www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 9 Issue 02 February 2020, Page No. 24948-24952
ISSN: 2319-7242 DOI: 10.18535/ijecs/v9i02.4435

Disease prediction using Machine Learning

Deepika joshi, Renu kant, Sachin Shakya
Computer Science and Engineering Student, HMR Institute of Technology And Management
Delhi, India

Abstract-
This paper comprises of efficient machine learning algorithms used in predicting disease through symptoms. As, the health industry has a huge amount of data for various fields so, we want to make a system where we can use various other applications of machine learning on health industry. This all had been done to make the better medical decisions and also for rise in the accuracy. As accurate analysis of the early prediction of disease helps in the patient care and the society services. These all challenges can be easier by the help of various tools, algorithms and framework provided by the machine learning. In addition to all these predictions we are making a chatbot for all that where patients can add the symptoms that are helpful to predict the disease and also check their diabetes status through the various information provided to system by the patients.

Keywords: Decision tree classification, NLTK, tokenization.

I. Introduction

As a rise in the field of technology machine learning is widely used in various fields. Now it has various applications in the field of health industry. It works as a helping hand for the field of health industry. Now through the help of machine learning algorithm patient can get early diagnosis, due to which early treatment can be provided to the patient. Now doctor has to work only for treatment not on diagnosis due to help of these technologies. By the help of various machine learning algorithms, we are providing various resources to the healthcare industry for the betterment of society. Here we are using the ml approach while in traditional approach we have less number of symptoms for the prediction. Through the help of various machine learning algorithms prediction models are designed to predict the result according to the symptoms with better accuracy.

Here in our project, we are making a chatbot for the patients where we can help the patients to get early diagnosis based on the symptoms they are noticing. So that they can get the better treatment earlier. Here we are using machine learning algorithm to predict the disease with the help of the symptoms provided by the patient. The patient will answer yes or no for the symptoms asked by the system. And based on the test and train data, the system will generate the model for prediction and based on the patient symptoms it will predict the disease. Here we are using decision tree classification algorithm for the prediction of the disease, because decision tree classification model has better accuracy than all other prediction models in the dataset of disease and symptoms which we have taken from Kaggle. Here in our dataset we have various attributes of symptoms and based on which disease is predicted. In nodes of decision tree are symptoms and in leaves are the values of the diseases. So, based on the model the prediction is done.

II. Literature Review

[1] In Oana Frunza.et.al, "A Machine Learning Approach For Identifying Disease-Treatment Relations In Short Texts" It involves ML methodology for building a model to predict healthcare information. Through the help of various paper, it extracts diseases and treatments, and identifies relation between diseases and the

treatments. After semantic extraction various validation rules are applied to distinguish actual relation.

[2] In L. Hunter and K.B. Cohen, "Biomedical Language Processing: What's Beyond Pubmed" In this Natural language processing technique is used to make the relation between the diseases and treatments. In this user enters the name of the disease and through the help of the NLP, it gives the solution which is stored in the database.

[3] In Pravin Shinde and Prof. Sanjay Jadhav, "Health Analysis System Using Machine Learning" In this Natural language processing technique and various machine learning techniques is used to make the relation between the diseases and treatments according to the short text. In this user enters the name of the disease and through the help of the Natural language processing, it gives the solution which is stored in the database. It also represents healthcare diagnosis treatment and prevention of the illness or injury in human.

[4] In Marimuthu Muthuvel and Deivarani Sivaraju, "Analysis of Heart Disease Prediction using Various Machine Learning Techniques" It involves heart disease prediction using the machine learning algorithms. Here the information from user such as blood pressure, hypertension, diabetes and other inputs are required. Here KNN, Naïve Bayes, SVM and Decision tree algorithms are used for the prediction. There accuracy through each of the model is calculated and then best model is taken for prediction.

[5] In V.Kirubha and S.Manju Priya, "Survey on Data Mining Algorithms in Disease Prediction " It involves the analysis of the application of the data mining in health industry and of various techniques used in the health prediction. Different algorithms were used there for disease diagnosis.

[6] In Harini D K and Natesh M, "Prediction OfProbability Of Disease Based On Symptoms Using Machine Learning Algorithm " In this machine learning algorithms are used for prediction and sometime there is difficulty of missing data for this latent factor model is used. New convolutional neural network model is used for the disease risk prediction. In this both structured as well as unstructured data is used from the hospital for the better accuracy of the prediction.

[7] In Asir Antony Gnana Singh Danasingh, "Diabetes Prediction Using Medical" In this machine learning algorithms are used for diabetes prediction using the user information and learning model.

[8] In K.Priyadarshini and Dr.I.Lakshmi, "A Survey on Prediction of Diabetes Using Data Mining Technique" In this various data mining techniques are used for the prediction of diabetes using the dataset of a patient. In this various machine learning algorithms are used for overall survey related to the diabetes prediction.

III. Overview of Method

In our project we work on applications of machine learning in healthcare industry. The prediction of the disease with the help of the symptoms. In this of the applications we have use decision tree classification algorithm as a prediction model as it has better accuracy as compared to the other models. In this study of disease prediction system we have used Anaconda to perform disease prediction on our training and testing dataset from Kaggle, the dataset consists of 132 symptoms and based on the symptoms we have 41 types of diseases to predict. As early and accurate diagnosis can help to identify the disease on time so that best treatment can be provided to the patients. So here we are using decision tree algorithm to get the accurate results.

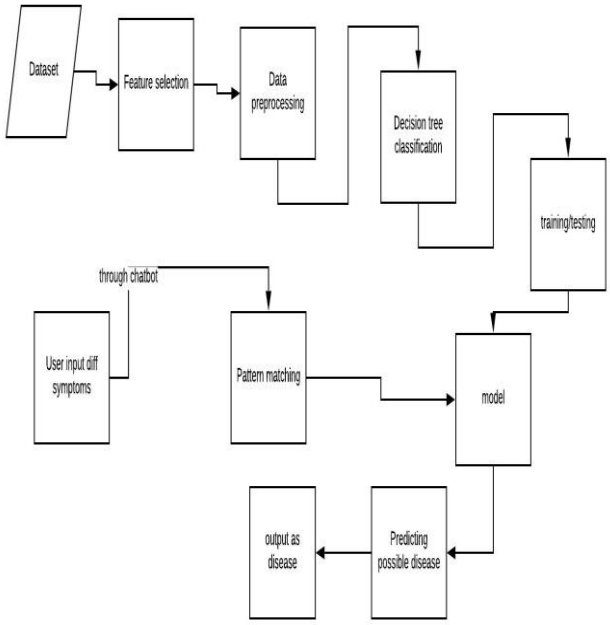

Fig.1 Flowchart of disease prediction system using symptoms.

Model building through decision tree has given step.

A. Data Pre-processing: Disease prediction dataset consists of total 132 symptoms sense by the patient and based on which we have 41 disease from which patient is suffering based on the different patients record which is used in data preprocessing. The symptoms there works as attributes, based on which the disease from which patient is suffering is predicted. There in dataset 1 indicates that following symptoms indicates the following disease and 0 indicates that these symptoms are not sense by the patient.

B. Classification modeling: Here we are using decision tree classification algorithm for the model building. The clustering of the dataset is done through the attributes called as symptoms and the decision tree features we have. In decision tree the nodes of the tree represent the symptoms provided in the dataset and leaves represents the value 0 0r 1, through which diseases are predicted.

Here firstly import some of the python packages which are sklearn, pandas, numpy, from sklearn.tree we import decision tree classifier. At beginning we take our training set as the root of the decision tree. Firstly, we map our strings to the numbers using the label encoder. Then the decision tree classifier is applied to each clustered x_train and y_train dataset to check the performance of the model.

C. After the building of the model using decision tree classification algorithm then we need a system which engage to the user for that we use chatbot. We have made a chatbot using retrieval-based model in which we had given some list of responses to the chatbot from the library. Using the help of predefined responses chatbot deals with patients, and through the help of our model predict the resulting disease based on the patient symptoms. For building of the chatbot we require NLTK from using NLTK stem we import WordNetLemmatizer. The steps followed in building chatbot using NLTK.

