



Medical Chatbot (Medibot)

Aadhil Biju¹, Nikhil.S.Menon², Sharon Lijo Joseph³, Darrel Lopez⁴, Sinijoy PJ⁵

¹Department of Computer Science and Engineering, AISAT, Kalamassery, India, aadhilbiju07@gmail.com

²Department of Computer Science and Engineering, AISAT, Kalamassery, India, immikhilmenon@gmail.com

³Department of Computer Science and Engineering, AISAT, Kalamassery, India, brave.sharon@gmail.com

⁴Department of Computer Science and Engineering, AISAT, Kalamassery, India, darellopez555@gmail.com

⁵Department of Computer Science and Engineering, AISAT, Kalamassery, India, sinijoypj@aisat.ac.in

ABSTRACT

In recent years, medical services are getting more open to a more extensive gathering of individuals with the help of innovation. The ideas of computerized reasoning, machine learning and neural organizations have given generous help with the field of medical services. In the present quick moving world, individuals will in general disregard their wellbeing which may result in a basic issue. Such an issue can be kept away from by utilizing the symptom-driven chatbot. Our project focuses on giving the clients accurate and precise prediction of the disease based on their symptom, The disease envisaging chatbot is created utilizing a Decision tree. For the forecast of sicknesses, we have utilized two arrangement calculations specifically, Decision tree and KNN (K-Nearest Neighbours). The presentation of these strategies are analyzed and dependent on their accuracy, the best model is chosen. According to our project, the accuracy of the Decision Tree and KNN are 95% and 92.5% respectively. Our medical chatbot also provides precautions and measures for the envisaged disease. Our medical chatbot also has a text-to-speech feature., for providing a human element to the conversation, rather than it being an automated conversation, which may feel mundane to the user.

Keywords: Chatbot, Medical Chatbot, Machine Learning, Decision Tree, Classifier, Tkinter

1. INTRODUCTION

Since beyond a few many years, human beings had been tirelessly running day in and time out that they fail to prioritize their fitness on a normal foundation. In the longer run, this hassle ends in jeopardizing a pleasant life. Nevertheless, with the aid of machine learning, we will now offer fitness care offerings to people at their comfort at reasonable expenses. One of the biggest blessings we possess is a healthy body. As the saying goes - "Health is Wealth", a healthy body and a more desirable lifestyle is something each one of us strives to achieve.

It's a common fact that most people are quite oblivious to most diseases and lack knowledge of them, by not usually being aware of all the symptoms of the said disease. So for small health problems, either they resort to the dangerous practice of self-medication, which is condemned by most medical practitioners, or personally go to hospitals for check-ups which are often time-consuming for daily commuters & for people in an area which faces transportation difficulties. Such people may choose to ignore, which may have deep repercussions. This is mainly because of their hectic schedule and not because they chose to abandon their health. Such a problem can be solved by using a medical chatbot.

2. LITERATURE SURVEY

Machine learning makes use of algorithms that examine an appropriate variety of input facts to make predictions. ML algorithms optimize their operations. optimal results are key to working for any practical application. The accuracy of those algorithms will increase with the input of easy information. Different supervised algorithms are used for disease prediction. With the assistance of a labelled training dataset, the supervised machine learning algorithms are trained first. To categorize the unlabeled dataset into similar groups, training a set of rules is done to the dataset. Supervised learning algorithms can be used for both classification and regression. Support Vector Machines (SVM) is a supervised machine learning. SVM separate features into multiple classes using an optimal separating line or using a kernel function. In order to gain a notion of the use of SVM within the healthcare subject, [1] we analysed a research report based totally on the prediction of Chronic Kidney Disease via the manner of Almansor et al. found that Artificial Networks (ANN) had more accuracy than SVM. SVM has higher accuracy values, as compared to other machine-learning models. However, Support Vector Machine has one foremost downside; despite the fact that the accuracy rate is immoderate, it takes a copious amount of time for its training phase. The Naive Bayes Algorithm, which is an extremely good ML algorithm, is based totally on the Bayes theorem. Which makes use of a probabilistic classifier. Probabilistic reasoning is used while the records treated through the device are massive and complex. [2] Research on Bayes Algorithms is finished via Pattekari&Parveen. Based on diverse input attributes, prediction of coronary heart disease executed the use of ML and records

