

Human Disease Prediction Based On Symptoms

Uday Kiran Kommineni
Dept.of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College
Vijayawada,India
udaykirankommineni123@gmail.com

Dhatri Phani Priya Kowthavarapu
Dept.of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College
Vijayawada,India
dhatridhatri111@gmail.com

Guna Sri Manjunadh Polukonda
Dept.Of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College
Vijayawada,India
guna.polukonda@gmail.com

Kalyan Chakravarti Yelavarti
Dept.of Information Technology
Velagapudi Ramakrishna Siddhartha
Engineering College
Vijayawada,India
klyn.518@gmail.com

Abstract— We have chosen this project as there were new diseases discovered in the 21st century which have identical and additional symptoms to the previous diseases but these new diseases are far more dangerous than the previous ones and have some additional symptoms. So, In this project we have identified these new diseases by taking the input from the user and identifying the disease and displaying it on the users smart screen display. Our project aim is to reduce the deaths of people by identifying the disease and cure it in advance. In this project we used Random Forest and Naive Bayes for predicting the disease. This prediction is done by considering and comparing the accuracies of both the algorithms and outputting the predicted disease which has best accuracy . Its implementation is completed via python programming language.

Keywords—Random Forest, Diagnosis of disease, Prediction, Machine Learning Algorithm, Database.

I. INTRODUCTION

This chapter discusses the origin of the problem, the problem description, basic definitions, and object detection applications.

1.1 Origin Of the Problem

Now a days health care department plays a critical role in curing the diseases of the patients. So, it is the responsibility of the healthcare domain to ensure the maximum safety and in addition to its benefits for the person. Since the world is running behind time and money many of the people are not much focused on their health and 40% of people forget the diseases which may lead them to dangerous disorders later. Nowadays scientific medical doctors are adopting many clinical technologies and methods for identifying and diagnosing the disease. The success treatment is continuously gained via a way of proper and correct prognosis.

But in case if he (or) she can't go to the hospital or any other health care clinics, then just by entering the symptoms and all other useful data of the person, they can get to understand the disorder from which he or she is suffering from. The health organizations can gain more help from this project via just asking the symptoms from the patient and by entering in the device either on website or application. Later the device can inform the accurate diseases. This project is based on the symptoms which completely uses the machine learning and python programming language. The GUI was made by using the Python T-kinter. By using the previous data

set which is available in hospitals we can easily train the machine to predict the disease.

According to analysis, we are having seventy percentage of peoples in India who suffers from popular sickness and twenty five percentage of people are facing the loss of life because lack of early detection of the disease and this is the reason to develop this mission. People can use this at their handy place and function an examination in their workplace or at home by using the GUI. GUI is designed in this sort of way so that everyone can effortlessly perform by using it and can check their health.

1.2 Basic Definitions and Background

1.2.1 Random Forest:

Random forest, it is a supervised M.L Method which is mostly used in Regression problems and classification issues. It chooses the random samples by using a sampling technique and constructs a decision tree for each sample and then considers the majority votes among all the decision trees. One of the most essential capabilities of the RF Algorithm is its set of rules as it can handle all the facts which contain the continuous values as well as the numerical values and categorical values as part of classification. It performs better effects for classification models

Operating of Random Forest algorithm:

- o **Step-1:** First, begin with selecting of random samples from a given dataset.
- o **Step-2:** Subsequent, this set of rules will construct a choice tree for every selected random pattern. Then it'll get the predicted result from each selection tree .
- o **Step-3:** On this step, vote casting may be finished for every anticipated result.
- o **Step-4:** At closing, pick the maximum voted prediction give up result because the final prediction result.

1.2.2 Naïve Bayes:

Naïve bayes is the word that is used for classifying the data based on the probabilities of the given classes which are mainly derived from the Bayes Theorem. It is used for the data which are not having the class labels, These are the algorithms which are used to determine the results without having the predefined class labels in the training dataset and using the probabilities

Working of Naïve Bayes Classifier:

- **Step-1:** Convert the given dataset into frequency tables.
- **Step-2:** Generate probability desk with the aid of finding the opportunities of given capabilities.
- **Step-3:** Now use the Bayes theorem to calculate posterior probability statement.

