonesia. The design of the ANN method for making chatbots uses hundreds of sentences to be processed into a database of information about the types of batik and explanations about the batik you want to find. Database processing with epochs reaching 200 produces an accuracy rate of 1.0 (100%) which means the accuracy rate is very good and the error rate reaches 0.03 (0%) which means the error rate is very low. The results of a very good level of accuracy and a very low error rate indicate that this AI can function properly and is feasible to use. It is hoped that with information technology like this, it will make it easier for people to find information about batik and can increase people's literacy and knowledge about batik in all regions of Indonesia.

Correspondece:
Taufik Maulana Tanjung,
Department of Electrical Engineering,
Universitas Singaperbangsa Karawang,
Email :
taufik.maulana19148@student.unsika.ac.id

INTRODUCTION

Information technology plays an important role in the progress of a country in order to massively disseminate information and make it easier for the public to obtain information. Every time changes, innovation from information technology continues to change, as in the current era of digitalization. Information technology in society is developing very rapidly, starting from the continued growth of internet users, appearances that make information technology easier to use, media and more reliable, advances in the field of artificial intelligence to physical appearances that are getting smaller and more portable [1]. In fact, according to the Association of Indonesian Internet Organizing Agencies (APJII) there are 196.71 million active internet users for information search in Indonesia out of Indonesia's total population of 266.91 million people, or the equivalent of 73% in 2020 [2].

With this convenience through information technology, people can search for various kinds of information, such as for example searching for information about explanations of batik in Indonesia. With this, of course, it will increase literacy culture and add insight and can be part of preserving Indonesian batik culture. Culture is the heritage of the state, which should be preserved and upheld [3]. The introduction of AI-powered technology, especially Chatbot systems, has ushered in a batch of new opportunities for a variety of industries [4]. Searching for information about batik on the internet using search engines is a common thing to do, to further attract people's interest in finding information about batik, a chatbot application was created that can provide information through communication with bots. Chatbot itself is an application that can be used to practice conversations in Indonesian and English by means of Text-to-text. This application will be a friend to interact using Indonesian and English like a human [5]. The use of Chat- bots in education increase connectivity, efficiency, and reduce uncer- tainty in interactions [5]. A Chatbot is an

intelligent agent capable of interacting with a user to answer a series of questions and provide the appropriate response [6].

This chatbot works as a computer program that can fulfill a function or provide information via chat/text messages. A “robot” that can respond to chats and understand your questions or requests via chat. This is possible because chatbots leverage artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) technologies so that they can simulate conversations (or chats) with users in natural language through messaging apps, websites, mobile apps, or by phone [5]. The research method used in making this chatbot web application is to use the NLP domain. Natural language processing (NLP) is a subset of AI technology that gives machines the ability to read, understand and interpret human language [6].

Natural Language Processing (NLP) is also a field of computer science, artificial intelligence and language (linguistics) related to the interaction between computers and human natural language. In principle, natural language is a form of conveying information from one user to another. Natural language can be represented in the form of sound or text. The way the chatbot works begins by receiving input from the user in the form of a text message, the system then uses NLP to process the input to analyze, identify and interpret the meaning intended by the user [7]. The response generated from NLP is a pattern matching on the knowledge base using the Boyermoore algorithm and produces the response response that is considered the most suitable. Boyer moore algorithm can only handle the right keywords according to the knowledge base. In addition, a Levenshtein distance algorithm is used as a spelling correction if a typing error occurs in the user's input by comparing two strings and taking the smallest difference value [8].

The design of a chatbot with NLP capabilities focuses on AI technology using an artificial neural network (ANN) algorithm that collects data from frequently asked questions in search of information [9]. So that with NLP, chatbots can analyze and process data that has been designed in the database and then matched with questions from users. Therefore, the bot can respond based on the questions submitted to the bot. Then with the chatbot web application, users can exchange languages quickly and flexibly search for information that we want to know and can also be used as learning material to be more active in literature [10]. We hope that this chatbot can help the community in making it easier to find information, especially information about batik in Indonesia and can be useful to add insight.

