HealthCareChatbot

Likitha Chimakurti
Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopeta,India

chimakurthilikitha@gmail.com

Annem Bhargavi
Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopeta,india

bhargaviannem528@gmail.com

Meerabi Shaik
Computer Science and Engineering
Narasaraopeta Engineering College
Narasaraopeta,India

meerabishiak782@gmail.com

1. Abstract-- Developing a healthcare chatbot represents a promising avenue for enhancing access to medical knowledge and reducing healthcare expenditures. The chatbot could analyze symptoms provided by users using NLP algorithms to offer potential diagnoses and basic information about diseases, thus potentially minimizing unnecessary doctor visits and saving both time and money. Additionally, it could provide users with details of healthcare providers specializing in the diagnosed condition for necessary consultations, alongside tailored dietary recommendations. This approach aims to empower users by furnishing reliable medical information, facilitating proactive health management and informed decision- making. Continuous learning from user interactions and feedback would further refine the chatbot's accuracy and information quality over time. Ensuring compliance with data protection regulations and safeguarding user privacy is pivotal to building trust and encouraging widespread .

Keywords -Healthcare,Chatbot,Disease Prediction,Natural language processing,Text to speech,Random forest classifier,TF-IDF,Cosine-Similarity.

2. INTRODUCTION

The importance of prioritizing healthcare for a healthy lifestyle is evident, yet accessing medical consultation can be challenging at times. To address this, a proposed solution involves developing a healthcare chatbot utilizing Natural Language Processing (NLP), a component of Artificial Intelligence. This chatbot aims to diagnose diseases and provide essential medical guidance, thereby reducing healthcare expenses and improving accessibility to medical information. Some chatbots serve as medical reference guides, empowering patients with insights into their conditions and aiding in health improvement. The effectiveness of such a chatbot lies in its comprehensive ability to diagnose various illnesses and provide necessary information. The system is designed to provide text or voice assistance, catering to users in their preferred language. By analyzing user symptoms, the chatbot can identify potential diseases, recommend appropriate physicians, and offer dietary suggestions. This enables individuals to gain insights into their health status and take proactive measures to safeguard their well- being. Chatbots operate through Machine Learning (ML) and Artificial Intelligence (AI), utilizing Natural Language Processing (NLP) techniques such as NLTK for Python to analyze speech and deliver intelligent responses, thus simulating human-like interactions.[9]

To fulfill the project's goals, an extensive dataset containing patient symptoms and corresponding diagnoses will be gathered and processed. This dataset will serve as the basis for training the machine learning algorithms integrated into the chatbot system. The effectiveness of the developed healthcare chatbot system will undergo evaluation through a series of experiments and performance metrics. Essential metrics such as accuracy, precision, and F1-score will be employed to gauge the chatbot's diagnostic accuracy and efficacy in providing suitable recommendations. Additionally, testing will be conducted to assess the usability and user experience of the chatbot system. It is imperative to comprehend patients' perceptions, preferences, and feedback as it is pivotal in refining the chatbot's functionality and enhancing its overall performance.[10]

3. LITERATURE SURVEY:

The solution proposed entails the development of a versatile healthcare chatbot capable of diagnosing diseases based on user-provided symptoms. Furthermore, the chatbot responds to user inquiries by leveraging TF-IDF and Cosine Similarity techniques to compute sentence similarity, selecting the most suitable response from its knowledge repository. With its multilingual capabilities supporting English, Hindi, and Gujarati, the chatbot is well-suited for deployment in rural India. Employing Natural Language Processing principles, the chatbot engages users in conversation and offers support for both speech-to-text and text-to-speech conversion, facilitating communication via voice. Through analysis of five distinct ML algorithms for disease prediction, the Random Forest emerged as the top performer, boasting an impressive accuracy rate of 98.43%, thereby serving as the core classifier for the system [1].

