INTERNATIONAL JOURNAL OF APPLIED ARTIFICIAL INTELLIGENCE AND ROBOTICS
E-ISSN: 0000-0000, DOI: 00.000/XXXX.YYYY.II.NNN
Vol. 1 No. 1 2025 pp. 01-10

Chatbot for Gastric Disease Consultation to Recommend a Healthy Diet Using KNN and NLP Methods

Rafli Abdul Aziz[1], Hendrawaty[2*], Azhar[3]

[1,2,3] Department of Information Technology and Computers, Lhokseumawe State Polytechnic
Jln. B.Aceh Medan Km.280 Buketrata 24301 INDONESIA

[1*]rafliabdulaziz89@gmail.com
[2]waty.hendra@yahoo.com
[3]azhar.x64@gmail.com

Article info: Received15/09/2025, Revised 01/10/2025, Accepted 07/10/2025

Abstract

Gastric disease is one of the most common health problems and requires special treatment, including proper dietary arrangements. This research aims to design and build a healthy diet recommendation system using *chatbots* that utilize Natural Language *Processing* (NLP) and K-Nearest Neighbor (KNN) methods This system is designed to help people with gastric diseases by providing appropriate dietary recommendations based on the symptoms they are experiencing. In the design stage, the system *identifies* the symptoms mentioned by the *user* through a conversation with the *chatbot*. NLP is used to extract relevant information from the *user*'s input text, while KNN is used to classify symptoms and provide appropriate dietary recommendations. The system was tested using a *dataset* that had been adjusted to cases of gastric diseases. The test results show that this *chatbot* system has a classification accuracy of 96.36%, with a confusion matrix that shows good performance in identifying symptoms and providing recommendations. The system is able to understand the context of the conversation well through NLP, while KNN provides accurate classification based on the available *datasets*. With these results, this system is expected to be an effective tool for gastric disease sufferers in managing their diet better and appropriately.

Keywords: *Gastric disease, Healthy diet, Chatbot, Natural Language processing, K-Nearest Neighbor*

1. Introduction

The stomach is one of the human body's organs within the digestive system that functions to digest food and absorb the nutrients contained in it. The stomach contains enzymes such as rennin, pepsin, and hydrochloric acid. It grinds and breaks down food until it becomes completely soft, like porridge [1]. There are many factors that cause gastric diseases. These diseases vary in symptoms—some differ while others are similar. Gastric diseases are among the most common illnesses experienced by the public and are often caused by poor eating habits [2]-[3].

Eating patterns refer to a person's behavior when consuming food or beverages with complete nutritional intake according to the body's needs. A healthy diet reflects a person's daily behavior by regularly consuming food and drinks with balanced nutrients to ensure benefits for the body [4]. Eating patterns are correlated with the occurrence of gastric diseases, which are often linked to unhealthy eating behaviors [5]. Therefore, it is necessary to provide dietary recommendations for gastric patients and to develop innovations, particularly in the field of technology, to address these problems. The application of technology can help facilitate early prevention and treatment of diseases, one of which is through the implementation of a chatbot.

A chatbot is a part of machine learning and an AI-based program designed to provide human-like

conversations with users [6]. The existence of a chatbot can support the implementation of healthy dietary recommendations for gastric disease patients. Previously, patients had to consult doctors directly, which required considerable time and cost. Therefore, an innovation is needed to make the process more efficient and convenient through a dietary recommendation chatbot for gastric patients, developed using machine learning methods or algorithms.

There are many types of machine learning methods; this study applies two of them—Natural Language Processing (NLP) and K-Nearest Neighbor (KNN). NLP is a branch of artificial intelligence (AI) that studies the interaction between humans and computers using natural language approaches [7]-[9]. Meanwhile, KNN is a simple and widely used machine learning model designed to solve classification and regression problems [10].

Previous studies have utilized various methods and algorithms related to the development of chatbots, especially health-related chatbots capable of detecting and handling different types of diseases. However, there has been no research specifically focusing on a dietary recommendation chatbot for gastric patients using KNN and NLP methods. Therefore, this research aims to apply NLP to understand the user's natural language input and KNN to classify gastric diseases based on their symptoms, in order to develop a chatbot for healthy diet recommendations for gastric disease patients using KNN and NLP methods.