1.3 Problem Statement with Objectives and Outcomes

1.3.1 Problem Statement:

In health industry there are a few troubles associated with machines or devices in an effort to supply unaccepted results, with the intention to keep away from the ones consequences and get the appropriate and desired results we're building a program or mission so as to give the correct predictions based totally on facts given by means of the person and primarily based on the datasets that are available in that system. This industry could be a very huge enterprise which has a lot of labour to be done. So, with the assistance of all the algorithms, strategies, and methodologies we have performed this assignment so that it will assist the peoples who're within the need.

We are making the undertaking with a purpose to assist all the patients who are in need to recognize the situation in their health, and from time to time if the character has been looking at few symptoms and he/she isn't sure approximately the about disease that he/she is encountered with. So, this may lead to various illnesses in destiny. So, to avoid that and get to know the ailment in early stages, the signs and symptoms this sickness prediction will assist loads to the numerous human's starting from kids to young adults to adults and the senior citizens.

1.3.2 Objectives:

1. The primary purpose of this challenge, referred to as "Disease Prediction based symptoms" is to watch for the suitable sickness of the affected person by using all the symptoms or signs related to a particular person.
2. This challenge affords Graphical user interface for the user to enter the symptoms.
3. It additionally predicts the disease based on the signs given to it.

1.3.3 Outcomes

1. The primary goal of this mission known as "Human ailment Prediction based totally on signs and symptoms" is to predict the correct disease of the patient using all fashionable signs and symptoms.
2. This assignment provides Graphical user interface for the user to enter the symptoms.
3. It also predicts the expected illnesses together with the person information and symptoms.

1.4 Real time Applications of Proposed work

1. Using datasets, this project will make predictions about the patients' diseases based on their symptoms and other details.
2. With the use of disease prediction, it is possible to tell the patient's illness well in advance in an earlier state and can avoid the dangerous disorders.
3. To make users more comfortable using the program

rather than talking to others about their illness.

4. It provides the necessary options to choose from the attributes (Symptoms).

II Review of Literature

2.1 Description of Existing Systems

This literature review normally concentrates on the sources that support the Machine learning and hospital related approaches analysis. The analysis articles outlined fully however numerous algorithms and frameworks are perhaps accustomed to classify facts.

Authors used four machine learning procedures: Decision Tree, Naive Bayes, Random Forest and KNN. Their result is computed using in a way that when an end handler enters all the signs and simply click on the Random Forest button. This project tells that they have same accuracy from all the four algorithms. The accuracy mentioned was 0.976.

Authors made a disease prediction system using multiple ML algorithms. More than 230 diseases were present in the data set that was processed. The breakdown system identifies the disease as the output based on a person's symptoms, gender and age. The KNN algorithm has given the better results when compared with all the other algorithms. The accuracy mentioned by the weighted KNN algorithm was 93.5%.

In this Classification Decision tree, Random Forest, Naïve Bayes, SVM, KNN Algorithms is used to forecast diseases. System accuracy reaches 98.3%. Comparison between all the algorithms performance was done based on the accuracy of algorithm. Decision Tree gives 84.5%, Random Forest performs much better than all the algorithms 98.95%, SVM 96.49%, KNN 71.28%, Naïve Bayes 89.4%.

This paper discusses an algorithm that predicts a disease from its symptoms. On the user interface symptoms are provide in a drop down box from which the end user can select the signs of his disease. In addition to the disease the project can suggest the medication for the user based on his symptoms.

Authors made a prediction machine for the diseases like heart, kidney, liver, diabetes, breast cancer by using classification. A prediction accuracy probability of 95% is obtained on average.

Authors have taken into consideration a database of 106 attributes as symptoms, from it they're predicting the forty-three illnesses. After detecting the ability of predicting the ailment. They used KNN algorithm and Naïve bayes for their technique. This project embeds the ML predictive technique that abstracts information from existing information units to see and expect destiny consequences.

Authors are using machine learning algorithms like SVM Support Vector Machine and MLR Multilinear Regression that try to accurately predict possible diseases. They analyzed the accuracy of this system for 5 different diseases. The results generated by the proposed system have an accuracy of up to 87%.

