Performance analysis with the Combination of visualization and classification techniques for Medical chatbot

Kamalesh S
Department of Information Technology
Velammal College of Engineering and Technology
Madurai, Tamilnadu
skl@vcet.ac.in

Shajida M
Department of Information Technology
Velammal College of Engineering and Technology
Madurai, Tamilnadu
shajidabatcha4@gmail.com

Sakthiyadharshini N P
Department of Information Technology
Velammal College of Engineering and Technology
Madurai, Tamilnadu
sakthiyadharshiniprakash@gmail.com

Aswitha B
Department of Information Technology
Velammal College of Engineering and Technology
Madurai, Tamilnadu
aswithabalamurugan20@gmail.com

***Abstract*— Natural Language Processing (NLP) continues to play a strategic part in complaint discovery, medicine discovery during the current epidemic. This abstract provides an overview of performance analysis with Combination of visualization and Classification Technique of NLP for Medical Chatbot. Sentiment analysis is an important aspect of NLP that is used to determine the emotional tone behind a piece of text. This technique has been applied to various domains, including medical chatbots. In this we have compared the combination of Decision tree with Heatmap and Naïve Bayes with Word Cloud. The performance of the chatbot was evaluated using accuracy, and the results indicate that the combination of visualization and classification techniques significantly improves the chatbot's performance.**

Keywords*— *Sentimental analysis, NLP, medical chatbot, Decision tree, Heatmap, Naïve Bayes, Word Cloud.

I. INTRODUCTION

In the field of artificial intelligence, natural language processing (NLP) studies how to use language to communicate with computers and people. In order to extract valuable insights that can guide medical decision-making, NLP can be used to analyse massive volumes of medical data, such as electronic medical records. NLP can be used to improve the accuracy and efficiency of medical diagnoses and provide patients with personalized health information.

A supervised learning method used in machine learning and data mining is a decision tree. Its layout resembles a flowchart, with each internal node standing for a feature or attribute and each branch for a decision formula. Decision trees are used for classification and regression analysis.

Naive Bayes is a probabilistic algorithm used in machine learning for classification problems. The Bayes theorem, which asserts that the likelihood of a hypothesis given evidence is proportional to the likelihood of the evidence given the hypothesis, forms the basis of the algorithm and the prior probability of the hypothesis.

Heatmap is a graphical representation of data that uses color-coding to depict the values of a matrix. In a heatmap, each row and column of a matrix is represented by a rectangle or "cell", with the color of the cell indicating the value of the corresponding element in the matrix. A word cloud is a visual representation of text data where the size of each word represents its frequency or importance in the text. It is a popular way to quickly identify the most common or important words in a large body of text.

Sentiment analysis has become an important tool in the field of NLP, as it helps to determine the emotional tone behind a piece of text. Medical chatbots in NLP are used to process and understand human language, allowing the chatbot to respond to questions and provide relevant information about various medical topics, such as symptoms, treatments. NLP models may struggle to understand the context of a patient's symptoms or medical history, which can lead to incorrect diagnoses or recommendations.

There are four phases in medical chatbot. The first phase in chatbot is Gathering input. This phase involves collecting and processing the data that will be used to train the chatbot and provide accurate and relevant information to patients. Second phase is Pre-processing. The pre-processing phase involves cleaning, transforming, and preparing the data so that it can be used to train the chatbot. Third phase is Feature extraction. Feature extraction involves transforming raw data into a format that can be used to train the chatbot and the fourth phase is Sentimental classification. This is a task in Natural Language Processing (NLP) that involves determining the sentiment expressed in a text.

This study aims to investigate the use of sentiment analysis in medical chatbots, with a focus on improving their effectiveness in providing emotional support to patients. The results of the study will be used to improve the performance of medical chatbots and make them better equipped to provide emotional support to patients.

II. RELATED WORKS

Hiba Hussain et al. [1] has proposed a prediction model driven system that predicts accurate diseases based on the symptoms which the user provides. The design of an interactive chatbot to collect symptoms from the user is based on NLP principles. This prediction model is designed using Machine learning algorithms such as KNN and Decision Tree The same dataset was subjected to both algorithms, and the optimal model was chosen based on the accuracy and confidence rates. Our findings show that

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Decision Tree and KNN have accuracy rates of 92.6% and 95.74 percent, respectively.

