أكاديمية سدايا
SDAIA Academy

Cardio Disease

Project Proposal for Classification

Abstract:

Cardiovascular disease is the leading cause of death throughout the United States. So, we will use this data to build classification models and to detect the presence of cardiovascular disease.

Question/need:

The goal of this project is to build classification models that address a useful prediction of heart disease. Physicians and patients will benefit from this predicted data.

Data Description:

The cardiovascular disease dataset is found on Kaggle. The data consists of 70,000 patient records and 14 features:

- id number.
- age in days.
- age in years.
- gender (1 - women, 2 - men).
- height cm.
- weight kg.
- ap_hi (Systolic blood pressure).
- ap_lo (Diastolic blood pressure).
- cholesterol (1: normal, 2: above normal, 3: well above normal).
- gluc (1: normal, 2: above normal, 3: well above normal).
- smoke (whether patient smokes or not (0 = no, 1 = yes)).
- alco (0 = no, 1 = yes).
- active (0 = passive life, 1 = active life).
- Target variable: cardio (0 = no, 1 = yes).

Tools:

The tool that we need is Pandas packages to manipulate data, Scikit Learn Framework (model_selection, sklearn.metrics, train_test_split , sklearn.linear_model) we will import (LogisticRegression , KNeighborsClassifier , DecisionTreeClassifier) and visualization library (such as Seaborn and Matplotlib), Metrics library to see the model performance on data, and Jupyter notebook to execute the code.

By Ghadah Alharbi and Rahaf Alyousef