‘K-Bot’ Knowledge Enabled Personalized Healthcare Chatbot

Rahul Pradhan¹, Jaya Shukla² and Mani Bansal³
GLA University, Mathura UP 281406, India

¹raul.pradhan@glu.ac.in, ²jaya.shukla_cs18@glu.ac.in, ³mani.bansal_cs18@glu.ac.in

Abstract: In the present scenario when most of the things are just a click away, people consider visiting hospitals as the most effective, reliable, and convenient way for their routine check-ups or disease diagnosis. In the past few years, Machine Learning has been playing a major role in the domain of healthcare. The proposed system or approach focuses on creating an alternate using Decision Tree Algorithm where people can interact with chatbot and it will identify other symptoms and predict the disease along with the confidence level and thus recommending a specialized doctor. Using the above framework can help people save their time and money as well.

Keywords: Diagnosis, Machine Learning, Healthcare, Chatbot, Symptoms, Decision Tree Algorithm, Disease

1. Introduction

We are in better surroundings when people around us are in good physical and mental health as well. As it is said that a healthy mind resides in a healthy body [1]. It has come to the light that over 95% of the population is affected by some major or minor health issues [2]. Recent studies have shown that more than 50% of people visiting a doctor are suffering from simple small-scale diseases such as cold and cough, flu, headache, fever, nausea, sore throat, etc. Most of which can be cured by simple home remedies without even paying a visit to a doctor [3].

In the past few years, Machine Learning has played a major role in the field of healthcare because of its ability to learn and analyse from the patterns and the experiences over a period of time [4].

The fast-pacing world around us is focusing on finding ways on how to live a healthy lifestyle to lower their risk of getting prone to any disease. In such a situation, a chatbot primarily focused on the medical domain can play a major role in monitoring a person’s health [2].

There a number of chatbots available with different functionalities in various domains such as entertainment, market, healthcare, e-commerce, etc.

Two of the broadly used chatbots are ‘Alexa’ and ‘Google Assistant’ which are the forefronts of personalized virtual assistants [3].

Few of the most innovative chatbots in the healthcare sector are:

- ‘Endurance’ that deals with suspected patients with Alzheimer’s and other forms of Dementia.
- ‘Casper’ helps people in dealing with Insomnia.
The basic functionality of a chatbot is to direct and assist the user in disease detection and thus help in leading a healthy lifestyle.

In this fast-pacing world, people tend to forget to take possible measures or precautions to maintain a healthy life. They do not pay much attention to their health. Most of the working section of society compromises with their health. Today's generation is more likely addicted to social media but is not concerned for their personal health [1].

Consider a scenario that you know someone who is a doctor with whom you can have a random conversation about your health and he diagnoses you for a disease by asking a few more questions. But not everyone knows a friend who is a medical consultant, so they end up not paying attention to their health.

In today's era, chatbots have become the most efficient, advanced, and time-saving technology. People consider visiting a doctor as the most convenient way to keep a check on their health. This paper aims at providing an alternate to this traditional way of visiting a doctor to get the treatment [1].

User / Patient needs to register or sign in to interact with the bot just like they interact with other human beings and through a number of questions, the chatbot will recognize the other related symptoms with respect to that particular symptom provided by the user and thus, predict the disease and confidence level and the recommended specialized doctor with a link that will redirect them to the website so that they can consult the doctor and book an appointment on the basis of the reviews and ratings provided by the other users.

This proposed framework can help people save their time and money and focus on making a chatbot that is free of cost and is available 24*7.

This paper aims at creating a chatbot that is highly trained on a dataset that contains the various symptoms and their respective prognosis (disease), and the dataset being split into training and testing dataset. The primary focus is on applying the Decision Tree Algorithm which is a Supervised Learning ML Algorithm. Here, the output is based on datasets on which it is trained.

The main objective is to build a Health Care Domain Chat-Bot for a ‘pragmatic approach for diagnosis’ using Machine Learning and TKinter (GUI for Python).