Vytautas Raulinaitis et al. [2] has presented a thesis which explores the effectiveness and usability of chatbots in the healthcare industry. The thesis analyzes existing research and evaluates the performance of two healthcare chatbots, Ada and Babylon, using a set of criteria. The study finds that both chatbots are reliable and offer accurate health advice, but their usability varies, with Ada being more user-friendly than Babylon. The thesis also discusses the potential benefits and challenges of chatbots in healthcare and concludes with recommendations for future research and improvements to chatbot design.

An innovative method of disease prediction presented by Sanjay Chakraborty et al. [3] makes use of a chatbot that uses artificial intelligence (AI). The significance of early disease diagnosis in infectious diseases as well as the shortcomings of current methods are covered in this essay. The creation of an AI-based chatbot model that can forecast the likelihood of infectious diseases based on a set of symptoms is then described. Data from patients with infectious disorders and healthy people were used to evaluate the model. The findings show that the chatbot model was highly sensitive and specific in its ability to predict the presence of infectious illnesses in patients. The promise of AI-based chatbots for disease prediction is highlighted in the paper's conclusion, along with the demand for more study in this field.

Himanshu Gadge et al. [4] have presented an Artificially Intelligent Chat-bot using applications of Deep Learning to fightCOVID-19 including various viral diseases faced by human being in day Today life. They offer advice on how to use Deep learning in CureBot to help people combat various diseases. She show how the Data is given as input to the Deep Neural Network and how task is constructed as learning problem. She has implemented speech to text conversation type for better use of Chat-bot.

Maria V. Vasileiou ET AL. [5] investigates the role of chatbots in telemedicine as a means of providing remote support to patients. The paper reviews the literature on the use of chatbots in healthcare and discusses the potential advantages they offer, including improved accessibility and reduced healthcare costs. It also describes the development of an intelligent dialog system for remote support using a chatbot. The system was designed to assist patients with chronic conditions and was evaluated through a user study. The results suggest that the chatbot was perceived as helpful and informative by users and had the potential to reduce the burden on healthcare providers. The paper concludes by highlighting the potential of chatbots in telemedicine and the need for further research to improve their functionality and user acceptance.

ManalAlmalki et al. [6] has identified and characterised these developing technologies and their use for treating Corona illness, as well as to highlight challenges encountered in doing so. This was done by analysing recent literature on chatbots used in healthcare related to COVID-19.

AniketPatole et al [7]. has reviewed about the development of a conversational chatbot named as psykh which uses natural language processing inorder to understand the user inputs. This chatbot ask a few questions to understand the user's current mental state. The answer to those questions gets stored in the Journal called the Happiness Journal and the bot chats through with the user to help him overcome his difficulties. This chatbot has got trained using the RASA architecture.

AyanouzSoufyane et al.'s [8] creation of a medical chatbot is very beneficial for individuals' personal use as well as for medical institutions seeking information about any ailment via text or voice. They combined NLP and the TF-IDF algorithm, which will raise the chatbot's quality. And the retrieval-based paradigm on which this chatbot is built. A chatbot's effectiveness can be improved by utilising more databases and word combinations, which would enable the chatbot to effectively address all ailments.

Nikita Vijay Shinde et al. [9] explored a healthcare chatbot that uses artificial intelligence and can engage with people to answer simple questions about health metrics. This chatbot will compile the users' symptoms and offer medical advice in accordance with them. Natural language processing is used by the chatbot when interacting with users. The chatbot works using the user's input; it uses sentence keywords to make decisions about how to resolve the user's query and provides an appropriate response. TF-IDF, Stemming, n-grams, and cosine similarity are used to execute some of these computations, such as Rank calculation and sentence similarity.

Papiya Mahajan et al. [10] states, For human-machine interaction, a chatbot is an excellent tool. The programme was designed to elicit a prompt response from the bot, so it provides the user with the appropriate response right away. Any person who knows how to text in their native language may use a chatbot, it has been determined. Customized diagnoses with accompanying symptoms are provided by a chatbot.

