The Covid-19 Chatbot Application Using A Natural Language Processing Approach

Cahyo Prianto [1], *Nisa Hanum Harani* [2]

[1,2]*Politeknik Pos Indonesia, Bandung, Indonesia*

[*]*cahyoprianto@poltekpos.ac.id*

Abstract

Cases exposed to the Covid-19 virus in Indonesia until June 2021 continue to experience a spike in increases, to handle it, various government policies continue to be rolled out and the public needs to be given correct, precise and fast information so that mutual awareness can be built to suppress cases exposed to COVID-19. With this background, this study aims to design and build a COVID-19 chatbot system based on artificial intelligence based on the Natural Language Processing algorithm. This chatbot is expected to be a place to ask questions about all things related to covid-19 so that it can become a personal assistant with two-way communication that can be accessed quickly for 24 hours. This chatbot system was built using the Python programming language, Node.js server and MariaDB as the database. As a client, this chatbot is integrated with the popular instant messaging application in Indonesia, namely WhatsApp. The data set used to train the chatbot was 369 question data and spread into 46 question tags. Testing the chatbot system using blackbox testing, and to test the expected output, the chatbot was tested using 350 testing data and the accuracy rate of the chatbot in answering reached 54%.

Keywords: Covid-19, Whatsapp, Natural Language Processing

1. Introduction

Covid-19 is a virus that first appeared in the city of Wuhan, China in December 2019 and continues to spread to various countries around the world, including Indonesia. On March 11, 2020, WHO has declared COVID-19 as a pandemic [1]. Based on data on May 17, 2021, which was taken from the worldometer, the total number of infected COVID-19 cases worldwide has reached 163,694,333 and of that number, 3,392,634 people died. The countries that contributed the highest number of cases were the United States, India and then Brazil [2]. In Indonesia, based on data from covid19.go.id on July 13, 2021, there were 2,567,630 confirmed cases of COVID-19.

To prevent the spread of the COVID-19 virus more widely, many policies have been rolled out by the central government and local governments, such as tightening entry and exit of several areas, socializing the 5M recommendation, banning going home during the Eid al-Fitr 1442H holiday, policies to enforce restrictions on community activities (PPKM), and others. The important thing about these policies and recommendations is that they can be socialized and conveyed to the wider community. Information related to covid-19 itself is very important for the public to know, what is covid-19, how is the prevention process, how to self-isolate, what are the symptoms of covid-19, information on referral hospitals and various other questions whose answers people need to know quickly.

Currently, various information related to covid-19 is widely spread through various methods such as the distribution of posters, banners, mass media, electronic media, social media, to various websites that contain information about covid-19, updated and precise information is very helpful for the community in respond to activities to prevent and treat cases of the COVID-19 virus. However, there is a problem where the information cannot
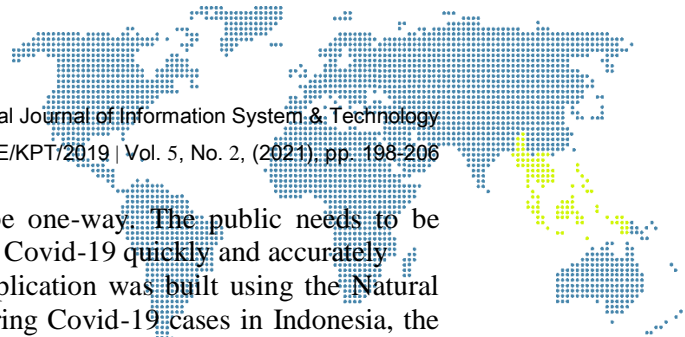
always be accessed easily, quickly and tends to be one-way. The public needs to be facilitated by a media to ask various things related to Covid-19 quickly and accurately

To answer this problem, a Covid-19 chatbot application was built using the Natural Language Processing approach. In the midst of soaring Covid-19 cases in Indonesia, the urgency is that this chatbot application can act as a personal assistant for the public who can be used to ask various questions related to this covid-19 and can be accessed for 24 hours.

