BITS Pilani Pilani | Dubai | Goa | Hyderabad Work Integrated Learning Programmes	Introduction to Data Science M.Tech Data Science and Engineering	Start Date : 21-06-2021 End Date: 21-08-2021

Heart Attack Analysis

Introduction

A heart attack occurs when an artery supplying your heart with blood and oxygen becomes blocked. A blood clot can form and block your arteries, causing a heart attack. This Heart Attack Analysis helps to understand the chance of attack occurrence in persons based on varied health conditions.

Objective

Follow the Data Science Methodology that we studied in Module 3. Analyze the data set and identify the most relevant parameters that leads to heart attack. Answer the questions like

- Increasing in age has any effect towards heart attack.
- Does increase in cholesterol level in body have any effect towards the heart attack
- Increase in blood pressure have any relation with heart attack
- Does family history have any effect towards heart attack

Dataset

The dataset is Heart_Attack_Analysis_Data.csv. It has been uploaded to canvas.

This dataset contains data about some hundreds of patients mentioning Age, Sex, Exercise Include Angia(1=YES, 0=NO), Chest Pain Type(Value 1: typical angina, Value2: atypical angina, Value 3: non-anginal pain, Value 4: asymptomatic), ECG Results, Blood Pressure, Cholesterol, Blood Sugar, Family History (Number of persons affected in the family), Maximum Heart Rate, Target -0=LESS CHANCE , 1= MORE CHANCE

Tasks in this assignment

1) Write a Data Science Proposal for achieving the objective mentioned.
2) Perform exploratory analysis on the data and describe your understanding of the data.
3) Perform data wrangling / pre-processing (Module 8).
    a. E.g., missing data, normalization, discretization, etc.
4) Apply any two filter feature selection engineering techniques
5) Compare the two selected feature engineering techniques.
6) Provide a high-level description of Machine Learning models – Gradient Boosting classifier and Random Forest to predict.
7) Compare the performance of the two classifiers – Gradient Boosting classifier and Random Forest .
8) Present the conclusions/results in the format shared.

BITS Pilani Pilani | Dubai | Goa | Hyderabad Work Integrated Learning Programmes	Introduction to Data Science M.Tech Data Science and Engineering	Start Date : 21-06-2021 End Date: 21-08-2021

Expected Submissions

Two files are expected as the assignment submission.

1. The summary of the work in the template provided. (you may fill only the boxes relevant to this problem statement)
2. The executed ipynb file with clear subdivision of the codes and brief description of the purpose of respective code. All the executed tables or graphs and results should be present in the ipynb file. The ipynb file maybe submitted as a single .pdf file.