Project Review

Alex Brown

10-26-2023

The Project:

Data Exploration and Preparation:

The data required thorough research to properly classify each feature. Some features were removed since the dataset was assumed to contain only patients from the United States. This was purely part of the project and under other circumstances would have been included as part of the data exploration.

Data Analysis:

The target of our analysis was the relationship between all features collectively and their effect on predicting the “Heart Attack Risk”, which in our case has been defined as at least one previous heart attack event in that patient’s past. After preprocessing the data, we selected three models for testing. Below are the results of the initial Train Test Split and the following Cross Validation on each of the three models:

A graph with lines and numbers

Description automatically generated with medium confidence

A graph with lines and dots

Description automatically generated

With default hyperparameters, Logistical Regression had a mean-squared error (MSE) of 0.36, which gave it a distinct advantage over the Neural Network and Random Forest models, which tested between 0.39 and 0.41.

Data Summary:

Phenomenon and Detection:

With the information provided, we will be able to predict the likelihood of a previous heart attack and inform patients with similar backgrounds of their risk.

Assessment of Product Accuracy:

We will be able to inform patients with high-risk and advise changes to their lifestyle. Overall we will be able to identify high-risk patients with 99.64% accuracy.

Results: