**Project Title:** Heart Disease Prediction Web Application

**Statement About the Problem:**

According to the World Health Organization, every year 12 million deaths occur worldwide due to heart disease. The load of cardiovascular disease is rapidly increasing all over the world from the past few years. 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 reduce the complications. The overall objective of my work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease. Decisions are often made based on doctors’ intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.

**Why is the particular Topic chosen?**

The motivation for doing this project was primarily an interest in undertaking a challenging project in an interesting area of research. While developing this website many challenges may arise and this Topic looks appealing and exciting. This area is possibly an area that we might study more deeply in future. The major motivation behind this research-based project was to explore the feature selection methods, data preparation and processing behind the training models in the machine learning.

**Methodology:**

In this project we will use various data sets. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. 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. We will use Logistic various machine-learning algorithm to train our system. We will obtain some information from the user like age, cholesterol, etc. to determine whether the user has any heart disease or not.

We will use the following steps to predict the heart disease

* Understanding the Problem
* Reading and understanding the data
* Exploratory Data Analysis and visualisation
* Modelling
* Generate Insight

This Project will be divided into two parts:

1. Exploring the dataset and training the model.
2. Building and hosting a web app.

With accurate predictions, we can solve the unnecessary trouble. Besides, if we can apply our machine learning tool into medical prediction, we will save human resource because we do not need complicated diagnosis process in hospitals. (Though it is a very long way to go.) The output will be a binary number 1 or 0. 1 indicates the patient has heart disease and vice versa.

There are 14 features in the dataset, which are described below.

* age: age in years
* sex: sex (1 = male; 0 = female)
* cp: chest pain type
  + Value 0: typical angina
  + Value 1: atypical angina
  + Value 2: non-anginal pain
  + Value 3: asymptomatic
* trestbps: resting blood pressure (in mm Hg on admission to the hospital)
* chol: serum cholesterol in mg/dl
* fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
* restecg: resting electrocardiographic results
  + Value 0: normal
  + Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  + Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
* thalach: maximum heart rate achieved
* exang: exercise induced angina (1 = yes; 0 = no)
* oldpeak = ST depression induced by exercise relative to rest
* slope: the slope of the peak exercise ST segment
  + Value 0: upsloping
  + Value 1: flat
  + Value 2: downsloping
* ca: number of major vessels (0-3) coloured by fluoroscopy
* thal: Thalassemia
  + 0 = normal
  + 1 = fixed defect
  + 2 = reversable defect
* condition: 0 = no disease, 1 = disease

**Hardware and Software to be used:**

**Hardware:**

• Processor (CPU) with 2 gigahertz (GHz) frequency or above

• A minimum of 2 GB of RAM

• A minimum of 20 GB of available space on the hard disk

• Internet Connection Broadband (high-speed) Internet connection with a speed of 4 Mbps or higher

**Software:**

• Google Chrome

• Visual Studio Code

* Jupyter notebook

**What contribution would the Project make:**

As time is passing, a lot of research data and patients records of hospitals are available. There are many open sources for accessing the patient’s records and researches can be conducted so that various computer technologies could be used for doing the correct diagnosis of the patients and detect this disease to stop it from becoming fatal. Nowadays it is well known that machine learning and artificial intelligence are playing a huge role in the medical industry. We can use different machine learning and deep learning models to diagnose the disease and classify or predict the results. A complete genomic data analysis can easily be done using machine learning models. Models can be trained for knowledge pandemic predictions and also medical records can be transformed and analysed more deeply for better predictions. Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment. The early prognosis of cardiovascular diseases 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.

**Conclusion:**

The major challenge in heart disease is its detection. There are instruments available which can predict heart disease but either they 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. Since we have a good amount of data in today’s world, we can use various machine learning algorithms to analyse the data for hidden patterns. The early prognosis of cardiovascular diseases can aid in making decisions on lifestyle changes inhigh risk patients and in turn reduce the complications, which can be a great milestone in the field of medicine.