



Disease prediction model using machine learning algorithms

Prachi Lakde¹, P.S.Kokare², Deshmukh Aakanksha³, More Rutuja⁴, B. A. Patil⁵

Dept. of E&TC, SKNCOE, Pune-411041, India.

¹prachilakde2001@gmail.com ²premskokare@gmail.com

³aakanshad234@gmail.com ⁴rutujamore200@gmail.com

⁵bhoomi.patil_skncoe@sinhgad.edu

Abstract— Hospitals are the most broadly used capability by using the usage of which a sick character receives scientific check-ups, health problem analysis and remedy recommendation. This has been a workout by means of the use of almost all the human beings over the world. People suppose it as the most reliable skill to test their fitness status. The proposed machine is to create an alternative to this conventional technique of visiting a fitness facility and making an appointment with a medical doctor to get a diagnosis. In this machine we are making use of the principles of computing gadget and reading to create a chatbot application. People can engage with the chatbot simply like they do with each and every other human and via a sequence of queries, the chatbot will predict the disease. This device can be of great use to people in conducting day with the aid of day check-ups, makes human beings conscious of their fitness recognition and encourages people to make suitable measures to proceed to be healthy. According to this research, such a gadget is no longer broadly used and human beings are much less aware of it. Executing this proposed machine can assist human beings preserve away from the time-consuming approach of journeying hospitals by means of the usage of this free of value application, at any place they are.

Keywords — Medical chatbot, Machine Learning, Disease Prediction, Random Forest, Decision Tree, Naïve Bayes, KNN.

I. INTRODUCTION

An affluent society is when its complete humans are healthy. It is important to preserve fitness if one desires to be happy. Only a wholesome body can have a healthy thought and it has an effective influence on the performance of people. Nowadays, humans are less aware of their health. In their busy life, they forget to take suitable measures to preserve their health and are much less aware of their fitness status. In this proposed system, a scientific chatbot is built to be a selection-based agent that motivates customers to select symptoms about their health troubles and based on the signs and

symptoms supplied by them, the chatbot returns the diagnosis. This chatbot gadget will be capable of becoming aware of signs from consumer interaction. Using these extracted symptoms, chatbot predicts the disease. The computing device gaining knowledge of algorithms employed here honestly shows that a clinical chatbot can extremely precisely diagnose sufferers with simple symptom analysis and a selection-based approach.

The fundamental objective of the proposed machine is to have the significance of health in lifestyles reach out to human beings and motivate humans to comply with measures to hold fitness by making the chatbot handy to all. Chatbot and fitness have a record of working well together. By the usage of such a system, we can predict disease and the usage of machine learning algorithms. Machine learning algorithms hire a range of statistical, probabilistic and optimization strategies to learn from previous rides and realize beneficial patterns from large, unstructured and complicated datasets. Machine learning algorithms have been a dominant approach in the statistics mining field. Disease prediction the usage of fitness facts has these days proven a workable software area for these methods. This study aims to discover the key traits among distinct sorts of supervised machine learning algorithms, and their overall performance and usage for disease threat prediction.

The motivation behind our project is that these days people are less aware about their health because of their hectic schedule this system will help them to maintain their health. If people want immediate help they can use this system. People can enter the symptoms and based on that system will predict the disease. This project is designed to predict diseases using machine learning algorithms. In this we are using four machine learning algorithms that are Decision Tree and Random Forest, KNN and Naïve Bayes We are using GUI for building labels, message box, title and option menu.

In this paper, a generic structure has been proposed for predicting sickness in the healthcare industry. This system was once experimented with decreased set points of the Diseases dataset using an increased machine learning algorithm, and additionally this system has been compared with different machine learning algorithms such as Random Forest and Decision tree. The overall performance of all these machine learning algorithms has been evaluated with accuracy, misclassification rate, precision, sensitivity and specificity. From the experiment results, the machine learning algorithms produce accuracy of 97.5%, 97.5%, 96.8%, and 95.3% in Naive Bayes, Random Forest, Decision Tree, and KNN respectively.

