Brett Geunes

CMSE 492

Project Proposal: Predicting Heart Attacks Using Various Factor

Introduction:

Heart disease is the leading cause of death for men and women in the US. Predicting the occurrence of heart attacks is crucial in saving lives and learning preventative measures for future cases. This project aims to train a model to predict heart attacks using various health factors.

Data:

The data for this project is obtained from Kaggle's Heart Attack Analysis & Prediction dataset, which provides various demographic, physical, and lifestyle factors along with the presence or absence of heart disease.

Methodology:

Exploratory Data Analysis (EDA) will be performed using Pandas to ensure the data is clean and usable. Seaborn's Pairplot functionality will be utilized for the visualization of the correlations between the different factors. Next, multiple feature selection tools will be used to identify the strongest features in the dataset. Principal Component Analysis (PCA) will then be applied for feature reduction if proven to be useful with the data.

Automated Machine Learning (autoML), such as the Tree-Based Pipeline Optimization Tool (TPOT), will be used to create a baseline model. Other models, such as Support Vector Machines (SVM) and Random Forest, will be tested and compared against the TPOT model. The model will be fine-tuned using the baseline model as the starting point, and the best model will be identified. Ensemble methods will then be explored to determine if combining models will maximize the model's robustness.

Deployment:

The model will be deployed using Github/Binder for easy use and reproducibility. A paper report on methods and findings will be written, providing new insights into heart disease and its causes to improve medical screenings in the future.

Conclusion:

The proposed project aims to predict heart attacks using various factors to save lives and prevent heart disease. The dataset obtained from Kaggle's Heart Attack Analysis & Prediction dataset will be explored

using EDA and different feature selection methods, autoML, and ensemble models to develop a robust model.

Data Link:

https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset

Supporting Visuals: