DISEASE PREDICTION USING MACHINE LEARNING

Srilatha Yelamati Assistant Professor, Dept of CSE, Vignan's Institute of Information Technology, Visakhaptnam.
Sowndarya Lahari Chintalapudi Assistant Professor, Dept of CSE, Vignan's Institute of Information Technology, Visakhaptnam.
N V S Sowjanya Assistant Professor, Dept of IT, Pragati Engineering College, Surampalem

Abstract

We live in a fast-paced world where people do not give enough time for their health and well-being. As a result, diseases are on a rise these days. Factors like pollution are also playing a huge role in people getting sick and prone to some serious medical issues. In some cases, people even lose their lives if they aren't given treatment on time. Hence prediction of diseases at early stages is very vital. It becomes difficult for even highly qualified doctors to predict the exact disease with given symptoms. The project aims to provide an application for the users where they can enter their symptoms and based on the data provided, the machine learning model incorporated into the application will predict the disease and the risk factor involved. If the risk involved is very high, then the application will suggest nearby hospitals depending on the geographic location of the user. If the risk involved ranges from low to medium, the users will be suggested nearby chemist shops and home remedies to overcome the disease.

Keywords: Supervised learning, Machine Learning, Predictive Models

Introduction

Disease Prediction using Machine Learning is a system that predicts the disease based on the information provided by the user. It also predicts the disease of the patient or the user based on the information or the symptoms he/she enters into the system and provides accurate results based on that information. If the patient is not very serious and the user just wants to know the type of disease, he/she has been through. It is a system that provides the user the tips and tricks to maintain the health system of the user and it provides a way to find out the disease using this prediction. Now a day's health industry plays a major role in curing the diseases of the patients so this is also some kind of help for the health industry to tell the user and also it is useful for the user in case he/she doesn't want to go to the hospital or any other

clinics, so just by entering the symptoms and all other useful information the user can get to know the disease he/she is suffering from and the health industry can also get benefit from this system by just asking the symptoms from the user and entering in the system and just a few seconds, they can tell the exact and up to some extent the accurate diseases. This project is previously done by many other organizations but we intend to make it different and beneficial for the users who are using this system. This Disease Prediction Using Machine Learning is completely done with the help of Machine Learning and Python Programming language for it and also using the dataset that is available previously by the hospitals using that we will predict the disease. Nowadays doctors are adopting many scientific technologies and methodology for both identification and diagnosing not only common diseases but also many fatal diseases. Successful treatment is always attributed to the right and accurate diagnosis. Doctors may sometimes fail to make accurate decisions while diagnosing the disease of a patient, therefore disease prediction systems that use machine learning algorithms assist in such cases to get accurate results. The project disease prediction using machine learning is developed to overcome general disease in earlier stages as we all know in the competitive environment of economic development mankind has been involved so much that he/she is not concerned about health. According to research, 40% of people ignore general disease which leads to harmful disease later. The main reason for ignorance is laziness to consult a doctor and time concerns. According to research, there are 70% of people in India suffering from general diseases and 25% of people face death due to early ignorance. The main motive to develop this project is that

a user can sit at their convenient place and have a check-up of their health. The UI is designed in such a simple way that everyone can easily operate on it.

Statement of the Problem

Nowadays in the health industry, there are various problems related to machines or devices which will give wrong or unaccepted results, so to avoid those results and get the correct and desired results we are building a program or project which will give accurate predictions based on information provided by the user and also based on the datasets that are available in that machine.. So, with the help of all those algorithms, techniques, and methodologies we have done this project which will help the people who are in need. The problem here is that many people go to hospitals or clinic to know how their health is and how much they are improving in the given days, but they have to travel to get to know their answers and sometimes the patients may or may not get the results based on various factors such as doctor might be on leave or some weather problem, he might not come to the hospital and so on. To avoid all those reasons and confusions, we are making this project which will help all those patients who are in need to know the condition of their health, and at times if the person has been observing few symptoms and he/she is not sure about the disease he/she is suffering from. For us to know the disease in the early stages of the symptoms, this disease prediction will help a lot to various people ranging from children to teenagers to adults and also the senior citizens.

Proposed Method

The proposed system of disease prediction using machine learning is that we have used many techniques and algorithms and all other various tools to build a system which predicts the disease of the patient using the symptoms and by taking those symptoms we are comparing with the system's dataset that is previously available. By taking those datasets and comparing with the patient's symptoms we will predict the accurate percentage disease of the patient. The dataset and symptoms go to the prediction model of the system where the data is pre-processed for the future references and then the feature selection is done by the user where he will enter the various symptoms. Then the classification of those data is done with Naive Bayes algorithm. Then the data goes in the recommendation model, there it shows the risk analysis that is involved in the system and it also provides the probability estimation of the system such that it shows the various probability like how the system behaves when there are n number of predictions are done and it also does the recommendations for the patients from their final result and also from their symptoms like it can show what to use and what not to use from the given datasets and the final results. Here we have combined the overall structure and unstructured form of data for the overall risk analysis that is required for doing the prediction of the disease. Using structured analysis, we can identify the chronic types of disease in a particular region and particular community. In unstructured analysis we select the features automatically with the help of algorithms and techniques. This system takes symptoms from the user and predicts the disease accordingly based on the symptoms that it takes and also from the previous datasets, it also helps in continuous evaluation of viral diseases, heart rate, blood pressure, sugar level and much more which is in the system and along with other external symptoms its predicts the appropriate and accurate disease.

