HEART DISEASE PREDICTION WEBSITE

A PROJECT SUBMITTED

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

BACHELOR OF TECHNOLOGY

in COMPUTER SCIENCE AND ENGINEERING

by

NAVEEN PAL (1819310052), PAWAN KUMAR (1819310058),

PRASHANT SRIVASTAVA(1819310062)

UNDER THE SUPERVISION OF

DR. RAVI PRAKASH CHATURVEDI SIR

TO THE

FACULTY OF UNITED COLLEGE OF ENGINEERING AND RESESACH,
GREATER NOIDA,

DR. APJ ABDUL KALAM TECHNICAL UNIVERSITY
LUCKNOW (June-2022)

DECLARATION

We hereby declare that the work presented in this report entitled “HEART DISEASE PREDICTION WEBSITE ", was carried out by us. we have not submitted the matter embodied in this report for the award of any other degree or diploma of any other University or Institute. we have given due credit to the original authors/sources for all the words, ideas, diagrams, graphics, computer programs, experiments, results, that are not my original contribution. we have used quotation marks to identify verbatim sentences and given credit to the original authors/sources.

We affirm that no portion of my work is plagiarized, and the experiments and results reported in the report are not manipulated. In the event of a complaint of plagiarism and the manipulation of the experiments and results, We shall be fully responsible and answerable.

Name :

NAVEEN PAL (1819310052)

PAWAN KUMAR (1819310058)

PRASHANT SRIVASTAVA (1819310062)

(Candidate Signature)

CERTIFICATE

Certified that **NAVEEN PAL, PAWAN KUMAR, PRASHANT SRIVASTAVA (1819310052, 1819310058, 1819310062**) respectively has carried out the research work presented in this project entitled **"HEART DISEASE PREDICTION WEBSITE"** for the award of "**BACHELOR OF TECHNOLOGY"** from Dr . A.P.J Abdul Kalam Technical University, Lucknow under supervision. The thesis embodies results of original work, and studies are carried out by the student himself/herself (print only that is applicable) and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Signature

(Name of Supervisor):

DR. RAVI PRAKASH CHATURVEDI

(Designation):

FACULTY TO CSE

Abstract

Heart-related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need fora reliable, accurate, and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart-related diseases. Heart is the next major organ comparing to the brain which has more priority in the Human body. It pumps the blood and supplies it to all organs of the whole body. Prediction of occurrences of heart diseases in the medical field is significant work. Data analytics is useful for prediction from more information and it helps the medical center to predict various diseases. A huge amount of patient-related data is maintained on monthly basis. The stored data can be useful for the source of predicting the occurrence of future diseases. Some of the data mining and machine learning techniques are used to predict heart diseases, such as Artificial Neural Network (ANN), Random Forest and Support Vector Machine (SVM).Prediction and diagnosing of heart disease become a challenging factor faced by doctors and hospitals both in India and abroad. To reduce the large scale of deaths from heart diseases, a quick and efficient detection technique is to be discovered. Data mining techniques and machine learning algorithms play a very important role in this area. The researchers accelerating their research works to develop software with the help of machine learning algorithms which can help doctors to decide both prediction and diagnosing of heart disease. The main objective of this research project is to predict the heart disease of a patient using machine learning algorithms.

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our project guide DR. RAVI PRAKASH CHATURVEDI, Assistant Professor, Department of Computer Science and Engineering, UCER, GREATER NOIDA, for her guidance with unsurpassed knowledge and immense encouragement. We are grateful to DR. SAMEER ASTHANA, Head of the Department, Computer Science and Engineering, for providing us with the required facilities for the completion of the project work. We are very much thankful to the Principal and Management, UCER, GREATER NOIDA,for their encouragement and cooperation to carry out this work.We express our thanks to Project Coordinator DR. DAYA SRIVASTAVA**,** for his continuous support and encouragement. We thank all teaching faculty of Department of CSE, whose suggestions during reviews helped us in accomplishment of our project. We would like to thank our parents, friends, and classmates for their encouragement throughout our project period. At last but not the least, we thank everyone for supporting us directly or indirectly in completing this project successful.

