By Abdullah Alsubaie and Hesham Doom

Project Report- Heart Disease

Description

In this project, we used Machine Learning to predict whether a person is suffering from a heart disease. After importing the data, we analyzed it using plots. Then, we did generated dummy variables for categorical features and scaled other features. We then applied four Machine Learning algorithms, `Logistic Regression`, `Random Tree Classifier` and `K Neighbors Classifier`. We varied parameters across each model to improve their scores. In the end, `K Neighbors Classifier` achieved the highest score of `87%` with `8 nearest neighbors`.

Steps:

Import Modules

First let's import the libraries which we will use in the project such as `numpy` and `pandas` to start with. we will use `pyplot` subpackage of `matplotlib` to make visualization. for machine learning , wewill use `sklearn` library.

Import dataset

Now that we have all the libraries we will need, we can import the dataset and take a look at it. The dataset is stored in the file `dataset.csv`. I'll use the pandas `read_csv` method to read the dataset.

Visualization

It's always a good practice to work with a dataset where the target classes are of approximately equal size. Thus, let's check for the same.

Data Processing

After exploring the dataset, we observed that we need to convert some categorical variables into dummy variables and scale all the values before training the Machine Learning model.

Machine Learning

We'll now import `train_test_split` to split our dataset into training and testing datasets. Then, we'll import all Machine Learning models I'll be using to train and test the data.