III. Existing Methodology

Though medical chatbots have proven to be an effective tool in improving patient access to healthcare information, there are some drawbacks to their use. One major concern is the risk of misdiagnosis or incorrect medical advice being given to patients. This can lead to serious health consequences and legal liabilities for healthcare providers. Another concern is the lack of personalization in chatbot interactions, which can lead to a poor user experience and reduced patient engagement. Medical chatbots use different algorithms depending on the specific application or task. Several algorithms can be compared to determine the most effective one for a particular task. For instance, SVM and KNN algorithms can be compared for their ability to classify medical conditions based on patient symptoms. Another comparison that can be made is between Decision Trees and Random Forest algorithms to predict disease outcomes based on patient data. Rule-based algorithms and NLP can be compared to interpret natural language input from patients and provide personalized advice. Finally, deep learning algorithms and traditional ML algorithms can be compared for their ability to analyze complex medical data, such as medical images, and provide accurate diagnoses.

IV. Proposed Methodology

In this paper, we are mainly comparing the performance of combination of NLP techniques such as classification and regression to analyse which technique can improve the performance of the medical chatbot.

Here we are focusing on the fourth phase of the medical chatbot. The bottlenecks of Sentimental Classifications are data quality, Bias, contextual understanding, and explainability. The bottleneck we are focusing on is contextual understanding. Sentiment analysis models may struggle with understanding the context of a user's message, which can lead to incorrect sentiment predictions. In the medical domain, understanding the context is particularly important as it may involve medical conditions, treatments, and emotions. To overcome this challenge, a combination of natural language processing (NLP) techniques can be used. Classification and visualization techniques can be used in conjunction to help medical chatbots overcome context problems in patient queries. Some of the classification techniques are Decision tree, Support vector Machine (SVM), K-Nearest Neighbours (KNN), and Naïve Bayes, and Some of the visualization techniques are Heatmap, word clouds, histograms, Bar graphs, charts, and box plots. Though classification techniques can provide better results while evaluating their performance, the combination of classification and visualization techniques can enhance the performance and provide effective results. Here we are comparing the combination of decision tree and heatmap with Naïve Bayes and word cloud.

A. *Decision tree with Heatmap*

Decision trees are a classification technique that can help chatbots classify patient queries into different categories based on predefined criteria. Heatmaps can then be used to visualize the decision tree and help medical professionals understand how the chatbot is classifying patient queries. This can help identify patterns and relationships in the data and identify areas where the chatbot may need improvement. Decision trees and heatmaps can be combined to create a powerful tool for a medical chatbot. This combination can help the chatbot make more accurate diagnoses and provide more personalized recommendations to patients.

The steps involved in building a decision tree with a heatmap in finding the context of patients in a medical chatbot are as follows:

1. Data Collection:

Data collection which is shown in fig 1 involves gathering data from various sources, such as chatbot conversations, user inputs, and responses. The collected data may include medical data, user demographics, and other relevant information.

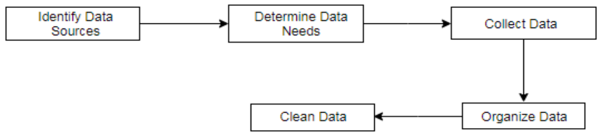

Fig. 1. Data Collection

2. Data pre-processing:

Data pre-processing which is shown in fig 2 is the process of cleaning and transforming raw data into a format that is suitable for analysis. In the context of a medical chatbot, this may involve removing irrelevant information, handling missing values, and ensuring consistency and accuracy.

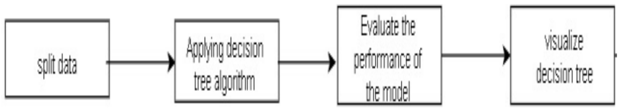

Fig. 2. Data pre processing

3. split the data:

Machine learning requires splitting the available data into training and testing sets, which is a crucial step in the process. The testing set is used to assess the performance of the machine learning model on fresh, unexplored data, whereas the training set is used to develop and train the model.

4. Build a decision tree:

Building a decision tree involves several steps. First, the data is split into training and testing sets. The decision tree algorithm is then applied to the training set to create the tree, with each node representing a decision based on a feature or attribute of the data. Generation of decision tree is shown in fig 3.