TABLE OF CONTENTS

Chapter 1

Introduction

According to the World Health Organization, every year 12 million deaths occur worldwide due to Heart Disease. Heart disease is one of the biggest causes of morbidity and mortality among the population of the world. Prediction of cardiovascular disease is regarded as one of the most important subjects in the section of data analysis. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. Many researches have been conducted in attempt to pinpoint the most influential factors of heart disease as well as accurately predict the overall risk. Heart Disease is even highlighted as a silent killer which leads to the death of the person without obvious symptoms. The early diagnosis of heart disease plays a vital role in making decisions on lifestyle changes in high-risk patients and in turn reduces the complications. Machine learning proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the health care industry. This project aims to predict future Heart Disease by analyzing data of patients which classifies whether they have heart disease or not using machine-learning algorithm. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can say that this technique can be very well adapted to do the prediction of heart disease.

MOTIVATION FOR THE WORK

The main motivation of doing this research is to present a heart disease prediction model for the prediction of occurrence of heart disease. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient.

This work is justified by performing a comparative study and analysis using three classification algorithms namely Naïve Bayes, Decision Tree, and Random Forest are used at different levels of evaluations. Although these are commonly used machine learning algorithms, the heart disease prediction is a vital task involving highest possible accuracy. Hence, the three algorithms are evaluated at numerous levels and types of evaluation strategies. This will provide researchers and medical practitioners to establish a better.

Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data [12]. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: “How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?” This is the main motivation for this research.

PROBLEM STATEMENT

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either it are expensive or are not efficient to calculate chance of heart disease in human. Early detection of cardiac diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients everyday in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time and expertise. 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.

Chapter 2

LITERATURE SURVEY

With growing development in the field of medical science alongside machine learning various experiments and researches has been carried out in these recent years releasing the relevant significant papers. [1] Purushottam ,et ,al proposed a paper “Efficient Heart Disease Prediction System” using hill climbing and decision tree algorithms .They used Cleveland dataset and pre processing of data is performed before using classification algorithms. The Knowledge Extraction is done based on Evolutionary Learning (KEEL), an opensource data mining tool that fills the missing values in the data set.A decision tree follows top-down order. For each actual node selected by hill-climbing algorithm a node is selected by a test at each level. The parameters and their values used are confidence. Its minimum confidence value is 0.25. The accuracy of the system is about 86.7%. [2] Santhana Krishnan. J ,et ,al proposed a paper “Prediction of Heart Disease Using Machine Learning Algorithms” using decision tree and Naive Bayes algorithm for prediction of heart disease. In decision tree algorithm the tree is built using certain conditions which gives True or False decisions. The algorithms like SVM, KNN are results based on vertical or horizontal split conditions depends on dependent variables. But decision tree for a tree like structure having root node, leaves and branches base on the decision made in each of tree Decision tree also help in the understating the importance of the attributes in the dataset. They have also used Cleveland data set. Dataset splits in 70% training and 30% testing by using some methods. This algorithm gives 91% accuracy. The second algorithm is Naive Bayes, which is used for classification. It can handle complicated, nonlinear, dependent data so it is found suitable for heart disease dataset as this dataset is also complicated, dependent and nonlinear in nature. This algorithm gives an 87% accuracy. [3] Sonam Nikhar et al proposed paper “ Prediction of Heart Disease Using Machine Learning Algorithms” their research gives point to point explanation of Naïve Bayes and decision tree classifier that are used especially in the prediction of Heart Disease. 3 Some analysis has been led to think about the execution of prescient data mining strategy on the same dataset, and the result decided that Decision Tree has highest accuracy than Bayesian classifier. [4] Aditi Gavhane et al proposed a