RESEARCH METHOD

In this study, the authors designed a chatbot using the artificial neural network (ANN) method using the Google Collab platform to run the syntax code from the chatbot and to analyze the database that was created so that it remains accurate and functional. An Artificial Neural Network (ANN) is a computational algorithm that can solve complex problems by imitating the brain and consists of neurons or artificial nodes arranged in layers and interconnected by synaptic weights/connections. The ANN architectural model consists of a network of 3 (three) layers of interconnected nodes, namely the input layer, the hidden layer, and the output layer [11].

Data used

Data collection techniques are carried out by collecting data through articles available in the browser. The data needed is only based on questions that have been made regarding information about batik in Indonesia. The data obtained reached 20 articles from each of the 20 types of batik and the data collected became a dataset for chatbots in this study. Then the dataset used in this study was entered manually and stored in a file with JSON (JavaScript Object Notation) format. The dataset has a structure including:

1. Intents, are part of the collection of all input and output data used to train the chatbot.
2. Patterns, is part of the list of questions that might be asked by the user.
3. Responses, the part that contains answers from the chatbot that will be given to users.
4. Tags group similar text data together and are used as targeted output to train a neural network.

Experimental Research

At this stage, research data that has been collected will be processed using Natural Language Processing (NLP) and Natural Language Toolkit (NLTK) intelligence. Before the data is used in the artificial neural network (ANN) algorithm, the data collected contains long text and will be processed through several stages, namely:

1. Case Folding’s, converts all letters to lowercase and removes characters other than letters [12].
2. Tokenization, the process by which text strings are converted into tokens [13].
3. Stemming, which is the process of removing unnecessary words so that the

bot issues answers with appropriate and neat words [13].

In addition, because this study uses an artificial neural network (ANN) algorithm in which there is an activation function in the algorithm. This activation function will determine the output of a neuron in the form of linear or nonlinear. In this algorithm, there is an activation function. This activation function will determine the output of a neuron in the form of linear or nonlinear. The activation function has several activation functions, such as softmax, ReLU (Rectified Linear Unit), Sigmoid, and others. The activation functions used in this study include:

1. ReLU activation function

   The ReLU function (Rectified Linear Unit) is an activation function that has simple calculations, where if there is an element with negative energy, the value will be changed to 0, and there will be no exponential, multiplication or division operations. The formula for the ReLU activation function is formulated in equation (1) which can be seen below:

$$f(x) = ma\,x(0, x)$$

or

$$f(x) = \begin{cases} 0 \; untuk & x \leq 0 \\ x \; untuk & x > 0 \end{cases} \quad (1)$$

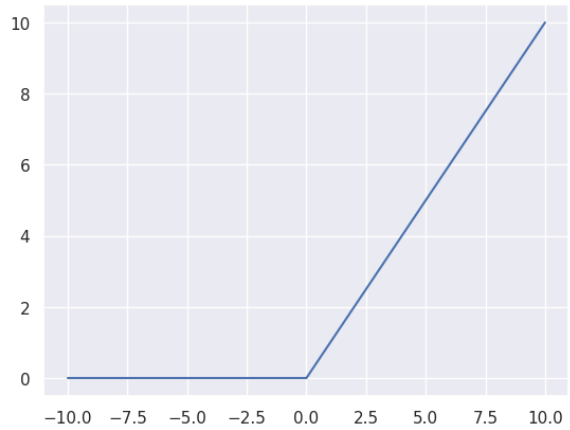

Figure 1. ReLU graph

2. Sigmoid activation function

   The sigmoid activation function used in this study is the binary sigmoid activation function, because the activation function has a value in the range 0 to 1. The formula for binary sigmoid activation is formulated in equation (2) below:

$$y = f(x) = \frac{1}{(1 + e^{-x})} \quad (2)$$

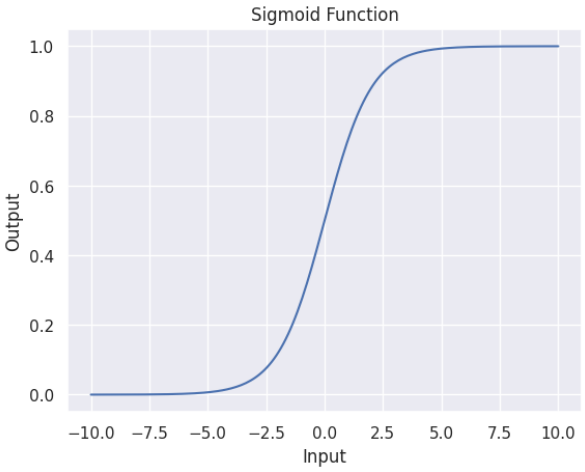

Figure 2. Binary Sigmoid Graph

Taking a sigmoid function logistically, we can evaluate the value of the function at several key points to understand the form of the function. When we take x (input) = 0, the function of the sigmoid logistics is evaluated to be:

