

Contents

| Topic | Pg No. |
|---|--------|
| Problem Description | 2 |
| The Datasets are provided as cited below for Predicting the “Heart disease and it's severity” | 2 |
| Main Tasks: | 2 |
| Primary Accuracy Metric | 3 |

Predicting the “Heart disease and it's severity”

Problem Description:

To predict the heart disease with either of the levels None , MildClass1, MildClass2, Moderate, Severe

In the age of increasing life expectancy, Cardio vascular disease continues to be the biggest contributing cause to human deaths. While lot of research has gone into the risk factors (like life style, unhealthy diet, tobacco consumption and lack of exercise) and the related socio-economic determinants, it continues to attract the attention of Societies/Governments, people and other stakeholders alike.

The societies and individuals are keen to predict the disease and plan prevention and/or appropriate interventions within the constraints of income and health care provisions. While the focus of the medical profession and healthcare industry shifted to prevention and wellness management, the focus of modern research shifted to address information gaps and refinement of knowledge. The new information arising out of predicting the heart disease severity (None , MildClass1, MildClass2, Moderate, Severe) may help plan appropriate interventions to reverse and or control the disease and bring the affected person to normal conditions or reduce potential disability thereby saving insurance and social costs.

Thus, predictive algorithms have been of interest and importance to medical or health care practitioners, researchers and public health managers to go beyond the clinical data and mine the data for developing predictive models.

You are expected to create an analytical and modelling framework to predict the heart disease condition, if so, else none based on the quantitative and qualitative features provided in the dataset while answering other questions too cited below.

The datasets are provided as cited below:

1. **AttributeDescription.csv** - Depicts attributes description
2. **Data :**
 - i. Train.csv (To build the model and tune the model & visualization)
 - ii. Test.csv (Model will be tested on this dataset in SCT tool)

Missing Values : Denoted as “”, “NA”, “-9”.
Target attribute : “Severity”

Main Tasks:

1. Exploratory Data Analysis using visualizations in R Notebook or Jupiter notebook format (Use only Train data for this task)
2. Based on domain and data understanding, which attributes do you intuitively think would impact the severity of the disease?
3. Learning curves :what is your observation based on the learning curves? Is there any bias or variance problem in the data or none?
4. Based on the learning curves observation, which model do you think is suitable for the data and why?
5. Using the ML methods, which attributes do you find as important and do you agree with the observations as per (2) in the Main Tasks? Explain.
6. You are expected to build a framework that predicts the heart disease severity (The target attribute "Severity").
7. Viva

Primary Accuracy Metric:

Accuracy