**Advance Sustainable Science, Engineering and Technology (ASSET)**
**Vol. 6, No.1, January 2024, pp. 0240106-01 ~ 0240106-10**
**ISSN: 2715-4211 DOI:** https://doi.org/10.26877/asset.v6i1.17590

# Implementation Of A Web-Based Chatbot Using Machine Learning For Question And Answer Services In Universities

**Airlangga Satria Dewantara[1*], Joko Aryanto[1]**

[1] Faculty of Sains & Technology, Universitas Teknologi Yogyakarta, Jl. Siliwangi Jl. Ring Road Utara Daerah Istimewa Yogyakarta 55285, Yogyakarta, Indonesia

*airlanggass02@gmail.com

**Abstract**. Advances in communication technology in line with information technology, Chatbot is an innovation that combines communication technology and information technology, is an application that can communicate with humans like a virtual assistant who can respond and answer every question asked. A university must already have a website that can be accessed by the general public so that information about the college can be accessed by everyone anywhere and anytime. To make it easier to get information on the website, chatbots can be the solution because most prospective students and students who are on the campus feel reluctant to browse further into the website that has been provided and usually only open the main homepage page of the website. Parents of students also find it difficult to find out what is on campus if a lot of information is provided in certain tabs of the website. In this study, I utilized Chatbot technology which is a Machine Learning that can process every text that inputted then analyze it and conduct machine training using the Neural Network algorithms that have been provided. This research uses a case study methodology, with Yogyakarta University of Technology as the subject, to develop a chatbot website that incorporates machine learning to facilitate the processing of user input questions.

**Keywords**: Chatbot, Machine Learning, Question and Answer, Neural Network, Python.

***(Received 2023-11-30, Accepted 2023-12-06, Available Online by 2023-12-10)***

## Introduction

Educational institutions have an important role in the development of the nation's children, especially in terms of science and personality, especially in a lecture institution which requires students (students) to be active, creative, and have critical thinking. A university must already have a website that can be accessed by the public so that information about the university can be accessed by everyone anywhere and anytime. To support these things, the institution also needs to have an adequate system in terms of services, especially the availability of information services that can be accessed easily anytime anywhere such as Chatbot. Chatbot is a simple system to facilitate conversations by using "pattern matching rules" to have conversations with users, but chatbots have limited language understanding capabilities [1]. In its application, chatbots can be built independently or use chat applications that provide services to

develop chatbots [2]. Chatbot in conducting conversations (understanding and answering chat) requires certain methods commonly referred to as applying natural language processing methods.

For several journal reviews related to the creation of chatbots which are the basis for this research, there is a research study by Agung[3]. This research uses the AutoResponder Application for WhatsApp and WhatAuto - Reply App to create its chatbot and uses Google Sheets for its database so that the chatbot can only be used for the WhatsApp application. The next research is by Teddy Wijaya[4] . The research discusses creating a Chatbot on a web platform using the PHP programming language and using the MySQL database to create it, in the database there are several keywords and bot replays that the system uses to automatically reply to chats to users. The next research is by Ahmad Cucus[5]. This research uses NLP (Natural Language Processing) for text processing and also uses a database to store data, the implementation of the program is carried out on the Android Platform. From several journals mentioned, chatbots that are made vary in type, some of which use the WhatsApp platform, Website, and Android, the features provided are also different, such as those using WhatsApp can only be used by users who are only registered in the WhatsApp application and only provide a choice of questions that have been provided, they cannot type their imputed questions.

This research implements the Python programming language is used by novice programmers as well as by highly skilled professional developers. It is being used in academia, at web companies, in large corporations, and in financial institutions[6]. The Chatbot that will be created has several features such as machine learning to process the questions that have been given and the user does not need to register to be able to access the chatbot created because the chatbot will be on the UTY website that is already operating which can be accessed by everyone. There are many methodologies related to the generation of chatbots ranging from simple pattern matching to complex neural networks and other heuristic solutions[7], And Machine learning can substantially reduce the computational costs and shorten the development cycle [8]. Based on these problems, the authors are interested in taking the title "Implementation of Web-Based Chatbot Using Machine Learning for Question and Answer Services in Higher Education". The application of this Chatbot system uses Yogyakarta University of Technology for a case study[9] which is expected to be able to overcome the problems that occur by utilizing machine learning that has been made.