paper “Prediction of Heart Disease Using Machine Learning”, in which training and testing of dataset is performed by using neural network algorithm multi-layer perceptron. In this algorithm there will be one input layer and one output layer and one or more layers are hidden layers between these two input and output layers. Through hidden layers each input node is connected to output layer. This connection is assigned with some random weights. The other input is called bias which is assigned with weight based on requirement the connection between the nodes can be feedforwarded or feedback. Avinash Golande et al, proposed “Heart Disease Prediction Using Effective Machine Learning Techniques” in which few data mining techniques are used that support the doctors to differentiate the heart disease. Usually utilized methodologies are k-nearest neighbour, Decision tree and Naïve Bayes. Other unique characterization-based strategies utilized are packing calculation, Part thickness, consecutive negligible streamlining and neural systems, straight Kernel self arranging guide and SVM (Bolster Vector Machine). [6] Lakshmana Rao et al ,proposed “Machine Learning Techniques for Heart Disease Prediction” in which the contributing elements for heart disease are more. So, it is difficult to distinguish heart disease .To find the seriousness of the heart disease among people different neural systems and data mining techniques are used. Abhay Kishore et a lproposed “Heart Attack Prediction Using Deep Learning” in which heart attack prediction system by using Deep learning techniques and to predict the probable aspects of heart related infections of the patient Recurrent Neural System is used. This model uses deep learning and data mining to give the best precise model and least blunders. This paper acts as strong reference model for another type of heart attack prediction models Senthil Kumar Mohan et al, proposed “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” in which their main objective is to improve 4 exactness in cardiovascular problems. The algorithms used are KNN, LR, SVM, NN to produce an improved exhibition level with a precision level of 88.7% through the prediction model for heart disease with hybrid random forest with linear model(HRFLM). Anjan N. Repaka et al, proposed a model stated the performance of prediction for two classification models, which is analyzed and compared to previous work. The experimental results show that accuracy is improved in finding the percentage of risk prediction of our proposed method in comparison with other models. Aakash Chauhan et al, proposed “Heart Disease Prediction using Evolutionary Rule Learning”. Data is directly retrieved from electronic records that reduce the

manual tasks. The amount of services are decreased and shown major number of rules helps within the best prediction of heart disease. Frequent pattern growth association mining is performed on patient's dataset to generate strong association.

Chapter 3

Methodology

3.1 Existing System

Heart disease is even being highlighted as a silent killer which leads to the death of a person without obvious symptoms. The nature of the disease is the cause of growing anxiety about the disease & its consequences. Hence continued efforts are being done to predict the possibility of this deadly disease in prior. So that various tools & techniques are regularly being experimented with to suit the present-day health needs. Machine Learning techniques can be a boon in this regard. Even though heart disease can occur in different forms, there is a common set of core risk factors that influence whether someone will ultimately be at risk for heart disease or not. By collecting the data from various sources, classifying them under suitable headings & finally analysing to extract the desired data we can conclude. This technique can be very well adapted to the do the prediction of heart disease. As the well-known quote says “Prevention is better than cure”, early prediction & its control can be helpful to prevent & decrease the death rates due to heart disease.

3.2 PROPOSED SYSTEM

The working of the system starts with the collection of data and selecting the important attributes. Then the required data is pre processed into the required format. The data is then divided into two parts training and testing data. The algorithms are applied and the model is trained using the training data. The accuracy of the system is obtained by testing the system using the testing data. This system is implemented using the following modules.

1.) Collection of Dataset

2.) Selection of attributes

3.) Data Pre-Processing

4.) Balancing of Data

5.) Disease Prediction

3.2.1 Collection of dataset

Initially, we collect a dataset for our heart disease prediction system. After the collection of the dataset, we split the dataset into training data and testing data. The training dataset is used for prediction model learning and testing data is used for evaluating the prediction model. For this project, 70% of training data is used and 30% of data is used for testing. The dataset used for this project is Heart Disease UCI. The dataset consists of 76 attributes; out of which, 14 attributes are used for the system.

3.2.2 Details of the attributes of Dataset:

1. Age.
2. Sex: 1 = Male, 0 = Female.
3. (cp) chest pain type (4 values – Ordinal), 1st value: typical angina, 2nd value: atypical angina, 3rd value: non-anginal pain, 4th value: asymptomatic.
4. (trestbps) resting blood pressure.
5. (chol) serum cholesterol.
6. (Fbs) – fasting blood sugar > 120 mg/dl.
7. (restecg) – resting electrocardiography results.
8. (thalach) – maximum heart rate achieved.
9. (exang) – exercise-induced angina.
10. (oldpeak) – ST depression caused by exercise relative to rest.
11. (slope) – the slope of the peak exercise ST segment.
12. (ca) – the number of major vessels colored by fluoroscopy.
13. (thal) – maximum heart rate achieved (Ordinal), 3 = normal, 6 = fixed defect, 7 = reversible defect.

