A PROJECT REPORT

ON

Disease Prediction using Machine Learning

Submitted in partial fulfillment of the requirement for the IV semester

Bachelor of Technology

By

Saksham Aggarwal 2016981

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

GRAPHIC ERA DEEMED UNIVERSITY

DEHRADUN

2021-2022

CERTIFICATE

This is to certify that the project report entitled "Disease Prediction using Machine Learning" is a Bonafede project work carried out by Saksham Aggarwal, roll no- 2016981. in partial fulfillment of award of degree of B- tech of Graphic Era Deemed University, Dehradun during the academic year 2021-2022. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated. The project has been approved as it satisfies the academic requirements

associated with the degree mentioned.

I would like to express our sincere gratitude to **Dr. Devesh Pratap Singh,** Head of Dept. of Computer Science, for providing a congenial environment to work in and carry out our project.

Problem Statement:

This Project aims to implement a robust machine learning model that can efficiently predict the disease of a human, based on the symptoms that he/she posses. It involves the use of machine learning algorithms such as Naïve Bayes. With the use of Python programming language this model is implemented. Training and testing data set are provided for the training the model on various different data values.

Doctors usually decide the disease with some common knowledge. When common knowledge is lacking, studies are summarized after some number of cases have been studied. But this process takes time Hence we will use machine learning which decide the result in the basic of some patterns. It requires the large amount of huge amount of data.

Motivation

The main motivation of doing this research is to present a disease prediction model for the prediction of occurrence of different diseases.

In medical science to determine the risk of building up a disease the prediction models are used so that it can enable early treatment or prevention of that disease. To markers of future disposition to a disease multiple or single analyses are used. Predicting the disease by analyzing the symptoms is the most integral part of treatment. With sufficient data prediction of disease by an algorithm is very easy and cheap.

The patients don't have to wholly rely on hospital or health clinic just for knowing about their disease from doctors and frequent visit would not be necessary as the disease can be prediction by patients from their homes easily.

Introduction

Organization of the medical data is always a challenging task for betterment of modern healthcare system. It is evident that mostly the medical data is exclusive only for the healthcare organizations which is only practised by elite hospitals.. This Data can be used as an excellent source for the collection of data of the common symptoms and the underlying disease. The data can be used for various research activities and to generate new real-life data Therefore, we should develop a machine learning model will take the symptoms from the user as input and predicts the possibility and risk of the disease affected or the development of such diseases in an individual. The history of the patient can also be saved in the system for future reference.

To improve the accuracy from a large data, the existing work will be done on unstructured and textual data. Firstly, we will train the model on some data values present in the Training data set ad then we will test the model using data set present in the testing data set. Here we have used Naïve Bayes algorithm. And we will also check the accuracy for every algorithm for finding which algorithm provide us the best results.

Since we have a good amount of data in today's world, we can use various machine learning algorithms to analyze the data for hidden patterns. The hidden patterns can be used for health diagnosis in medicinal data.

Tools and Software used:

1.Tkinter:- Tkinter is the standard GUI library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. It provide us with various widgets like buttons, labels, text boxes, Frame, message box and title which we have used in this project.

2.Numpy:- NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more

3.Pandas:- It is the most popular Python library. It is open source that help it to work with relational or labeled data both easily and intuitively. It is built on top of NumPy. It mainly provides two data structures for manipulating data which are Series and Data-Frame.

4.sklearn:- Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

Methodology

In this project machine learning algorithm is being used:
a. Naïve Bayes

1.Naive Bayes:- It is a probabilistic model based on Bayes Theorem which is used to calculate the probability $P(A/B)$ of an event A occurring, when we are given some prior knowledge B. It makes an assumption that all features are independent of each other.
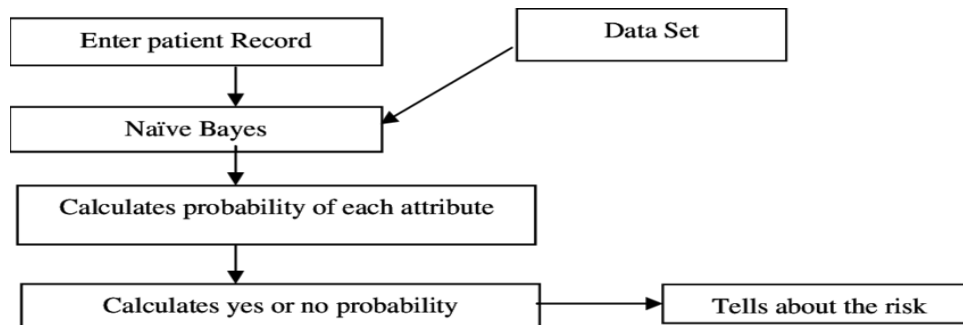
$$P(A/B)=(P(A)P(B/A))/P(B)$$

This algorithm is used in this project and is providing us ~100% accuracy.

