Heart Disease Prediction

Milestone: Project Proposal

Group 3

Mugdha Sanjay Parbat - 002142372

Pranav Chandrakant Pulkundwar - 002121679

Telephone

+1 (617) 901-8417

+1 (617) 901-8418

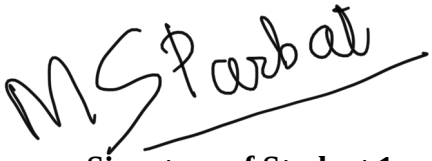
Email ID

parbat.m@northeastern.edu

pulkundwar.p@northeastern.edu

Percentage of contribution by Student 1: 50%

Percentage of contribution by Student 2: 50%



Signature of Student 1:

Signature of Student 2:

Submission Date: 16th February 2022

Problem Setting:

About 659,000 people in the United States die from heart disease each year—that's 1 in every 4 deaths. Heart disease or 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.

A major challenge faced by the healthcare industry is the difficulty to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate, and many other factors. Due to such constraints, scientists have turned towards modern approaches like Data Mining and Machine Learning for predicting the disease. Machine learning (ML) proves to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry.

Problem Definition:

In clinical data analysis prediction of cardiovascular disease is one of the most critical and highly important subjects. Recent studies show that heart disease is one of the leading causes of death regardless of the sex of a person.

In this project, different Machine Learning approaches are to be applied and compared to determine whether a person is suffering from heart disease or not.

Data Sources:

This database 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 names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

The dataset has been obtained from the following website :-

https://archive.ics.uci.edu/ml/datasets/heart+disease

The following are the creators and donors:-Creators:

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi, M.D.
2. University Hospital, Zurich, Switzerland: William Steinbrunn, M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.

4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano, M.D., Ph.D.

Donor:

David W. Aha (aha '@' ics.uci.edu) (714) 856-8779

Data Description:

This data is the collection of information from the Cleveland Heart Disease dataset taken from the UCI repository, formed by taking into consideration some of the information of 779 individuals. The dataset consists of 303 rows and 14 columns of data. The 14 features or columns considered in this analysis are as follows,

1. Age
2. Sex
3. Chest-pain type
4. Resting Blood Pressure
5. Serum Cholesterol
6. Fasting Blood Sugar
7. Resting ECG
8. Max heart rate achieved
9. Exercise induced angina
10. ST depression induced by exercise relative to rest
11. Peak exercise ST segment
12. Number of major vessels colored by fluoroscopy
13. Thal (Thalassemia)
14. Diagnosis of heart disease