3.2.3 Pre-processing of Data

Data pre-processing is an important step for the creation of a machine learning model. Initially, data may not be clean or in the required format for the model which can cause misleading outcomes. In pre-processing of data, we transform data into our required format. It is used to deal with noises, duplicates, and missing values of the dataset. Data pre-processing has the activities like importing datasets, splitting datasets, attribute scaling, etc. Pre processing of data is required for improving the accuracy of the model.

3.3.4Splitting the Data in training and testing set:

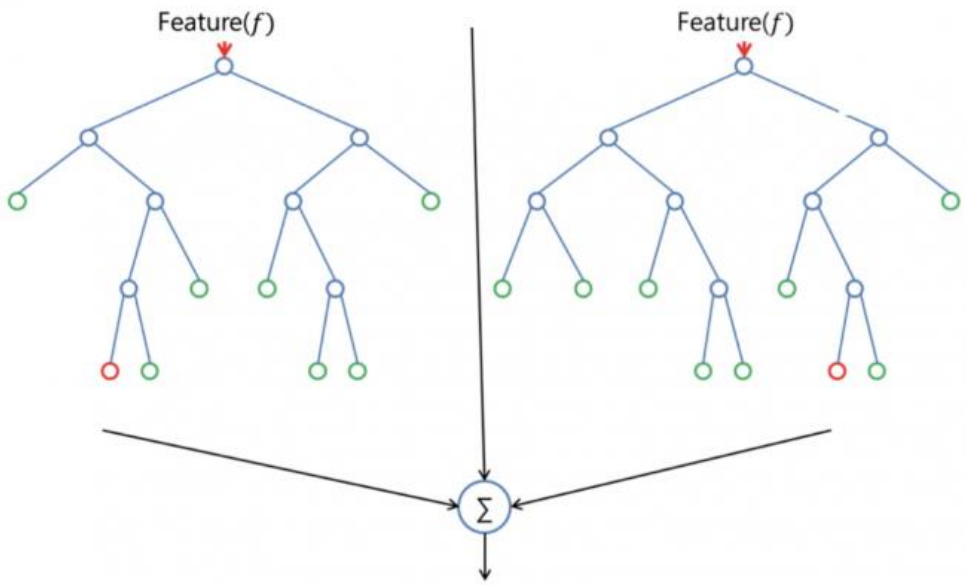
In this process we divide our data in training and testing using the train test split method of python. This process divides the data in different sets which are used for training and testing which further helps in validating our model accuracy.

3.3.5Algorithm Used:

3.3.4.**1 Random Forest ALGORITHM**

The Random forest or Random Decision Forest is a supervised Machine learning algorithm used for classification, regression, and other tasks using decision trees.

The Random forest classifier creates a set of decision trees from a randomly selected subset of the training set. It is basically a set of decision trees (DT) from a randomly selected subset of the training set and then It collects the votes from different decision trees to decide the final prediction.



3.3.6 Fitting the model:

As the last step we have to fit our training set to the logistic regression which gives out values of 0 and 1 i.e if a person don't have heart disease then the Logistic Regression will give out a value of 0 else it will give out a value of 1.

3.3.**7 Checking the Accuracy and results of Model:**

After all the steps are concluded we need to use certain methods to find out if our model is behaving right or not.To determine this we use certain factors like confusion matrix and accuracy score of our model.

```
In our model we attained an accuracy score of
0.8524590163934426
```

The confusion matrix of our model was as follows:

```
array([[25,  4],
       [ 5, 27]], dtype=int64)
```