Chatbot is an automatic answering system that can help provide information based on questions from users through text messages[3]. The chatbot application is built using the Natural Language Processing (NLP) approach which is part of the branch of artificial intelligence (AI) that focuses on natural language processing[4]-[7]. This chatbot is implemented in a Whatsapp application, WhatsApp is an instant messaging application platform that is popular in many countries, currently WhatsApp users have penetrated as many as two billion users [8]. In Indonesia alone whatsapp is a very popular application as recorded in the 2019 digital report data from "we are social hootsuite" it was recorded that 83% of internet users in Indonesia are whatsapp users, this number makes Indonesia one of the countries with the most users in the world. [9]. The Indonesian government through the Ministry of Communication and Information on March 19, 2020 has launched the whatsapp covid-19 chatbot via the number wa.me/6281133399000, which in less timeover 2 weeks have sent a total of 46 million messages to nearly 2.3 million users [10]. The Kominfo chatbot uses a pattern matching approach where the user will use a certain pattern of rules then the chatbot will choose a predetermined answer. While in this research to be carried out, the chatbot will be developed in Indonesian using a natural language processing algorithm approach where users can ask questions freely without pattern limitation to follow

Making a chatbot application that is applied to the WhatsApp application will really help many people to get information quickly. With this background, in this study the author takes the research title "The Covid-19 Chatbot Application using the Natural Language Processing approach".
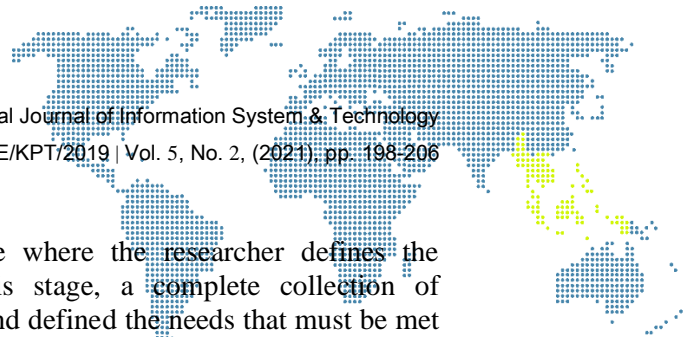
2. Research Methodology

The research method used in this study is divided into several stages so that the process carried out is more focused. In general, the research steps taken to create a work order are as follows in figure 1 [11]-[13]:
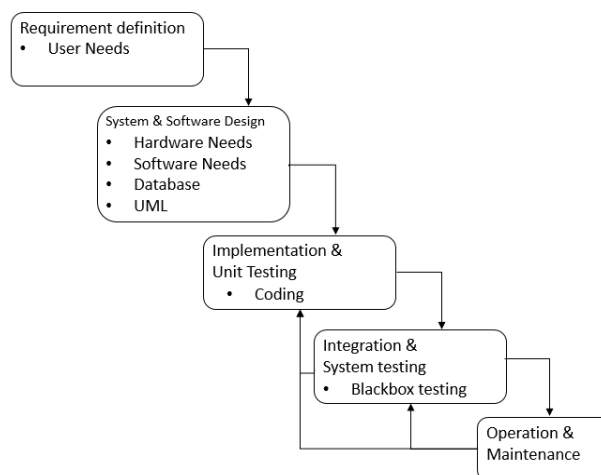
Figure 1. Research Methodology Flowchart

2.1. Requirement Definition

The requirements definition stage is the stage where the researcher defines the requirements in building the system[14]. At this stage, a complete collection of requirements is carried out which is then analyzed and defined the needs that must be met in the program. Data collection at this stage can be done through an interview, as well as a literature study. Researchers collect data used starting from data outlets, machines, types of services, and so on through interviews conducted by the office to clients and also through literature studies in the form of journals. The data is then analyzed to produce a user requirement document. This document will then be used as a reference by researchers in designing system requirements.