II. OBJECTIVES AND PROPOSED INNOVATION

A. Objective

The primary goal of the proposed system is to have the significance of health in lifestyles attain out to human beings and inspire humans to observe measures to maintain health through making the chatbot handy to all. Chatbot and fitness have a record of working properly together. It creates desirable human-like conversational surroundings for interaction between the person and the system. In this system, the user talks about their health and it is a great way for the users to adjust to a healthy

lifestyle.

An essential aspect of this system is that speaking with a non-human entity presents an experience of safety mainly when it comes to intellectual fitness as it remains as a private assembly with the diagnosis being reachable only to the user. This system is supposed to help and deliver on the spot actions where humans cannot attain due to timing or budget as it is easily accessible and free of cost. It allows the person to have free medical check-up primarily based on the symptoms where the user's fitness problem is easily identifiable.

B. Proposed Innovation

Chatbot is useful for every person who needs to get the grasp of something about wellbeing. The user can interact with chatbot and can count on them to get timely diagnosis. This framework helps customers with existing signs that affect their wellbeing. The true welfare of the chatbot is to motivate the common population by giving appropriate direction in regards to the excellent and sound living. The chatbot allows users to select symptoms for disease prediction.

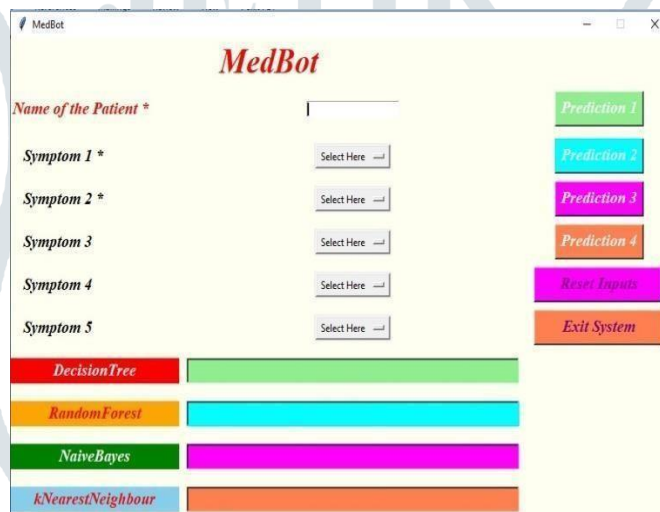


Figure 1: Default GUI

User selects symptoms in the application. They want to put up some personal details which will be confidential. User then interacts with the device and the symptoms are used as inputs for various machine learning algorithms to predict the disease.

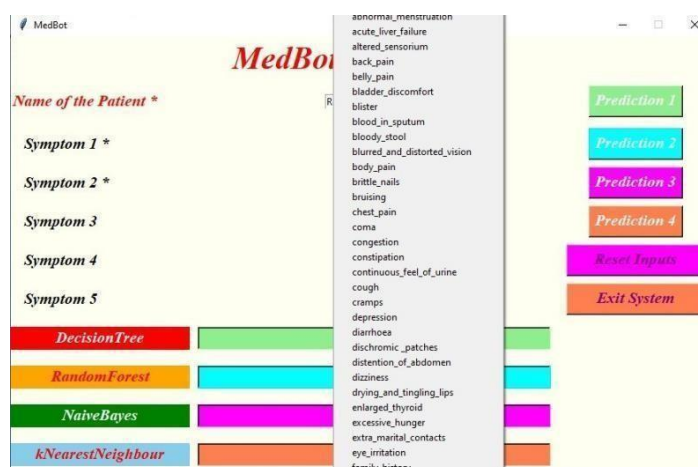


Figure 2: Symptoms selection list.