1. Text processing: A ml need numerical feature vector while we have text data so due to this pre-processing is done.
   - Tokenization: Here we convert normal text strings into the tokens. Sentence tokenizer converts to list of sentences and word tokenizer converts to list of word.

     sent_tokens = nltk.sent_tokenize(raw)# converts to list of sentences

     word_tokens = nltk.word_tokenize(raw)
   - Lemmatization: For this we import WordNetLemmatizer from NLTK.stem. So here we take lemmas the actual words. Here in our code we lemmatize our tokens.
2. Preprocessing: Here firstly we define a function called as Lemtokens which has tokens as the attributes, which will take lemmatize tokens as the input and returns the normalized tokens which is the list of words after removing the punctuations.
3. Keyword matching: As we are building a retrieval-based model chatbot. So for this we have to define a function for chatbot inputs and responses. We give a list of inputs and responses to our chatbot and then chatbot uses keyword matching for the system.
4. Generating response: Now the important part is there tp generate response from the chatbot for the patient's input. So, for this we use some of the modules such as, TFidf vectorizer which is used to convert the raw documents collection into the TF-IDF feature matrix and cosine similarity module which is used to find similarity between the words entered by the patient and our bot.

   Firstly, we define a function which takes the patient's response then

use our tokens and check for similarity, it flattens the vals and sort it. If the input matches then bot is ready to work on the patient's input.

For the response from the bot we use our decision tree classification model which we had built previously. Using the nodes and the features the algorithm will able to predict the patient's disease which works as the response of the bot.

So, by the help of modules the bot able to predict the patient's disease through the help of their symptoms. For this model building we have use two basic algorithms of Machine learning one is decision tree classification and another is NLTK for chatbot building.

IV. Results and Evaluation

The prediction model is developed for the disease prediction system using symptoms with chatbot and the accuracy is also good. The decision tree algorithm is used for the better accuracy.

On the basis of the dataset of the symptoms we have and the input entered by the user the disease is predicted. The chatbot helps as a interactive system for the user through which the diagnosis of the patient can be done earlier.