2.2. System dan Software Design

The system and software design stage is the stage where researchers develop processes, data, and relationships between data to be able to carry out business processes and meet needs according to the results of the needs analysis that has been done previously. The system and software design stage focuses on data structures, software architecture, interface representation, and procedural details (algorithms).
The researcher designed the database as the data structure, the interface as a liaison between the system and the user, and UML as the system modeling. This stage will produce software requirements documents and hardware requirements documents as a reference for building or creating a system

2.3. Implementation and User Testing

The implementation and unit testing stage is the implementation of making a system design that has been made into a complete application form and can run as needed. Researchers make functions in python by coding and making chatbots according to the results of the design that has been done previously by combining the python programming language and nodejs[14],[15].

2.4. Integration and System Testing

The integration and system testing stage is the stage where the researcher integrates the application and tests the system that has been made in order to get good and optimal results. Researchers tested the system using a specification-based blackbox approach. System test results will be obtained through the output and input given for existing functions regardless of how the process is to get the output. If the input and output are in accordance with what was designed, then the result is that the system has been running as needed.

2.5 Operation and Maintenance

The operation and maintenance stage is the stage where the researcher operates the system, and performs system maintenance related to hardware, software, and other media. System maintenance is needed so that system performance remains stable in use.

3. Results and Discussion

3.1. Ongoing System Analysis

Analysis of the system that was running when the user wanted to find information about covid-19. The current system analysis will be explained in the following flowmap flow in figure 2.

3.2. Analysis Of The System To Be Built
3.2.1. Analysis of the procedures to be built

In building the system it is necessary to make a design. The design of this system is intended so that in the formation of the system a system can be produced that functions
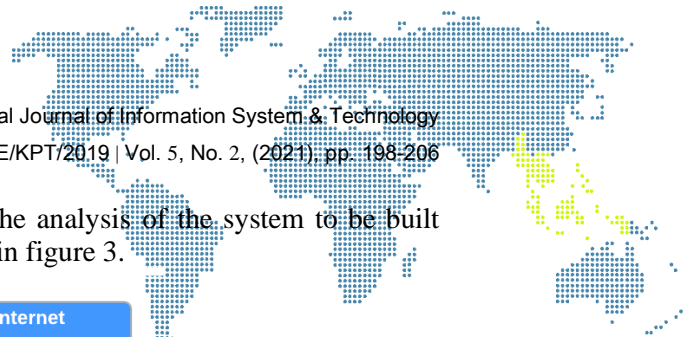
optimally and provides convenience for its users. The analysis of the system to be built will be explained in detail in the following flowmap in figure 3.
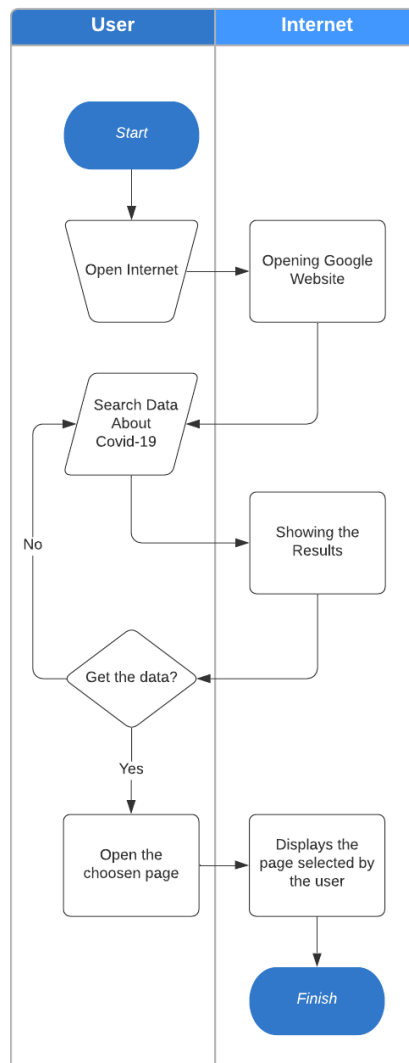