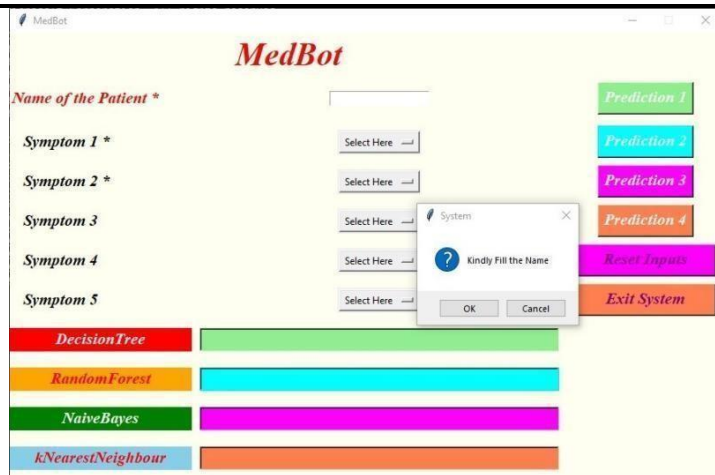


Figure 3: Input Checking.

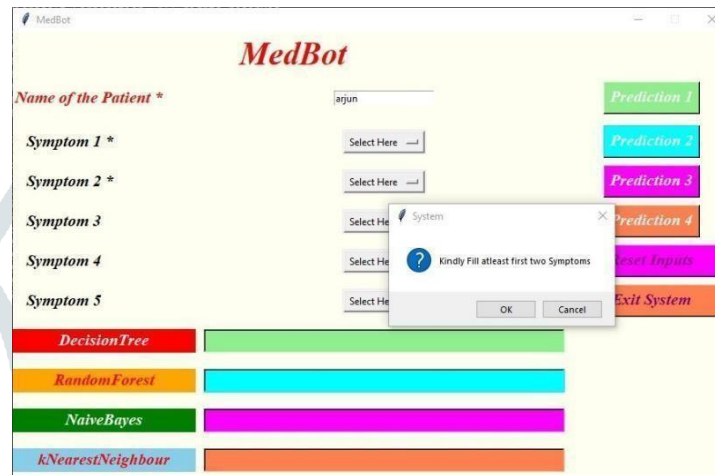


Figure 4: Input Checking for symptoms

There is an admin who controls the chatbot application. The admin views the small print of all the customers which are stored in the database and can even delete the data. The chatbot is skilled on symptoms-disease dataset. From the signs recognized by the user, KNN algorithm along with other 3 machine learning algorithms can predict the disease, depending on the dataset. The machine acknowledges the disease.

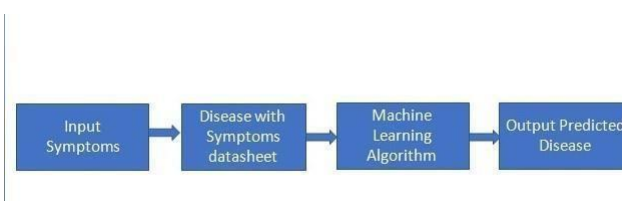


Figure 5: Working model of the proposed system.

Figure 3 shows the flow diagram of the proposed model. The chatbot helps to encourage patients to discuss their medical issues and provides a suitable diagnosis. Here the disease prediction is done using machine learning algorithms. Initially the user selects their symptoms through a dropdown list of symptoms. It uses the machine learning algorithm KNN, Naïve Bayes, Decision tree, Random

forest. Firstly they are trained on a training data set which is 70% of the actual dataset.

III. RELATED WORKS

The proposed thinking of the paper “An empirical evaluation of machine learning strategies for chronic kidney disease prognosis,” via B. Khan [1] is to construct by means of the use of computing device learning algorithms that can assist human beings discover the acceptable remedy for their disease. There are many treatments on hand for a unique ailment and no one can maybe understand the acceptable remedy which is nice and appropriate for their disease. The predominant focal point is on persistent disease. In this paper, we present experiential evaluation of ML methods for classifying the kidney patient dataset.