$$S(0) = \frac{1}{(1 + e^{-x})}$$

$$= \frac{1}{(1 + 1)}$$

$$= 0.5$$

This is useful for the interpretation of sigmoid as probability in logistic regression models, as it indicates that zero input produces 0.5 output.

At with x = 1, you can find a slightly larger output value:

$$S(1) = \frac{1}{(1 + e^{-x})}$$

$$= \frac{1}{(1 + 0.36)}$$

$$= 0.73$$

And with x = 10, where x is the number of test questions and the result is that the value

of the sigmoid function becomes very close to 1.

$$S(10) = \frac{1}{(1 + e^{-x})}$$

$$S = \frac{1}{(1 + 0.0134)}$$

$$= 0.999954$$

It can be seen from the calculation results using the sigmoid formula, which is with an input value of 10, the result of the calculation is close to 1, which means that the data is perfect.

RESULTS AND DISCUSSION

The creation of the batik chatbot web was carried out through three stages, such as dataset creation, chatbot trials through Google Colabs, and deployment using visual studio code software. In these three stages also consists of several processes, which will be explained below.

Datasets

After searching the data, then the data is managed through Visual studio code software with the .JSON (JavaScript Object Notation) format. The dataset consists of 170 inputs, 170 tags, and 170 responses. This data can be seen in figure (3), which is a cut image of the dataset from making a chatbot.

.

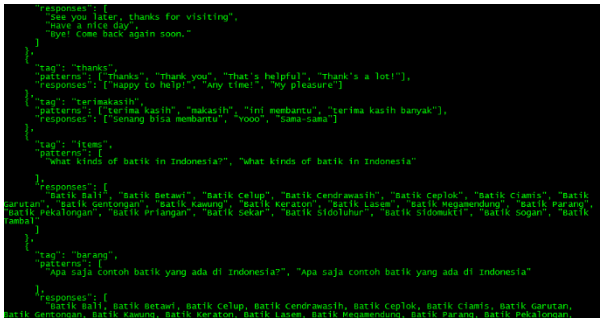

Figure 3. Chatbot Dataset Slices

Preprocessing

Before preprocessing the dataset, the dataset is imported first, with the syntax code import. Then, in the preprocessing stage, data exploration is carried out in this chatbot using tokenizer and lemmatization. The tokenization and lemmatization processes are carried out using the syntax code, which can be seen in figure (4).

```
#removing punctuations
import string
data['inputs'] = data['inputs'].apply(lambda wrd:[ltrs.lower() for ltrs in wrd if ltrs not in string.punctuation])
data['inputs'] = data['inputs'].apply(lambda wrd: ''.join(wrd))
#tokenize the data
from tensorflow.keras.preprocessing.text import Tokenizer
tokenizer = Tokenizer(num_words=2000)
tokenizer.fit_on_texts(data['inputs'])
train = tokenizer.texts_to_sequences(data['inputs'])
```