2.2 Summary of Literature Study:

We have determined in this chapter that Machine learning algorithms are performed for predicting the diseases. Similarly, the frameworks used are state-of-the-art. This research concluded that a methodology that takes less time to educate and a framework that is less difficult to expand is

required. We located that the strategies formerly proposed require more time to train a version. In addition, it gives less accuracy. There are many algorithms compared with every different one and the higher one has greater accuracy was taken as their final prediction algorithm.

III PROPOSED METHOD

3.1 Design Methodology

The Disease Prediction based on signs is developed to overcome standard sickness in advance degrees as all of us realize in the serious stage of the problem and to avoid the death rate in society due to lack of awareness among the people about the deadly diseases. The project "Disease Prediction based on signs and symptoms" is implemented using python completely. Even the interface of this assignment is carried out by the usage of python's library interface called T-kinter. This prediction is basically finished with the assistance of the Random Forest algorithm and Naive Bayes algorithm with a set of rules. Right here first the user needs to enter his name, age, gender and blood group then choose the symptoms from the given drop-down menu. After entering the symptoms user can click the predict button to predict the disease which is related to given signs. The user needs to enter the signs to expect the disease, for better accurate result the user wishes to enter all of the given symptoms, then the machine will offer the correct end result along with its accuracy between 0 and 1. To reset the values there is clear button and to exit the navigation or user interface we have provided an exit button.

3.2 System Architecture Diagram

Disease prediction using machine learning predicts the presence of the disease for the person based on various signs and symptoms. The structure of the Disease prediction Architecture by using machine learning encompasses various datasets via which we will examine the symptoms of the person and predict it. In between we are having a preprocessing step in which we clear all the noisy data, missing data and incomplete data while collecting the data from the various hospitals and different persons with different symptoms.

Preprocessing can be done by various techniques such as tuple can be ignored if it has more missing values or automatically filling, manually which is a difficult task so it can be done by automatically filling the values by mode or it can also be done by binning, regression and clustering. Then datasets are converted into the smaller units by column wise and from there it gets categorized based on the classification algorithms in a while the labelled statistics is then processed into the thru which the records receive processed and goes in to the disease prediction model using all the inputs from the person that is stated above. After taking the inputs it will convert the input data into binary format. It will consider the 0s tuple of each disease and then it will replace the given symptoms values with 1s. Then it compares with all the records in the dataset and it will find the similar tuple or nearly matchable tuple. Then after the user gets into the above facts and overall processed information combines and compares inside the prediction version of the system and subsequently predicts the disorder.

A structure diagram is a graphical illustration of a set of principles, which might be part of an architecture, consisting of their principles, elements, and components. The diagram explains about the system software in the belief of review of the device. An ensemble of machine learning techniques is used for class, regression and other tasks that operates to build many selection trees at training time.

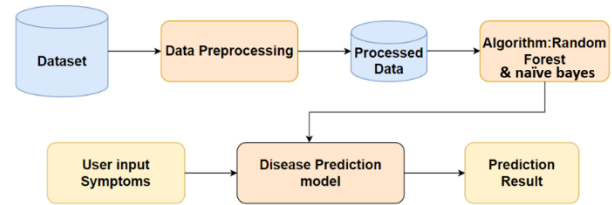


Fig 3.1 Architecture Diagram

3.3 Description of Algorithm

3.3.1 Random Forest Algorithm

An approach to ensemble learning for classification, regression, and other tasks that builds many decision trees during training.

Algorithm 1: The RF algorithm.[8]

Training Segment:

- p
- D : Training the dataset with m s occurrences, features, and aim variable.
 - K : Number of class labels in target variable.
 - N : Number of classifiers in RF.

Process:

For $m = 1$ to N

1. Create a random samples D_i from the training dataset D .
2. Construct a tree for each random feature from each random sample.

Consider a node n

- (i) Randomly choose $q \approx \sqrt{p}$ or $q \approx p/3$ feature.
- (ii) Choose a splitting feature using the best splitting method.
- (iii) Similarly follow the same process for the further nodes till the tree was generated.

Repeat the same process from I to iii until rules are met to the given condition.

3. Construct trained classifiers C .

Test Phase:

Use a simple majority vote to combine the classifiers developed using N . The projected Class label from the classifiers C is provided as follows for a test instance x :

$$C_B(x) = \underset{j}{\operatorname{argmax}} \sum_{b=1}^B I(C_b(x) = j), \text{ for } j = 1, \dots, K$$

Working of Random Forest algorithm

Step1 – First, begin with the selecting the random samples from a given dataset.

