HEART DISEASE PREDICTION

PROBLEM STATEMENT

Heart disease describes a range of conditions that affect your heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias) and heart defects you’re born with (congenital heart defects), among others.

The term “heart disease” is often used interchangeably with the term “cardiovascular disease”. Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart’s muscle, valves or rhythm, also are considered forms of 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 clinical data analysis.

The aim is to build a predictive model that predicts whether a person has a heart disease or not.

DATASET

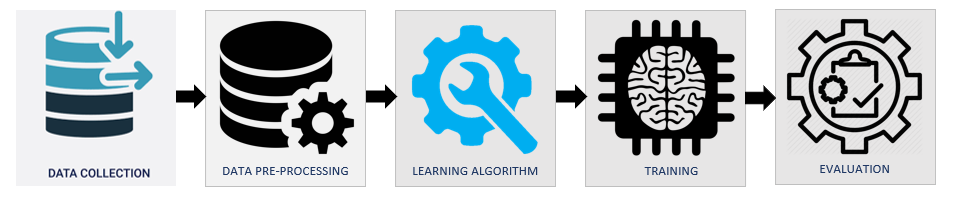
The dataset contains 76 attributes, but all published experiments refer to using a subset of 14 of them. In particular, the Cleveland database is the only one that has been used by ML researchers to this date. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4.

* Age: displays the age of the individual.
* Sex: displays the gender of the individual using the following format :
  + 1 = male
  + 0 = female
* Chest-pain type: displays the type of chest-pain experienced by the individual using the following format :
  + 1 = typical angina
  + 2 = atypical angina
  + 3 = non — anginal pain
  + 4 = asymptotic
* Resting Blood Pressure: displays the resting blood pressure value of an individual in mmHg (unit)
* Serum Cholestrol: displays the serum cholesterol in mg/dl (unit)
* Fasting Blood Sugar: compares the fasting blood sugar value of an individual with 120mg/dl. If fasting blood sugar > 120mg/dl then : 1 (true)

else : 0 (false)

* Resting ECG : displays resting electrocardiographic results
  + 0 = normal
  + 1 = having ST-T wave abnormality
  + 2 = left ventricular hyperthrophy
* Max heart rate achieved : displays the max heart rate achieved by an individual.
* Exercise induced angina :
  + 1 = yes
  + 0 = no
* ST depression induced by exercise relative to rest: displays the value which is an integer or float.
* Peak exercise ST segment :
  + 1 = upsloping
  + 2 = flat
  + 3 = downsloping
* Number of major vessels (0–3) colored by flourosopy : displays the value as integer or float.
* Thal : displays the thalassemia :
  + 3 = normal
  + 6 = fixed defect
  + 7 = reversible defect
* Diagnosis of heart disease : Displays whether the individual is suffering from heart disease or not :
  + 0 = absence
  + 1, 2, 3, 4 = present.

PROPOSED SYSTEM



* Data Collection – UCI Machine Learning repository
* Data pre-processing – Treating missing values, Feature scaling, Feature selection, Label Encoding etc..
* Algorithm selection – kNN
* Training
* Evaluation