Figure 4. Tokenization process

In the Tokenization, process is a process where the string text is converted into a list of tokens. In addition, there are two tokens, namely the words tokenizer that is used to search for a list of words in a string. Then stem is also done, which is to delete words that are not needed so that the bot issues answers with appropriate and neat words. The result of preprocessing is that there are 81 tokens or unique words and 51 labels or categories and these results can be seen in figure (5).

```
Tags :
['barang' 'batik' 'batikbali' 'batikbalikenglish' 'batikbetawi'
 'batikbetawikenglish' 'batikcelup' 'batikcelupenglish' 'batikcendrawasih'
 'batikcendrawasihkenglish' 'batikceplok' 'batikceplokenglish'
 'batikciamis' 'batikciamisenglish' 'batikenglish' 'batikgarutan'
 'batikgarutanenglish' 'batikgentongan' 'batikgentonganenglish'
 'batikkawung' 'batikkawungenglish' 'batikkeraton' 'batikkeratonenglish'
 'batiklasem' 'batiklasemenglish' 'batikmegamendung'
 'batikmegamendungkenglish' 'batikparang' 'batikparangenglish'
 'batikpekalongan' 'batikpekalonganenglish' 'batikpriangan'
 'batikprianganenglish' 'batiksekar' 'batiksekarenglish' 'batiksidoluhur'
 'batiksidoluhurenglish' 'batiksidomukti' 'batiksidomuktienglish'
 'batiksogan' 'batiksoganenglish' 'batiktambal' 'batiktambalenglish'
 'funny' 'goodbye' 'greeting' 'items' 'salam' 'selamattingal'
 'terimakasih' 'thanks']
Kata-kata Unik :
["'s" 'a' 'ada' 'anyon' 'apa' 'apakah' 'are' 'bagaimana' 'bali' 'banyak'
 'batik' 'betawi' 'bye' 'celup' 'cendrawasih' 'ceplok' 'ciami' 'contoh'
 'day' 'dengan' 'di' 'dimaksud' 'do' 'doe' 'funni' 'garutan' 'gentongan'
 'good' 'goodby' 'halo' 'hello' 'help' 'hey' 'hi' 'how' 'ikat' 'in'
 'indonesia' 'ini' 'is' 'itu' 'jagad' 'joke' 'jumpa' 'kabarmu' 'kasih'
 'kawung' 'keraton' 'kind' 'know' 'lasem' 'later' 'lot' 'makasih' 'me'
 'mean' 'megamendung' 'membantu' 'of' 'orang' 'parang' 'pekalongan'
 'permisi' 'priangan' 'saja' 'sampai' 'see' 'sekar' 'selamat' 'sidoluhur'
 'sidomukti' 'sini' 'sogan' 'someth' 'tambal' 'tell' 'terima' 'thank'
 'that' 'there' 'tinggal' 'what' 'yang' 'you']
```

Figure 5. Preprocessing Results

Modeling

The next stage is architectural modeling in this chatbot that uses a neural network (NN) or also known as artificial neural networks (ANN). The modeling stage is carried out using two design applications, namely modeling through Google Colabs and through Visual Studio Code. In the artificial neural network (ANN) architecture, this chatbot consists of 4 layers, namely, input layer, hidden layer 1 (ReLU), hidden layer 2 (Sigmoid), and output. The input layer is a layer that contains the input text from the dataset and has passed the preprocessing stage. Then there is hidden layer 1 with 81 neurons that use ReLU activation. In addition, there is a hidden layer 2 with 51 neurons that use sigmoid activation. In addition, the last layer is output to provide output

to the user. From the ANN architecture layer, it can be seen in Figure (6).

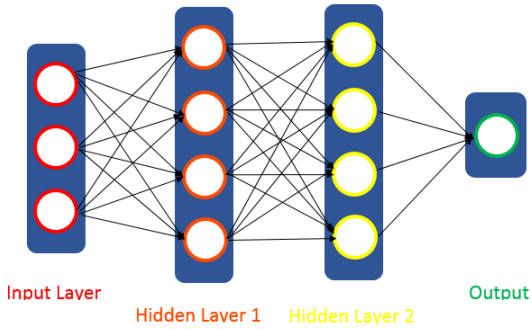

Figure 6. ANN architecture [17]

At the modeling stage, the data is tested repeatedly and gradually to get the highest accuracy value, which is worth one (1) and the lowest error value, which is zero (0). The process of testing the data at this modeling stage is that the data is processed per epoch with the model to be compiled using the SGD (Stochastic Gradient Descent) optimizer and the "Categorical Cross entropy" loss function. Where the results of the data testing process can be seen in Figure (7) below.