3.3.8 **Deploying the Machine learning model using Django Framework:**

3.3.8.1 Django Framework:

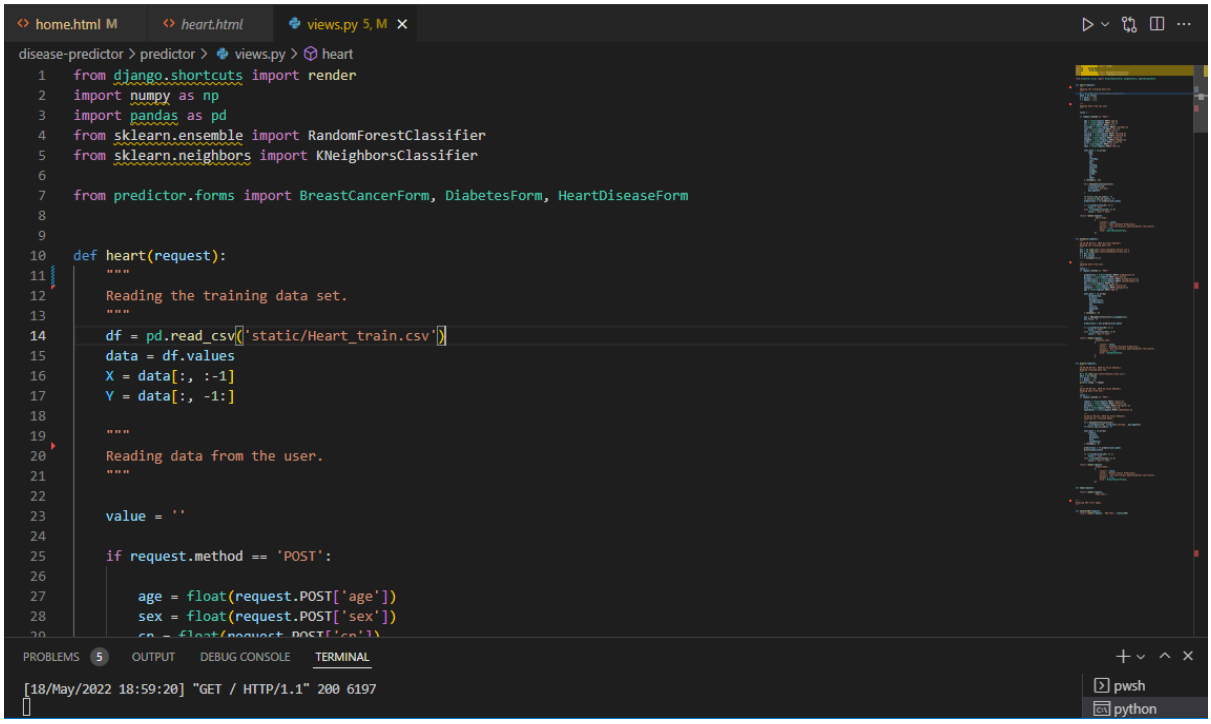
Django is a Python framework that makes it easier to create web sites using Python.

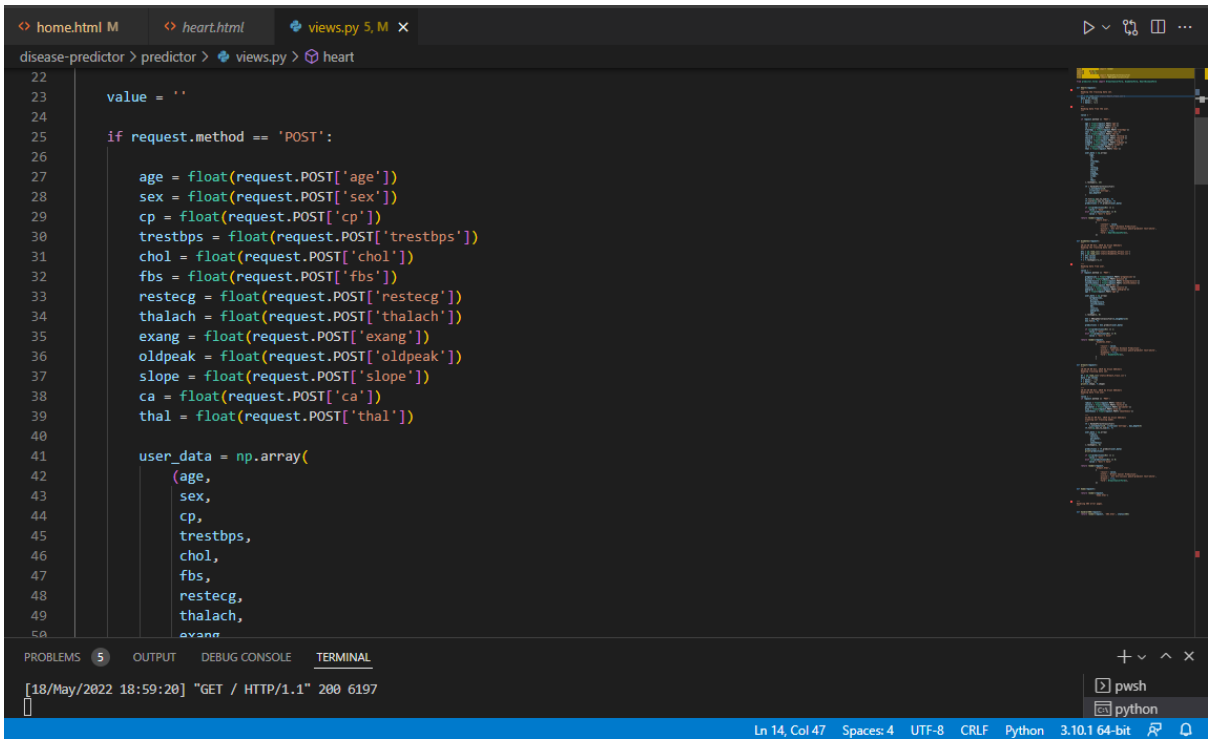
Django takes care of the difficult stuff so that you can concentrate on building your web applications.

