Predicting the risk factors of cardiology diseases

Abstract

The objective of this project is to use classification models to formulate cardiology disease risk factors, including age and gender, chronic diseases, smoking, obesity, high cholesterol level, and sedentary lifestyle.

Predicting the risk factors of cardiology diseases to help diagnose patients early, start treatment, and prevent complications.

Design

Modifiable and nonmodifiable risk factors. non-modifiable risk factors including age, gender, and chronic diseases. modifiable risk factors as treating high cholesterol levels, smoking caseation, exercise, and diet to lose weight.

With this data, we are increasing social awareness and preventing further complications by detecting the disease as early as possible.

Data

The dataset contains 70,000 data points with 14 features for each.

include age, gender, weight and height, chronic diseases (diabetes mellitus, hypertension, hypercholesteremia), smoking, and a sedentary lifestyle.

Algorithms

Models:

Logistic regression, k-nearest neighbors, and random forest classifiers were used before settling on decision Tree classifier as the model with strongest cross-validation performance.

*Final decision Tree 5-fold CV scores:

Accuracy 0.63

Predicting the risk factors of cardiology diseases

-F1 score 63

-precision 0.63 macro avg.

-recall 0.63 macro avg.

Tools

-Numpy and Pandas for data manipulation.

-Scikit-learn for modeling. sklearn.metrics

-Matplotlib and Seaborn for plotting.

-Tableau for interactive visualizations.