2. Methods

A. Data and Data Collection

Data are a collection of raw facts that require further processing. In developing a system, it is essential to use relevant and verifiable data without manipulation. In addition to data, information gathering is also necessary. This process is carried out to serve as a reference material and to strengthen the research being conducted.

The data used in this study consist of symptom data and healthy food recommendations according to the classification of gastric disease types diagnosed by the system. The data were obtained from a medical professional through an interview with Dr. Jauza RM, who was serving at RSUD Cut Meutia during the data collection process. In addition, datasets used in this research were sourced from Disease and Symptoms and Healthy Diet datasets available on Kaggle, which were utilized for the development of the Gastric Disease Consultation Chatbot for Healthy Diet Recommendations Using KNN and NLP Methods.

B. System Requirements Analysis

1) Functional Requirements Analysis

Functional requirements describe all entities involved in the system and the processes performed by those entities. The functional requirements of this system are as follows:

1. Perform registration
2. Perform login
3. Monitor chat activity
4. Conduct consultations
5. View consultation history
6. Perform CRUD (Create, Read, Update, Delete) operations on data
7. Perform logout

2) Non-Functional Requirements Analysis

Non-functional requirements consist of hardware and software components. Hardware refers to the physical components of a computer that can be seen and touched, while software refers to a set of instructions given to a computer to perform specific tasks.

The non-functional requirements for the Gastric Disease Consultation Chatbot for Healthy Diet Recommendations Using KNN and NLP Methods in this research are as follows:

a) Hardware

The hardware used in this study consists of one laptop used to complete the entire process of system development and report preparation.

The hardware specifications are as follows:

- One Lenovo laptop
- CPU: Intel Core i5-1135G7, 2.4 GHz

b) Software

The software used for system development includes:

1. Operating System: Windows 11 Pro 64-bit, as the server operating system.
2. Web Browser: Google Chrome Version 114.0.5735.199, for accessing the system.
3. Text Editor: Visual Studio Code Version 1.79, used to develop the system.
4. Design Tool: Figma Version 93.4.0 (online), for designing the user interface.
5. Framework: Django Version 4.1.13, as the framework used in system development.
6. Programming Language: Python Version 3.12.3.
7. Database Server: MongoDB Compass Version 1.43.3.
8. Diagram Tool: Draw.io Version 21.6.3 (online), used to create activity diagrams and system designs.

C. System Design

The system design in this study consists of Use Case Diagram and Use Case Activity Diagram. These designs were created to provide a clear illustration of the workflow and process within the chatbot consultation system.

1) Use Case Diagram Design

A Use Case Diagram is used to visually represent the interactions between actors and the main functionalities within a system. This diagram is essential in system design as it helps map how actors—whether system users or other interacting entities—are involved in the processes that occur within the system.Through a Use Case Diagram, it becomes easier to understand how each actor is related to various system functionalities and what roles they play in operating the system's features.In this study, the system being developed involves two main actors, namely User and Admin. The User is the individual who will conduct consultations regarding gastric disease and receive healthy diet recommendations. The Admin is responsible for managing data, algorithms, and ensuring that all system functionalities operate properly. The actor design for the *Gastric Disease Consultation Chatbot for Healthy Diet Recommendations Using KNN and NLP Methods* can be seen in Figure 1.