Figure 2. Ongoing System Analysis Flowmap

3.2.2. Use Case Diagram

Use case diagram is a construction to describe the relationships that occur between actors and activities contained in the system. The target of use case modeling is to define functional and operational requirements of the system by defining usage scenarios that are agreed upon between the user and the developer. From the analysis of existing application users, the use case diagram for the Covid-19 Information Chatbot System uses Natural Language Processing. Use case diagram can be seen in figure 4.
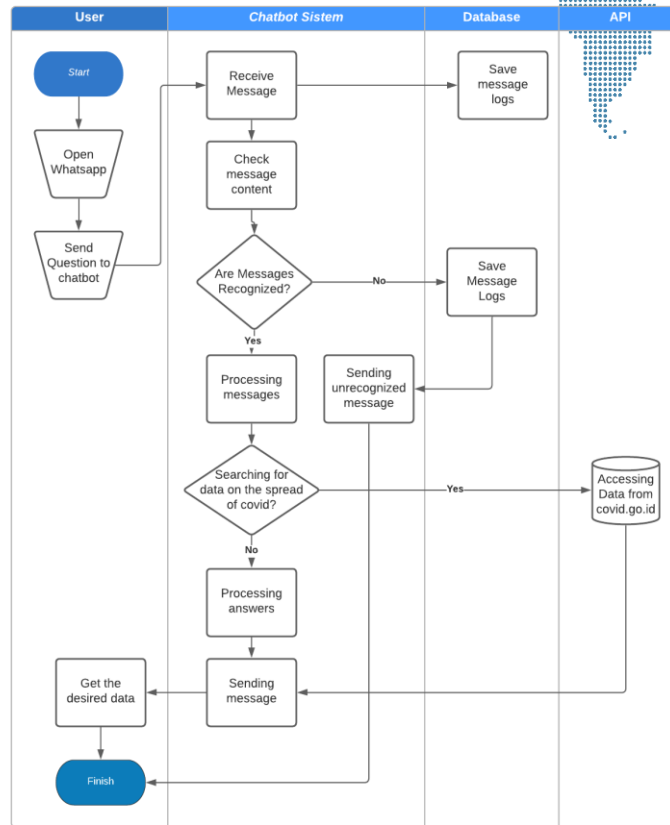
Figure 3. Flowmap Of The System To Be Built
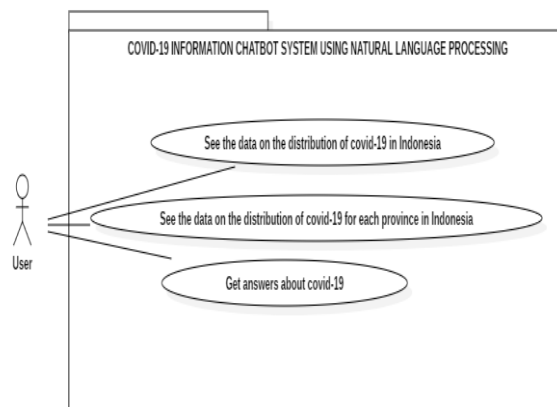
Figure 4. Use Case Diagram

3.3. System Software and Hardware Architectural Design
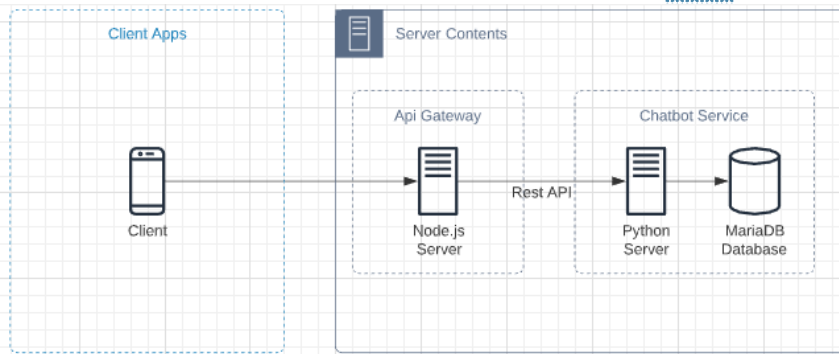### 3.3.1. Software Architecture