## Methods

The methodology of this research is shown in Figure 1, which describes the flow of the research. It starts with the identification of Problem, Data Collection, Designing System, Implementation System, Testing System, and conclusion.

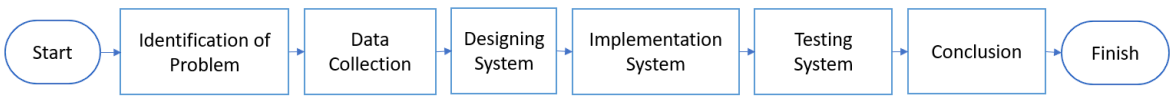


Figure 1 Research Methodology

### *2.1 Identification of Problem*

The analysis of the system that runs at the University of Technology Yogyakarta can be seen in Figure 2.


Figure 2 Current System


Figure 3 Proposed System

From Figure 2 it can be concluded that when students or parents of students have questions about anything related to UTY there are 2 ways. The first way is that parents or students can enter the UTY SIA web and search through all the tabs to get the information they want. The second way is to contact the UTY admin and ask questions about the desired information but have to wait for the admin to answer first to get the information. The analysis of the proposed system in the study can be seen in Figure 3 which can be concluded when parents or students want to find out the desired information about UTY, they only need to enter the UTY website after that ask questions with the chatbot that has been provided in real-time and get answers to the questions asked so that parents or students do not need to search thoroughly on the UTY website or wait for the admin to answer the question.

This research also has problem limitations that include the following:

1. Making this Chatbot uses data available on the UTY website[10] page and the UTY PMB website[11] provided by the University of Technology Yogyakarta.
2. The data that has been obtained is entered into a JSON format file.
3. System design is done using the Python programming language.
4. Using the Neural Network Algorithm to process incoming conversations and send answers that match the User's questions.

### *2.2 Data Collection*

This stage involves collecting the data needed to create a Chatbot for prospective new students, parents or guardians, and students of Yogyakarta University of Technology. At this stage, problems related to the registration of prospective new students are sought and solutions are formulated to overcome them. This research collected data directly through interviews with students to obtain data relevant to the purpose of this research. In addition, we also developed a questionnaire that will be distributed to each respondent to record their responses, which will be the main source of data in this study.

The data used in this study was also sourced from the official website of Yogyakarta University of Technology, which can be accessed through the URLs UTY website[10] and UTY registration website[11]. The accuracy of this data was also ensured by consulting with staff and lecturers working at the Yogyakarta Technological University. The data collection process is carried out on an ongoing basis starting from November 2022 to June 2023. Data was collected by accessing and recording information from the websites into a Word document. The data was then used to formulate questions and corresponding answers. Furthermore, this question-and-answer data is compiled in JSON format[12].

*2.3 System design*

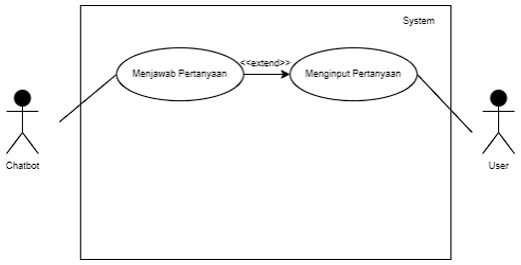


Figure 4 Use Case Diagram.