They proposed an app leveraging Artificial Intelligence to aid in diagnosing various diseases and providing pertinent information regarding a patient's illness. Employing techniques from NLP , the application operates within human language to offer feedback akin to responses provided by medical professionals. Its objective is to address user inquiries by processing input, utilizing specific methods, and furnishing results based on symptoms. Initially, input undergoes processing converting entire sentence to lowercase , ensuring uniformity and consistency in the algorithm's treatment of words, regardless of their casing. This preprocessing step aims to mitigate discrepancies arising from variations in word case. The output demonstrates a notably low cosine similarity, implying that responses may not consistently match the input precisely.[2]

Ashwini Shangrapawar and colleagues developed a chatbot aimed at assisting individuals suffering from various health issues by recommending suitable medications. The chatbot operates continuously, providing round-the-clock availability by utilizing a hardware boot as a server for clients via programming. Specifically, this medical chatbot offers guidance on common ailments such as fever, cold, typhoid, malaria, and jaundice. The system incorporates input speech signals captured through a microphone, which are then compared with the database stored in a Raspberry Pi, enabling responses directly from the database. In cases where information regarding a particular ailment or medication is unavailable in the database, the chatbot retrieves responses from Google servers. A limitation of this approach is its reliance on hardware, necessitating physical transportation from one location to another.[3]

This paper introduces a dataset and a healthcare chatbot built on the RASA framework, employing NLP techniques for disease detection and medical guidance provision.It underwent validation and refinement with input from health care professionals. Evaluation of the healthcare chatbot's performance utilized RASA test methods and validation, yielding on the collected data. Additionally, human evaluation resulted in an accuracy rate of 56.50%.[4]

In 2021, they suggested a "Healthcare Chatbot" aimed at diagnosing diseases and furnishing fundamental information about the respective conditions prior to consulting a medical professional. The adoption of such a medical chatbot is anticipated to lower healthcare expenses and enhance medical information accessibility for a wider populace. Utilizing natural language, chatbots serve as computer programs facilitating user interaction. Within this system, users can engage with the chatbot via SMS, receiving responses in both speech and text formats. The chatbot, upon user interaction, identifies relevant illnesses and suggests specialized experts to address user concerns, offering recommendations for resolution. Notably, the system accommodates multiple users concurrently without latency issues. The primary objective of this endeavor is the swift and accurate prediction of diseases based on symptoms presented by consumers. To achieve this, a decision tree algorithm is employed for disease prediction. Through the provision of predictive diagnoses, chatbots hold substantial potential for transforming the healthcare sector.[5]

The objective is to develop a medical chatbot utilizing Neural Networks capable of providing information, diagnosing diseases, and offering basic guidance on when and where to seek medical consultation. The efficacy of these medical chatbots relies on Natural Language Processing techniques enabling users to articulate their health concerns and queries. Users have the flexibility to inquire about personal health matters via the chatbot, eliminating the need for physical presence at a clinic or hospital. This approach is proposed to reduce costs and enhance access to medical information via the utilization of chatbots. Developmental roadmap involves assessing customer sentiments as part of the program development process.[6]

Various documents undergo content verification , representation. A TF-IDF matrix is then generated, producing matrices S, U, and V, followed by calculating cosine similarity by multiplying the three matrices.[7]

The system functions as a communication platform utilizing Natural Language Processing (NLP) for user interaction, essentially operating as a computer program.Through NLP, the chatbot processes userinput, extracting sentence keywords to address queries effectively.Calculations suchas Rank and sentence similarity are executed using techniques like TF-IDF, Stemming, n- grams, and cosine similarity.Machine learning techniques are integrated into the system to resolve healthcare queries,employing appropriate algorithms. Usershave the capability to create profiles, detailing symptoms and receiving recommendations for doctors and dosage reminders. This chatbot serves as an advisor in emergency situations, providing primary care advice prior to visiting a physician, , and in some instances, acting as a substitutefor minor health issues. Additionally, itextends support to those in urgent need of solutions. Through symptom input, users canidentify potential illnesses accurately.[8]

4. PROPOSED SYSTEM

Our proposed system allows users to interactwith the bot via voice or text to addressqueries, leveraging an expert system for responses. Users can access information about available doctors specializing in specific diseases. Designed for multipleusers, the system facilitates online counseling sessions. A pattern-template structure is used in the database to store chatbot data. Furthermore, the bot provides food advice and recommendations for analgesics. tailored to the user'sdisease.