Software architecture in figure 5 is how each software is connected to each other and the software architecture in this system is:

1) Client: That is the whatsapp application that will be used, and chatbots can be contacted via the following number wa.me/6281322206359
2) Servers:
   (a) Node.js server which will run the nodejs system which will later be connected connecting the system to the whatsapp system
   (b) Python server that will run a system python that takes care of every message that comes to the chatbot and provides the answer

(c) MariaDB which functions as a database that stores messages received by the chatbot
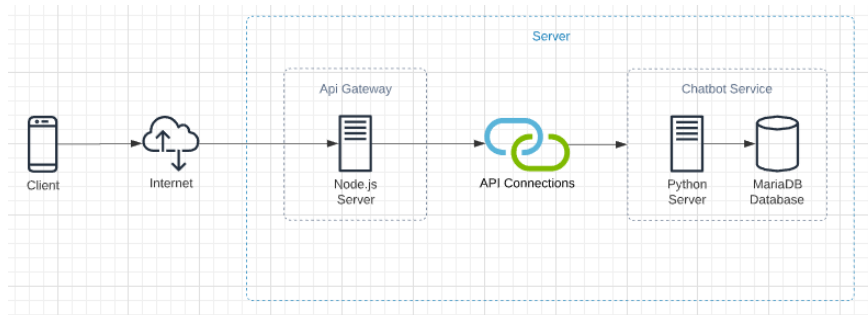

Figure 5. Hardware Architecture

3.3.2. Hardware Architecture

Hardware architecture in figure 6 is how the hardware is interconnected in this system. The following is an explanation of the hardware used in this system:

1) Client: Client is a user who uses the whatsapp application on a smartphone such as android or iphone
2) Internet: Every hardware device is connected to the internet in order to communicate with each other
3) Server: Server is the place where this system will be run which contains the nodejs, mariadb and python system applications that have been created in this research


Figure 1. Hardware Architecture

3.4. Implementation of Natural Language Processing

The implementation of Natural Language Processing is carried out on a python-based system using the Natural Language Tool Kit (NLTK) Library to create text, sastrawi to process words served by the natural language toolkit for Indonesian and pytorch to perform model training. In its implementation, the NLTK Library will process incoming sentences from the chatbot into several initial steps of text such as tokenization, stemming, and bag of words for the filtering process. This step is shown in table 1 below.

Table 1. Text Processing

No	Example of sentences	Tokenisasi	Filtering	Stemming
1	Bagaimana cara pencegahan covid 19	Bagaimana	-	-
		Cara	Cara	Cara
		Pencegahan	Pencegahan	Cegah
		covid	Covid	Covid
		19	-	-
2	Berapa jumlah kematian akibat covid 19	Berapa	-	-
		Jumlah	Jumlah	Jumlah
		Kematian	Kematian	Mati
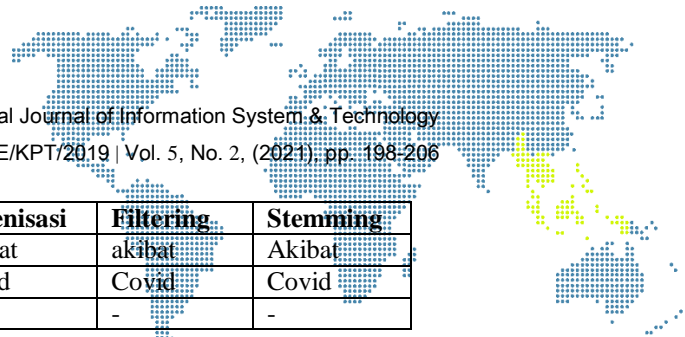
No	Example of sentences	Tokenisasi	Filtering	Stemming
		Akibat	akibat	Akibat
		Covid	Covid	Covid
		19	-	-

After making the process to do text mining, then create json data that will be used for training data by creating questions according to the specified tag. This step can be seen in table 2 below. The training data used in this study amounted to 369 data questions that can be answered by the chatbot and spread into 46 tags, some of which can be seen in table 3.

