

Heart Disease Prediction

TEAM 8

Guoshan Yu, Jiwoo Suh, Kyuri Kim

Problem Statement

Heart disease is a leading cause of death in the US, affecting people of most races. Detecting and preventing the key risk factors for heart disease is crucial in healthcare. This project aims to develop a machine-learning model to predict a patient's likelihood of having heart disease based on key indicators such as high blood pressure, high cholesterol, smoking, diabetic status, obesity, physical activity, and alcohol consumption.

Dataset

We will use the Personal Key Indicators of Heart Disease dataset from Kaggle:
'<https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease>'
which contains 319,795 samples with 18 attributes, with the target variable "HeartDisease" indicating whether the respondent had heart disease or not.

Methodology

The neural network used in this project is a Multi-Layer Perceptron (MLP), a type of neural network that is commonly used in classification tasks. In addition to MLP, machine learning algorithms such as logistic regression, SVM, decision tree, and random forest may be used for comparison and ensemble learning. The scikit-learn library in Python will be used for implementing these algorithms.

Evaluation

The performance of the neural network will be evaluated using various metrics such as accuracy, precision, recall, F1 score, and ROC AUC score. Since the dataset is imbalanced, techniques such as oversampling, undersampling, or weighting will be applied to balance the classes and improve the model's performance.

Schedule

The project is expected to be completed by the last week of April. The schedule is as follows:
Week 2: Data preparation and preprocessing
Week 3: Model development and optimization & Model evaluation and comparison with other algorithms
Week 4: Report writing and presentation preparation

References

The following reference material will be used to obtain sufficient background on applying MLP to the specific problem of heart disease detection:

1. "Neural Network Design" by Martin T. Hagan