ASTEEC Conference Proceeding: Computer Science
3rd International Conference on Information Science and Technology Innovation (ICoSTEC)
July 27, 2024, Yogyakarta, Indonesia

Application of Natural Language Processing Technology For Making Agricultural Media Chatbots

Karuniaman Buulolo[1], Rika Rosnelly[2], Zakarias Situmorang[3]

[1,2,]*Magister of Computer Science, Potensi Utama University*
JL. KL. Yos Sudarso Km. 6.5 No. 3-A, Medan Indonesia
[3]*Departement of Computer Science, Katolik Santo Thomas University*
JL. Setia Budi, kp. Tengah, Kec. Medan Tuntungan, Medan, Indonesia
[1]karuniaman12@gmail.com
[2]rika@potensi-utama.ac.id
[3]zakarias65@yahoo.com

Abstract — **Agriculture is the use of resources by humans to produce food raw materials and manage the surrounding environment. With the development of industry today, many fields use artificial intelligence, including agriculture. The aim of the research carried out is to help farmers find out some important information related to agriculture based on current relevance. Implementation of the application of Natural Language Processing (NLP) using the Dialogflow Platform to design and integrate conversational user interfaces for chatbot web applications. Chatbots have the ability to answer questions and provide solutions more quickly and efficiently compared to human-to-human interaction. The contribution of this research is that Indonesian agriculture is expected to be able to implement chatbots to improve the quality of information services and for users to make it easier to obtain information related to Indonesian agriculture. Based on the results of the evaluation of model analysis with confusion matrix which was carried out by testing the reliability method, the results of an accuracy level of 85% were obtained from the correct number of test data divided by the total dataset, recall was 100%, precision was 85% and F-measure was 57.14%. This shows that the Chatbot is very feasible and effective to use.**

Keywords — **Natural Language Processing, Chatbot, Agricultural Media, Confusion Matrix, Web System**

I. INTRODUCTION

Artificial intelligence is a form of technology that is experiencing very rapid development in this modern era. Artificial intelligence allows machines to think and make their own decisions, one of which is chatbot technology [1]. Chatbot or conversation with a robot is an artificial intelligence application that can simulate intelligent conversations between humans based on the knowledge provided [2]. Chatbots are intelligent agents that can imitate humans' ability to communicate via text messages [3]. Agriculture is the use of biological resources carried out by humans to produce food, industrial raw materials, energy sources, and manage the surrounding environment [4]. The majority of the Indonesian population makes their living in the agricultural sector, since the colonial era until now it cannot be separated from the agricultural and plantation sectors because these sectors have a very important meaning in determining the social economic reality of society in various regions of Indonesia [5].

One form of application of artificial intelligence is chat robots or chatbots, a technology that processes input in the form of text and then obtains keywords to provide answers or responses. Then, to keep the conversation going continuously, it is necessary to design a dialogue system [6]. The chatbot trend is increasing rapidly in number because of the ease and more benefits of using the software [7]. Natural language processing (NLP) has now been developed in various fields, including agriculture. The development of Artificial Intelligence in the agricultural sector is quite rapid, starting from planning, planting to maintenance, all of which are usually carried out by artificial intelligence [8]. In agriculture and plantations, farmers often experience obstacles, namely the lack of knowledge of farmers to recognize events or the correct treatment of the plants being planted [9]. Natural Language Processing or often abbreviated as NLP is a field of computer science, artificial intelligence and language (linguistics) which is related to the interaction between computers and natural human language. Natural language is a language that humans can understand [10].

[11] conducted research using an agricultural application entitled AgronomoBot. a smart answering Ohatbot applied to agricultural sensor networks Environmental control and automation for swine housing View project. This research produces applications that provide convenience and practicality in general communication. Emerging applications bring automatic, continuous and intelligent features for communication via messaging applications using web robots called Ohatbots[12]. For agricultural purposes, it is important that data about field conditions, such as air and soil temperature, air relative humidity, soil moisture, rainfall, wind speed and other related variables, is quickly and easily available for use by agricultural management systems, by specialists, or farmers

itself in the decision-making process. Chatbot applications that use Natural Language Processing (NLP) are able to simulate human-like conversations that can provide responses to a question [13].