2. Related Work

Table 1 will show the brief summary with the basic idea and identified research gap in current state of art in field of healthcare chatbots.

| Author      | Title                                      | Approach/Model/Basic Idea | Dataset Used                                                                 | Pros                                                                 | Cons                                                                 |
|-------------|--------------------------------------------|----------------------------|------------------------------------------------------------------------------|----------------------------------------------------------------------|----------------------------------------------------------------------|
| Mathew et. al. [1] | Disease Prediction and Treatment Recommendation Chatbot | K-nearest neighbor Algorithm (KNN) | They used symptoms-disease dataset which contains general information about symptoms and diseases. | It analyzes symptoms of different users and then predicts the disease to the user. | Symptoms and disease identified cannot be made into a report.         |
| Rahman et. al. [2] | Bangla Healthcare Chatbot using Machine Learning | Support Vector Machine Algorithm (SVM) | Customized Bangla datasets and a dataset containing information from several hospitals, doctors. | Maintains record of the health of the user, performs related events | Incapable of taking sufficient symptoms that may give an |
| Authors               | Title                                                                 | Language Model                                      | Dataset Description                                                                 | Predicts Health Issues | Additional Features                                                                 |
|----------------------|-----------------------------------------------------------------------|-----------------------------------------------------|-------------------------------------------------------------------------------------|------------------------|-------------------------------------------------------------------------------------|
| Dharwadkar et al. [5]| A Medical ChatBot                                                     | Natural Language Processing (NLP) and Support Vector Machine Algorithm (SVM) | Heart Disease Dataset for 4 different cities of different sizes is used. Users can ask dosage related queries by voice. The system speaks out and display medicine. Cannot be used as a voice or face recognizer and cannot interact with the patient at deep levels. |
| Jasnani et al. [6]   | Contextual Chatbot                                                    | Natural Language Processing (NLP) and Bag of Words Model | JSON file data containing the intents will be denoted by tags that contain formats in which the user will give input. User can understand problems they are facing and get a prediction about the disease they might have. User cannot use any other language interface except text. |
| Nalinipriya et al. [7]| BayMax: A Smart Healthcare System                                     | Logistic Regression Algorithm                       | The dataset contains diseases and their respective symptoms and basic advice to improve the health of the user. Predicts the possible health issues and necessary measures to be taken to cure it. Does not supports data pipelining that can automate all the processes and provide deep understandings of the input data. |
| Amato et al. [8]     | Chatbots meet eHealth: automatizing healthcare                         | Ensemble (Random Forest Algorithm)                  | Dataset provided by C.M.O. Centre which contains 13 different diseases and a record of a total of 16733 patients. The proposed framework is adaptable to different clinical scenarios and medical tasks. Does not adapts to the user’s need, human behavior and characteristics such as age, gender, |
Bharti et al. [9] | Conversational A.I. Chatbot for Delivering Tele-Health | Rule-based grammar matching Algorithm and ML matching Algorithm | Data issued by the National Health Mission in the mid of COVID-19. | Reduces the gap between the rural population by understanding their needs and providing them the proper information and the possible precaution to be taken. | Does not cope with assistance for an anti-depression system that will focus on providing music therapy and mental health tests.

Chen et al. [10] | Disease Prediction over Big Data from Healthcare Communities | Convolutional Neural Network (CNN) based unimodal disease risk prediction (MDRP) algorithm | Hospital data collected from Central China which contains EHR, medical image and gene data of 20320848 records. | It works on both structured as well as unstructured data of the hospital dataset. | There is a large amount of missing data because of human error which needs to be imputed.

3. Proposed Approach
A chatbot is a virtual assistant that automates the interaction with the user. A chatbot aims in making a conversation between a human and a machine without the involvement of any other human being. Chatbots are powered by artificial intelligence using machine learning techniques and file handling.

The primary objective of this paper is to predict the disease a person is suffering from based on the symptoms present and predicted.

In the proposed system, we have tried to change the functioning of a chatbot from being only used as a way of communication to predict disease based on the symptoms present. To achieve this, the system uses the Decision Tree Algorithm. The proposed approach is as follows:

- **The user interacts with the chatbot:**
  The user starts a conversation with the chatbot by registering or logging in into the system and then providing input in the form of Yes/No or yes/no through buttons. It is just as simple as starting a conversation with any human being over text. The conversation between the user and the chatbot is completely text-based.
  The registration and login page for the user where he can either log in into his account or register for a new account if he is not registered is shown in the figure 1.
Figure 1. Registration and login page.

The registration page for a new user where he can register using username and a password is shown in the figure 2.

Figure 2. New user registration page.

