Abstract

Due to big data progress in biomedical and healthcare communities, accurate study of medical data benefits early disease recognition, patient care and community services. When the quality of medical data is incomplete the exactness of study is reduced. Moreover, different regions exhibit unique appearances of certain regional diseases, which may results in weakening the prediction of disease outbreaks. Our web application “Self-Diagnosis” provides machine learning algorithms for effective prediction of various disease occurrences in disease-frequent societies. It experiment the altered estimate models over real-life hospital data collected. Using structured and unstructured data from hospital it use one of Random forest, SVM Classifier, Logistic regression, K-Neighbours classifier, Decision Tree and Map reduce algorithm to diagnose the problems. So we have decided to make a context based web app “Diagnose me” emphasizes on working on the big data models taken from various hospitals and doctors so that diseases can be diagnosed with maximum efficiency. Any sort of python IDE e.g. Anaconda will be used to develop and design the web app with flask framework.

Table of Contents

List of Figures

Chapter 1: Introduction

1.1. Background

With the exponential increase in medical data collection and management, more devotion has been paid to disease expectations from perception of big data inquiry. Numerous experiments and strategies have been conducted to classify and improve the condition of disease recognition involving a large number of medical personals. The majority of these wellness programs include an annual screening to detect individuals with the highest risk of developing a chronic disease. However, the percentage of successful recognition in a curable time is very low. A very significant number of patients, despite receiving a proper treatment, die due to the delay. How can big data analysis expertise be used to recognize the disease and generate a better diagnosis method?

These problems of disease prediction based on the symptoms can be solved with the following strategies - first the system will use Decision tree mapping algorithm to generate the pattern and causes of disease. It clearly shows the diseases and sub diseases. Second, we increase the operational efficiency by using an algorithm to divide the data into partitions.

1.2. Objectives

- Predict diseases and sub diseases with a certain level of accuracy based on the symptoms from patients.
- To get familiar with Data Analysis and Machine Learning using Python Development Environment and its libraries.
- To make initial diagnosis of diseases available to people of all economic status despite of limited medical resources in many countries

1.3. Motivation and Significance

At present, a patient has to wait for a number of days and go through a long queue, paying expensive charge to diagnose a disease which is not feasible for people of every economic status.

There are numerous cases of people dying due to late identification of disease despite receiving proper treatment, all because of inaccuracy in disease diagnosis and recognition. Proper Treatment early can be the difference to life and death. Existing Machine Learning algorithms are able to predict diseases on the basis of a structured data, which are broad and ambiguous. First, prevailing systems are dearer only rich people could pay for to such calculation systems. Second, the guess systems are nonspecific and indefinite so far. Prevailing systems are able to predict only the disease that the symptoms are pointing towards but our proposed algorithm and diverging techniques will be able to predict not only diseases but also sub diseases based on decision tree algorithm. This context based web application will be able to increase prediction accuracy following Map-reduce algorithm and will be open for all free. Prior to the existing prediction system, this web application will be a positive step forward towards right diagnosis solely based upon patients symptoms, providing immense aid and ease in medical diagnosis before time, which will be a major factor in disease treatment and possibly towards saving life. Easy to access and free of cost will be a major boost up for this application, analysing between a thousands of sets of data, aiming for maximum efficiency.

Chapter 2: Related Works/ Existing Works

WebMD

WebMD has created an organization that we believe fulfils the promise of health information on the Internet. It provides credible information, supportive communities, and in-depth reference material about health subjects that matter to you. They are a source for original and timely health information as well as material from well-known content providers.

The WebMD content staffs blends award-wining expertise in journalism, content creation, community services, expert commentary, and medical review to give their users a variety of ways to find what we are looking for.

And that, we believe, requires dedicated, full-time staff professionals with state-of-the-art expertise in:

- Health news for the public
- Creating and maintaining up-to-date medical reference content databases
- Medical imagery, graphics, and animation
- Communities
- Live web events
- User experience
- Interactive tools

Mayo Clinic

Mayo Clinic is a non-profit organization committed to clinical practice, education research, providing expert, whole-person care to everyone who needs healing. Mayo Clinic's mission is to inspire hope and contribute to health and well-being by providing the best care to every patient through integrated clinical practice, education and research. Our primary value is "The needs of the patient come first. Mayo Clinic works with hundreds of insurance companies and is an in-network provider for millions of people. In most cases, Mayo Clinic doesn't require a physician referral. Some insurers require referrals, or may have additional requirements for certain medical care. All appointments are prioritized on the basis of medical need, and the team members who will care for you or your family have the expertise and skills to provide the best care possible.

Chapter 3: System Requirement Specification

3.1. Software Specification

3.1.1. Front End Tools:

- HTML + CSS + JS+ flask framework

3.1.2. Back End Tools:

- Python

3.1.3. Software:

- Anaconda

3.1.4. Text editor:

- Jupyter Notebook

3.1.5.. Browsers:

- IE 11 above, Safari 11 above, Firefox 61.0.1 above, Chrome 67.0 above

Chapter 4: Project Planning and Scheduling

Task	1	2	3	4	5	6	7	8	9	10	11	12
Research and Study												
Web Page Designing												
Core Programming												
Testing and debugging												
Documentation												

Fig 4.1: Gantt Chart

References

1. Drew Conway and John Myles White. (2018). *Machine Learning for Hackers.* O'Reilly.
2. edX. (2018, May 03). *edX*. Retrieved october 02, 2018, from www.edx.org: https://courses.edx.org/courses/course-v1:GTx+CS7641+1T2018/course/
3. Symptom Checker. (n.d.). *Symptom Checker*. Retrieved October 1, 2018, from MedicineNet: https://www.medicinenet.com/symptoms_and_signs/symptomchecker.htm#introView
4. Wikipedia. (2018, September 26). *Wikipedia*. Retrieved October 03, 2018, from www.wikipedia.com: https://en.wikipedia.org/wiki/Decision_tree_learning