Step2 – next, this set of rules will assemble a decision tree for every pattern. Then it will get the prediction product from every selection tree.

Step3 – Here, vote casting might be executed for every expected result.

The subsequent diagram will illustrate its operating:

Fig 3.2 Represents the Random Forest illustration

In our project we are using dataset which was taken from GitHub which is already preprocessed. It consists of patient's symptoms and their corresponding diseases which has been taken from various hospitals and preprocessing was done on it. We are using the dataset which gives the better accuracy for the desired model. Dataset was in a binary format. After collecting the data, we have to change it into

Fig 3.3: Dataset

Fig 4.1 GUI Interface

Fig 4.2 Random Forest accuracy

Fig 4.3 Naïve Bayes accuracy

Our undertaking mainly focuses on early detection of the disorder of an affected person by taking five symptoms of the patient. This allows treatment of the disease earlier and the affected individual's existence can be increased. To pick out this disease we will take at a maximum of 5 symptoms from the affected person. The tool will expect the disease based on symptoms taken from the user and evaluation done among the accuracies of random forest and naïve bayes algorithms and then the comparison is done between the accuracies and results the best one among the both algorithms.

V CONCLUSION AND FUTURE WORK

5.1 Conclusion

This project is to determine the technique of predicting the disease based on the symptoms by using machine learning. We are using two algorithms i.e Random Forest and Naive bayes. When the user selects the symptoms form the interface and submit it then the machine will predict the disease based on comparison between accuracies of Random Forest and Naive bayes Algorithms.

5.2 Future Work

This system is based on the GUI, so the user can access the system from anywhere in the world. It is used to know the disease before consulting a doctor and take necessary precautions to avoid further growth of disease. For example, some people live in distant places which are far away from big cities and they don't have sophisticated equipment and doctors to treat the disease, so by using this they can take necessary actions until they reach the doctor. By using this, we have a chance to eliminate the disease at an early stage. The technology has been improved rapidly and this development can lead this equipment to reach everyone.

REFERENCES

- [1] M Vamshi Krishna Reddy, G V P Sai Abhijith, K Sai Nath, Mangali Sathyanarayana(2022) Disease Predictor Based on Symptoms Using Machine Learning.
- [2] Rinkal Keniya, Aman Khakharia, Vruddhi Shah,Vrushabh Gada, Ruchi Manjalkar, Tirth Thaker, Mahesh Warang, Ninad Mehendale.(2020) Disease Prediction from various Symptoms using Machine Learning .
- [3] Dr C K Gomathy, Mr. A. Rohith Naidu (2021) The prediction of disease using machine learning
- [4] Priyanka J. Panchal, Shaezrah A. Mhaskar, Tejal S. Ziman, "Disease Prediction using Machine Learning" , 2020.
- [5] Palle Pramod Reddy, Dirisinala Madhu Babu, Hardeep Kumar and Dr.Shivi Sharma, "Disease Prediction using Machine Learning", 2021.
- [6] Anjaiah. A, Sravya. V, Reddy. E. V, Vidya.K., Ankit. B. V. R., "Health Disease Prediction Based on Symptoms using Machine Learning", 2022.
- [7] Md. Ehtisham Farooqui, Dr. Jameel Ahmad , Disease Prediction System using Support Vector Machine and Multilinear Regression ,2022.
- [8] Sunwoo Han, Hyunjoong Kim. "On the Optimal Size of Candidate Feature Set in Random forest", Applied Sciences, 2019
- [9] Dhiraj Dahiwade, Prof. Gajanan Patle, Prof. Ektaa Meshram (2020) Designing Disease Prediction Model Using Machine Learning Approach.
- [10] Pahulpreet Singh Kohli, Shriya Arora (2022) Application of Machine Learning in Disease Prediction.
- [11] Sneha Grampurohit and Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms", International Conference for Emerging Tehnology (INCET), 2020.
- [12] Raj H. Chauhan, Daksh N. Naik, Rinal A. Halpati, Sagarkumar J. Patel and A.D. Prajapati, "Disease Prediction using Machine Learning", International Research Journal of Engineering and Technology, vol. 7, no. 5, pp. 2000-2002, 2020.