- **Decision Making:**
  As soon as the conversation starts between the chatbot and the user, the user is asked to respond to the symptoms provided in the form of Yes/No. Based on the responses of the user, the chatbot decides if it has to continue the conversation or to stop the conversation and provide the result of the predicted disease based on the symptoms.
  
  The conversation between the chatbot and user where user responds to the questions asked by the chatbot related to the various symptoms is shown in the figure 3.

Figure 3. Conversation window between user and chatbot.

- **Response from chatbot:**
  The chatbot responds on the basis of the decision making in the previous stage. Then, it starts asking about different symptoms and then finally predicts a disease on the basis of the
symptoms present and given. The user can visit the website using the link generated and consult the recommended doctor from there. All the response from the chatbot is text-based. The output or the prediction made by the chatbot which includes the disease and the symptoms present and given and the confidence level for the particular disease is shown in the figure 4.

Figure 4. Disease prediction on the basis of symptoms and its confidence level.

4. Datasets used
For the proposed system, we have used two datasets. These are:

- **Dataset containing the symptoms and the corresponding diseases:**
  The dataset (figure 5) being split into training and testing dataset for building the model, contains the various symptoms and their respective prognosis (disease).

Figure 5. Dataset containing the symptoms and the corresponding diseases.
Dataset containing the doctor’s information:
The data for this dataset (figure 6) is scrapped from the website Practo using web scraping, and contains information such as doctor’s name and a link to the website where the user can get all the information about the doctor for a particular city.

Figure 6. Dataset containing the doctor’s information.

5. Algorithm used
The algorithm on which the proposed system is built is the Decision Tree Algorithm.
Decision Tree Algorithm belongs to the deterministic family of classifiers and is a model-based technique. It creates a tree-like structure from the training dataset. This tree is then traversed to generate the predictions. After the creation of the tree, it is extremely easy for the algorithm to generate the predictions for the new instances [11].

The advantage of the Decision Tree Algorithm is that it can deduce numerous relationships in the training data and attempts to always get an accuracy of around 100%. But it can also be thought of as a disadvantage of the Decision Tree Algorithm.

However, a number of questions remained unanswered. These are:

- How does the algorithm decide which attribute and threshold to choose for a particular node?
- How does the algorithm decide what should be the optimal value for the depth of the tree?

Decision Tree Algorithm is the most powerful Machine Learning algorithm to have existed. Despite all its flaws, it can work well with moderately large and complex datasets with 'n' different classes.

The Decision Tree implementation of sci-kit learn will continue creating the tree with more and more nodes unless and until all instances are correctly classified. This phenomenon is known as overfitting.

Decision Tree Algorithm uses an impurity metric to create an optimal split. In other words, to divide a dataset as such that instances of different should be separated, i.e., the two most common metrics of impurity used by Decision Tree Algorithm are:

- Gini Impurity
- Entropy or Information Gain
\[ G_i = 1 - \sum_{k=1}^{K} p_{ik}^2 \]  
\[ H_i = -\sum_{k=1}^{K} P_{ik} \log(P_{ik}) \]  

Here, \( P_{ik} \) is the ratio of ‘k’ class instances at \( i \)th node.

After deducing \( G_i \) (Gini Index) or \( H_i \) (Entropy), the decision tree algorithm then attempts to minimize a cost function to get the optimal pair of \((k, t_k)\) where \( k \) is attribute name and \( t_k \) is its threshold value for a particular node.

\[ J(k, t_k) = \left( \frac{m_{left}}{m} \right) G_{left} + \left( \frac{m_{right}}{m} \right) G_{right} \]  

Here, \( J \) is the cost function whose value is calculated for every pair of \((k, t_k)\) from the training dataset. The pair that yields the smallest ‘\( J \)’ value becomes the root node.

Thus, the Decision Tree Algorithm works for moderately large datasets not extremely large datasets. The algorithmic implementation of Decision Tree Algorithm is shown in the figure 7.

\[ \text{Figure 7. Decision Tree Algorithm Implementation} \]

6. Result Analysis

From the analysis of the model by using different algorithms such as Logistic Regression, K Nearest Neighbours, Support Vector Machine and Decision Tree Classifier, the Decision Tree Algorithm proved to be the best suitable algorithm for the system. This model is trained to predict the prognosis(disease) for the symptoms present and given on the basis of the user input for the trained features. The predicted outcome or result was found to give an accuracy of around 0.94 (94%).

