

Dataset description

Supervised machine learning algorithms operate on a dataset that is a collection of labeled examples which consist of features and a label i.e. in our case *target* is a label and other columns are features, and each row is an example.

This dataset consists of 11 features and a target variable. It has 6 nominal variables and 5 numeric variables. The detailed description of all the features are as follows:

- 1. Age:** Patients Age in years (Numeric)
- 2. Sex:** Gender of patient (Male - 1, Female - 0) (Nominal)
- 3. Chest Pain Type:** Type of chest pain experienced by patient categorized into 1 typical, 2 typical angina, 3 non-anginal pain, 4 asymptomatic (Nominal)
- 4. resting bps:** Level of blood pressure at resting mode in mm/HG (Numerical)
- 5. cholesterol:** Serum cholesterol in mg/dl (Numeric)
- 6. fasting blood sugar:** Blood sugar levels on fasting > 120 mg/dl represents as 1 in case of true and 0 as false (Nominal)
- 7. resting ecg:** Result of electrocardiogram while at rest are represented in 3 distinct values 0 : Normal 1: Abnormality in ST-T wave 2: Left ventricular hypertrophy (Nominal)
- 8. max heart rate:** Maximum heart rate achieved (Numeric)
- 9. exercise angina:** Angina induced by exercise 0 depicting NO 1 depicting Yes (Nominal)
- 10. oldpeak:** Exercise induced ST-depression in comparison with the state of rest (Numeric)
- 11. ST slope:** ST segment measured in terms of slope during peak exercise 0: Normal 1: Upsloping 2: Flat 3: Downsloping (Nominal)

Target variable:

- 12. target:** It is the target variable which we have to predict 1 means patient is suffering from heart risk and 0 means patient is normal.