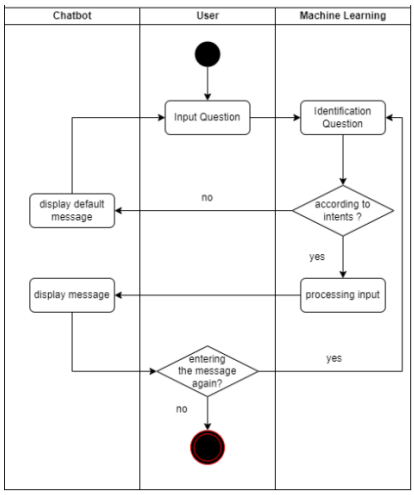


Figure 5 Activity Diagram

The detailed design stage is used to determine the system to be built including making use case diagrams, activity diagrams, and flowcharts. Figure 4 shows the use case diagram[13] of this system. Activity Diagram[14] describes the workflow or activity of a system business process or menu in the software. Figure 5 shows the machine learning activity diagram implemented on the website. The system flowchart[15] is built to be able to provide a complete picture of the existing data requirements. The flowchart of how machine learning is implemented on the website is shown in Figure 6.

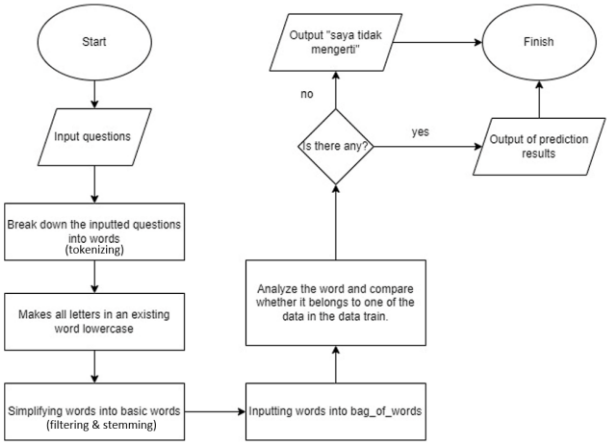


Figure 6 Flowchart.

*2.4 Implementation Phase*

The development of this website[16] adopts Python programming language for machine learning with neural network algorithm[17] and combines HTML[18], CSS, and JavaScript[19] to connect machine learning with the website being created. It also uses several libraries available in Python, such as Sastrawi[20], Flask[21], NumPy, etc. The detailed implementation of Chatbot can be divided into several stages, namely the preprocessing stage, the transformation of existing data into data.JSON, the data training stage, the "bag of words" model building stage, and the final stage of full connection.

1. Text preprocessing

   The stage of text preprocessing[22] is the stage where the system selects the data that is processed in

each document. The preprocessing process includes 3 stages, tokenizing, filtering, and stemming.

2. Transformation of existing data into data.JSON
   Data transformation into data.JSON is needed for the system to be able to read each question pattern that appears along with the target answer. The way to transform data into data.JSON is by separating each word in the data into several parts, namely:
   a. Tags (category) is a knowledge or category that becomes a reference for the system in determining the response.
   b. A Pattern is a series of letters that are expected to match or match one or even more with user input.
   c. Responses (output) are the results of answers that will be issued based on index tags and patterns determined by the system.
3. Data training stage
   After the data is collected, the system tests the data to get a good and satisfactory level of output accuracy. The data training[23] steps are as follows:
   a. Transform the pattern form in the data.JSON into matrix form[24].
   b. Flattering or inserting each row of the matrix into the input layer nodes will signal to the hidden layer.
   c. Assessing the output after completing the calculation from the input layer to the hidden layer which will be given each weight by the system.
4. "bag of words" model building stage
   At this stage, the bag of words value is transformed by transforming the input text from the user into a binary number[25].
5. Full connection
   After getting a binary value from the input text, the binary value is entered into the input layer which will signal the hidden layer of the neural network architecture.

### *2.5 Testing phase*

The performance measurement of the Chatbot application made in this study is based on the results of the confusion matrix by calculating the accuracy, precision, and recall levels.