SYSTEM ARCHITECTURE:

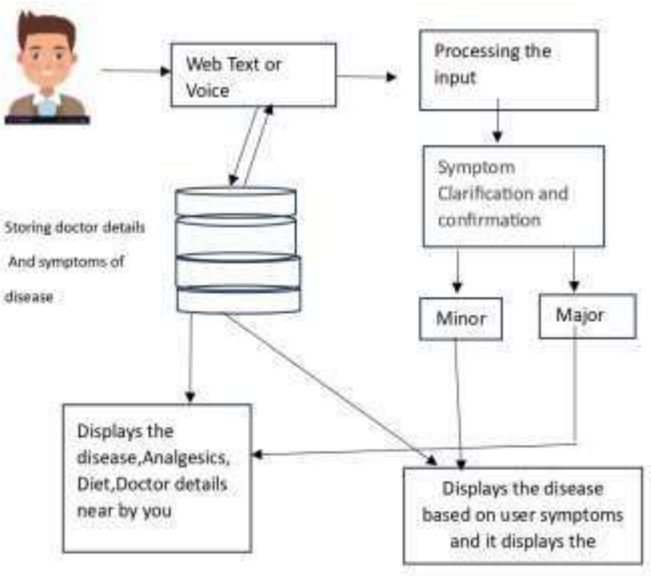

Fig1: system architecture

The depicted system(Fig1) allows users to engage with a chatbot in a user-friendly manner to discuss their health concerns, which are subsequently stored for future reference. The chatbot conducts an inquiry into the user's symptoms through a series ofquestions to confirm their nature. Dependingon the severity, the condition is classified as either minor or major. If deemed major, the chatbot suggests local physician information for more support additionally providesinformation on suitable pain relief medication. Additionally, it offers dietary recommendations to aid recovery. The user interface is designed to facilitate easy interaction, promoting the use of the chatbotfor minor health issues, thereby reducing unnecessary visits to hospitals

5. Data Flow Diagram

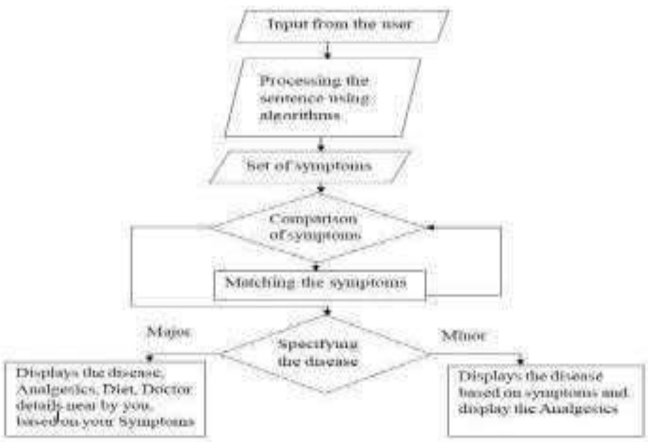

Fig-2flow diagram of chatbot

The chatbot functions by receiving user inputand processing it using specialized algorithms. These algorithms analyze the input in conjunction with a database of symptoms to comprehend the user's condition accurately. Through a series of systematic inquiries After confirming the user's symptoms, the chatbot classifies the illness as minor or significant. After making this decision, the chatbot informs the user of the severity of their condition. In instances where a major disease is identified, the chatbot offers recommendations for nearby doctors for further consultation. Additionally,it presents information about suitable pain relief medications and suggests dietaryadjustments to expedite recovery.[Fig2]

Data Preprocessing

The original Kaggle dataset, comprising 4920 records and encompassing 41 distinct diseases along with their respective symptoms and precautions, was initially in a raw format. To ensure data integrity, thorough checks for inconsistencies were conducted[11]. Subsequently, employing the count vectorizer technique facilitated the transformation of unstructured categorical data into structured numerical form. In the resultant dataset, each symptom was delineated as a distinct column, with diseases represented across rows. A binary encoding scheme was adopted, where the presence of a symptom for a particular disease was denoted by a value of 1, while its absence was marked by 0. This encoding strategy effectively captured the disease-symptom relationships, facilitating comprehensive analysis and modeling endeavors.

