Heart Disease Writeup

Prepared by

Rania Almneie
Hanadi Alshahrani
Najd Alqahtani
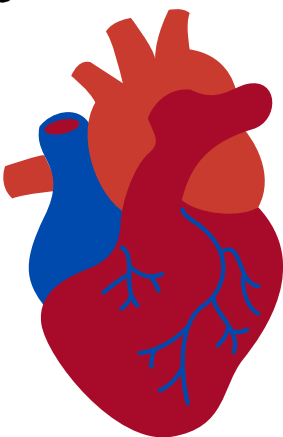
Abstract

We all know that early detection of heart disease is critical for human life . So we wanted to build a classification model to predict whether the patient had heart disease or not.

We obtain the data from the Kaggle website , then apply the EDA to this data, and it is now ready to predict our target using a classification model for more details, refer to the following components:

"A healthy heart is the main source of your strength."

Design

We wanted as a team to analyze the heart disease data to predict whether the patient had heart disease or not.

Algorithms

The first step to achieve our goal was EDA, which included data exploration, cleaning, handling the outliers and and data visualization with seaborn, matplotlib and tableau and plotty.
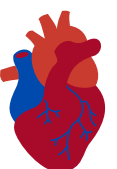After that we built a classification model, by used sk-learn.

Data

The dataset was downloaded from the Kaggle website and consisted of 1026 observations. The predictor Y (Positive or Negative diagnosis of Heart Disease) is determined by 14 features (X):
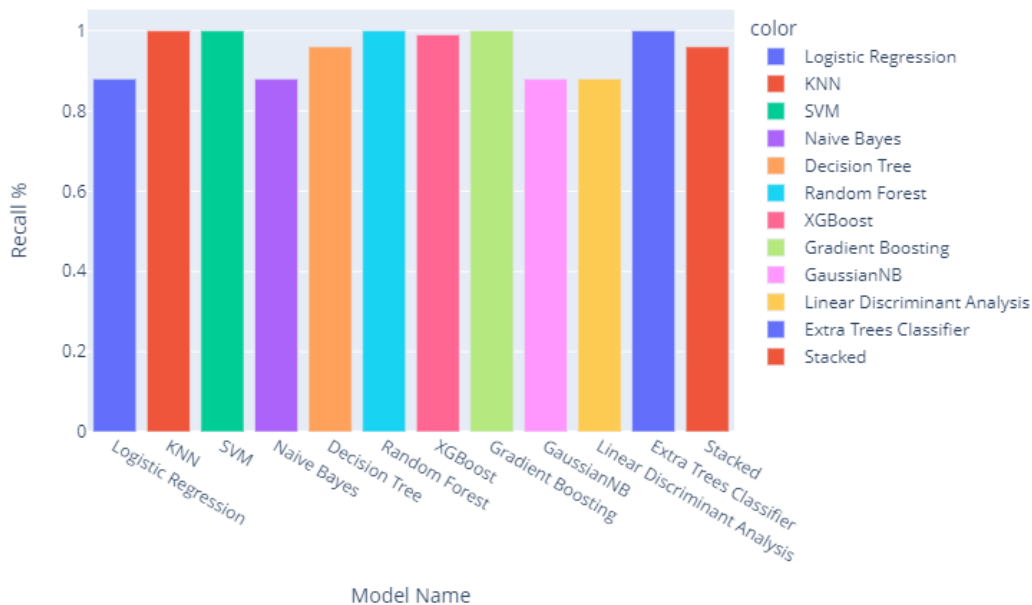
- age: age in years
- sex: male, female
- cp: chest pain type
- trestbps: resting blood pressure
- chol: serum cholesterol
- fbs: fasting blood sugar
- restecg: resting electrocardiographic results
- thalach: maximum heart rate achieved
- exang: exercise induced angina
- oldpeak: ST depression induced by exercise relative to rest
- slope: the slope of the peak exercise ST segment
- ca: number of major vessels
- thal: number of defect type
- target: disease, no disease

Communication

Comparision of Classification Models

Depending on our goal, we tried to reduce the number of incorrect predictions for people with heart disease (Recall). The classifiers that we considered to be the best in terms of recall are depicted in the plot (KNN, SVM, Random Forest, Gradient Boosting, and Extra Trees) that have 1 recall.

Tools:

- Pandas
- NumPy
- Heroku
- Flask
- Seaborn
- Plotly
- matplotlib
- Sklearn
- SQL Alchemy