**TASK 1**

**Problem to solve**

We chose category healthcare from Sustainable Development Goals(SDG) topic for this project. A classification of heart disease patients is a great importance in cardiovascular disease diagnosis where this system is expected to aid healthcare professionals in the diagnosis of heart disease. But however, the diagnosis system requires a various features of dataset of every possible class of the disease in order to make the system reliable. There are a few problems on developing the system:

1. Clinical decision are often made based on doctor’s insight and experience rather that knowledge rich data hidden in the dataset.
2. Classified using data based on doctor’s insight leads to unwanted biases, errors and excessive medical cost which affect the providing services quality to patients.

**Objective**

The objective of this project are:

1. To predict the presence of heart disease in the patients.
2. To get the actual accuracy rate of the dataset.
3. To classify the heart disease based on patients symptoms.

**Goals**

The prediction system is expected to provide appropriate results and accuracy and making effective decisions on medical parameters data such as age, sex, blood pressure and obesity level for prediction. Significant factors like relationship between medical factors related to heart disease and pattern help in predict the likehood of patients getting a heart disease.

The knowledge that has been implemented will give the expert to get a better quality and reduce the extent of adverse medicine effect.

**Question**

1. How can we optimize the prediction using various data of patients?
2. Do we have a reliable source to make sure this dataset is fit for the project?
3. How long this system will be relevance due to the rapid growth of technology?

**Success**

The system can be finalized as successful when the classification performs a very good accuracy in term of prediction as well as generating a proper classified disease.

**Measurement**

The main measurement of this system is the accuracy of the system. An accuracy of 90.32% percent an above is expected.

**Data Sources**

Our data that were used to conduct the experiments which can be found as an open source dataset at Kaggle. The dataset consist of various features that affect the patients preferences on determine some sickness of the patients where most of them were contributed by online community that works on various data science project.