In the paper by Rashmi Dharwadkar [2], the scientific chat- bots functioning depends on Natural language processing that helps users to put up their health. The person can ask any private query related to fitness care through the chatbot except physically being available to the health facility by means of the usage of Google API for voice-text and text-voice conversion. Query is sent to the chatbot and the person gets a related answer which is displayed on the android app. This system typically focuses on analyzing customers' sentiments.

In the paper by Gergo Pinter [4], Methods and Modeling Strategy In the existing study, modeling is carried out by using desktop learning methods. Training is the groundwork of these methods, as well as many artificial brain (AI) techniques . According to psychological and social literature, creatures engage with their environment through the trial and error approach to reap the most useful overall performance . Based on this idea and the use of the capacity of computers to repeat a set of instructions, these prerequisites can be provided for pc applications to engage with the surroundings through updating values and optimizing functions, accord.

In the paper by Bendi Venkata Ramana [6], In this study, popular Classification Algorithms were regarded for evaluating their classification overall performance in terms of

The frequent attributes for AP liver statistics and Taiwan records are Age, SGOT, SGPT, ALP, Total Bilirubin, Direct Bilirubin, Total Proteins and Albumin are critical in determining liver status. With the selected dataset, KNN, Back propagation and SVM are giving higher outcomes with all the feature set combinations.

The proposed concept of the paper "A Survey on machine learning techniques for the diagnosis of liver disease" [7] by Golmei Shaheamlung the reason of this paper is to give a survey and comparative evaluation of the complete computing device studying strategies for prognosis and prediction of liver ailment in the scientific area, which has already been used for the prediction of liver disorder through a range of authors and the analysis are based on Accuracy, Sensitivity, Precision, and Specificity. With this survey and found out that it has absolutely located and

determined that some computer mastering algorithms such as Decision tree, J48 and ANN provide better accuracy on detection and prediction of liver disease. An extraordinary algorithm has distinctive overall performance based totally on different situations but most importantly the dataset and function determination is also very essential to get higher prediction results..

IV. DESIGN AND METHODOLOGY

The said chatbot application is built with the help of four algorithms and graphical user interface. These algorithms are namely KNN, naive bayes, decision tree and random forest.

K nearest neighbor-

It is a supervised learning algorithm. It is a basic yet essential algorithm. It finds extensive use in pattern finding and data mining.

It works by finding a pattern in data which links data to results and it improves upon the pattern recognition with every iteration.

Naive Bayes-

This algorithm is a family of algorithms based on naïve bayes theorem. They share a common principle that every pair of predictions is independent of each other. It also makes an assumption that features make an independent and equal contribution to the prediction.

Decision tree-

Decision trees are classified as a very effective and versatile classification technique. It is used in pattern recognition and classification for images. It is used for classification in very complex problems due to its high adaptability. It is also capable of engaging problems of higher dimensionality. It mainly consists of three parts: root, nodes and leaf. Roots consist of the attribute which has the most effect on the outcome, leaf tests for the value of a certain attribute and leaf gives out the output of the tree.

Random forest -

Random Forest Algorithm is a supervised learning algorithm used for both classification and regression. This algorithm works on 4 basic steps –

1. It chooses random data samples from the dataset.
2. It constructs decision trees for every sample dataset chosen.
3. At this step every predicted result will be compiled and voted on.
4. At last most voted prediction will be selected and be presented as a result of classification.