6. ALGORITHMS:

Decision Tree

Svm

Knn

Random Forest

Text preproceesing

Tokenization: The text is parsed word by word, with separation occurring whenever specified characters are encountered and punctuation is removed accordingly. This process sets the stage for subsequent steps.

Here the below diagram(Fig3) tokenizes the text data in the specified columns of the DataFrame

.

Fig3tokenizes the text data

TF-IDF:

Term frequency(tf):

Typically, when constructing a model to comprehend text, it's common practice to eliminate all stop words. Alternatively, another approach involves determining the the terms' relative importance using TF-IDF. With each document having its own distinct term frequency, TF-IDF is the number of times a word appears divided by the total number of words in that text.

$$\mathbf{tf_{ij} = n_{ij} / \sum_k n_{ij}}$$

INVERSE DATA FREQUENCY (IDF):

The definition of the Inverse Document Frequency (IDF) is the logarithm of the ratio of the total number of documents to the number of documents containing a particular word..IDF assigns weight to unique words across all documents in the corpus based on their occurrence frequency.

$$Idf(w)=\log(N/dft)$$

Here the below(Fig4) initializes a TF-IDF vectorizer (tfidf_vectorizer).Fits and transforms the symptoms data using TF-IDF, resulting in a TF-IDF matrix(tfidf_matrix).Converts the TF-IDF matrix to a DataFrame (tfidf_df).Prints the TF-IDF DataFrame.

	_patches	abdominal_pain	abnormal_menstruation	acidity	\
0	0.517716	0.0	0.0	0.0	
1	0.546610	0.0	0.0	0.0	
2	0.543606	0.0	0.0	0.0	
3	0.605124	0.0	0.0	0.0	
4	0.000000	0.0	0.0	0.0	
...	...	...	...	...	
4915	0.000000	0.0	0.0	0.0	
4916	0.000000	0.0	0.0	0.0	
4917	0.000000	0.0	0.0	0.0	
4918	0.000000	0.0	0.0	0.0	
4919	0.000000	0.0	0.0	0.0	

Fig4:shows the term significance within the document

Cosine similarity:

By measuring the cosine of the angle that separates two non-zero vectors in an inner product space, cosine similarity determines their similarity. In the realm of data mining, this method is also used to gauge cluster cohesion.Finding the distance in n dimensions between two vectors is all that the cosine distance represents.

Here below(Fig5) calculates the cosine similarity matrix for diseases based on their TF-IDF representation of symptoms

Disease	Fungal infection	Fungal infection \
Disease		
Fungal infection	1.000000	0.947140
Fungal infection	0.947140	1.000000
Fungal infection	0.952373	0.891420
Fungal infection	0.855553	0.776283
Fungal infection	0.681132	0.559602
...	...	...
(vertigo) Paroymsal Positional Vertigo	0.000000	0.000000
Acne	0.098174	0.103654
Urinary tract infection	0.000000	0.000000
Psoriasis	0.083380	0.080033
Impetigo	0.095923	0.101276

Fig5:calculates the cosine similarity matrix for diseases based on their TF-IDF representation of symptoms

N-gram:N-Grams aid machines in comprehending words within content, contributing to a deeper understanding of language. An N-gram represents a neighboring sequence of N-items extracted from a text sample, where N can be adjusted to consider two items (bigrams), three items (trigrams), and so forth. These contiguous sequences assist in predicting subsequent words in a sentence., The value of 'N' can be adjusted based on the context of sentence

Here this below(Fig6) generates n-grams for each symptom, and then prints the generated n-grams.

