**Problem Statement**

**Introduction**

A heart attack, medically known as a myocardial infarction, occurs when blood flow to the heart is severely reduced or blocked, typically due to a buildup of fat, cholesterol, and other substances forming a plaque in the coronary arteries. This blockage can lead to the formation of a blood clot, which disrupts the oxygen supply to the heart muscle, potentially causing damage or death to the affected heart tissue.

This analysis leverages data-driven approaches to identify patterns and correlations that contribute to heart attack risks. By applying machine learning algorithms, healthcare providers can estimate the probability of a heart attack for a specific individual and provide personalized recommendations for preventive measures.

The goal is to enable early detection and proactive management of heart disease, ultimately reducing the prevalence of heart attacks and improving patient outcomes. This problem is crucial for advancing preventive cardiology and empowering individuals to take control of their cardiovascular health.

**Dataset**

The dataset is Heart\_Attack\_Analysis\_Data.csv.

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

1. **Import Libraries/Dataset** 
   1. Download the dataset.
   2. Import the required libraries.
2. **Data Visualization and Exploration [1M]** 
   1. Print at least 5 rows for sanity check to identify all the features present in the dataset and if the target matches with them.
   2. Print the description and shape of the dataset.
   3. Provide appropriate visualization to get an insight about the dataset.
3. Do the correlational analysis on the dataset. Provide a visualization for the same. Will this correlational analysis have effect on feature selection that you will perform in the next step? Justify your answer. **Answer without justification will not be awarded marks.**
4. **Data Pre-processing and cleaning [2M]** 
   1. Do the appropriate pre-processing of the data like identifying NULL or Missing Values if any, handling of outliers if present in the dataset, skewed data etc. Mention the pre-processing steps performed in the markdown cell.
   2. Apply appropriate feature engineering techniques. Apply the feature transformation techniques like Standardization, Normalization, etc. You are free to apply the appropriate transformations depending upon the structure and the complexity of your dataset. Provide proper justification. **Techniques used without justification will not be awarded marks**. Explore a few techniques for identifying feature importance for your feature engineering task.
5. **Model Building [5M]**
   1. Split the dataset into training and test sets. **Answer without justification will not be awarded marks.** [1M]
      1. Train = 80 % Test = 20%
      2. Also, try to split the dataset with different ratios of your choice.
   2. Build model using Logistic regression and decision tree [4 M]
      1. Tune hyperparameters (e.g., number of trees, maximum depth) using cross-validation. Justify your answer.
6. **Performance Evaluation [2M]**
   1. Compare the performance of the Logistic Regression and Decision Tree models using appropriate evaluation metrics.
   2. Provide insights into which model performs better and why. **Answer without justification will not be awarded marks.**

**Instructions for Assignment Evaluation**

1. Organise your code in separate sections for each task. Add comments to make the code readable.
2. Deep Learning Models are strictly not allowed. You are encouraged to learn classical Machine learning techniques and experience their behaviour.
3. Notebooks without output shall not be considered for evaluation.

***For clarifications, contact Rajesh Pandey <rpandey@wilp.bits-pilani.ac.in>***