Table 2. Training Data

No	Tag Type	Question
1	Perkenalan Bot	Data apa saja yang uta saya minta ?
		Apa kegunaan bot ini ?
		Nama kamu siapa ?
		Siapa yang membuat kamu ?
2	Gejala Covid	Apa saja gejala covid-19 ?
		Gejala seseorang covid-19 ?
		Gejala covid ?
		Ciri seseorang covid-19 ?

Table 3. Example of Tag Data

Tag	Total
Perkenalan bot	12
akibat-covid	3
akurasi-vaksin	1
Herd immunity (kekebalan-kelompok)	3
Isolasi mandiri	2
Vaksin	2
Gejala covid-19	24
Coronavirus dan covid-19	14
Imunitas tubuh	3
Terinfeksi covid-19	5
data-covid-meninggal	18
data-covid-positif	13
data-covid-sembuh	13
genose-tes	2
utase-covid	4
perkenalan-bot	9
proses-vaksin	3
rapid-tes	2
rumah-sakit-rujukan	6
salam	1
seberapa-bahayanya-covid-19-ini	7
siapa-saja-yang-berisiko-terinfeksi-covid-19	5
swab-tes	1

3.5. Test and Test Results
3.5.1. Test

The test is carried out in two stages, the first is blackbox testing on the system built by looking at the response from the chatbot, and the second is measuring the accuracy of the answers given to questions from the user.
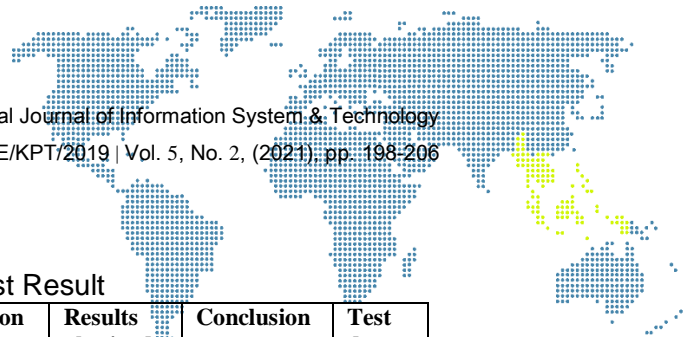
3.5.2. Blackbox Test Results

Table 4. Blackbox Test Result

Tested Class	Test Procedure	Input	Expected Output	Evaluation criteria	Results obtained	Conclusion	Test date
View Indonesian distribution data	Send a message to the chatbot	Keyword "minta data covid"	Chatbot replies to messages, with data on the distribution of covid-19 in Indonesia	The chatbot replies to messages with the contents of the data on the distribution of covid-19 in Indonesia to users	In accordance with the expected output	Accepted	04/6/2021
Send inquiry		Keyword "nama kamu siapa"	Chatbots reply to messages, explain themselves and their duties	The chatbot replies to messages to the user what he can do	In accordance with the expected output	Accepted	04/6/2021

3.5.2. Chatbot Answer Accuracy Test Results

To find out whether the system that has been built can correctly answer the questions given, then testing is needed to find out. Testing using 350 data sets of questions, some of which are as shown in table 5, this test produces a chatbot accuracy rate of 54%.