0	[(itching, skin_rash), (skin_rash, nodal_skin_...
1	[(skin_rash, nodal_skin_eruptions), (nodal_ski...
2	[(itching, nodal_skin_eruptions), (nodal_skin_...
3	[(itching, skin_rash), (skin_rash, dischromic)...
4	[(itching, skin_rash), (skin_rash, nodal_skin_...
	...
4915	[(vomiting, headache), (headache, nausea), (na...
4916	[(skin_rash, pus_filled_pimples), (pus_filled_...
4917	[(burning_micturition, bladder_discomfort), (b...
4918	[(skin_rash, joint_pain), (joint_pain, skin_pe...
4919	[(skin_rash, high_fever), (high_fever, blister...

Fig6:generates n-grams for each symptom

RANDOM FOREST:

A well-liked machine learning technique for classification and regression applications is called Random Forest. This approach of ensemble learning creates a large number of decision trees during training and outputs the class mode. The average prediction of the individual trees.we have various algorithms like svm,knn,decision treeand random forest .But random forest has produced more accuracy,precision,f1-scores compared to all algorithms**.**

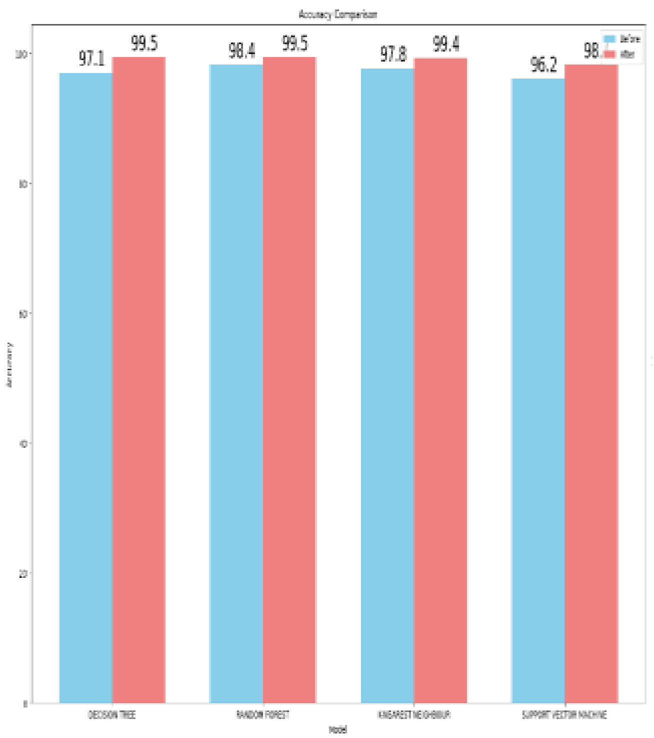

Fig7:Accuracy analysis of algorithms

Here the above figure(Fig7) indicates the accuracy of decision tree,random forest,knn,svm and for existing work (before) and proposed work(after)