1. Accuracy
   Accuracy[26] is the ratio of correct predictions (positive and negative) to the overall data. Calculation of accuracy using Equation.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\%$$

2. Precision
   Precision[27] is the ratio of positive correct predictions compared to the overall results that are predicted to be positive. Calculation of precision using Equation.

$$Presisi = \frac{TP}{FP+TP} * 100\%$$

3. Recall
   Recall[27] is the ratio of true positive predictions compared to the overall true positive data. Recall is calculated using Equation .

$$Recall = \frac{TP}{FN+TP} * 100\%$$

## Results and Discussion

### *3.1 Results*

The results of the implementation of the design of the Chatbot website for questions and answers about the Yogyakarta University of Technology are the main page for conducting these question-and-

answer activities. By using machine learning to answer questions inputted by users, the answers given will be given automatically and output that matches the question if the question is in the existing system, otherwise the default answer will be provided. The chat page has an initial appearance that can be seen in Figure 7 it shows if the chatbot button is clicked it will display a conversation window that will be used to conduct questions and answers.

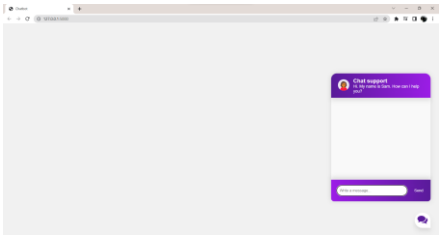


*Figure 7 Chatbot Window*

In the text preprocessing stage, the first step is to perform case folding. In this research, case folding is used to convert all letters in the document into lowercase letters. The next step is tokenizing, which is the process of separating the input string based on the words that make it up. An example of the tokenizing result can be seen in Figure 8.


*Figure 8 Tokenizing*


*Figure 9 Filtering*

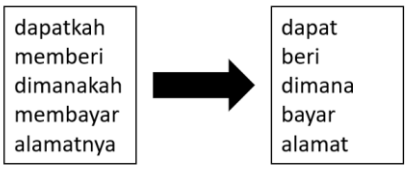


Figure 10 Stemming.

The next stage in text preprocessing is filtering. At this stage, a stop list is used to remove less important words or a wordlist to keep relevant words. An example of filtering can be seen in Figure 9. Stemming is another important stage needed to reduce the number of different indices in a document, as well as to group words that have similar roots and meanings, despite having different forms. Figure 10 shows an example of the stemming result. After data collection is complete, the system performs data testing to achieve an optimal and satisfactory level of output accuracy. Table 1 is an example of the training data used in this study.

*Table 1 Sample Data*

| tag | patterns | | | responses |
|---|---|---|---|---|
| Salam | Hai | Selamat pagi | Hallo | Ada yang bisa dibantu ? |
| Alamat | Dimana Alamat kampus? | Kampus dimana? | Kampus di dekat mana? | Alamat kampus ada di … |
| Selesai | Terima kasih | thanks | bye | Sampai jumpa kembali |

Table 1 is an example of data in JSON format that will be used in the data training process. This data training is done so that the machine can recognize patterns or patterns of user input text, which will later be used to help answer questions from users in the Chatbot system. The steps in data training are as follows:

1. Converting data patterns in JSON format into matrix form, as shown in Figure 11.
2. Perform flattening or take each row of the matrix and put it into the input layer which will send signals to the hidden layer. An example of the flattening result can be seen in Figure 12.
3. The output value of the training data can be seen in Figure 14.

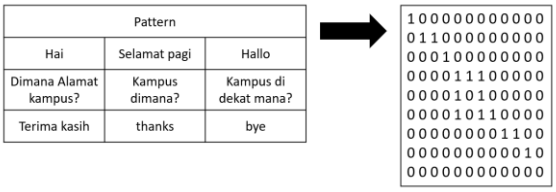


Figure 11 Converting data into a matrix

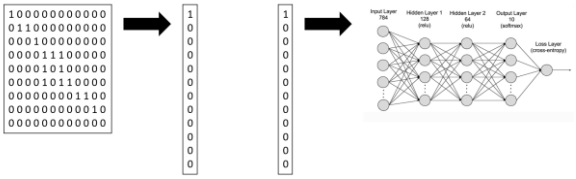

Figure 12 Filtering result

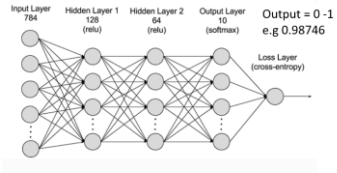


Figure 13 Training data output

The values in the hidden layer cannot be entered directly by the user as they are automatically determined by the system and used as the accuracy level based on existing patterns. The next step is the transformation of the "bag of words" value. An example of the results of the "bag of words" value transformation can be seen in Figure 14.

