Smart Health Disease Prediction

Submitted in partial fulfillment of the requirements for the award of degree of
BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE & ENGINEERING

Submitted to:
Prof. Arshleen

Submitted By:
Shobhit Aryan Lal:18BCS1099
Himanshu Dubey:18BCS1160

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERIN
Chandigarh University, Gharuan
August 2020

Introduction:

The present system of healthcare services is costly as well as time consuming. One has to physically visit doctor for the diagnosis and identification of disease and which is sometime not possible if the doctor is out of reach. So, the above problem statement can be solved using automated programs which can save time as well as money and it could be easier for the patient as well.

The proposed solution is to built a system that can predict the disease and its intensity on the basis of symptoms using machine learning algorithms and data mining. Smart Health Disease prediction is a web as well as android based application that predicts the disease of the patient with respect to the symptoms given by user. The datasets of disease and their symptoms has been collected from different health related websites. With the help of this smart health disease prediction system patient will be able to know the probability of the disease with the given symptoms and also the precautionary measures to take. It will help in the early diagnosis of the disease and the medicinal practices can be then started.

This system is based on machine learning using naïve bayes algorithm which will be self-implemented and it's performance will be compared with that already existing systems and algorithms from sklearn library(a ML python package). For easy and user-friendly access the system will be developed in both web and android versions. The algorithm will be fully implemented in Python programming language. The web version will use languages like HTML, CSS, Javascript, React.js, Flask and android version will use kotlin and other android related tools.

The naïve bayes algorithm is purely based on Bayesian theorem in probability with strong and naïve independence assumptions. It simplifies learning by assuming that features are independent of given class. Bayesian classifier assign the most likely class to a given example described by its feature vector. It is relatively simple algorithm to understand and build. It is faster to predict classes using this algorithm than many other classification algorithms. It can be easily trained using a small dataset.

Feasibility Study:

Technical Feasibility: The project is technically feasible as it can be built using the existing available technologies like machine learning, data mining, web development, android development. Designing, developing and implementation of the algorithm is little bit difficult but it will be managed after reviewing the research papers related to the algorithm.

Economic Feasibility: The project is economically feasible as the cost of the project is involved only in the hosting of the project. As the data samples increases, which consume more time and processing power. In that case better processor might be needed.

Operational Feasibility: The project is operationally feasible as the user must be having basic knowledge about computer, mobile and Internet. Disease Predictor is based on client-server architecture where client is users and server is the machine where datasets are stored.

Need and Significance:

The project will sufficiently fullfill the prevailing gaps between the patient and health care systems and services by providing an easy and early access of diagnosis of disease.

Methodology

Disease Prediction systems has been already implemented but in a very specific way like heart disease or kidney diseases using different techniques like Neural Network, Decision Tree, Random Forest, Naïve Bayes or other classifier algorithms. From the analysis it was found that Naïve Bayes is more accurate than other techniques. So, this project also uses Naïve Bayes for the prediction of different diseases. The steps to be followed is mentioned below:

1. **Data collection**: From different health related websites the data has been collected which is unstructured data. Converting unstructured data to structured data will help a lot in extracting meaningful information.
2. **Data preprocessing**: Preprocessing of data is a very important step in which all the null values are removed, the data is normalized, and various other operations are performed on the data to make it suitable for fitting the model.
3. **Implementation of algorithm**: The mathematical derivation of naïve bayes algorithm has to be implemented in python programming language considering the different hyperparameter tunings. So that the data doesn't overfit or underfit and provide better accuracy.
4. **Creating API**: The API is going to be created using Flask (a python framework) which will help to connect our model with the front-end design.
5. **Website development**: Using HTML, CSS, JavaScript, React.js, a fully responsive and user-friendly website has to be created for the easy access by the patience. It will have multiple modules like for prediction of disease, recommended precautions for a particular disease, contact information of nearby hospitals, clinics, doctors, etc.
6. **Android app development**: Using Kotlin a similar app will be developed with similar features.

Module and Team member-wise distribution of work:

The project leader Shobhit Aryan Lal (18BCS1099) will do the requirement gathering phase i.e, collecting and pre-processing of data, implementation of the machine learning algorithm naïve bayes. He will create the required APIs and the full stack front-end development of website version of the project with proper hosting.

Himanshu Dubey (18BCS1160), the team member, will do research work on the existing system and take care of how our product will look different from others. He develop the fully functional android version of the project with proper hosting and will do the software testing part by providing the corner test cases and remove the bugs present, if any.

All the documentations will be the result of joint work.

Innovations in project:

Previous research works and projects in this domain are mostly depends on one particular type of disease like heart disease prediction, diabetes classifier or a kidney disease prediction system. But this project will cover almost every type of human disease with high accuracy and low error.

The recommendation of medicines and precautions would be innovative part which will make the software more reliable and helpful. Integration of database management system for the patient's medical record will help the doctor in easy and precise treatment for a particular disease.

Software and Hardware requirements:

Software tools:

- Anaconda Navigator (Jupyter Notebook, spyder) or Google colab
- VS code / Notepad++ / Sublime Text editor
- Android Studio
- Browser (preferably google chrome)

Programming Languages:

- Python
- HTML , CSS, Javascript, react.js

- Flask
- Kotlin

Hardware Required:

- Processor- intel core i5
- Ram- 8GB
- HDD- 1 TB and above
- Monitor
- Keyboard
- Mouse

Bibiliography

- http://www.ijcstjournal.org/volume-5/issue-2/IJCST-V5I2P13.pdf
- https://ieeexplore.ieee.org/document/8819782
- https://ieeexplore.ieee.org/document/6357980
- https://www.researchgate.net/publication/323946641_Short_Survey_on_Naive_Bayes_Algorithm