Results and Discussion

The project Disease Prediction using Machine Learning is developed to overcome general disease in earlier stages as we all know in the competitive environment of economic development the mankind has involved so much that he/she is not concerned about health according to research there are 40% of peoples how ignores about the general disease which leads to harmful disease later. The Project "Disease Prediction using Machine Learning" is implemented using python completely. Even the interface of this project is done using python's library called Django along with bootstrap in Html and CSS. Here first the user needs to register into the system in order to use the prediction, user needs to register with a username, email-id, and password. All these values are stored into the file system respectively, the user has the option to move forward or leave, then the user needs to log in to the system using the username and password which he/she provided during the time of registration.

If he/she enters an incorrect username and correct password then the error message will prompt stating an incorrect username and if he/she enters the incorrect password and correct username then the error message will prompt stating incorrect password, so both username and password are necessary in order to login to the system. After logging in the user needs the name and needs to select the symptoms from the given drop-down menu, for a more accurate result the user needs to enter all the given symptoms, then the system will provide the accurate result. This prediction is basically done with the help of an algorithm called Naïve Bayes. When the user selects all the symptoms then he needs to press the predict button which will give the desired output. The project is designed user-friendly. and also secure to use, every user requires authentication to enter into the system after which it provides the result based on the user input.

- Once the user opens the system to log in user needs to register by clicking on the register/signup button
- After which user needs to provide some basic details of signup and then the details of the user are saved in the system
- Then the user needs to log in to have a checkup of his/her health
- When a user tries to log in if he provides a wrong user name the system will provide a prompt message stating that the user is not found
- And if a user tries to enter the wrong password the system will prompt stating that the password is incorrect hence the user needs to enter the correct user id and password to get into the system
- After the user enters the system user has to provide the symptoms which he/she is going through based on which we have several algorithms which predict the disease and also display the percentage of accuracy
- The user needs to select all the fields of symptoms to get an accurate result.
- Data collection and dataset preparation will involve the collection of medical information from various sources like hospitals, then pre-processing is applied to the dataset which will remove all the unnecessary data and extract important features from data.
- Developing a probabilistic model for Disease Prediction will run effectively on extensive databases of healthcare.
- Training and experimentation on datasets: The Disease Prediction model will be trained on the dataset of diseases to do the prediction accuracy and produce a Confusion matrix. In this
- prediction, Naïve Bayes Algorithm is used.
- Deployment and analysis on the real-life scenario the trained and tested prediction model will be deployed in a real-life scenario made by the human experts & will be leveraged for further improvement in the methodology.

NAIVE BAYES ALGORITHM
- It is used to predict the categorical class labels.
- It classifies the class data based on the training set and the values in a classifying attribute and uses it in classifying new data.
- It is a two-step process Model Construction and Model Usage.
- This Bayes theorem is named after Thomas Bayes and it is a statistical method for classification and supervised learning method.
- It can solve both categorical and continuous values attributes.
- Bayes theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes theorem is stated mathematically as the following equation.

$$P(A/B) = P(B|A) \, P(A)/P(B)$$

	OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY GOLF
0	Rainy	Hot	High	False	No
1	Rainy	Hot	High	True	No
2	Overcast	Hot	High	False	Yes
3	Sunny	Mild	High	False	Yes
4	Sunny	Cool	Normal	False	Yes
5	Sunny	Cool	Normal	True	No
6	Overcast	Cool	Normal	True	Yes
7	Rainy	Mild	High	False	No
8	Rainy	Cool	Normal	False	Yes
9	Sunny	Mild	Normal	False	Yes
10	Rainy	Mild	Normal	True	Yes
11	Overcast	Mild	High	True	Yes
12	Overcast	Hot	Normal	False	Yes
13	Sunny	Mild	High	True	No

Naive Bayes Dataset

- The given dataset is divided into two parts namely feature matrix and response vector.
- Feature matrix contains all the vectors, which means rows of the dataset in which each vector consists of the values of dependent features. In the above dataset, features are outlook, temperature, humidity, and windy.
- Response vector consists of values of class variables for each row of the feature matrix. In the above dataset, the class variable name plays golf.
- The fundamental naive based assumption is that each feature makes an independent and equal contribution to the outcome.

Registration Page   Login Page

Login Successful Page

Symptoms Selection Page

Conclusion

This paper - disease prediction using machine learning is very much useful in everyone's day to day life and it is mainly more important for the healthcare sector, because they are the one that daily uses these systems to predict the diseases of the patients based on their general information and their symptoms that they have been through. Nowadays health industry plays major role in curing the diseases of the patients so this is also some kind of help for the health industry to tell the user and also it is useful for the user in case he/she doesn't want to go to the hospital or any other clinics, so just by entering the symptoms and all other useful information the user can get to know the disease he/she is suffering from and the health industry can also get benefit from this system by just asking the symptoms from the user and entering in the system and in just few seconds they can tell the exact and up to some extent the accurate diseases. If the health industry adopts this project then the work of the doctors can be reduced and they can easily predict the disease of the patient. The Disease prediction is to provide predictions for the various and generally occurring diseases that when unchecked and sometimes ignored can turn into fatal disease and cause a lot of problems to the patient and as well as their family members.

References

1. S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in *IEEE Access*, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

2. M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1329-1333, doi: 10.1109/ICICT50816.2021.9358597.

3. K. G. Dinesh, K. Arumugaraj, K. D. Santhosh and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 2018, pp. 1-7, doi: 10.1109/ICCTCT.2018.8550857.

4. M. Ganesan and N. Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models," *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019, pp. 1-5, doi: 10.1109/ICSCAN.2019.8878850.

5. R. J. P. Princy, S. Parthasarathy, P. S. Hency Jose, A. Raj Lakshminarayanan and S. Jeganathan, "Prediction of Cardiac Disease using Supervised Machine Learning Algorithms," *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 570-575, doi: 10.1109/ICICCS48265.2020.9121169.