Documentation of the Dataset

Name: Cardio Vascular Disease Prediction.

Introduction: The dataset contains the information of the people on different aspects. Cardio Vascular Disease is very common to world - wide. Cardiovascular Diseases (CVDs) are the leading cause of death globally, taking an estimated 17.9 million lives each year. CVDs are a group of disorders of the heart and blood vessels and include coronary heart disease, cerebrovascular disease, rheumatic heart disease and other conditions. More than four out of five CVD deaths are due to heart attacks and strokes, and one third of these deaths occur prematurely in people under 70 years of age. This dataset has information of people that have been collected from:

Citation of the dataset:

1. Dataset1: We take all the data from this site on 1 to 304 in our dataset.

2. Dataset2: From 305 to 632 in our dataset have been developed by us.

3.Dataset 3 : The dataset consist 1190 instances. We only take the first 365 data from 634 to 999 in our dataset.

Size of the Dataset: This dataset has 999 rows and 13 columns in total.

Description of the features: There are 13 features in each row of the dataset.

They are:

1. Age: Displays the age of the individual in years. Age can't be zero or negative.

2.Gender: Displays the gender of the individual with the following format:

1= male

0 = female

3. Exang (exercise induced angina): Angina is usually felt in the center of chest but may spread to either or both of the shoulders, back , neck , jaw or arm. The pain or discomfort associated with angina usually feels tight, gripping or squeezing, and can vary from mild to severe.

1 = Yes (The person has angina)

0 = No

4.Weight: Displays the weight of the individual in kilogram (kg).Weight can't be zero or negative.

5.Trestbps(Resting Blood Pressure): Displays the resting blood pressure value of an individual in mmHg(unit). Over time, high blood pressure can damage arteries that feed the heart. High blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, increases the cardio vascular risk even more. Resting blood pressure of anything above 120 – 140 is a matter of concern.

6.Thalac(Maximum Heart Rate): Displays the max heart rate achieved by an individual. People with a maximum heart rate of over 140 are more likely to have cardio vascular disease. The increase in cardiovascular risk, associated with the acceleration of heart rate, was comparable to the increase in risk observed with high blood pressure. It has been shown that an increase in heart rate by 10 beats per minute was associated with an increase in the risk of cardiac death by at least 20%, and this increase in the risk is similar to the one observed with an increase in systolic blood pressure by 10 mm Hg.

7.Cholesterol: A high level of low-density lipoprotein (LDL) cholesterol (the "bad" cholesterol) is most likely to narrow arteries that ups the risk of heart attack. However, a high level of high-density lipoprotein (HDL) cholesterol (the "good" cholesterol) lowers your risk of a heart attack. Cholesterol greater than 200 is of concern.

8.FastingBS(Fasting Blood Sugar): Compares the fasting blood sugar value of an individual with 120 mg/dl. If fasting blood sugar > 120 mg/dl then 1(true , the person has diabetes). Else: 0 (false).

9.Slope: The slope of the peak exercise ST segment —

0: downsloping

1: flat

2: upsloping

A treadmill ECG stress test is considered abnormal when there is a horizontal or down-sloping ST-segment. Exercise ECGs with up-sloping ST-segment depressions are typically reported as an 'equivocal' test.

10.ChestPain: Displays the type of chest-pain experienced by the individual using the following format:

1 = typical angina

2 = atypical angina

3 = non – angina pain

4 = asymptopic

11.Resting ECG: Displays resting electrocardiographic results.

0 = normal

1 = having ST-T wave abnormality

2 = left ventricular hypherthrophy

12 . Thal (thalassemia): Displays the thallasemia of an individual.

1 = normal

2 = fixed defect

3 = reversible defect

13 . Cardio: The target value that defines if one will have cardio vasucular disease or not.

1 = Yes

0 = No