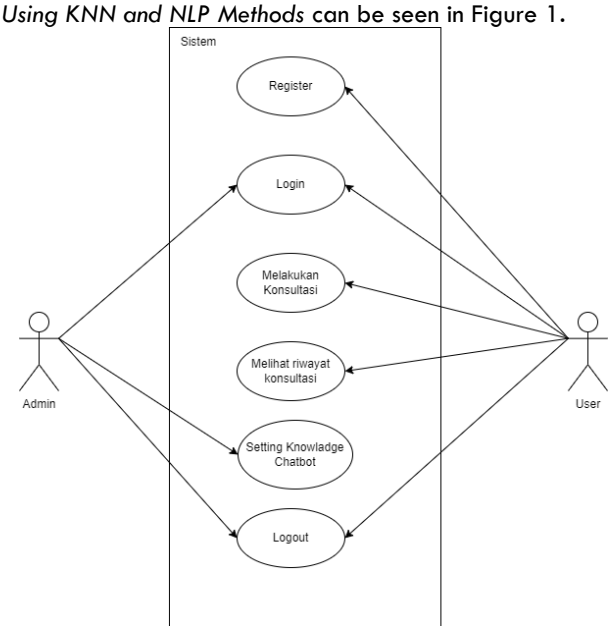

Figure 1 Use Case Diagram

Fig. 1 illustrates the Use Case Diagram of the *Gastric Disease Consultation Chatbot for Healthy Diet Recommendations Using KNN and NLP Methods*, which will be developed in this study.

2) Activity Diagram Design

An Activity Diagram is one type of diagram in UML (Unified Modeling Language) that is used to describe the workflow or sequence of activities occurring in a process. This diagram is particularly useful for modeling complex system functions and behaviors in a way that is easier to understand.

In the context of software development, an Activity Diagram helps to illustrate how different system components interact, as well as how data and control flow throughout the system.

The Fig.2 presents the Consultation Activity Diagram of the system, which can be seen in the next illustration.

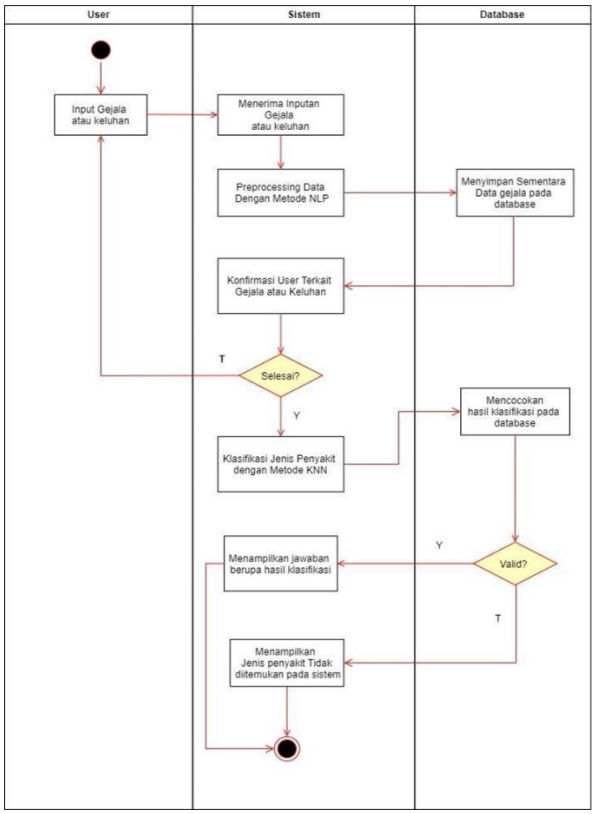

Figure 2 Consultation Activity Diagram

Figure 2 illustrates the flow of initiating a conversation, in which the user interacts with the system by starting a query and entering symptoms or relevant information, which is then processed by the system. Subsequently, the Gastric Disease Consultation Chatbot for Healthy Diet Recommendations Using KNN and NLP Methods performs a check within the database. This figure explains the consultation process flow for gastric disease, starting from the user inputting symptoms to the system performing classification. The classification results are then displayed and stored within the chatbot, which is connected to the existing database.