Chatbots have human-like capabilities that can answer questions asked by users, so that many conveniences will be obtained in dealing with agricultural problems [14]. Artificial Intelligence capabilities that provide virtual intelligence to the chatbot enable it to assess the correct answer to questions asked by the user [15]. So in this research we will apply the use of NLP in the agricultural sector which can be used by farmers to find out any information about agriculture from care to harvest. This research combines the use of NLP and the time relevance of the questions asked by using a combination of error and rule based methods so as to get answers that match the questions with the appropriate validity period.

It is hoped that this research can help farmers by providing information about various kinds of problems in agriculture. In this research, answering user questions is based on the relevance of the applicable time. So the results of this research can be used for a long period of time, as long as the information/questions and answers relevant to agriculture are always updated at all times.

II. Research Methods

The flow of research carried out in this research is depicted in Figure 1, On study This explain Suite test try And evaluation performance to research conducted . Based on say key information provided into the existing system database through preprocessing stage . Following stage process implementation application Natural Language Processing algorithm for manage input and Naïve Bayes Classifier as classification question user to in category / class on chatbots.

In application implementation system as well as testing based on scenario that has been designed . Furthermore testing that has been done obtained given discussion And analysis from every tests carried out with objective For get results from study.

Implementation Natural Laguage Processing algorithm and Naïve Bayes Classifier based Chatbots can processing questions asked by user into the Language natural , and look for say relevant key with desired information user:

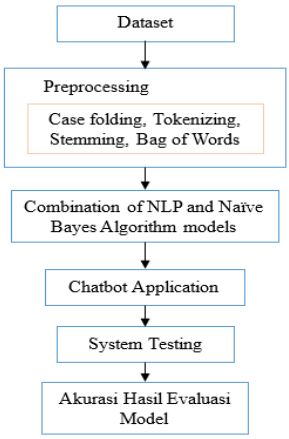

Fig. 1 Stages of research methodology

From Fig.1, the research stages that will be carried out in each process are explained in the following sub-chapters.

A. *Data Collection Stage*

The data collection stage was carried out to obtain the information needed to achieve the research objectives. In this research, the dataset was obtained from a questionnaire form, the aim was to find out what questions are usually asked by users related to agriculture. A collection of data that has been obtained is then entered manually and saved in a file in JSON (JavaScript Object Notation) format. This dataset has a structure including:

1. Intents, a collection of input and output data to train the chatbot.
2. Tags, to group similar text data as targeted output for neural network training
3. Input, this section contains keyword data or input patterns that the user wants
4. Responses, this section contains output pattern data generated by the chatbot and will be sent to the user.

B. *Input Dataset Preprocessing Stage*

At the preprocessing stage, the Natural Language Processing approach is used with several process stages as follows:

1. Case Folding

   The Case Folding stage is the process of changing all the letters in the sentence given by the user to lowercase. In the case folding process, use the lower function to change all letters to lower case. The case folding results can be seen in Table 1 below:

TABLE 1
CASE FOLDING

Question	Case Folding Results
How do you apply fertilizer to RICE plants?	how to apply fertilizer to rice plants

Table.1 explains examples of input provided by users with capital letters. Then after carrying out this step, the input changes to lowercase.

2. Tokenizing

   The tokenizing stage is breaking down the sentence into its constituent words. With input provided by the user, the input is then described based on the words that make it up. Table.2 shows how the input is decomposed into its constituent words:

TABLE 2
CASE FOLDING

Question	Tokenizing Results
how to apply fertilizer to rice plants	"how", "way" "giving", "fertilizer", "on", "plants", "rice", "?"

Table 2 explains the results of the case folding stage to the tokenizing stage after the input has been changed

to lower case. Then in the tokeninzing stage, the input is broken down into the words that make it up.

3. Stemming

Stemming is the process of removing affixes from words to produce basic words. Basically, in a base word there is more than one affix, therefore the process of eliminating the correct affix is needed so that the base word can be found. The following is an example of the results of the stemming process.

TABLE 3
CASE FOLDING

Question	Score
"how", "way" "giving", "fertilizer", "on", "plants", "rice", "?"	how method giving pupul on plant paddy ?