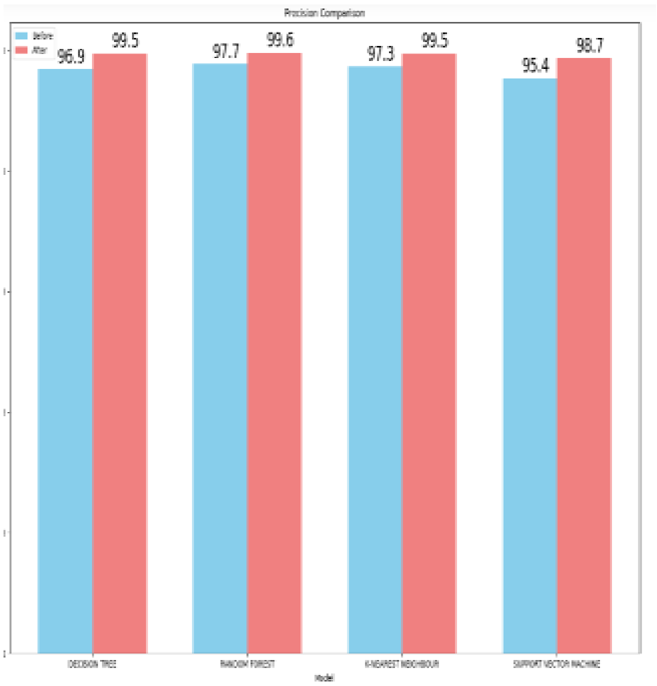

Fig8:Precision analysis of algorithms

Here the above figure(Fig8) indicates the precision of decision tree,random forest,knn,svm and for existing work (before) and proposed work(after).

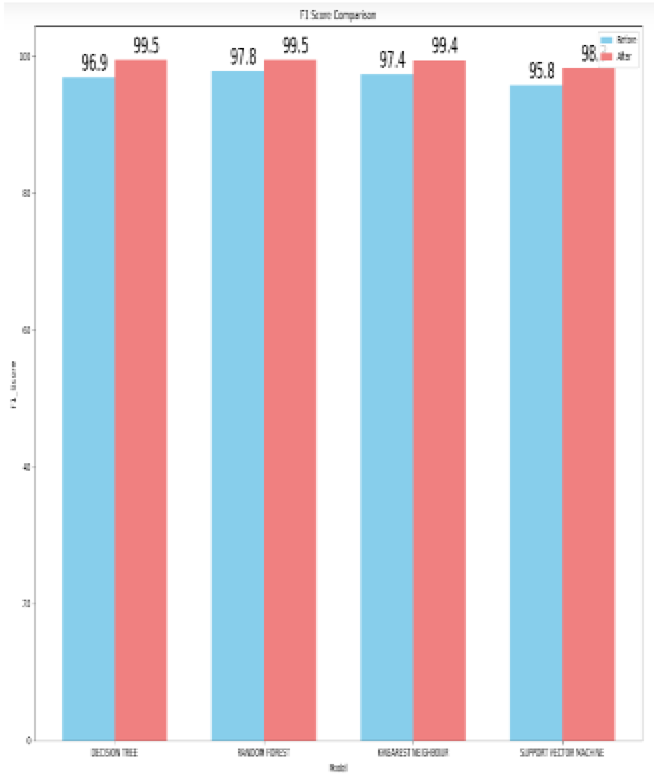

Fig9:f1-analysis analysis of algorithms

Here the above figure(Fig9) indicates thef1- score ofdecision tree,random forest,knn,svm and for existing work (before) and proposed work(after).

Classification Algorithms	Exsisting [1]work values	Proposed work values
Random Forest	97.15	99.45
Decision Tree	96.41	99.00
SVM	96.00	98.33
KNN	96.88	99.35

Table1: Results of k-fold validation

here the above table (1) indicates the k-fold validation scores produced by different algorithms for existing and and proposed system the k-fold score for random forest is high compared to all algorithms

Classificati of Algorithm s	Exsisting work[1] values			Proposed work values		
	Accu racy	preci sion	F1 Scor e	Accu racy	preci sion	F1 Scor e
Random Forest	98.4 3	97.7 4	97.8 1	99.5 1	99.6 1	99.5 2
Decision Tree	97.1 2	96.9 3	96.9	99.5 1	99.5 9	99.5 1
SVM	96.2 2	95.4 7	95.8 2	98.4 1	98.7 2	98.3 6
KNN	97.8 8	97.3 1	97.4 9	98.6 2	98.1 5	99.0 1

Table2.Algorithm evalution metrices

Here the above table (2) shows the evalution metrices such as accuracy,precision and f1- score for exisisting and proposed systems of different algorithms as shown in above table2.here the scores of random forest are very high hence we have choosen random forest compared to all algorithms

7. Communication page:

Fig10: chatbot interacts with user

Here in the above figure(Fig10) it shows how chatbot interacts with user and it predicts and provide preventive measures