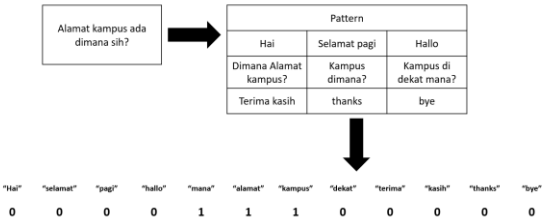


*Figure 14 Transformation result of "Bag of Words" value.*

After obtaining the binary value of the input text, the next stage is the fully connected stage, where this binary value is input into the input layer which will then send signals to the hidden layer in the neural network architecture as seen in Figure 15. The output value is determined by the system after performing calculations from the input layer to the hidden layer so that the Chatbot system can determine the right answer. After getting the output values of the various patterns, the next step is to take the highest value to respond to the question, as shown in Figure 16.

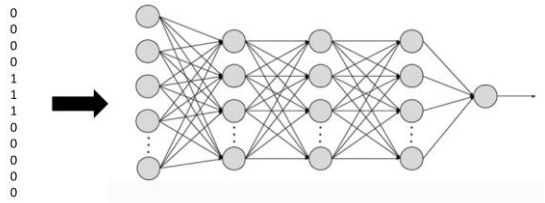

*Figure 15 Full connected phase*

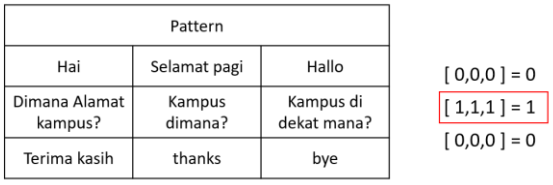


*Figure 16 Example of response determination*

## *3.2 Testing*

The test will perform a series of inputting questions that follow different patterns, writing with typos, and asking questions that do not have available data. The purpose of this test is to examine the chatbot's capabilities in the face of diverse input variations. This test aims to evaluate the extent to which the chatbot can provide relevant and useful responses to the user, regardless of constraints such as typos and questions without available data.

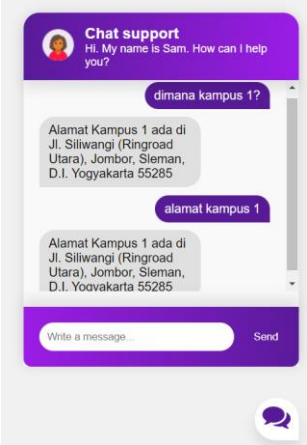


*Figure 18 Testing Chatbot 1*


*Figure 19 Testing Chatbot 2*

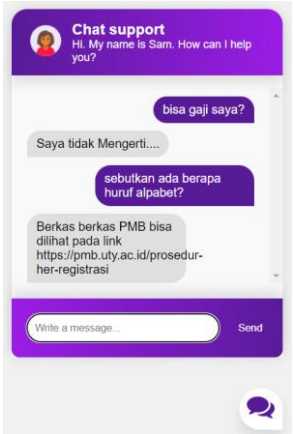


*Figure 20 Testing Chatbot 3*

From Figure 18, the chatbot can provide answers that match the question about campus address 1, even though there are several different question patterns. This shows that the chatbot has the flexibility to understand variations of questions related to campus address 1, so that users can receive relevant answers to the questions asked. From Figure 19, it can be concluded that even though there are several words with typos, the chatbot is still able to provide answers to user expectations. This shows that the chatbot can recognize and understand the intent of the message received, even when there are writing errors. From Figure 20, when there is a question, whose pattern does not exist in the data prepared, the chatbot will answer I don't understand. Meanwhile, if the question deviates but there is a slight allusion to the pattern chatbot gives a random or inappropriate answer. From the tests that ware carried out with the patterns described above, this study conducted 50 trials to calculate the accuracy, precision, and recall of the chatbot that has been made. After conducting the test, the results can be seen in Table 2.