4. Bag of Words

Bag of Words is the process of counting the occurrences of each word. This process begins by saving training data into the system, which will then be given a value of 1 if the word is in the word bag and a value of 0 if it is not in the word bag. The following is an example of the Bag of Words process:

TABLE 4
BAG OF WORDS

Say	Bag of Words						
	How	method	giving	fertilizer	on	plant	paddy
How	1	0	0	0	0	0	0
method	0	1	0	0	0	0	0
giving	0	0	0	0	0	1	0
fertilizer	0	0	1	0	0	0	0
on	0	0	0	1	0	0	0
plant	0	0	0	0	0	0	1
paddy	0	0	0	0	1	0	0

C. NLP Model Analysis Stage

System model analysis is used to study the interaction between computers and human (natural) language, at the model analysis stage with NLP it is described to describe the system architecture method of flow design which is used as a solution to overcome problems. The following is a description of the system architecture used to apply the NLP method:

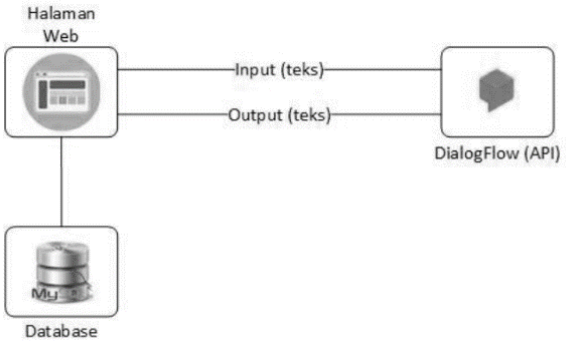

Fig. 2 System architecture

From Fig.2 above, it shows the stages of the chatbot application system implemented in designing the system being created. The analysis stage is in the form of coding in accordance with the results of the software design that has been created.

The resulting program is a chatbot that can be accessed via the website. The program will then be connected to the Dialogflow API using the Firebase server as a coordinator in capturing end-user expressions (text input) and providing responses from Dialogflow.

D. System Chatboth GUI

User interface design on chatbot which is processed with natural language, text input is processed to extract keywords.

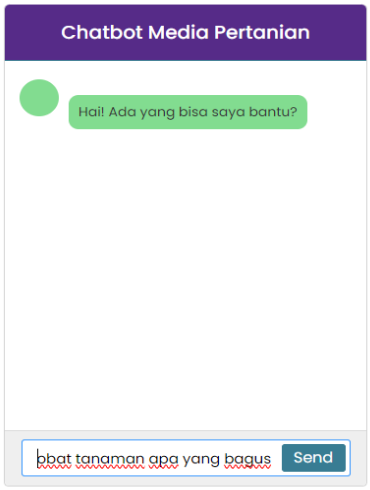

Fig. 3 Chatbot gui

From Fig.3 The algorithm for checking sentence similarity (NLP) in chatbots is applied to the modified input to check its similarity to questions from a predefined set of questions, the answers to which are available.

Based on keywords, the information required by the user is understood and the information is provided from the database.

Fig. 4 Implementation of system database

E. *Testing Phase*

System testing is used as a step to ensure the quality of the system that has been built and check the harmony between the system components to be implemented.

TABLE 5
FUNCTIONAL TESTING BASED ON TEMPLATES

No.	Inputs	Results
1.	how to control tungro disease in rice	planting simultaneously and setting planting times
2.	the types of diseases that most often attack rice	bacterial leaf blight, stem rot, tungro disease, leaf spot disease
3.	how to increase plant productivity naturally	One natural way to increase plant productivity is to use organic fertilizer such as compost and manure, as well as implementing sustainable agricultural practices
4.	what types of fertilizer are good	compost, manure, guano fertilizer etc
5.	what are the benefits of drip irrigation systems in agriculture	Drip irrigation systems can increase water use efficiency, reduce the risk of excess water, and deliver nutrients directly to plant roots, thereby increasing crop yields

F. *Visualization Analysis Results*

At this stage, system evaluation is carried out directly to users using a questionnaire or survey form regarding user satisfaction in using this chatbot website:

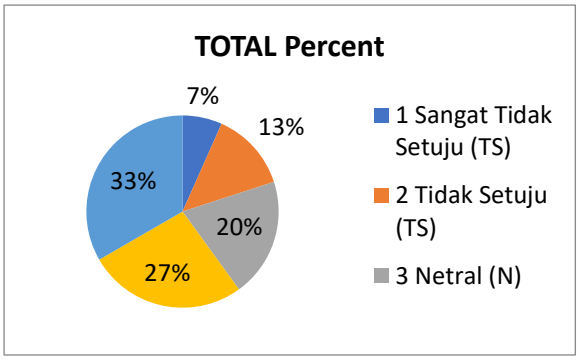

Fig. 5 Test diagram

The diagram above is a visualization of the total analysis results of the answers consisting of 6 questions used by the user.

