ReadMe

Authors: Avivartta Krishna, and Binnu Ponugoti

The aim of his project is to compare and contrast the different machine learning classifiers such as, SVM, Naive Bayes, Decision Tree, Logistic Regression, Random Forest and KNN.

We have also included a neural network testing multiple optimizers and added Explainability using SHAP, LIME and ELI5.

Heart Disease Dataset

Source: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset
The dataset contains the different features that patients with a heart disease may have.

- age
- sex
- chest pain type (4 values)
- resting blood pressure
- serum cholesterol in mg/dl
- fasting blood sugar > 120 mg/dl
- resting electrocardiographic results (values 0,1,2)
- maximum heart rate achieved
- exercise induced angina
- oldpeak = ST depression induced by exercise relative to rest
- the slope of the peak exercise ST segment
- number of major vessels (0-3) colored by fluoroscopy
- thal: 0 = normal; 1 = fixed defect; 2 = reversible defect

The last column addresses whether the patient meets the Target, that is, whether they have a heart disease or not.

How to Run the Code

We recommend starting with the Elementary Data Analysis of both, the big and small dataset. The code entails a preliminary statistical analysis and the different machine learning classifiers. It includes the accuracy, precision and F1 score of each classifier, in addition to their learning curves.
We then suggest running the neural network file where we have compared the accuracy and loss generated by optimizers such as Adam, Adagard, SGD, and Ftrl .
Finally, we suggest running the file named SHAP, LIME, and ELI5.