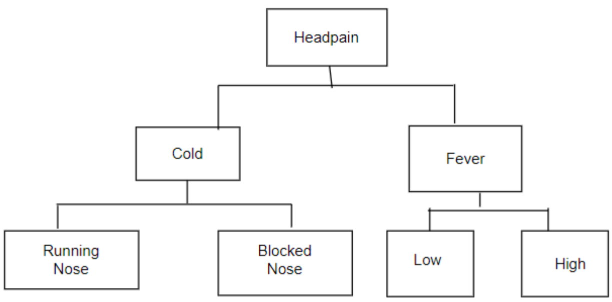

Fig. 3. Generation of heatmap

5. Calculate Accuracy:

To calculate the accuracy score, we first need to split the data into training and testing sets. Then, we apply the decision tree algorithm to the training set to create the model. The accuracy score can be calculated by comparing the predicted context (such as symptoms or conditions) of the patient by the chatbot with the actual context provided by the patient.

Mathematical formula:

$$\text{Accuracy}=\frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Predictions}$$

6. Generate heatmap:

The values of a matrix are depicted as colours in a heatmap, which is a graphical depiction of data. Heatmaps are often used to visualize the distribution of values across two or more dimensions, such as time or location. To generate a heatmap, the data is first organized into a matrix format, with rows representing one variable and columns representing another variable. Each cell in the matrix represents a value, which is then mapped to a color scale.

Pseudocode for generating the heatmap:

6a. Preprocess the patient data to ensure it is in the appropriate format for heatmap generation.

6b. Identify the relevant features for the heatmap generation. These features may include patient symptoms, medical history, demographic information, and any other relevant data points.

6c. Create a matrix to represent the patient data, where each row represents a patient query and each column represents a feature.

6d. Compute a similarity metric between each pair of patient queries. This can be done using techniques such as cosine similarity or Euclidean distance.

6e. Use the similarity matrix to generate the heatmap, where each cell represents the similarity between two patient queries. Cells with higher similarity scores will have a darker color, while cells with lower similarity scores will have a lighter color.

6f. Visualize the heatmap using a library such as matplotlib or seaborn, and customize the visualization as needed (e.g., adjusting color schemes, adding labels, etc.).

6g. Analyze the heatmap to identify patterns and clusters in the patient data. This can help to identify common themes and topics in patient queries and enable the chatbot to provide more personalized and relevant responses.

6h. Use the insights gained from the heatmap analysis to improve the performance of the medical chatbot and provide better support to patients.

7. Evaluate the Decision Tree:

Evaluating a decision tree involves assessing its performance in accurately predicting the labels of the testing set. This can be done by comparing the predicted labels to the actual labels and calculating metrics such as accuracy, precision, recall, and F1 score. Based on the evaluation results, we can make necessary adjustments to the model or dataset to improve its performance.

8. Refine the Decision Tree:

Refining a decision tree is shown in fig 4, which is used for finding the context of the patients in a medical chatbot involves making changes to the model or dataset to improve its performance. One way to refine the decision tree is by adding or removing features from the dataset, which can improve the accuracy and interpretability of the model.

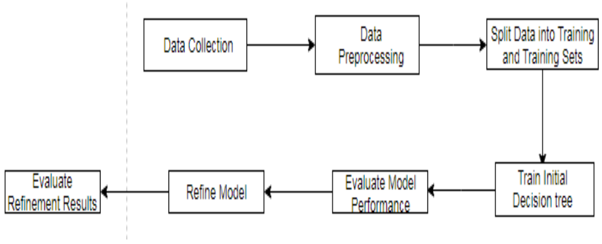

Fig. 4. Refine the decision tree

B. *Naïve Bayes with Word Cloud*

Naive Bayes is a classification technique that can help chatbots classify patient queries based on the likelihood of certain words or phrases occurring in different categories. Word clouds can then be used to visualize the most common words or phrases in each category, helping medical professionals understand the context and meaning of patient queries and identify any potential misclassifications.