The explanation of the activity diagram is as follows:

1. The user inputs symptoms or complaints into the system to start the consultation process.
2. The system receives the input in the form of symptoms or complaints from the user.
3. The system then performs data preprocessing using the NLP method. In this process, the input sentence is processed to extract a single keyword representing the symptom or complaint experienced by the user.
4. The preprocessed data is temporarily stored in the database.
5. The system then confirms with the user by asking, “Are there any other symptoms or complaints you are experiencing?” to determine whether the process should continue to the next stage.
6. The user answers the question posed by the system. If there are additional complaints, the process returns to step 1. If not, the process proceeds to the next stage, where the system continues to respond to the user’s queries.
7. Once no further complaints are entered, the system performs a classification process to determine whether the user’s symptoms or complaints fall into the category of gastric diseases.
8. The system then compares the classification results with the data stored in the database to check for a match.
9. If a match is found, the system displays the results in the form of a gastric disease diagnosis, healthy eating recommendations, foods to avoid, and suggested meal menus.
10. If no match is found, the system displays a message indicating that the disease related to the user’s symptoms or complaints is not found in the system.

D. Database Design

The database is designed to manage, store, and provide access to the data required by the system. The purpose of this database design is to organize data efficiently so that it can be easily utilized by users. The system uses a NoSQL database, which consists of several collections or tables, including the dataset table, classification table, user table, knowledge table, and chat history table.

E. User Interface Design

The user interface design aims to create a display that serves as a guide for users in operating the system. This system includes several pages, such as the consultation page, classification page, symptom page, dataset page, user page, knowledge page, and KNN settings page, as well as several other supporting pages. Figure 3 shows the design of the consultation page.

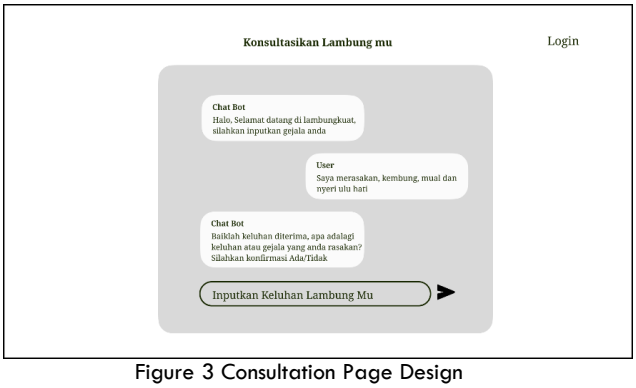

Figure 3 Consultation Page Design

The consultation page is the interface displayed to the user after pressing the Get Started button. On this page, users can interact with the system through a chat feature by inputting their symptoms. Based on the provided input, the system will later generate appropriate health and dietary recommendations.

F. Testing Design

Testing is one of the most crucial stages in the development of a system. The system designed in this study is a gastric disease consultation chatbot that provides healthy diet recommendations using the K-Nearest Neighbor (KNN) and Natural Language Processing (NLP) methods. The purpose of the testing process is to evaluate the system's performance and identify any weaknesses that may exist. The user interface testing for this chatbot system utilizes the Black Box Testing method, which focuses on examining the functionality of the system without delving into its internal code structure.

3. Result and Discussions

A. System Implementation

The implementation of the designed system can be described as follows:

1. Chat Page
   The chat page serves as the main interface of this system, where users can perform consultations with the chatbot. This page can be accessed from the main page by clicking the "Get Started" button, which will then redirect the user to the chat page. The appearance of this page is shown in the Fig. 4 illustrating the consultation interface.

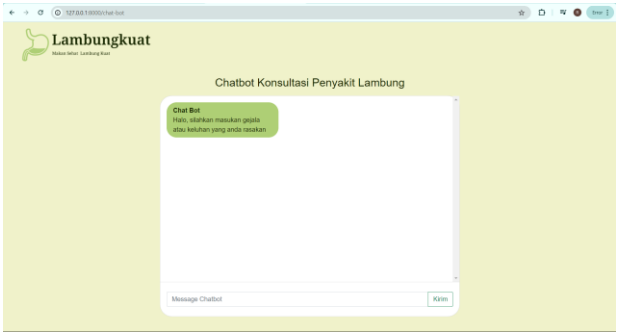

Figure 4 Chat Page

2. Chat History Page
   When the user selects the consultation history option, the system will display the results and records of previous consultations, as shown in the Fig. 5.

Figure 5 Chat History Page

3. Classification Page

The classification page is displayed to the admin upon logging into the system. This page presents the classification of various types of gastric diseases available in the system. The classification page is shown in Fig. 6.