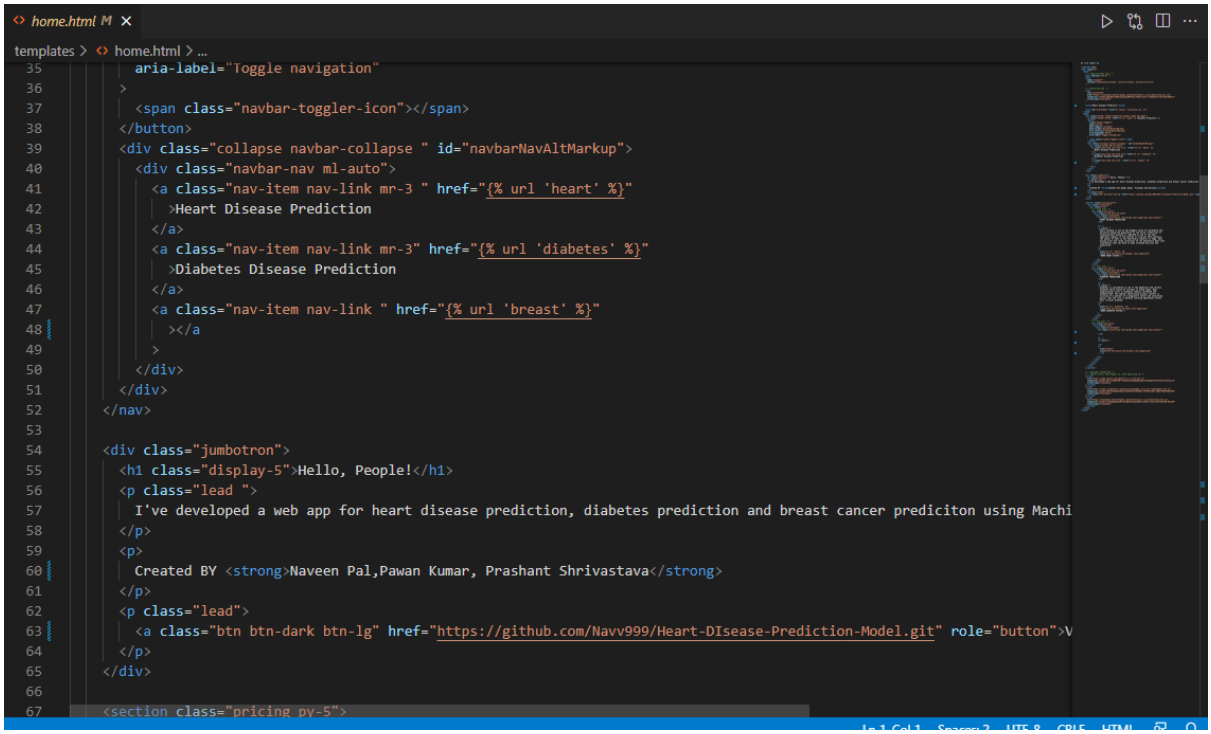
Django emphasizes reusability of components, also refereed to as DRY (Don't Repeat Yourself), and comes with ready-to-use features like login system, database connection and CRUD operations (Create Read Update Delete).

