

A healthcare organization together with a couple of government hospitals in a city has collected information about the vitals that would reveal if the person might have a coronary heart disease in the next ten years or not. This study is useful in early identification of disease and have medical intervention if necessary. This would help not only in improving the health conditions but also the economy as it has been identified that health performance and economic performance are interlinked. Given the data, we need to develop appropriate models to identify/predict if the person is likely to have heart disease or not.

Data:

1. The total number of records is 34281 with 24 independent attributes and the 25th column is the target which needs to be predicted.
2. The variables are masked, and we get little information from their names.
3. Missing values are represented as NA in some columns and as -99 in some other columns.

Analysis and report:

1. Analysis is expected in R (Submit detailed commented code - both Rmd and html files)
2. Generate any visualizations that provide insights about the data.
3. Prepare a presentation which should explain the process of implementation along with visualizations and results interpretation.

Evaluation Metric:

Considering the given case, decide a metric which you should be focussing on for improvement. (Accuracy/Recall/Precision/F1 Score)

File Naming Convention for submission in SCT:

Naming conventions for all the files (mandatory):

<FirstName>_<LastName>_B56.<extension>

Example:

Seema_Basantani_B56.Rmd

Seema_Basantani_B56.html

Seema_Basantani_B56.pptx

Seema_Basantani_B56.zip