**Project Proposal**

**Cardiovascular Disease**

**Munirah Abdullah Alraqibah**

**Abstract:**

Healthcare expenses are overwhelming national and corporate resources due to asymptomatic diseases, including cardiovascular diseases. Therefore, there is an imperative need for primeval discovery and treatment of such diseases. Cardiovascular diseases (CVDs) are a group of disorders of the heart and blood vessels. The aim of this project is predicting heart disease of the patient through some predictors, listed in the description, by using classification techniques.

**Description of the dataset:**

The dataset contains about 70,000 observations “patients” with 13 Features each, those input features have 3 types: Objective: factual information; Examination: results of medical examination; and Subjective: information given by the patient. The predictors are Gender, Age, Height, Weight, Systolic blood pressure, Diastolic blood pressure, Cholesterol, Glucose, Smoking, Alcohol intake, and Physical activity. The response variable is a binary variable with a patient having a value of ‘0’ if they do not have heart disease and a value of ‘1’ if they do.

Data Source References: <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>

**Tools:**

Logistic Regression and KNN are classification techniques that is going to be use in this project along with pandas, numpy and seaborn environments. Moreover, maybe later is going to be use more environments as long as it will be useful to the data.

**MVP Goal:**

* Having clean data.
* Doing exploratory data analysis (EDA).
* Fitting the data with right model.
* Using visualization to understand the data more