The combination of Naive Bayes and word cloud could be a powerful tool for analyzing large amounts of unstructured data in a medical chatbot and could help to identify patterns and insights that might not be apparent from a simple analysis of the raw data.

steps involved in using Naive Bayes with word cloud in finding the context of the patient in a medical chatbot and displaying the accuracy of the performance:

a. Data Collection: Data collection involves gathering data from various sources, such as chatbot conversations, user inputs, and responses. The collected data may include medical data, user demographics, and other relevant information.

b. Data Preprocessing: Preprocess the data by cleaning, transforming, and encoding it into numerical form.

c. Text Preprocessing: Convert the text data into a format that can be used for analysis, such as tokenization, stop-word removal, and stemming.

d. Feature Extraction**:** Extract features from the text data using techniques such as bag-of-words and TF-IDF.

e. Data Splitting: Machine learning requires splitting the available data into training and testing sets, which is a crucial step in the process. The testing set is used to assess the performance of the machine learning model on fresh, unexplored data, whereas the training set is used to develop and train the model.

f. Model Training and Testing: Train a Naive Bayes classifier on the training data and testing data and calculate the accuracy score.

g. Generate Word Cloud: Generate a word cloud to visualize the most frequently occurring words in the dataset. The important term or frequently occurred word is displayed as a collection of words.

Pseudocode:

1. Concatenate all the messages sent by the patient to form a single string
2. Remove stop words from the string
3. Tokenize the string into individual words
4. Perform lemmatization or stemming on the words
5. Create a frequency distribution of the words
6. Create a word cloud using the frequency distribution
7. Display the word cloud to the user

Fig. 5. Generation of word cloud

8. Display Results: Display the accuracy score and the word cloud to show the most common words in the dataset.

V. RESULT AND DISCUSSION

In this study, we compared the performance of two algorithms, namely Decision Tree and Naive Bayes, in finding the context of patients in a medical chatbot. The comparison was based on their accuracy, precision, recall, and F1-score metrics. The dataset used in this study contained information on patients' age, gender, and three symptoms. The symptoms were encoded as categorical variables, and the data was preprocessed before feeding it into the algorithms.

The results show that the Decision Tree algorithm outperformed Naive Bayes in terms of accuracy, precision, recall, and F1-score metrics which is shown as in table1. The accuracy of the Decision Tree algorithm was 97.98%, while that of Naive Bayes was 66.67% which is shown in table 1, 2 and fig 6.

TABLE I. EVALUATION METRICS FOR DECISION TREE

	Accuracy	Precision	Recall	F1 score
Class 1	100%	1.0	1.0	1.0
Class 2	97.98%	0.0	0.0	0.0
Class 3	97.98%	1.0	1.0	1.0

TABLE II. EVALUATION METRICS FOR NAÏVE BAYES

	Accuracy	Precision	Recall	F1 score
Class 1	100%	1.0	1.0	1.0
Class 2	66.67%	0.0	0.0	0.0
Class 3	66.67%	0.0	0.0	0.0

The performance of our proposed method is compared with other NLP techniques. These models result in an accuracy score of 66.67% for Naïve Bayes, 95% for SVM and Decision tree for 97.98% for Decision tree which is shown graphically in fig and table3 represents Performance comparison of proposed method with Naïve Bayes and SVM

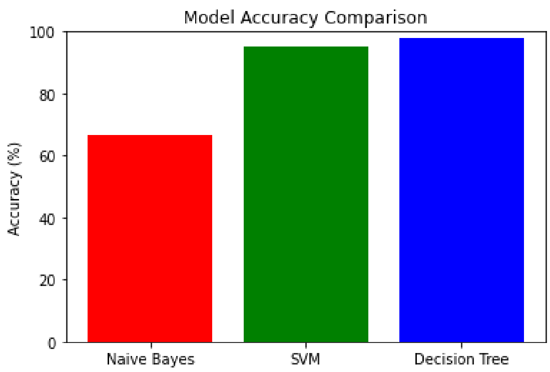

Fig. 6. Model Accuracy Comparison

TABLE III. PERFORMANCE COMPARISON OF PROPOSED METHOD WITH NAÏVE BAYES AND SVM

	Naïve Bayes	SVM	Decision tree
Accuracy	66.67%	95%	97.98%
Performance	33% less efficient compared to Decision tree	4% less efficient compared to Decision tree	Outperforms naïve bayes and SVM by 4-33%

To further analyze the performance of the algorithms, we also created a confusion matrix to visualize the distribution of the predicted and actual classes. The confusion matrix revealed that the Decision Tree algorithm had a higher true positive rate and a lower false positive rate than Naive Bayes. This indicates that the Decision Tree algorithm was better at correctly identifying the patients' context.