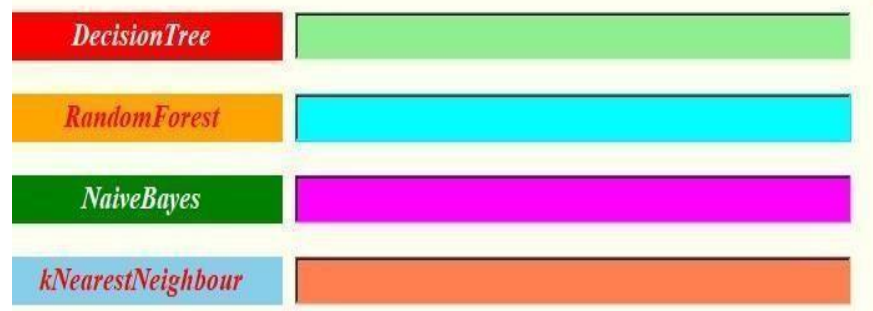


Figure 6: Algorithms to be used.

Graphical user interface-GUI made for this project is a simple tkinter GUI consisting of labels, message box, button, text, title and option menu. Labels are further used for different sections. Option Menu is used to create drop down menu. Buttons are used to give functionalities and predict the output come of models also two utility buttons namely exit and reset are also created. Message box are used at three different places, one- to restrain the to enter name second to choose at least two symptoms. If the user did not choose any symptoms, then it will prompt a message like “kindly fill at least two symptoms”. If user wants to exit from the system, then there is option “EXIT SYSTEM”. By Clicking at that option, a message will prompt like “Do you want to exit the system”.

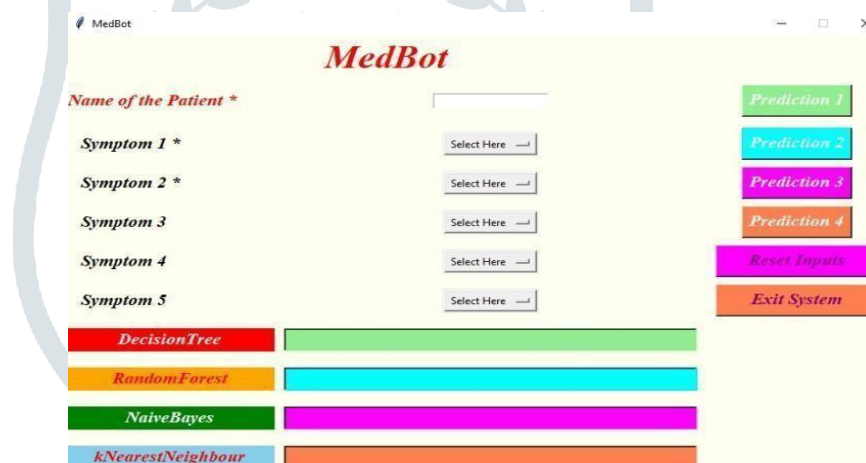


Figure 7: User friendly GUI which contains buttons like exit, reset, prediction.

System Flow

Step1: First user will have to enter the name of the patient.

Firstly the user will have to enter the name; if the user will not enter the name then the system will prompt a message as “Kindly Fill the Name”. **Step 2:** User will enter symptoms as input. There are five symptoms out of which two symptoms are compulsory and users will have to enter at least two symptoms.

Step 3: Providing dataset to the model. Our dataset contains one thirty one symptoms and forty one diseases and we are providing this dataset to the model.

Step 4: Prediction of disease using machine learning algorithm on the basis of symptoms that user have entered. We are using two machine learning algorithms that are Random Forest and Decision Tree.

Step 5: With the help of Graphical User Interface we are displaying the output. GUI is a simple tkinter it is standard library of python.GUI consisting of labels, message box, button, title and option menu.

Step 6: Name of the patient and symptoms will be given as an input to GUI.

Step 7: After pressing the prediction button we will get the output as predicted disease.

V. RESULTS

After performing the experiments and by using four different algorithms based on classification and prediction which are random forest, decision tree, KNN, naive bayes the prediction of disease is happened. Dataset provided here played a major role. Four predictions of diseases are done with the help of four different algorithms. As showed in snapshots above after carrying out some steps the system will give the output predicted disease. The system will demand for the name of the patient. Talking about the next step, the system will compulsorily need two symptoms on the basis of which disease will be predicted on the GUI built. Dataset provided to the system contains various kinds of symptoms and diseases which will be easy for users to make use of the said system.

