T5 Bootcamp

18 December 2021

Documentation

Abstract

The heart disease accounts to be the leading cause of death worldwide. It is difficult for medical practitioners to predict the heart attack as it is a complex task that requires experience and knowledge. The health sector today contains hidden information that can be important in making decisions. Data mining algorithms such KNeighbors Classifier, Decision Tree, Random Forest, Logistic, XGB, LGBM are applied in this project for predicting CVD. The project result shows prediction accuracy of 97%. Data mining enable the health sector to predict patterns in the dataset.

Design

The project was designed with medical history related input and trined to predict the medical states of certain patient regarding to CVD, in addition this project is one of the T5 Data Science BootCamp requirements.

Data

- Objective: factual information.
- Examination: results of medical examination.
- Subjective: information given by the patient.
- Data provided by (https://www.kaggle.com/sulianova/cardiovascular-disease dataset) has been used in this project.

Algorithms

- KNeighbors Classifier
- Decision Tree
- Random Forest
- LGBM
- Logistic

Tools

- Python visualization libraries (matplotlib).
- Data analysis in pandas for discovering the data and train a model.