*Table 2 Confusion matrix value*

| | Positive | Negative |
|---|---|---|
| Positive | 40 | 4 |
| Negative | 2 | 4 |

Performance measurement based on confusion matrix results in Table 2 by calculating the level of accuracy, precision, and recall as follows:

$$Accuracy = \frac{40+4}{40+4+4+2} = \frac{44}{50} = 88\%$$

$$Precicion = \frac{40}{40+2} = \frac{40}{42} = 95\%$$

$$Recall = \frac{40}{40+4} = \frac{40}{44} = 91\%$$

Based on the results of the chatbot performance calculation, it has an accuracy of 88%, precision of 95%, and recall of 91%. This test will perform a series of inputting questions that follow different patterns, writing with typos, and asking questions that do not have available data. The purpose of this test is to examine the chatbot's capabilities in dealing with a wide variety of inputs. This test aims to evaluate the extent to which the chatbot can provide relevant and useful responses to the user, regardless of constraints such as typos and questions without available data.

## Conclusions

Based on the results of testing the system that has been implemented in the research on Web-Based Chatbot with Machine Learning Utilization for Question-and-Answer Services in Higher Education, the author concludes several important points.

First, the developed system can effectively overcome the problems faced, where this system successfully responds to chats from users automatically. In other words, Chatbot can provide satisfactory responses to users. Secondly, the system has also proven to be able to provide correct and accurate answers when faced with questions that match the patterns that have been documented in the data, without experiencing errors. This shows the system's ability to understand and respond appropriately to user questions. In addition, the results of this study also indicate great potential in the development of machine learning based Chatbot to improve question and answer services in higher education, especially at Yogyakarta University of Technology.

As a student and lecturer, the author has great hopes for the future, which is that the implementation of the Chatbot website that has been developed can be officially integrated into the website of Universitas Technology Yogyakarta (UTY). The authors believes that this will bring many significant benefits, especially for new students and parents, to provide easier and faster access to the various information available on the UTY website.

With this Chatbot, it is hoped that prospective students and parents will feel more comfortable and efficient in finding the information they need, and this will enhance their experience in undergoing education at UTY.

## Acknowledgments

The author would like to thank the University of Yogyakarta especially the Informatics study program for supporting this project.

## References

[1] A. Augello, G. Pilato, G. Vassallo, and S. Gaglio, “Chatbots as Interface to Ontologies,” vol. 260, pp. 285–299, 2014.

[2] F. Ishlakhuddin, “Chatbot Berbasis Ontologi untuk Mendukung Pemantauan Kinerja dan Keamanan Server dengan Rule-Base,” 2020.

[3] Agung Siswanto Bayu Aji, “Membangun Chatbot Layanan Helpdesk Perpajakan Kpp Pratama Jakarta Setiabudi Satu,” *Sebatik*, vol. 26, no. 1, pp. 194–201, Jun. 2022, doi: 10.46984/sebatik.v26i1.1916.

[4] Teddy Wijaya, Muhammad Rusli, Erwin Syah Rany, and Harfebi Fryonanda, “Membangun Aplikasi Chatbot Berbasis Web Pada CV. Unomax Indonesia,” 2019.

[5] Ahmad Cucus, Robby Yuli Endra, and Tiya Naralita, “Chatter bot untuk konsultasi akademik di perguruan Tinggi,” *Jurnal Sistem Informasi dan Telematika*, vol. 10, no. 1, 2019.

[6] A. Menura Mukhiddinovna, “PROGRAMMING LANGUAGE PYTHON METHODOLOGY FOR CREATING AND USING DIDACTIC MATERIALS FOR STUDENTS,” 2022.

[7] A. Trivedi and Z. Thakkar, “Chatbot generation and integration: A review,” 2019. [Online]. Available: www.IJARIIT.com

[8] J. Wei *et al.*, “Machine learning in materials science,” *InfoMat*, vol. 1, no. 3. Blackwell Publishing Ltd, pp. 338–358, Sep. 01, 2019. doi: 10.1002/inf2.12028.