III. RESULTS AND DISCUSSIONS

Based on keywords, the information provided into the system database has gone through the preprocessing stage. The results of system implementation and testing are based on the scenarios that have been designed. Next, the tests that have been obtained are given a discussion and analysis of each test carried out with the aim of obtaining results from the research.

A. *Natural Language Processing Model Analysis Results*

The NLP model is described from a graphical representation between elements in the text, the following is an information model of entities extracted from questions and answers using a system diagram used in this data built with an entity histogram:

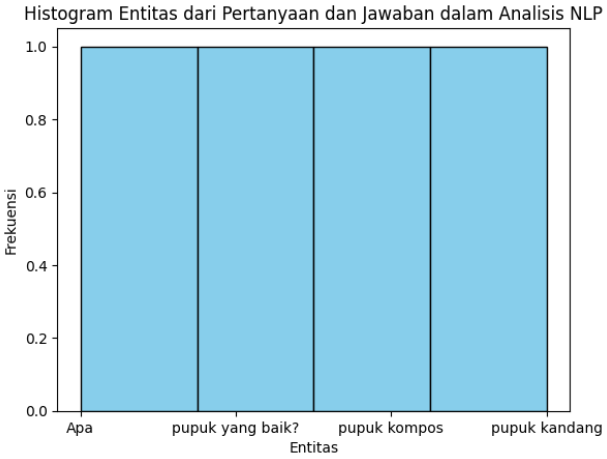

Fig . 6 NLP model analysis results

The visualization output provides an overview of the types of entities by the model in the context of questions and answers. The NLP model uses the " *en_core_web_sm* " model from spaCy, Histogram shows the frequency distribution of entities extracted from questions and answers in the NLP analysis text. Each point on the line represents the length of the entity, and

the labels on the x-axis indicate the origin of the entity (Question or Answer).

B. Chatbot Application Implementation Results

In this research, after creating the process results of the NLP model according to the needs, then implementing the analysis and design carried out, the results of the application being built can be seen as in Fig.7

Fig.7 Chatbot user interface design and implementation

In Figure 7, based on the results of the chat that was built, the initial process is where the user first starts the chatbot or gives a greeting to the chatbot, then the 'bot' will respond according to the training *phrases* . For example: 'Hi' or 'hello. The system is able to provide answers based on input from the user. Chatbot development using natural language processing methods has produced good results for agricultural media. The system can respond and answer questions from users. The following is the appearance of the answer from the 'bot' when the input from the user does not match what is in the system:

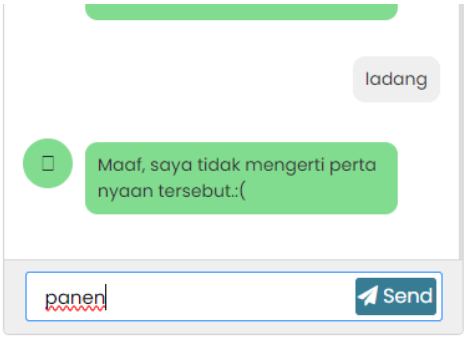

Fig.8 Display of chatbot response to user question (failed)

In Figure 8, the chatbot system built in this research provides a response when a typo or incorrect question occurs regarding agricultural information in the input entered by the user.

C. System Test Results

In the process after implementation, the chatbot is tested using previously labeled data, to find out to what extent it can answer questions accurately. So, system testing is carried out using the reliability testing method to assess the effectiveness and feasibility of a system that has been built.

Testing is carried out by seeing whether the questions asked by users get answers as expected. The calculations for this test use the following formula:

$$\text{Nilai Akurasi} = \frac{\textit{Jumlah jawaban benar}}{\textit{Jumlah pertanyaan}} \text{ x } 100\%$$

TABLE 6
RELIABILITY TEST RESULTS

No.	Question	Answer	Results
1.	what is organic fertilizer	Organic fertilizer is fertilizer that comes from natural ingredients such as compost, manure and plant waste	True
2.	how to deal with plant pests naturally	use natural predators such as predatory insects or apply organic solutions such as neem oil	True
3.	when is the right time to water the plants	Watering should be done in the morning or evening to avoid rapid evaporation of water	True
4.	what are the advantages of using a drip irrigation system	Drip irrigation systems increase water use efficiency and deliver nutrients directly to plant roots	True
5.	how to recognize plants that lack nutrients	Symptoms such as leaf discoloration or stunted growth can be signs of nutrient deficiencies	True
6.	what are the benefits of crop rotation in agriculture	Crop rotation helps reduce the risk of disease and pests and maintains soil fertility	True
7.	how to start organic farming at home	start by making compost from kitchen scraps and plants, and use organic fertilizer	True
8.	what types of plants are suitable for planting vertically	crops such as tomatoes, lettuce, and strawberries are suitable for vertical farming	False