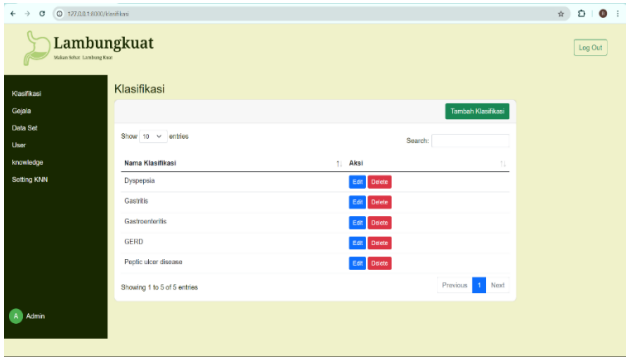

Figure 6 Classification Page

4. Symptoms Page

The symptoms page is displayed to the admin when selecting the “Symptoms” option from the sidebar in the admin interface. This page contains symptom data stored in the database. The implementation of the symptoms page is shown in the Fig. 7.

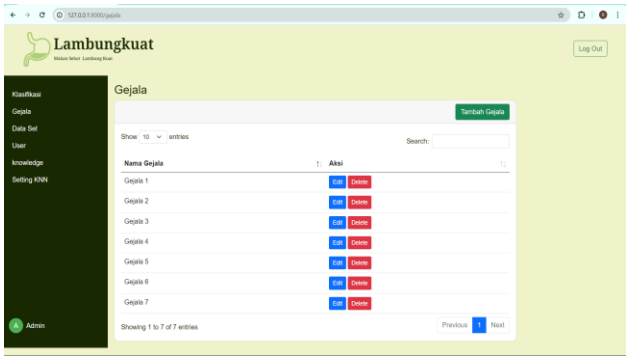

Figure 7 Symptoms Page

4. Dataset Page

The dataset page is displayed to the admin after clicking the “Dataset” button on the sidebar, which then redirects to this page. On the dataset page, the admin can view the dataset used in the system, which includes various types of gastric diseases that can be classified by the system along with their corresponding symptoms. Each record in the database is represented as G1 to G7, depending on the number of symptoms associated with each disease type. Additionally, a search feature is provided to help the admin easily find specific dataset entries within the dataset table. The implementation of the dataset page can be seen in the Fig. 8.

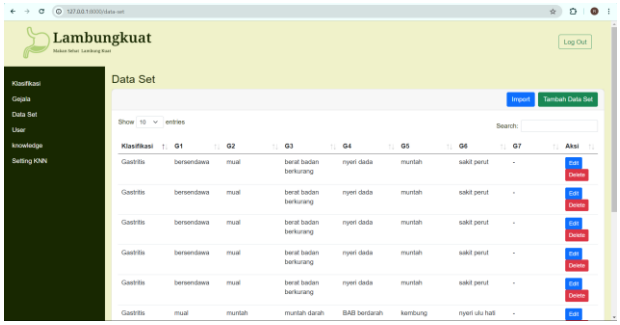

Figure 8 Dataset Page

6. User Page

The user page is displayed to the admin when clicking the "User" button on the sidebar, which then redirects to this page. On this page, the admin can view all users registered in the system, along with their corresponding data such as email and role, whether as an admin or a regular user. The implementation of the user page is shown in the Fig.9 .

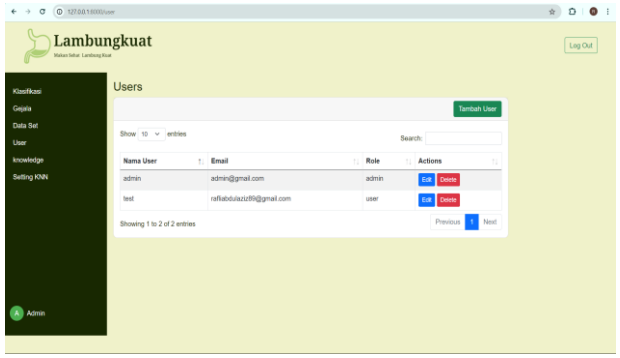

Figure 9 User Page

7. Knowledge Page

The knowledge page is displayed to the admin after clicking the "Knowledge" button on the sidebar, which then redirects to this page. On this page, the admin can view the system's knowledge table, which consists of raw knowledge containing the information input into the system, and processed knowledge representing the system's responses to the existing knowledge. The implementation of the knowledge page can be seen in Fig. 10.