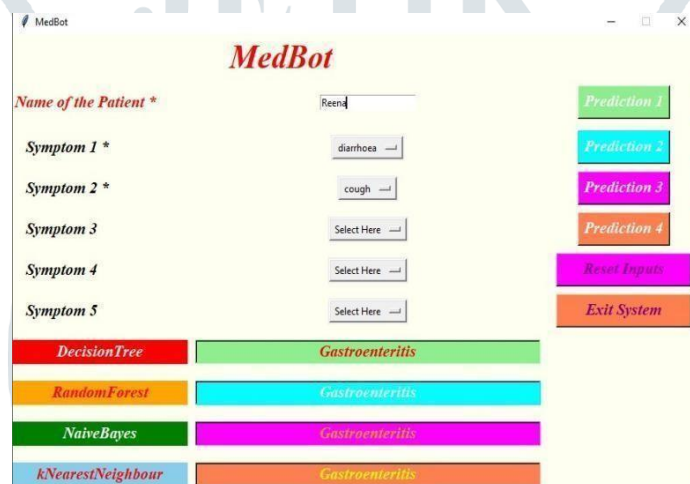


Figure 8: Final output.

VI. CONCLUSION AND FUTURE SCOPE

A. Conclusion

This paper explicates a chatbot which can be brought into play to change the traditional practice of disease detection. As soon as the patient reaches a doctor he can get help from the said chatbot. He does not have to compromise with his health till the doctor can see him. This medical chatbot can also assist the doctor in various aspects. By the comfort of home one can start taking healthy measures to protect his body from some major disease with the help of this chatbot. Cost effectiveness, reliability and accuracy are one of the biggest parameters on which the said system will work. The GUI of the said system is user friendly and it does not require sound knowledge of technology. By pressing the button a person can get to know about the health problem he is facing.

B. Future Scope

As the said medical chatbot does not recommend the treatment for this reason patients have to visit the hospital if he is not getting well by taking his own healthy measures, till then it is one of the best choices a sick person can go for. Specialized treatment will be obviously given by the doctor. This chatbot proves to be very supportive in the areas where hospitals and availability of doctors is a big issue. The patient after knowing the fact that he is not well and following the medical chatbot can schedule an appointment with the doctor as soon as possible.

REFERENCES

- [1] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, "An empirical evaluation of machine learning techniques for chronic kidney disease prophecy," *IEEE Access*, vol. 8, pp. 55012–55022, 2020.
- [2] Mrs. Rashmi Dharwadkar, Dr. Mrs. Neeta A. Deshpande "A Medical Chatbot" in *International Journal of Computer Trends and Technology (IJCTT)* V60(1):41-45, June 2018.
- [3] Dey, S. K., Hossain, A., & Rahman, M. M. (2018). "Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm". 2018 21st International Conference of Computer and Information Technology (ICIT).
- [4] G. Pinter, I. Felde, A. Mosavi, P. Ghamisi and R. Glaoguen, "COVID-19 Pandemic Prediction for Hungary; A Hybrid Machine Learning Approach", *MDPI*, vol. 8, no. 6, 2020.
- [5] H. Shee, K. W Cheruiyot and S. Kimani, "Application of k-Nearest Neighbour Classification in Medical Data Mining," *International Journal of Information and Communication Technology Research*, Volume 4, No. 4, April 2014 Ramana, Bendi Venkata, M. Surendra Prasad Babu, and N. B. Venkateswarlu. "A critical study of selected classification algorithms for liver disease diagnosis." *International Journal of Database Management Systems* 3.2 (2011): 101-114.
- [6] Jacob, Joel, Joseph Chakkalakal Mathew, J. Mathew, and E. Isac. "Diagnosis of liver disease using machine learning techniques." *Technol* 5, no. 04 (2018).