9.	how to prevent soil erosion	use ground covers, such as mulch, and plant cover crops	True
10.	what is the role of bees in agriculture	Bees act as plant pollinators, helping in the production of fruit and seeds	False
11.	how to choose plant varieties that suit the local climate	choose varieties that are resistant to local climatic conditions and suitable for the growing season	True
12.	What are the benefits of ground cover plants?	Cover crops help prevent erosion, improve soil health, and control weeds	False
13.	how to make organic pesticide at home	Mix soapy water, neem oil and water, then spray on plants as an organic pesticide	False
14.	when is the right time to prune plants	pruning should be done when the plant is not actively growing, such as in winter	False
15.	what is the difference between organic fertilizer and chemical fertilizer	Organic fertilizer comes from natural ingredients, while chemical fertilizer is made synthetically	True
16.	how to make a compost bin at home	use kitchen scraps and green-brown materials, such as leaves, to create a compost bin	True
17.	what is the role of earthworms in agriculture	Earthworms help improve soil structure and make it more fertile	True
18.	how to manage rainwater for agriculture	use rainwater collection systems and irrigation canals to manage water efficiently	True
19.	what is meant by sustainable agriculture	Sustainable agriculture is agricultural practices that pay attention to environmental, social and economic sustainability.	True
20.	how to recognize invasive plants	Invasive plants usually grow quickly, spread widely, and take resources from surrounding plants	false

From the reliability results in table 6, it can be seen that of the 20 testing intents that were tested, there were 3 (False) testing failures that occurred, the expected answers did not match the previously labeled questions, while 17 (True) testing questions in the dataset successful and according to the expected answer. Thus, the level of consistent and accurate response to the chatbot in this study obtained a success accuracy level of:

$$Success\ Accuracy = \frac{17}{20} \text{ x } 100\% = 85\%$$

So, the test results have a success accuracy rate of 85%. Shows that Chatbot is very feasible and effective to use.

D. Visualization Results of Model Analysis Evaluation

The evaluation was carried out to observe the model performance of the algorithm implemented from system testing data carried out using the reliability testing method of 20 test data, which was calculated using accuracy, precision, recall, f-measure, described in the form of a confusion matrix as follows:

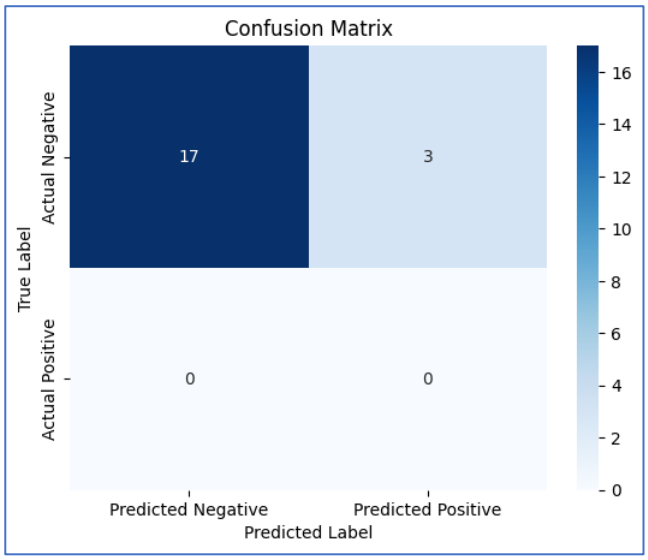

Fig .9 Confucin matrix visualization

From the results of the confusion matrix, an evaluation of the performance of the classification model on test data is produced, where the values produced by the model are compared with the actual values. The results of the confusion matrix are True Positive (TP) 17, True Negative (TN) 0, False Positive (FP) 3 and False Negative (FN): 0. The following are the results of the confusion matrix calculation, namely:

$$Accuracy = \frac{17}{20} \text{ x } 100\% = 85\%$$

$$\text{Recall} = \frac{17}{17+0} \text{ x } 100\% = 100\%$$

$$\text{Precision} = \frac{17}{17+3} \text{ x } 100\% = 85\%$$

$$\text{F-measure} = \frac{2*1*0{,}4}{1+0{,}4} \text{ x } 100\% = 57{,}14\%$$

Based on the results of the evaluation of model analysis with confusion matrix which was carried out using reliability testing, accuracy of 85% was obtained from the correct number of test data divided by the total dataset, recall was 100%, precision was 85% and F-measure was 57.14%.