mining techniques. As a conclusion of these studies, the proposed approach changed into proved to be the maximum reliable approach to are awaiting sufferers ill with coronary heart illness. However, it has no application in any real-global datasets. Decision tree is a renowned machine learning algorithm. A decision tree version classifies data properly right into a tree-like form. There are a couple of tiers and in each tier, there are more than one nodes. The top-most node is referred to as the root node and the internal nodes constitute input values of the tree. We have carried out a Decision tree in our illness prediction chatbot, however, we were given its accuracy greater than the KNN classifier. The K-Nearest Neighbours (KNN) set of guidelines is used for fine-tuning classification problems.[1] Based on measures of dissimilarity, it classifies new data instances. The KNN uses any of the measures of dissimilarity to calculate the distance between the testing and training phase samples inside the multidimensional area, choosing the nearest distance from it. According to the type of the sample points in training, the K points found. The KNN algorithm technique is a proven passive classification approach to data models. The KNN algorithm trains the sample information by means of itself, consuming extra time inside the classification thereby taking less time in the training phase. KNN is easy to implement.KNN is dependent on surrounding neighbouring samples.. Due to the essential benefits of the KNN algorithm, a variety of research has been carried out to remedy sickness prediction issues and the use of the KNN algorithm. In 2016, [3] a coronary heart ailment prediction system was proposed through the usage of Princy& Thomas the usage of techniques of statistics mining that centred at the KNN and ID3 algorithms. In KNN, the most important step is value selection for K. This is due to the fact it can have an enormous effect on accuracy. In our project, the accuracy of KNN is 92.5% and the accuracy of the decision tree is 95%.

3. PROPOSED SYSTEM

The chatbot in our project is used for information acquisition, i.e; retrieval of symptoms..Being a hybrid chatbot, It acquires the patient’s information along with the symptoms and the disease is predicted on the basis of the symptoms and does this on a multi-turn based level. The disease prediction chatbot is designed using the concepts of machine learning algorithm chatbot - a hybrid model which uses 2 machine learning algorithms. Upon receiving a single symptom from the user, multiple related questions are asked to the user regarding the first entered symptom. After getting a clear idea of the disease based on the symptoms entered by the user, The result, as well as the required precautions to take, is displayed. The dataset used in the project comprises a total of 40 diseases and 132symptoms.The project uses 5 CSV files(Training,Testing,Symptoms_discriptive,Symptoms_severity,Symptom_Precaution) .Decision Tree and KNN are the 2 machine learning algorithms that we have used for envisaging the type of disease. Both classifiers were trained and tested using training.csv and test.csv respectively. The chatbot is triggered by the user by entering a single symptom to the chatbot which he or she experiences and later on the chatbot ask some questions based on the symptom entered by the user and then it envisages the type of disease the user has and also provides some measures to be taken for recovering from the disease. The accuracy of the

Decision tree and KNN are 95% and 92.5% respectively. For creating our disease prediction medibot we have used jupyternotebook.The datasets were taken from Kaggle

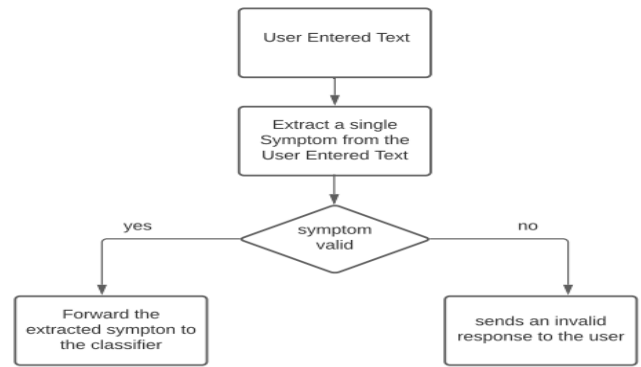


Figure 1: Flowchart of Chatbot

The chatbot checks whether the symptom entered by the user is valid or not. If the symptom is valid then the symptom is forwarded to the classifier. If the entered symptom is invalid then the chatbot sends an invalid response and asks the user to enter a valid symptom. Our chatbot uses 2 machine learning algorithms namely KNN and Decision Tree to envisage the type of disease the user has based on the symptom entered by the user. In KNN symptoms are scattered or plotted in a graph. New Data instance(the symptom extracted from the user text) is plotted in the graph.K nearest neighbours to the new data instance are found and the disease is predicted. For finding the K neighbours we are using Euclidean distance and the value of K we have chosen is 5. In the decision tree algorithm, the internal nodes are represented as symptoms and the leaf nodes are represented as diseases. The threshold value of the decision tree was used for creating the left and right subtree. The threshold value of the decision tree was 0.5.

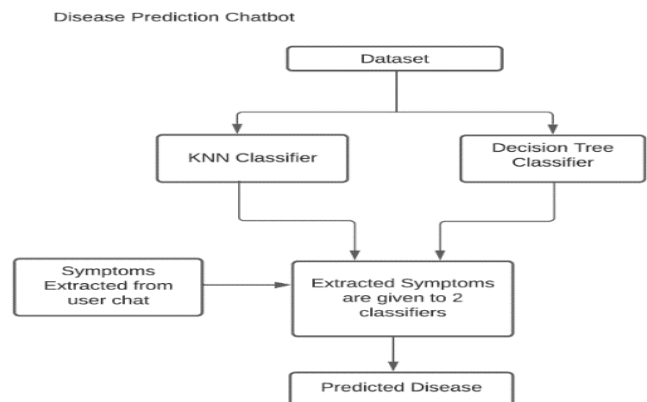


Figure 2: Block diagram of Disease Prediction System

The disease prediction dataset includes overall 132 symptoms and 40 diseases. The symptoms work as attributes,

based totally on which affected person’s suffering is anticipated and disease represents the record of the dataset. There in dataset 1 shows that the following symptom is present for the disease and 0 indicates that the symptoms aren't experienced for that disease.

