

A Project Report on

DISEASE PREDICTION USING

MACHINE LEARNING

IV SEMESTER PROJECT

B.TECH IN COMPUTER
SCIENCE AND ENGINEERING

Submitted by:

Amandeep Singh

University Roll No: 2016623

Batch: 2020-2024

Under the Supervision of:

Dr. Manoj Diwakar

Associate Professor



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GRAPHIC ERA DEEMED TO BE UNIVERSITY

DEHRADUN, UTTARAKHAND



GraphicEra
UNIVERSITY
University under section 3 of UGC Act, 1956
Accredited by NAAC with Grade A

ACKNOWLEDGEMENT:

I would like to take this opportunity to express my gratitude to entire faculty at Department of Computer Science and Engineering, Graphic Era Deemed to be University, who evaluated the project from time to time and gave me valuable suggestions as to how to improve the project.

I am grateful to Dr. Manoj Diwakar, Associate Professor, for his supervision, encouragement, inspiration and guidance. Working under him is being an enriched experience. In all, I found congenial work environment in Graphic Era Deemed to be University and this project completion will mark a new beginning for me in the coming days.

I am highly indebted to Graphic Era Deemed to be University for providing me the required infrastructure and facilities to accomplish the given task.

Amandeep Singh
B.Tech CSE
2020-2024
Graphic Era University

CONTENTS

- INTRODUCTION
- PROBLEM STATEMENT
- AIM
- METHODOLOGY
- MOTIVATION
- THINGS USED
- IMAGES FROM THE PROJECT
- CONCLUSION

SMART PARKING SYSTEM

INTRODUCTION:

This report is about the project chosen i.e., DISEASE PREDICTION SYSTEM USING MACHINE LEARNING. This report tries to give a brief overview of the objective of the project, Models and the libraries used, code, application and the Methodology followed to achieve this project.

This model simply aims to implement a machine learning model that can efficiently predict the disease of a human, based on the symptoms that he/she possess. Let us look into how we can approach this machine learning problem.

PROBLEM STATEMENT:

Health information needs are changing the information seeking behaviour and can be observed around the globe. Challenges faced by many people are looking online for health information regarding diseases, diagnoses and different treatments. If a recommendation system can be made for doctors and medicine while using review mining, it will save a lot of time. In the current system, the user face problem in understanding the heterogeneous medical vocabulary as the users are laymen. User is confused because a large amount of medical information on different mediums are available.

The idea behind recommender system is to adapt to cope with the special requirements of the health domain related with users. With the rise in number of patient and disease every year medical system is overloaded and with time have become overpriced in many countries. Most of the disease involves a consultation with doctors to get treated. With sufficient data prediction of disease by an algorithm can be very easy and cheap. So in order to increase the efficiency of the current medical system by reducing the workload on them, we have designed a disease prediction System.

AIM:

To implement a Disease Prediction System that can efficiently predict the disease of a human, based on the symptoms that he/she possess, using four different algorithms which will enhance the working of the medical System.

METHODOLOGY:

Prediction of disease by looking at the symptoms is an integral part of treatment. In our project, DISEASE PREDICTION USING MACHINE LEARNING, we have tried to predict the disease by looking at the symptoms of the patient. We have used 4 different algorithms for this purpose and gained an accuracy of 92-95%. Such a system can have a very large potential in medical treatment of the future. We have also designed an interactive interface to facilitate interaction with the system. We have also attempted to show and visualized the result of our study and this project.

The overall project making can be divided into few sections :

1. Data Collection

Dataset for this project was collected from a study of university of Columbia performed at New York Presbyterian Hospital during 2004.

2. libraries

1. tkinter: It's a standard GUI library of python. Python when combined with tkinter provides fast and easy way to create GUI
2. Numpy: Numpy is core library of scientific computing in python. It provides powerful tools to deal with various multi-dimensional arrays in python. It is a general purpose array processing package.
3. pandas : it is the most popular python library used for data analysis. It provides highly optimized performance with back-end source code purely written in C or python.
4. sklearn: Sklearn is an open source python library with implements a huge range of machinelearning, pre-processing, cross-validation and visualization algorithms. It features various simple and efficient tools for data mining and data processing.

3. Algorithms

There are four different kind of models present in our project to predict the disease these are:

1. Decision tree: is classified as a very effective and versatile classification technique. It is used for classification in very complex problems due to its high adaptability. It is also capable of engaging problems of higher dimensionality. Decision tree is the first prediction method i have used in our project. It gives us an accuracy of ~95%.
2. Random Forest Algorithm: is a supervised learning algorithm used for both classification and regression. In this project i have used random forest classifier with 100 random samples and the result given is ~95% accurate.
3. K Nearest Neighbour: is a supervised learning algorithm. It is a basic yet essential algorithm. It works by finding a pattern in data which links data to results and it improves upon the pattern recognition with every iteration. The result given is of ~92% accuracy.
4. Naïve Bayes algorithm: is a family of algorithms based on naïve bayes theorem. They share a common principle that is every pair of prediction is independent of each other. It also makes an assumption that features make an independent and equal contribution to the prediction. In this project we have used naïve bayes algorithm to gain a ~95% accurate prediction.

MOTIVATION:

The main motivation of this project comes after seeing the really high fees paid at hospital for getting ourselves checked for small diseases. A normal sickness can be treated at home but many people are unaware of the disease/sickness they are going through, and they rush to doctors for help. But what if these diseases can be predicated at home by checking of some symptoms that the patient is experiencing ? This will not only safe time of the patient to visit the doctor, but also will help the medical staff by reducing their workload. By using proper algorithms with the concept of machine learning we can predict the different diseases by going through the database collected and can successfully create a disease prediction modeland

THINGS USED:

Software Used: Jupyter Notebook

Language Used: Python

Algorithms Used: Decision tree ,Random Forest, K Nearest Neighbour, Naïve Bayes.

IMAGES FROM THE PROJECT

Smart Disease Predictor System

Disease Predictor using Machine Learning

Name of the Patient *

Symptom 1 *

Symptom 2 *

Symptom 3

Symptom 4

Symptom 5

DecisionTree

RandomForest

NaiveBayes

kNearestNeighbour

Prediction 1

Prediction 2

Prediction 3

Prediction 4

Reset Inputs

Exit System

Smart Disease Predictor System

Disease Predictor using Machine Learning

Name of the Patient *

Symptom 1 *

Symptom 2 *

Symptom 3

Symptom 4

Symptom 5

DecisionTree

RandomForest

NaiveBayes

kNearestNeighbour

Prediction 1

Prediction 2

Prediction 3

Prediction 4

Reset Inputs

Exit System

Amandeep Singh

fast_heart_rate

chest_pain

brittle_nails

fluid_overload

irritability

Dengue

Tuberculosis

Hyperthyroidism

Hyperthyroidism

CONCLUSION

We planned to create a system which can predict disease on the basis of symptoms given to it. Such a system can decrease the rush at hospitals and also can reduce the workload on medical staff. We were successful in creating such a system by using 4 different algorithms. On an average we achieved accuracy of ~94%. Such a system can be largely reliable to do the job. In Future we can we also add a way to store the data entered by the user in the database which can be used to help in creating better version of more such system. Our system also has an easy-to-use interface. It also has various visual representation of data collected and results achieved.

REFERENCES

- columbia.edu (Disease prediction Database)
- Google
- Youtube