IV. CONCLUSIONS

Based on the results of implementation and testing that have been carried out on chatbot applications using a natural language processing approach, the following conclusions can be drawn:

1. The average results of accuracy, precision, recall and f-measure provide a better picture of the performance of the NLP application model in agricultural media chatbot systems
2. Chatbot applications can provide information about needs and solutions to problems in agriculture, so that information search can be more efficient and effective for the community.
3. From the test results of the implementation of the natural language processing model in the agricultural media chatbot application, the conversation that occurs is like a user and an admin carried out via internet access so that there is no limit to time, place and where the information is needed.
4. The application of chatbots that use machine learning and are based on Natural Language Processing (NLP) in public services can provide major benefits, including efficiency in costs and time, increased productivity, and improved quality of services for community needs, especially for farmers.

REFERENCES

[1] A. Ahmad, "Getting to Know Artificial Intelligence, Machine Learning, Neural Networks, and Deep Learning," Islamic Light Foundation, 22 June 2017.

[2] A. Tasyah, D. Rahyu, DB Utomo, D. Fadillah, FF Nuruladhi, I. Ardiansyah, LMI Afriansyah, RD Yudhistira, S. Marfirah and SO Hariansyah, "Searching for Bali Regional Tourism Information Using Chatbot Technology," vol . 8, 2017.

[3] AD R, F. Imamah, YM Andre and A. , "Chatbot Application (Milki A. Tasyah, D. Rahyu, DB Utomo, D. Fadillah, FF Nuruladhi, I. Ardiansyah, LMI Afriansyah, RD Yudhistira, S Marfirah and SO Hariansyah, "Searching for Bali Regional Tourism Information Using Chatbot Technology," vol. 8, 2017.

[4] A. Ahmad, "Getting to Know Artificial Intelligence, Machine Learning, Neural Networks, and Deep Learning," Islamic Light Foundation, 22 June 2017.

[5] AD R, F. Imamah, YM Andre and A., "Chatbot Application (Milki Bot) Integrated with Web CMS for Customer Service in Small and Medium Enterprises," Cendikia Journal, vol. XVI, 2018.

[6] A. Ahmad, "Getting to Know Artificial Intelligence, Machine Learning, Neural Networks, and Deep Learning," Islamic Light Foundation, 22 June 2017.

[7] IMSIKGDP I Nyoman Satria Paliwahet, "Searching for Bali Regional Tourism Information Using Chatbot Technology," vol. 8, 2017.

[8] A. Sahi, "Online Web-Based LP3I Entrance Screening Selection Academic Potential Test Application using the CodeIgniter Framework," Communication Information Technology, vol. 7, 2020.

[9] F. Mustakim, F. and N. Hayati, "Artificial Neural Network Algorithm on Text-based Chatbot Frequently Asked Questions (FAQ) on the National University Lecture Web," Information and Communication Technology, vol. 5, 2021.

[10] F. Yusup, "Validity and Reliability Test of Quantitative Research Instruments," vol. 7, 2018.

[11] G. Gandi, MS Mustofa, and A. Luthfi, "Farmer Social Networks in the Ijon System in Agriculture in Pagenteran Village, Pulosari District, Pemalang Regency," Solidarity, vol. 6, no. 1, pp. 1—10,2017.

[12] H. Hani, DB Prasetyo and HC Rustamadji, "Automation of the Frequently Ask Questions Service Based on Natural Language Processing on Telegram Bot," Telematics, vol. 17, 2020.

[13] IMSIKGDP I Nyoman Satria Paliwahet, "Searching for Bali Regional Tourism Information Using Chatbot Technology," vol. 8, 2017.

[14] NT Anggraini, "Initiation of Natural Language Processing (Nlp) and Classification of Questions for Making 'Statbot' (StatistikaChatterBot," J.Ohem.Inf.Model.,vol.53,no.9,pp. 1689—1699,2019.

[15] N. . A. Purwitasari and M. Soleh, "Implementation of the Artificial Neural Network Algorithm in Creating Chatbots Using a Natural Language Processing Approach," Science and Technology, 2022.