For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability that this fruit is an apple and that is why it is known as 'Naive'.

- **Real time Prediction:** Naive Bayes is an eager learning classifier and it is sure fast. Thus, it could be used for making predictions in real time.
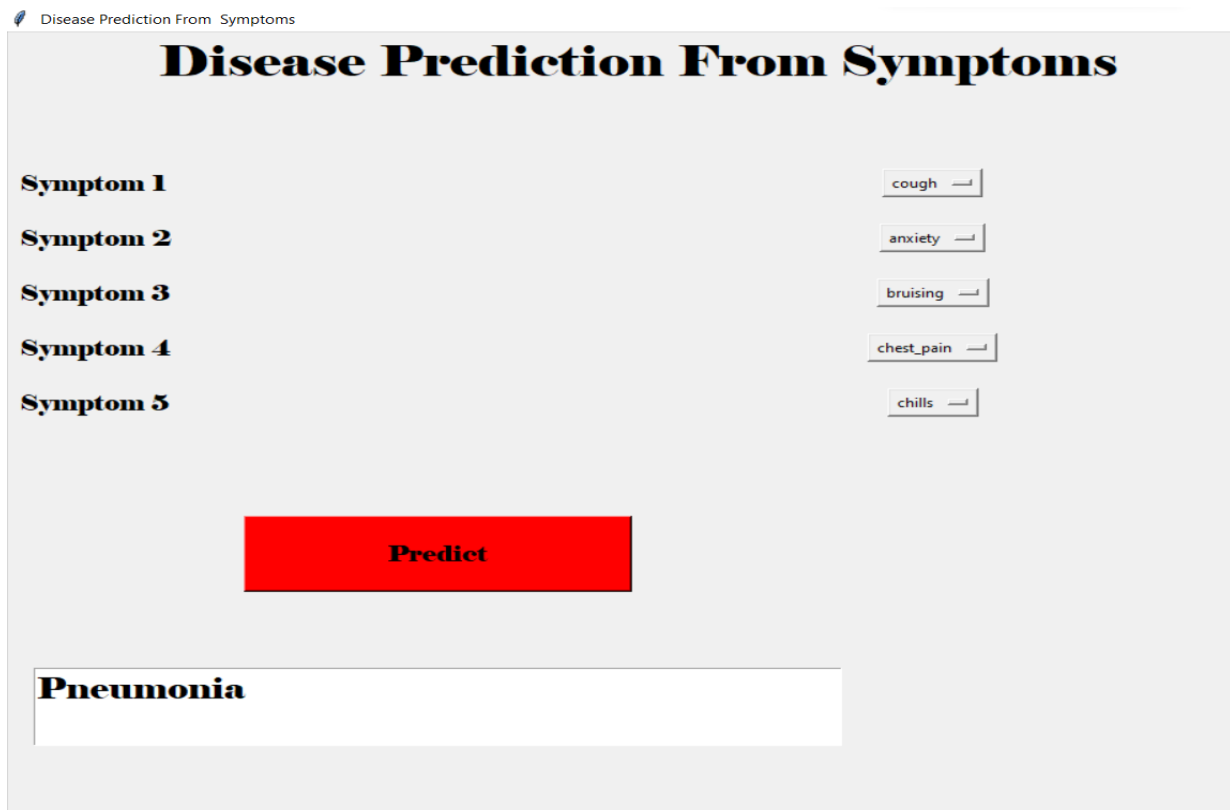
- **Multi class Prediction:** This algorithm is also well known for multi class prediction feature. Here we can predict the probability of multiple classes of target variable.

GitHub Code link:-
https://github.com/Saksham322002/Disease-prediction-using-ML.git

GUI

Conclusion:-

This project has been designed to predict the disease on the basis of the symptoms provided to it by the user which will benefit both the health department by not overcrowding hospitals and also the user who can easily check their health status whenever they need without making any prior appointments with the doctor. The user interacts with the Prediction Engine by filling a form which holds the parameter set provided as an input to the trained models. The Prediction engine provides an optimal performance compared to other state of art approaches. With more advancement we can make the health sector

flawless and the best.

References:-
Wikipedia.com
geeks for geeks
Kaggle