8. CONCLUSION:

The chatbot serves as an effective tool for facilitating conversations between humans and machines. It's designed to provide prompt and accurate responses without any delay, ensuring a seamless user experience. Overall, the chatbot is deemed user-friendly and accessible to individuals proficient in typing in their native language. Furthermore, it offers individualized diagnosis based on symptoms given by user

9. CONCLUSION:

The chatbot serves as an effective tool for facilitating conversations between humans and machines. It's designed to provide prompt and accurate responses without any delay, ensuring a seamless user experience. Overall, the chatbot is deemed user-friendly and accessible to individuals proficient in typing in their native language. Furthermore, it offers individualized diagnosis based on symptoms given by user

10FUTURE ENHANCEMENT:

This is the era of messaging applications, as people will use them for longer than any other app in the future. The adoption of customized care would effectively save many lives and increase public awareness of health issues. This medical chat can take place from any location. The only things they need are their basic desktop computer or smartphone and an active internet connection. By adding additional word combinations and utilizing databases more frequently, the medical chatbot's effectiveness can be raised to the point where it can handle any kind of illness.

11REFERENCES:

1.S. Badlani, T. Aditya, M. Dave and S. Chaudhari, "Multilingual Healthcare Chatbot Using Machine Learning," *2021 2nd International Conference for Emerging Technology (INCET)*, Belagavi, India, 2021,pp.1-6 doi10.1109/INCET51464.2021.9456304

2. Dinesh Kalla, Fnu Samaah,” Chatbot for Medical Treatment using NLTK Lib” IOSR Journal of Computer Engineering IOSR-JCE, Volume 22, Issue 1, Feb -2020.

3. Ashwini Shangrapawar, Ankita Ravekar,” Artificial Intelligence based Healthcare Chatbot System”, IRJET Vol:07, Issue-02:Feb-2020..

4. Mishra, Prateek, et al. "Personalized Healthcare Chatbot: Dataset and Prototype System." International Conference on Computational Intelligence in Communications and Business Analytics. Cham: Springer International Publishing, 2022.

5. Athulya N, Jeeshna K, S J Aadithyan, U Sreelakshmi, Hairunizha Alias Nisha Rose, “Healthcare Chatbot”, © 2021 IJCRT | Volume 9, Issue 10 October 2021 | ISSN: 2320-2882.

6. R. Goel, R. P. Goswami, S. Totlani, P. Arora, R. Bansal and D. Vij, "Machine Learning Based Healthcare Chatbot," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2022, pp. 188-192, doi:10.1109/ICACITE53722.2022.9823901

7. Mrs Rashmi Dharwadkar1, Dr.Mrs. Neeta A. Deshpande, A Medical ChatBot, International Journal of Computer Trends and Technology (IJCTT) – Volume 60 Issue 1-June 2018 [8] N-gram Accuracy Analysis in the Method of Chatbot Response, International Journal of Engineering & Technology. (2018)

8. N. V. Shinde, A. Akhade, P. Bagad, H. Bhavsar, S. K. Wagh and A. Kamble, "Healthcare Chatbot System using Artificial Intelligence," *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2021, pp. 1-8, doi: 10.1109/ICOEI51242.2021.9452902.

9. Mahajan, Papiya, et al. "Healthcare chatbot using natural language processing." 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions). 2020.

10. JAIN, BHAVIKA, and DR SUNIL MAGGU. "Health Chatbot." (2023)

11. https://www.kaggle.com/itachi9604/diseasesymptom -description- dataset

12. U. Bharti, D. Bajaj, H. Batra, S. Lalit, S. Lalit and A. Gangwani, "Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 870-875, doi: 10.1109/ICCES48766.2020.9137944.

13. L. Athota, V. K. Shukla, N. Pandey and A. Rana, "Chatbot for Healthcare System Using Artificial Intelligence," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 619-622, doi: 10.1109/ICRITO48877.2020.9197833.