

# Project Proposal: Early Detection of Cardiovascular Diseases Using Classification

## Team Members:

- ❖ Mohammad Jambughodawala(2022A7PS0044H)-f20220044@hyderabad.bits-pilani.ac.in
- ❖ G.B.Hemanth Karthikeya (2022A7PS0053H)-f20220053@hyderabad.bits-pilani.ac.in
- ❖ Anshul Gopal(2022A7PS0009H)-f20220009@hyderabad.bits-pilani.ac.in

## Data:

- We plan to utilize the publicly available [cardiovascular dataset](#), which contains anonymized records of 4000 patients. The dataset includes information on heartbeat, diabetes, smoking status, and other relevant factors.

## Problem:

- **Early detection of Cardiovascular disease** is often challenging due to its non-specific symptoms and the limitations of traditional tests. This can lead to delayed diagnosis, worse health outcomes, higher healthcare costs, and even death. By improving early detection and awareness, we aim to save lives and promote better heart health for everyone.

## Why is it interesting:

- Early detection significantly increases the success rate of treatment and reduces the severity of heart disease. This project aims to develop data-driven models that can **classify** individuals at high risk for developing cardiovascular disease at an earlier stage, potentially leading to:
  - Increased survival rates, reduced treatment intensity
  - Decreased healthcare costs

## Implementation and Learning:

- **Data Preprocessing:** We will perform data cleaning and feature engineering techniques to ensure data quality and identify the most relevant features for classification.  
**Feature subset selection** will be employed to reduce redundancy and improve model efficiency.

- **Classification Algorithms:** Each team member will implement and evaluate different data mining algorithms, focusing on specific aspects of early detection:
  - **K-Nearest Neighbours (KNN):** Classifies based on the majority class of similar neighbours.
  - **Decision Trees:** Uses a tree of rules to classify data by asking a series of questions about its features.
  - **Naive Bayes:** Classifies by calculating class probabilities based on independent features and prior knowledge.
- **Model Evaluation:** We will compare the performance of each algorithm using metrics like accuracy, precision, recall, and F1-score. This will help us identify the most effective approach for early detection.

#### **Expected Outcomes:**

- We anticipate our project will result in:
  - Development of accurate and interpretable models for early detection of cardiovascular diseases.
  - Identification of previously unknown risk factors or patterns associated with the disease.