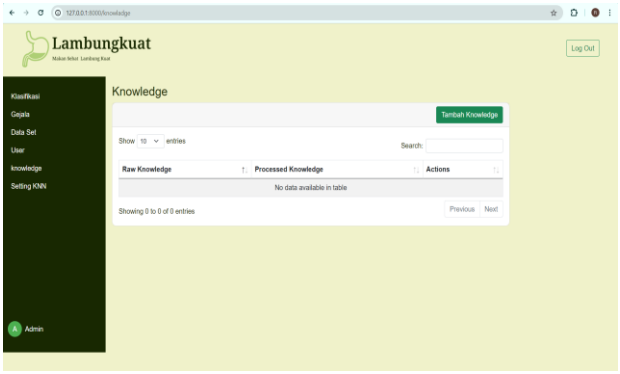

Figure 10 Knowledge Page

8. KNN Settings Page

The KNN settings page is displayed to the admin when clicking the "KNN Settings" button. This page functions to determine the value of K used in the K-Nearest Neighbors (KNN) classification process. The implementation of the KNN settings page can be seen in the Fig. 11.

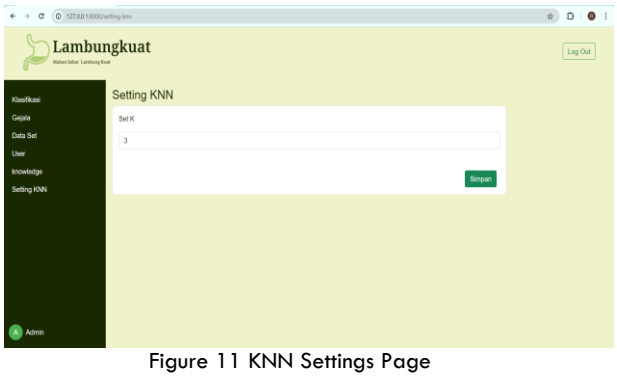

Figure 11 KNN Settings Page

B. Accuracy Testing

This test involves building a model to measure the accuracy used in the system. The process includes importing data from a CSV file (data.csv), separating features and labels, and splitting the dataset into training and testing sets with a ratio of 80:20. The features are normalized using 'StandardScaler'.

In the gastric disease consultation chatbot for healthy diet recommendations using the KNN and NLP methods, a function named 'train_and_evaluate_knn(metric)' is implemented to train the KNN model with three nearest neighbors and a specified distance metric (in this case, the Euclidean metric). The model is evaluated by predicting the labels of the test data and calculating the confusion matrix and accuracy. Both the confusion matrix and accuracy results are then displayed on the screen.

```
Confusion Matrix:
[[22  0  2  2  0]
 [ 0 35  0  0  0]
 [ 0  0 34  0  0]
 [ 2  0  0 33  0]
 [ 0  0  0  0 35]]
Accuracy: 0.9636363636363636
```

Figure 12 Output of Code Snippet for Method Implementation in the System

The evaluation results show an accuracy of 96.36%, with a precision score of 96.83, a recall score of 95.73, and an F1-score of 95.73 for the KNN method using the Confusion Matrix. To further validate the performance of the KNN model, a manual confusion matrix calculation was conducted.

C. Classification Testing Using the KNN and NLP Methods

Classification testing using the K-Nearest Neighbor (KNN) and Natural Language Processing (NLP) methods in the gastric disease consultation chatbot for healthy diet recommendations is an essential step to evaluate the effectiveness and accuracy of the algorithms in classifying data.

