

CSC111 Project Proposal: Am I Infected?

Aina Merchant, Avni Agrawal, Paarth Arya, Reeya Bansal

Tuesday, March 16, 2021

Problem Description and Research Question

Delays in treatment is a pressing problem in today's healthcare system. This leads to a two issues, the first one being that patients who need immediate medical care do not receive it in time, and the second that people often get discouraged to visit the hospital even when they need to. To address this issue and help patients and doctors understand the severity of a situation quickly, we wish to design a model that helps diagnose potential diseases that patients could potentially have based on a number of symptoms they exhibit along with the probability of having each of those diseases. This model works by using real-world data collected by medical professionals that lists the symptoms of various diseases. By letting a model assist in the process of diagnosis, we can speed up the system for doctors who can focus their attention on patients who have more severe ailments and also help patients realise how urgently they need to visit the hospital. The question we hope to explore is: **Can we predict what illness or disease one has based on the symptoms they are exhibiting?**

Disclaimer: We are in no way saying that the model should be allowed to make all the diagnoses or that it is 100% accurate. It is just a tool for assistance that needs to constantly be updated with real-world data and serves to provide a general idea to both doctors and patients.

Computational Plan

The dataset represents diseases and their common symptoms in a tabular form along with data for symptom severity, symptom descriptions and symptom precautions. This project will model the dataset into a tree which has all internal nodes as symptoms and the leaves as the diseases associated with a chain of symptoms. We will call this tree DiseaseTree. As a source of data for diseases and their associated symptoms, the Kaggle dataset on Disease Symptom Prediction will be used: Disease-Symptom Dataset To model the precautions based on health factors and test the software, we will use the Health Dataset

Disease Symptom_1	Symptom_2	Symptom_3
Fungal infection	itching	skin_rash
Jaundice	vomiting	fatigue
Typhoid	chills	vomiting

The project will be an interactive software that accepts the user's health data and the symptoms being experienced. The project will also model the person's health data and possible precautions into a graph. This graph will use a modified topology that contains internal and external nodes. The internal nodes will contain the data for a person's health (example: nutrient levels, hormone levels, sleep cycle, etc.) and the external nodes will contain the precautions. We will call this graph PrecautionGraph.

The project will aim at making two kinds of computations:

1. Based on the symptoms the person is experiencing, the project will calculate the likelihood of a disease in percentage. This will be done using the DiseaseTree. We will go down the tree based on an algorithm. Since it is possible that a person may be infected with a disease but not show all the symptoms related to it, we will recurse through certain subtrees even when symptoms do not match. We will keep track of all this data and finally calculate the percentage of that person being infected with different diseases and return the diseases a person is most prone to. We could choose to do this graphically.

2. Using the values calculated above, the project will calculate possible precautions for the diseases the user is most prone to. Based on the likeliness of the precautions to work with a particular health factor, we will find the best suitable precautions for the person for each disease.

The prone diseases will be returned in form of a graph along with their likelihood percentage. The precautions will be returned in text, sorted from the most important precautions to the least important precautions.

We will be using the Tkinter library to make the project's user interface. The interface is meant for the user to input their health information and symptoms they are facing. Once the prone diseases and apt precautions are calculated from the trees and graphs, we will use the user-interface to display the information and provide possible leads to take the precautions.

We may also consider using the Scrappy library to extract information from reliable websites if the user inputs symptoms or health data trends that are not already registered in the project's tree and graph. This is to ensure that all given data is addressed by our project.

References

<https://www.kaggle.com/datasets/itachi9604/disease-symptom-description-dataset>

https://figshare.com/articles/dataset/Data_development_of_hormonal_levels_over_time_from_Is_paternal_oxytocin_an_oxymoron_oxytocin_vasopressin_testosterone_oestradiol_and_cortisol_in_emerging_fatherhood/20025810

<https://docs.python.org/3/library/tkinter.html>

<https://www.geeksforgeeks.org/python-tkinter-tutorial/>