Also, by calculating the wall time, i.e., the total time required by a computer to execute a program, the decision tree algorithm proved to be better than other algorithms. The wall time for different algorithms is shown in table 2.
Table 2. Wall time for different algorithms.

| Algorithm                        | Wall Time (s/ms) |
|----------------------------------|------------------|
| Logistic Regression              | 1.02 s           |
| K Nearest Neighbours (KNN)       | 141 ms           |
| Support Vector Machine           | 659 ms           |
| Decision Tree Classifier         | 88.6 ms          |

Thus, this also proves that the Decision Tree Algorithm as compared to other algorithms is a good choice for the purpose.

7. Conclusion
This paper aims at providing a healthcare domain chatbot that can be used as an alternate to the traditional way of visiting a doctor and getting treatment. The user can respond to the symptoms and in response to that, the chatbot will predict the disease.

The dataset contains the symptoms and the respective disease(prognosis). After analyzing the symptoms, the chatbot finally predicts the disease and the confidence level to the user and provides a link to consult the recommended doctor from where the user can set up an appointment with the doctor based on their availability.

Our Healthcare Chatbot will have a great impact on the life of its users. It gives them the freedom to consult a doctor 24*7 and also can communicate with the doctor directly. For minor health queries, people won't have to hamper their schedule to consult a doctor. User can freely ask their health-related queries by a chatbot.

This chatbot allows users to select the symptom that they feel. This project is developed by keeping the idea in mind that the user saves their time for consulting the doctors related to their health issues.

8. Future Scope
The role of a chatbot can sometimes be beyond thinking. The chatbot can be beneficial for the user if the symptoms or the disease identified by the system can be generated into a report and forwarded to an available doctor where he/she can guide the user with more advice and preventive measures to maintain their health.

Also, it can be modified in such a way that it can be used as a voice or face recognizer and can interact with the patient or user at deep levels so, that the user can also use any other language interface except text or buttons to interact with the chatbot.

References
[1] Mathew R B, Varghese S, Joy S E and Alex S S 2019 Chatbot for Disease Prediction and Treatment Recommendation using Machine Learning Third International Conference on Trends in Electronics and Informatics
[2] Rahman M M, Amin R, Liton M N K and Hossain N 2019 Disha: An Implementation of Machine Learning Based Bangla Healthcare Chatbot 22nd International Conference of Computer and Information Technology 1–6
[3] Bhirud N, Tatale S, Randive S and Nahar S 2019 A Literature Review On Chatbots In Healthcare Domain International Journal of Scientific & Technology Research 8 225–31
[4] C. G P K, Ranjan S, Kumar V and Ankit T 2019 A Personalized Medical Assistant Chatbot: MediBot International Journal of Science Technology & Engineering 5 42–6
[5] Dharwadkar M R and Deshpande M N A 2018 A Medical ChatBot International Journal of Computer Trends and Technology (IJCTT) 60 41–5
[6] Kandpal P, Jasnani K, Raut R and Bhorge S 2020 Contextual Chatbot for Healthcare Purposes
(using Deep Learning) 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) 625–34

[7] Nalinipriya G, Priyadarshini P, S P S and RajaRajeshwari K 2019 BayMax: A Smart Healthcare System Provide Services to Millennials Using Machine Learning Technique IEEE 6th International Conference on smart structures and systems ICSSS 2019

[8] Amato F, Marrone S, Moscato V, Piantadosi G, Picariello A and Sansone C 2017 Chatbots meet eHealth: automatizing healthcare Workshop on Artificial Intelligence with Application in Health 40–9

[9] Bharti U, Bajaj D, Batra H, Lalit S, Lalit S and Gangwani A 2020 Medbot: Conversational Artificial Intelligence Powered Chatbot for Delivering Tele-Health after COVID-19 Fifth International Conference on Communication and Electronics Systems 870–5

[10] Chen M, Hao Y, Hwang K, Wang L and Wang L 2017 Disease Prediction by Machine Learning over Big Data from Healthcare Communities IEEE Access 5 8869–79

[11] Wikipedia contributors 2020 Decision tree learning Wikipedia, The Free Encyclopedia