This classification testing involves various user input scenarios to ensure that the system can handle variations in language and terminology used by users. Furthermore, the test aims to identify areas where the algorithms may face difficulties or produce classification errors. The results of this testing provide valuable insights into how the chatbot system can be further improved to achieve higher levels of accuracy.

The success rate of the method testing can be calculated using the following formula:

Success rate (%) = (Number of passed test cases / Total number of test cases) x 100 %

= (102/135) x 100%

= 75.55%

The calculation of the classification test results using the KNN and NLP methods shows a success rate of 75.55%. This means that out of 135 test cases, the model successfully provided accurate predictions in 102 cases. These results indicate that the model demonstrates fairly good performance but still requires further evaluation and optimization to achieve more optimal results.

4. Conclusion

The gastric disease consultation chatbot for healthy diet recommendations using the KNN and NLP methods produced several conclusions. The designed system successfully integrated the Natural Language Processing (NLP)

method to understand user inputs and the K-Nearest Neighbor (KNN) method to classify symptoms and provide healthy diet recommendations. The KNN model, with optimal parameter settings (optimal K value), was able to deliver accurate prediction results, achieving an accuracy of 96.36% in testing using the Euclidean distance metric.

REFERENCES

[1] S. M. Rachma, M. Nishom, and S. F. Handayai, "Gastric Disease Diagnostic Expert System Application Using the Fuzzy Mamdani Method," J. Informatics Inf. Syst. Softw. Eng. Appl., vol. 5, no. 2, pp. 104–114, 2023, doi: 10.20895/inista.v5i2.1057.

[2] [2] H. Sampul, P. Studi, T. Informatika, F. Teknik, D. A. N. Komputer, and U. P. Batam, "Penerapan Algoritma Naive Bayes Dalam Memprediksi Penyakit Lambung," vol. 12, no. 2, pp. 10–15, 2022.

[3] D. Parbagasan, K. Uluan, A. Martiasari, and S. M. Lisca, "Hubungan Pola Makan , Pola Istirahat Dan Pola Kebersihan Pada Anak Pra Sekolah Terhadap Perilaku Hidup Bersih Dan Sehat Di TK Santo Asih," vol. 00, 2023.

[4] A. Trilaksmi, D. Nazma, L. Panduwaty, and N. Prasetyaningsih, "Penyuluhan Pola Makan Yang Sehat Pada Pengemudi Ojek Online Untuk Mencegah Gastritis," vol. 2, no. 7, pp. 805–812, 2023.

[5] I. Ibrahim and A. Abdulazeez, "The Role of Machine Learning Algorithms for Diagnosing Diseases," J. Appl. Sci. Technol. Trends, vol. 2, no. 01, pp. 10–19, 2021, doi: 10.38094/jastt20179.

[6] P. Kandpal, K. Jasnani, R. Raut, and S. Bhorge, "Contextual chatbot for healthcare purposes (using deep learning)," Proc. World Conf. Smart Trends Syst. Secur. Sustain. WS4 2020, pp. 625–634, 2020, doi: 10.1109/WorldS450073.2020.9210351

[7] C. Pathak et al., "Chatbot based Disease Prediction and Treatment Recommendation using AI". 2021, doi http://dx.doi.org/10.2139/ssrn.3869072

[8] L. Xu, L. Sanders, K. Li, and J. C. L. Chow, "Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review," JMIR Cancer, vol. 7, no. 4, 2021, doi: 10.2196/27850.

[9] A. Rahmatulloh, A. Ginanjar, I. Darmawan, N. I. Kurniati, and E. Haerani, "Chatbot for Diagnosis of Pregnancy Disorders using Artificial Intelligence Markup Language (AIML)," Int. J. Informatics Vis., vol. 7, no. 1, pp. 77–83, 2023, doi: 10.30630/joiv.7.1.1595.

[10] A. J. T, D. Yanosma, and K. Anggriani, "Implementasi Metode K-Nearest Neighbor (Knn) Dan Simple Additive Weighting (Saw) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka," Pseudocode, vol. 3, no. 2, pp. 98–112, 2017, doi: 10.33369/pseudocode.3.2.98-112.