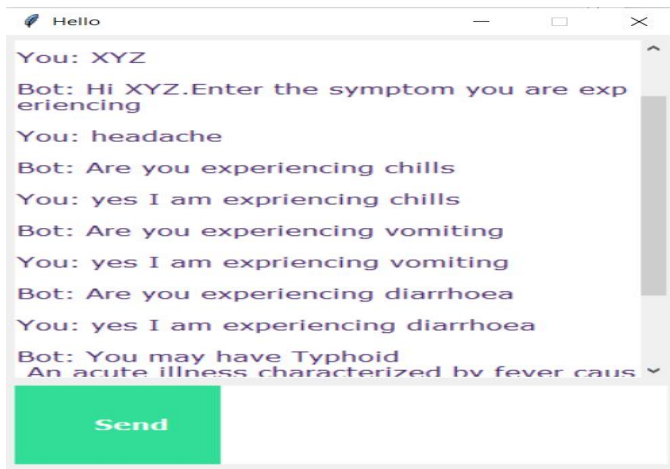


Figure 3: GUI of the MediBot

The GUI of our chatbot was created using Tkinter. The geometry() method in Tkinter was used to set the dimensions of the root window and is used to set the position of the main window on the user’s desktop. Theresizable() method in the Tkinter library was used to change the size of the root window. Using the get() method we were able to fetch the user-entered text from the Tkinter window and the response was printed in the Tkinter window using the insert() method. Our medical chatbot also has a text-to-speech feature., for providing a human element to the conversation, rather than it being an automated conversation, which may feel mundane to the user. For enabling the text-to-speech feature we have used the pyttsx3 module.

Table 1: List of diseases

Fungal infection	Alcoholic hepatitis
Allergy	Tuberculosis
GERD	Common Cold
Chronic cholestasis	pneumonia
Drug Reaction	Dimorphic haemorrhoids(piles)
Peptic ulcer disease	Heart attack
AIDS	Varicose veins
Diabetes	Hypothyroidism
Gastroenteritis	Hyperthyroidism
Bronchial Asthma	Hypoglycaemia
Hypertension	Osteoarthritis
Migraine	Arthritis
Cervical spondylosis	Vertigo
Paralysis	Acne
Jaundice	Urinary tract infection
Malaria	Psoriasis
Chicken pox	Impetigo
Dengue	Hepatitis B
Typhoid	Hepatitis C
hepatitis A	Hepatitis D

The above table shows the 40 diseases present in our dataset.

4. RESULT

Table 2:Algorithm Comparison

Table head	Comparison of Algorithm			
	Algorithm	Disease count	Answer accuracy	% Accuracy
1	DT	40	38	95%
2	KNN	40	37	92.5%

The accuracy for the classifier is calculated using the equation (answer accuracy/disease count)*100. In the above dataset, the total disease count is 40. Out of 40 diseases, KNN correctly predicted 37 diseases and Decision tree (DT) predicted 38 diseases. The accuracy of KNN is (37/40)*100 = 92.5% and the accuracy of the decision tree is (38/40)*100 = 95%.

Table 3: Algorithm accuracy and error rate.

Algorithm	Accuracy	Error Rate
KNN	92.5%	7.5%
DT	95%	5%

The accuracy of Decision Tree and KNN is 95% and 92.5% respectively. The error rate or the misclassification rate of the Decision tree and KNN is 5% and 7.5%. The error rate of the classifier is calculated using the formula (1-(accuracy/100))*100.

5. CONCLUSION

Our project represents a prediction model-driven system that predicts accurate diseases based on the symptoms as well as suggests a preventive/precautionary measure. An interactive chatbot enables it to retrieve symptoms provided by the user. The chatbot also has a text to speech functionality for more accessibility. The prediction model is designed using Machine learning algorithms such as KNN and Decision Tree. Both the algorithms were applied to the same dataset. Multiple questions regarding the user's symptoms are asked and then the disease or ailment that troubles the user is displayed, along with a layman's description as well as a precautionary measure, advising the user on the next step. As per our results, the accuracy of Decision Tree and KNN are 95% and 92.5% respectively. Hence, the decision tree gave better results in this dataset. The model is predominantly used to classify 40 diseases. We integrated classification in the system in order to provide an application which the patients can utilize effortlessly. We believe that this approach incorporated into existing strategies in the field of healthcare will provide assistance to doctors, nurses, researchers, health enthusiasts and most of all, the patients.

REFERENCES

- [1] Yu, Hong. (2019). **Experimental Disease Prediction Research on Combining Natural Language Processing and Machine Learning**. 145-1500.1109/ICCSNT47585.2019.8962507.
- [2] Akhil, Jabbar&Samreen, Shirina. (2016). **Heart disease prediction system based on hidden naïve Bayes classifier**. 10.1109/CIMCA.2016.8053261.
- [3] Enriko, I Ketut&Suryanegara, Muhammad &Gunawan, Dinda. (2016). **Heart disease prediction system using a k-Nearest Neighbour algorithm with simplified patient's health parameters**. 8. 59-65.