The formula of an artificial neural network is using the forward pass formula, where the forward pass is the process of carrying data at the input through each neuron in the hidden layer to the output layer, which is then calculated for errors [17].

$$dot_j = \sum_{i}^{3} w_{ji} x_i + b_j$$

$$h_j = \sigma(dot_j = \max(0, dot_j)$$

The equation above is an example of a forward pass in the first architecture (see architecture image above) that uses ReLU as an activation function. Where i is the node at the input layer (3 input nodes), j is the node at the hidden layer while h is the output node at the hidden layer [17].

```
6/6 [==============================] - 0s 8ms/step - loss: 0.0332 - accuracy: 1.0000
Epoch 195/200
6/6 [==============================] - 0s 6ms/step - loss: 0.0330 - accuracy: 1.0000
Epoch 196/200
6/6 [==============================] - 0s 6ms/step - loss: 0.0325 - accuracy: 1.0000
Epoch 197/200
6/6 [==============================] - 0s 6ms/step - loss: 0.0320 - accuracy: 1.0000
Epoch 198/200
6/6 [==============================] - 0s 6ms/step - loss: 0.0319 - accuracy: 1.0000
Epoch 199/200
6/6 [==============================] - 0s 6ms/step - loss: 0.0315 - accuracy: 1.0000
Epoch 200/200
6/6 [==============================] - 0s 5ms/step - loss: 0.0311 - accuracy: 1.0000
```

Figure 7. Data Trial Compile Results

Judging from the results of the compilation of the data trials, where the data was matched 200 times based on the number of epochs in the syntax code 200 times, resulting in high accuracy data that is worth (1) and the lowest error result that is worth (0.0311). Visualization of the accuracy and error results can be seen in Figures (8) and (9).

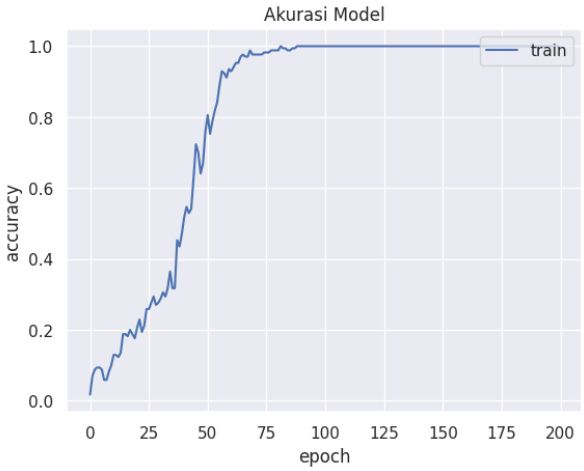

Figure 8. Data Accuracy Results

In figure 8, a graph of the results of database accuracy that has been carried out by various processes, where at the beginning of measuring the database accuracy level experienced instability when matching data for 0-75 times, where the range of accuracy values ranged from 0.1428 – 0.97, and after matching data more than 100 times the range of accuracy values began to stabilize at number 1 (perfect accuracy level).

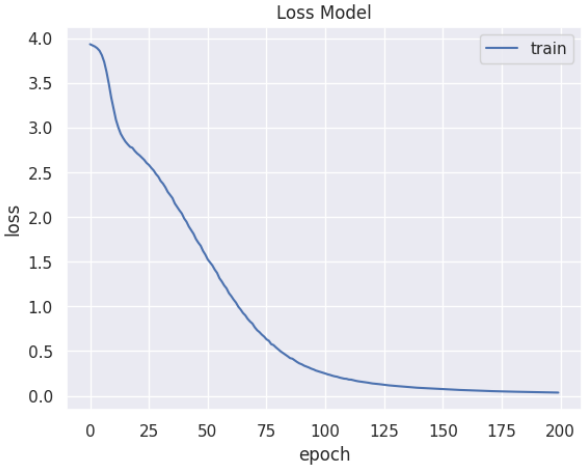

Figure 9. Data Error Results

In figure 9, a graph of the results of database loss that has been carried out by various processes, where at the beginning of the measurement the database loss level experienced instability when matching data for 0-75 times, where the range of loss values ranged from 4 –

0.6, and after matching data more than 100 times the range of accuracy values began to stabilize at 0.25 – 0.0311 (very low loss rate).