4. Conclusion

Based on the results of the analysis, implementation, and test results, it can be concluded as follows:

a) This research has succeeded in building a Covid-19 chatbot information system that is implemented in the WhatsApp application. The chatbots can be contacted via the following number wa.me/6281322206359

b) The Covid-19 Chatbot information system has been successfully built using the Natural language Processing approach with an accuracy of answering questions correctly by 54%.

References

[1]   WHO, march 2020. [Online]. Available: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. [Diakses 15 february 2021].

[2]   J. G. Rizal, "www.kompas.com," 17 may 2021. [Online]. Available: https://www.kompas.com/tren/read/2021/05/17/093407265/update-corona-dunia-17-mei-33-juta-orang-meninggal-dunia-karena-covid-19. [Diakses 11 july 2021].

[3]   E. Adamopoulou, "Machine Learning with Applications," *Elsevier.* 2020

[4]   E. Adamopoulou, "An Overview of Chatbot Technology," *IFIP International Federation for Information Processing,* pp. 373 - 383, 2020.

[5]   S. Raj, Building Chatbots With Python, Bangalore, India: Apress, 2019.

[6]   I. C. Education, "www.ibm.com," IBM, 2 Juli 2020. [Online]. Available: https://www.ibm.com/cloud/learn/natural-language-processing. [Diakses 15 Mei 2021].
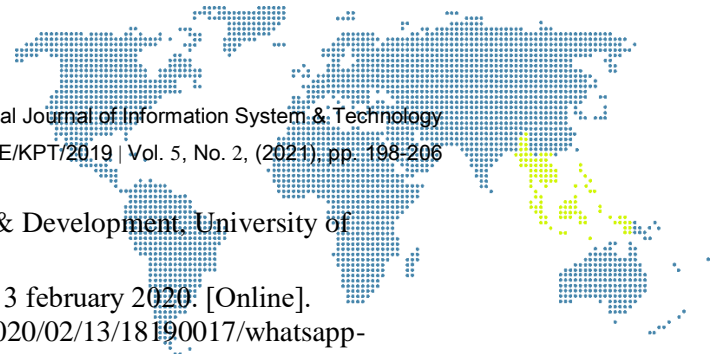
[7]   C. Jack, CHATBOT: Architecture, Design, & Development, University of Pennsylvania, 2017

[8]   W. K. Pertiwi, "www.tekno.kompas.com," 13 february 2020. [Online]. Available: https://tekno.kompas.com/read/2020/02/13/18190017/whatsapp-tembus-2-miliar-pengguna. [Diakses 10 july 2021].

[9]   A. S. Wardani, "www.liputan6.com," 18 november 2019. [Online]. Available: https://www.liputan6.com/tekno/read/4113678/83-persen-pengguna-internet-indonesia-pakai-whatsapp. [Diakses 11 july 2021].

[10]   B. Clinten, "www.tekno.kompas.com," 7 april 2020. [Online]. Available: https://tekno.kompas.com/read/2020/04/07/20030047/chatbot-covid-19-pemerintah-diakses-lebih-dari-2-juta-pengguna. [Diakses 15 januari 2021].

[11]   Eko Handoyo, M. Y. Ticketing Chatbot Service using Serverless NLP Technology. *5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE). IEEE,* 2018.

[12]   Urmil Bharti, D. B. Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19 . *Proceedings of the Fifth International Conference on Communication and Electronics Systems.* 2020.

[13]   Lucia vaira, M. A. MamaBot: a System based on ML and NLP for supporting Women. *proceeding of the 22nd international database engineering & aplication symposium.* 2018

[14]   K. Akshay, Natural Language Processing Recipes, Bangalore, India: Apress, 2019

[15]   J. Hunt, A beginner guide to python 3 program, Switzerland: Springer, 2019.

Authors

1st Author
Cahyo Prianto
Lecturer of Politeknik Pos Indonesia, Bandung, Indonesia
cahyoprianto@poltekpos.ac.id

2nd Author
Nisa Hanum Harani
Lecturer of Politeknik Pos Indonesia, Bandung, Indonesia
nisahanum@poltekpos.ac.id