Disease prediction based on symptoms

Problem statement:

Develop a medical diagnosis system based on machine learning algorithms for prediction of any disease.

- perform data exploration, preprocessing and visualization
- implement Classification algorithms using sklearn library
- evaluate the model using appropriate performance metrics
- develop the Disease and symptoms prediction system.

Introduction:

People are currently suffering from a variety of diseases. Many people are unsure if the symptoms they are experiencing are indicative of a certain disease, and hence they are unable to take the required safeguards. Anticipating the disease during the prodromal stage lowers the likelihood of complications. People will not be able to visit a doctor every time they experience a symptom. It may sometimes become a serious ailment if not treated. A model is suggested that uses a variety of symptoms as input to predict the illness. For disease prediction, the suggested method utilizes Decision trees, Naive Bayes, and Random forest classifiers. The ultimate result will be the mode of all these machine learning models. Users will be given a graphical user interface (GUI) to choose their symptoms.

About Dataset:

This Dataset is a knowledge database of disease-symptom associations generated by an automated method based on information in textual discharge summaries of patients at New York Presbyterian Hospital admitted during 2004. The first column shows the disease, the second the number of discharge summaries containing a positive and current mention of the disease, and the associated symptom. Associations for the 150 most frequent diseases based on these notes were computed and the symptoms are shown ranked based on the strength of association.

- 408 columns depicting the symptoms + 1 of frequency of occurence
- 134 rows columns representing diseases
- 0 denotes not that a symptom is not related to the disease. 1 denotes that the symptom is related to the disease