Later this model was deployed on website using Django framework.

Some Code Snippets:-






home.html M
templates > home.html > ...
aria-label="Toggle navigation"

</button>

<a class="nav-item nav-link mr-3 " href="{% url 'heart' %}"
>Heart Disease Prediction

<a class="nav-item nav-link mr-3" href="{% url 'diabetes' %}"
>Diabetes Disease Prediction

<a class="nav-item nav-link " href="{% url 'breast' %}"
>

</nav>

<h1 class="display-5">Hello, People!</h1>

I've developed a web app for heart disease prediction, diabetes prediction and breast cancer prediciton using Machi

Created BY Naveen Pal,Pawan Kumar, Prashant Shrivastava

<a class="btn btn-dark btn-lg" href="https://github.com/Navv999/Heart-DIsease-Prediction-Model.git" role="button">V

<section class="pricing py-5">
Ln 1 Col 1 Spaces: 2 UTF-8 CRLF HTML

Chapter 4

Details about the technologies used:

Machine Learning:

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.ning:

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

Django:

Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

5.1 SYSTEM CONFIGURATION

5.1.1 Hardware requirements:

Processer : Any Update Processer

Ram : Min 4GB Hard Disk : Min 100GB

5.1.2 Software requirements:

Operating System : Windows family

Technology : Python3.7,Django

IDE : Jupiter notebook, VS code

List of python libraries:

import pandas as pd

- pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool,built on top of the Python programming language

from sklearn.model_selection import train_test_split

- Scikit-learn is probably the most useful library for machine learning in Python. The sklearn library contains a lot of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction.

from sklearn.preprocessing import StandardScaler

- The idea behind StandardScaler is that it will transform your data such that its distribution will have a mean value 0 and standard deviation of 1.
- In case of multivariate data, this is done feature-wise (in other words independently for each column of the data).
- Given the distribution of the data, each value in the dataset will have the mean value subtracted, and then divided by the standard deviation of the whole dataset (or feature in the multivariate case).

from sklearn.ensemble import RandomForestClassifier

- Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

from sklearn.metrics import accuracy_score

- ccuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition: Accuracy = Number of correct predictions Total number of predictions.

from sklearn.metrics import confusion_matrix

- A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing.

from sklearn.metrics import classification_report

Chapter 5

CONCLUSION AND FUTURE WORK

Heart diseases are a major killer in India and throughout the world, application of promising technology like machine learning to the initial prediction of heart diseases will have a profound impact on society. The early prognosis of heart disease can aid in making decisions on lifestyle changes in high-risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine. The number of people facing heart diseases is on a raise each year. This prompts for its early diagnosis and treatment. The utilization of suitable technology support in this regard can prove to be highly beneficial to the medical fraternity and patients. In this paper, the seven different machine learning algorithms used to measure the performance are SVM, Decision Tree, Random Forest, Naïve Bayes, Logistic Regression, Adaptive Boosting, and Extreme Gradient Boosting applied on the dataset. The expected attributes leading to heart disease in patients are available in the dataset which contains 76 features and 14 important features that are useful to evaluate the system are selected among them. If all the features taken into the consideration then the efficiency of the system the author gets is less. To increase efficiency, attribute selection is done. In this n features have to be selected for evaluating the model which gives more accuracy. The correlation of some features in the dataset is almost equal and so they are removed. If all the attributes present in the dataset are taken into account then the efficiency decreases considerably.

LIST OF ABBREVATIONS:

ML Machine Learning

AI Artificial Intelligence

NN Neural Networks

SVM Support Vector Machine

XG Extreme Gradient

RF Random Forest

References-

- www.wikipedia.com
- www.kaggle.com
- www.analyticsvidhya.com
- www.medium.com