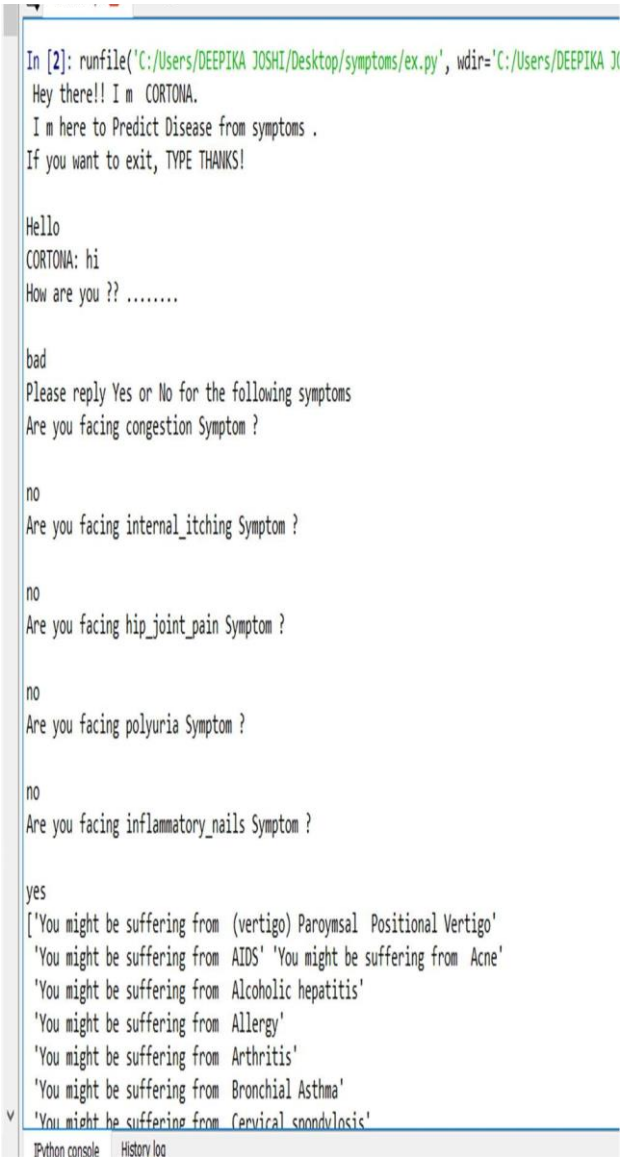

Fig 1. Output for chatbot of the disease prediction system through symptoms

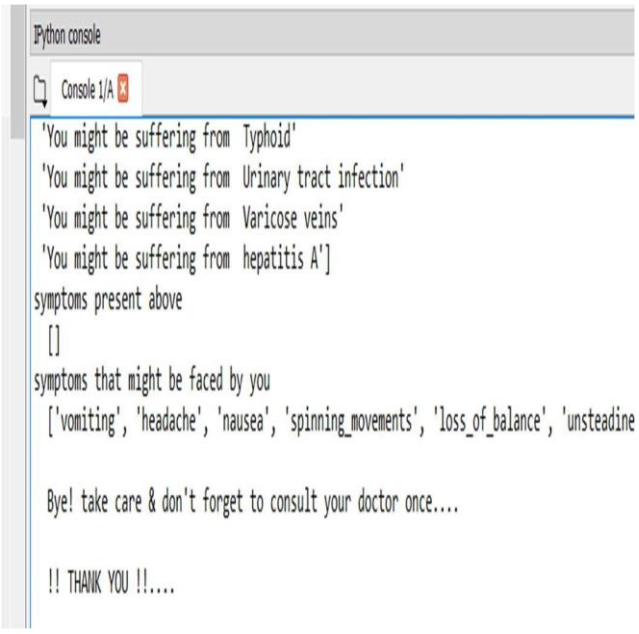

Fig 2. Output for the disease prediction system through symptoms

V. Conclusion

We implement this system for interactive and user friendly envoirnment to predict patient's disease through the help of symptoms sense by the patient using chatbot. The system we have is easy to access also for establishing real-time communication, using modern and updated technology and has better accuracy with respect to the other ones. Through this technique the time of doctors in diagnosis will be less required. Now the doctor can provide best and on time treatment to the patients. This paper represents the methodology for implementing disease prediction using symptoms with the technology of chatbot for the betterment of the healthcare industry using the recent methodology. As early and accurate diagnosis can help to identify the disease on time so that best treatment can be provided to the patients.

Acknowledgement

We would like to express our sincere gratitude to our guide Prof. Megha Gupta, CSE, HMRITM for giving us the opportunity and providing valuable guidance and timely suggestions during the entire duration of our research work. This research is made possible with the guidance and support of her. We would also like to convey our deep regards to all other faculty members of CSE department, who have bestowed their great effort and guidance at appropriate time.

We want to grate everyone who has supported us throughout the creation of this research.

References

[1] Oana Frunza, Member, IEEE "A Machine Learning Approach for Identifying Disease Treatment Relations in Short Texts" IEEE transactions on knowledge and data engineering, vol. 23, no. 6, june 2011.

[2] L. Hunter And K.B. Cohen, "Biomedical Language Processing:What's Beyond Pubmed?" Molecular Cell, Vol. 21-5, Pp. 589-594,2006.

[3] Pravin Shinde and Prof. Sanjay Jadhav, "Health

[4] Analysis System Using Machine Learning", International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014.

[5] Marimuthu Muthuvel and Deivarani Sivaraju, "Analysis of Heart Disease Prediction using Various Machine Learning Techniques", International conference on Artificial Intelligence,

[6] Smart Grid and smart city applications, (AISGSC 2019).

[7] V.Kirubha and S.Manju Priya, "Survey on Data Mining Algorithms in Disease Prediction ", International general of emerging trends and technology in computer science, August 2016.

[8] Harini D K and Natesh M, "Prediction Of Probability Of Disease Based On Symptoms Using ML Algorithm", International research journal of engineering and technology, May 2018

[9] Asir Antony Singh Danasingh,"Diabetes Prediction Using Medical", Journal of computational intelligence in bioinformatics, ISSN 0973-383X Volume 10 2017

[10] K.Priyadarshini and Dr.I.Lakshmi, "A Survey on Prediction of Diabetes Using Data Mining Technique", International Journal of innovative research in science, engineering and technology, Sept 2017.