Testing the system model and chatbot database produces a fairly high accuracy of 100% and 0% error. The test was carried out by testing 5 questions in English and Indonesian on the bot where the results of the test were that all questions were answered correctly by the bot. The accuracy and error calculation models can be calculated using the calculation formulas (3) and (4).

Accuracy Calculation Formulas

$$Accuracy = \frac{Correct\ number\ of\ tests}{Total\ number\ of\ tests} x\ 100\%$$

$$Accuracy = \frac{10}{10} x\ 100\%$$

$$Accuracy = 100\%$$

Error Calculation Formula

$$Error = \frac{Error\ number\ of\ tests}{Total\ number\ of\ tests} x\ 100\%$$

$$Error = \frac{0}{10} x\ 100\%$$

$$Error = 0\%$$

Evaluation

At this stage the bot will be tested whether it can provide an appropriate response or not based on the data that has been processed. Evaluation results can be seen in table (1).

Table 1. Chatbot Test Results in English

No	User Input	Respon Chatbot	Result
1	Hello	Hello, thanks for visitting	Succed
2	What is batik	Batik is the work of the Indonesian people, which is a blend of art and technology by the ancestors of the Indonesian nation.	Succed
3	What is batik Betawi	Betawi batik is a traditional craft of the people of Jakarta. Its manufacture began in the 19th century.	Succed
4	What is batik ceplok	This batik has a distinctive geometric shape resembling a fried egg. With a combination of roses, stars, and other small objects, the Ceplok Gempol looks beautiful when worn.	Succed
5	What is batik priangan	Priangan Batik is a term used to give identity to the various batiks produced and taking place in Priangan, an area in West Java and Banten where the population speaks and has Sundanese culture.	Succed
6	What is batik ciamis	This typical Ciamis batik motif is known as Sarian batik. Generally, the motif is in the form of a machete with the dominant colors being black and brown.	
7	What is batik cendrawasih	Cendrawasih is a typical Papuan bird that is very iconic. This bird is the inspiration for the Cendrawasih batik motif. Usually this motif combines images of birds of paradise with images of plants and flowers typical of Papua.	Succed
8	What is batik tambal	The tambal motif has the meaning of patching, meaning patching or repairing things that are damaged.	Succed
9	What is batik sidomukti	Sidomukti comes from the word (sido) which means so and (mukti) which means noble and prosperous. The meaning of the Sidomukti batik motif is the hope that the bride and groom will get glory and prosperity in building a household.	Succed
10	Oke good bye	Bye! Comeback again soon	Succed

Flowchart

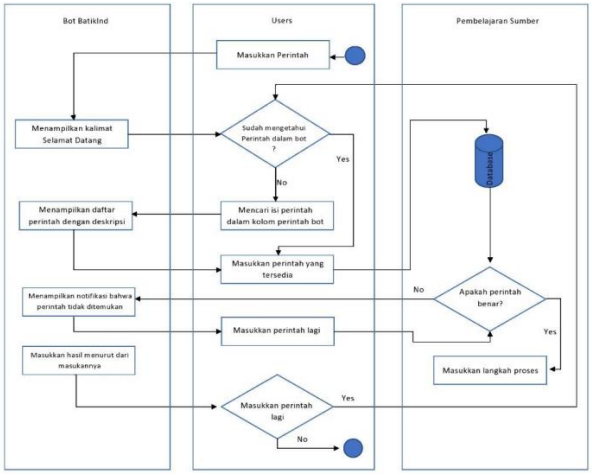

Figure 10. Flow chart Chatbot

In figure 10 is a flow diagram of chatbot work, where there are 3 parts on the flow chart, namely bot, user, source learning. In the flow diagram, it begins with the user to input commands, which then from the input will be processed by the bot, after that the bot will match the input with the database, after the bot finishes matching the data and there is compliance, the bot continues the process and can respond to the user according to user needs.

Deployment

In the chatbot deployment process, architectural design is used through visual studio code, because this design has problems when the Deployment uses design from google colabs, then designing through 2 applications, namely google colabs and visual studio code. Therefore, the Deployment is carried out using the design results from visual studio code using the flask framework to run the web chatbot. The deployment results can be seen in figure (11).

