CS545 Machine Learning
Final Project Proposal
Summer 2023

Team Members:
Cristian Ion
Madeleine Fenner
Leo Lu

Introduction

Heart disease is a leading cause of death around the world, and prevention through early prediction and diagnosis is an important task. To aid in this effort, machine learning models can be applied to large datasets curated with relevant patient records. However the choice of model is not a straightforward process, and previous research articles have aimed to compare the heart disease prediction performance for a variety of models such as Support Vector Machines (SVMs), Multilayer Perceptrons (MLPs), Logistic Regression, Naive Bayes, and XGBoost (Bhatt et al., 2023; Nandal et al., 2022).

Similar to previous research, this project aims to investigate the most effective heart attack prediction model by comparing the performance of two distinct algorithms, SVM and Logistic Regression, utilizing the Cardiovascular Disease Dataset sourced from Kaggle. The dataset contains a comprehensive collection of health attributes and cardiovascular disease outcomes, enabling us to build and evaluate the predictive capacity of the selected models (this will be accomplished using Python as the programming language). Using the results, we will conduct a comparative analysis with measurements of accuracy, precision, and recall, to determine the most effective algorithm for accurate heart attack prediction.

Dataset: [Cardiovascular Disease dataset | Kaggle](#)

Models: SVM, Logistic Regression

Citations
Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms. 2023; 16(2):88. https://doi.org/10.3390/a16020088

Nandal N, Goel L and TANWAR R. Machine learning-based heart attack prediction: A symptomatic heart attack prediction method and exploratory analysis [version 1; peer review: awaiting peer review]. F1000Research 2022, 11:1126. https://doi.org/10.12688/f1000research.123776.1