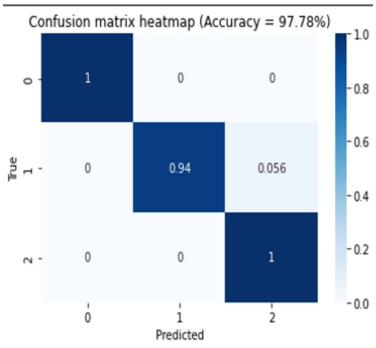

Fig. 7. Confusion matrix of decision tree with heatmap

Naïve bayes algorithm with word cloud analyses the dataset used and find the most frequent symptoms said by the user and displays it as a group of the words which is shown in fig 5.

The Decision tree has enhanced the performance of medical chatbot in finding the context of the patient by boosting the AUC of the model by 33.3%. AUC is an evaluation metric called accuracy score which evaluates the degree of correctness in predictions made by the model. $Accuracy = \frac{((TP+TN))}{TP+TN+FP+FN}$.

$Precision = \frac{TP}{TP+FP}$ is used to evaluate the precision metrics. Precision is the ratio of correct decisions made by the model to the total prediction made on the class instance. The Recall is represented by $Recall = \frac{TP}{TP+FN}$ and F-Score is represented by $F1\ score = \frac{((Precision*Sensitivity))}{Precision+Sensitivity}$

By comparing these techniques, we analyse that the Decision tree and heatmap provide an accuracy of 97.78% and a combination of Naïve Bayes and Word cloud provides the accuracy of 66% and SVM provide an accuracy of 95%. So, we conclude that the combination of a decision tree and Heatmap provides better accuracy.

VI. Conclusion

In conclusion, the combination of visualization and classification techniques can greatly enhance the performance of a medical chatbot. By using visualizations, such as graphs or charts, the chatbot can better present data to users and make it easier for them to understand their health information. Additionally, classification techniques can help the chatbot more accurately identify and diagnose medical conditions, allowing for more effective treatment recommendations. To improve the performance of a combination of decision tree and heatmap for a medical chatbot, one booster technique that could be used is ensemble learning. Ensemble learning involves combining multiple models to create a stronger overall model that can make more accurate predictions. By combining the strengths of the decision tree and heatmap models, the ensemble model may be able to make more accurate predictions than either model alone. Additionally, using a neural network as the final model may allow the chatbot to learn more complex patterns in the data.

References

[1] Hiba Hussain, Komal Aswani, Mahima Gupta, Dr. G.T.Thampi "Implementation of Disease Prediction Chatbot and Report Analyzer using the Concepts of NLP, Machine Learning and OR" IRJET, April 2020.

[2] Vytautas Raulinaitis, "Evaluation of chatbots used in healthcare," Haaga-Helia University of Applied Sciences, Degree Programme in Business Information Technology, Bachelor's Thesis, 2022

[3] Sanjay Chakraborty, Hrithik Paul "An AI-Based Medical Chatbot Model for Infectious Disease Prediction", Research Gate, January 2023.

[4] Himanshu Gadge, Vaibhav Tode, Sudarshan Madane "A Chatbot for Medical Purpose using Deep Learning" IJERT, May-2021.

[5] Maria V. Vasileiou, Ilias G. Maglogiannis "The Health ChatBots in Telemedicine: Intelligent Dialog System for Remote Support " Journal of Healthcare Engineering, Hindawi,2022.

[6] Manal Almalki, Fahad Azeez "Health Chatbots for Fighting COVID-19: a Scoping Review" ACTA Inform Med, December 2020.

[7] Aniket Patole, Vishvesh Dumbre, RitikKesharwani, Harmeet Khanuja "Mental Health Chatbot (Psykh)" IRJET, March 2021.

[8] Ayanouz Soufyane, Boudhir Anouar Abdelhakim, and Mohamed Ben Ahmed "An Intelligent Chatbot Using NLP and TF-IDF Algorithm for Text Understanding Applied to the Medical Field" IEREK, August 2021.

[9] Nikita Vijay Shinde; Aniket Akhade; Pranali Bagad; Harshit Bhavsar; S.K. Wagh; Amol Kamble "Healthcare Chatbot System using Artificial Intelligence" 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), June 2021.

[10] Papiya Mahajan, Rinku Wankhade, Anup Jawade, Pragati Dange, Aishwarya Bhoge "Healthcare Chatbot using Natural Language Processing" International Research Journal of Engineering and Technology (IRJET), November 2020.