[9] D. Assyakurrohim, D. Ikhram, R. A. Sirodj, and M. W. Afgani, “Metode Studi Kasus dalam Penelitian Kualitatif,” *Jurnal Pendidikan Sains dan Komputer*, vol. 3, no. 01, pp. 1–9, Dec. 2022, doi: 10.47709/jpsk.v3i01.1951.

[10] Universitas Teknologi Yogyakarta, “UTY Homepage,” 2023. https://uty.ac.id/ (accessed Sep. 20, 2023).

[11] Universitas Teknologi Yogyakarta, “PMB UTY,” 2023. https://pmb.uty.ac.id/ (accessed Sep. 20, 2023).

[12] B. Cecconi, C. K. Louis, X. Bonnin, A. Loh, and M. B. Taylor, “Time-frequency catalogue: JSON implementation and python library,” *Frontiers in Astronomy and Space Sciences*, vol. 9, Feb. 2023, doi: 10.3389/fspas.2022.1049677.
[13] O. Filipova and O. Nikiforova, “Definition of the Criteria for Layout of the UML Use Case Diagrams,” *Applied Computer Systems*, vol. 24, no. 1, pp. 75–81, May 2019, doi: 10.2478/acss-2019-0010.
[14] J. Lang and D. Spišák, “Activity Diagram as an Orientation Catalyst within Source Code,” 2021.
[15] J. Li, “Using Flowchart to Help Students Learn Basic Circuit Theories Quickly,” *Sustainability (Switzerland)*, vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127516.
[16] R. R. Rerung, M. Fauzan, and H. Hermawan, “Website Quality Measurement of Higher Education Services Institution Region IV Using Webqual 4.0 Method,” *International Journal of Advances in Data and Information Systems*, vol. 1, no. 2, pp. 89–102, May 2020, doi: 10.25008/ijadis.v1i2.185.
[17] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: analysis, applications, and prospects,” *IEEE Trans Neural Netw Learn Syst*, 2021.
[18] R. Tabarés, “HTML5 and the evolution of HTML; tracing the origins of digital platforms,” *Technol Soc*, vol. 65, p. 101529, 2021, doi: https://doi.org/10.1016/j.techsoc.2021.101529.
[19] A. Wirfs-Brock and B. Eich, “JavaScript: The first 20 years,” *Proceedings of the ACM on Programming Languages*, vol. 4, no. HOPL, Jun. 2020, doi: 10.1145/3386327.
[20] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, “Improving Text Preprocessing for Student Complaint Document Classification Using Sastrawi,” in *IOP Conference Series: Materials Science and Engineering*, Institute of Physics Publishing, Jul. 2020. doi: 10.1088/1757-899X/874/1/012017.
[21] V. Rama Vyshnavi and A. Malik, “Efficient Way of Web Development Using Python and Flask,” 2019.
[22] T. A. Koleck, C. Dreisbach, P. E. Bourne, and S. Bakken, “Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review,” *Journal of the American Medical Informatics Association*, vol. 26, no. 4, pp. 364–379, Apr. 2019, doi: 10.1093/jamia/ocy173.
[23] A. Rajšp and I. Fister, “A Systematic Literature Review of Intelligent Data Analysis Methods for Smart Sport Training,” 2020, doi: 10.3390/appxx010005.
[24] C. R. Harris *et al.*, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825. Nature Research, pp. 357–362, Sep. 17, 2020. doi: 10.1038/s41586-020-2649-2.
[25] Z. Huang *et al.*, “Binary tree-inspired digital dendrimer,” *Nat Commun*, vol. 10, no. 1, Dec. 2019, doi: 10.1038/s41467-019-09957-6.
[26] D. Chicco, N. Tötsch, and G. Jurman, “The matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation,” *BioData Min*, vol. 14, pp. 1–22, 2021, doi: 10.1186/s13040-021-00244-z.
[27] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.