Figure 11. Deployment Results

CONCLUSION

The results obtained from this entire research series, obtained several conclusions regarding the trial and implementation of the chatbot web system using the ANN method. The results of the trial of the chatbot design get high accuracy results, namely 100% and a very low error rate of 0%, so that the implementation of the ANN algorithm can work properly and the bot is able to respond to users. The level of accuracy and error is determined based on the number of data samples made, where the more datasets, the better the results of the trial.

In the deployment process, there were problems with the syntax code in the Google Colabs application, which could not convert the chatbot syntax code into a web file, so a visual studio code application was designed so that the chatbot could be deployed on the website. Then because this research still has shortcomings and there are no other studies that discuss this, my hope is that in the future this research can be developed again and can be continued.

REFERENCES

[1] T. Mildawati, "Teknologi dan Perkembangannya di Indonesia," *Ekuitas,* vol. 4, pp. 103-104, 2020.

[2] APJII, "Laporan Survei Internet APJII 2019-2020," *Asos. Penyelenggaraan Jasa Internet Indonesia,* vol. 1, pp. 1-146, 2020.

[3] Suryanti, "Pelestarian Batik Sebagai Warisan Budaya di Kalangan Siswa SMA Muhammadiyah 2 Surakarta," *Naskah Publikasi,* vol. 1, p. 1, 2013.

[4] S. S. P. R. &. K. D. R. Dsouza R, "Chat With Bots Intelligently : A Critical Review & Analysis," *International Conference on Advance in Computing Communication and Control (ICACS),* pp. 1-6, 2019.

[5] I. Afrianto, M. F. Irfan, and S. Atin, "Aplikasi Chatbot Speak English Media Pembelajaran Bahasa Inggris Berbasis Android," *Komputika J. Sist. Komput.,* vol. 8 , no. 2, pp. 99-109, 2019.

[6] S. P. M. &. H. D. Ondes, "How Chatbots Can Be Involved in The Education Process," *International Conference On Emerging eLearning Technologies and Application ,* pp. 575-580, 2019.

[7] P. P. F. L. M. P. F. &. S. D. Clarizia, "Chatbot : An Education Support System For Student," *International Symposium Education Conference (EDUCON),* pp. 1100-1107, 2019.

[8] J. T. Wibowo, "Proyek Implementasi Teknologi Chatbot Pada Layanan Customer Service," Dokumen Departemen Teknik Elektro Fakultas Teknik, vol. 1, pp. 130-141, 2020.

[9] K. Foote, "A Brief History Of Natural Language Processing (NLP)," 2019.

[10] D. W. Sari, "Implementasi Natural Language Processing pada Chatbot Peribahasa Indonesia," Naskah Jurnal Departemen Ilmu Komputer, vol. 1, pp. 141-148, 2018.

[11] N. B. V. J. C. Vincentius Riandaru Prasetyo, "Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya," Naskah Jurnal Universitas

Surabaya, vol. 1, pp. 114-121, 2021.

[12] F. K. N. H. Feri Mustakim, "Algoritma Jaringan Syaraf Tiruan Pada Chatbot Berbasis Teks Frequently Asked Question (FAQ) Web Kuliah Universitas Nasional," Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), vol. 5, no. 4, pp. 439-447, 2021.

[13] R. Y. E. a. T. N. A. Cucus, "Chatter Bot Untuk Konsultasi Akademik Di Perguruan Tinggi," Explor. J. Sist. Inf. dan Telemat, vol. 10, no. 1, 2019.

[14] e. a. Y. I. Al Barsh, "MPG Prediction Using Artificial Neural Network," International Journal of Academic Information Systems Research (IJAISR), pp. 7-16, 2020.

[15] R. G. d. F. A. D. A. C. Rachman, "Implementasi Text Mining Pengelompokkan Dokumen Skripsi Menggunakan Metode K-Means Clustering," vol. 11, no. 2, pp. 167-174, 2020.

[16] V. S. G. S. J. B. d. B. B. M. Uma, "Formation of SQL from Natural Language Query using NLP," International Conference on Computational Intelligence in Data Science (ICCIDS), pp. 1-5, 2019.

[17] S. Sena, "Pengenalan Deep Learning Part 1 : Neural Network," 28 October 2017. [Online]. Available: https://medium.com. [Accessed 15 Mei 2023].