Proposal For First Large Project

Title: Predicting Heart Attack Risk

Team Members: Navya Sadineni, Usama Ayub

Contributions:

Navya: Will perform data pre-processing and will find best hyperparameter for the models.

Usama: Will perform models implementation and models evaluation. Select best model for the dataset and perform prediction.

Problems to Solve:

We want to develop python program using machine learning models to predict either the patient have a risk of 'Heart Attack' or not. Using the data collected from patient consisting of various medical attributes. Using the statistical machine learning it can be made easy to assist healthcare professionals in early heart disease detection.

Implementation Plan:

We will first understand 'heart attack' dataset and perform data visualisation to get better insight. If needed we will clean the data by handling missing values, outliers, skewness and inconsistencies using transformation methods and using PCA. Perform exploratory data analysis to get better insights of predictor and target variable.

Split the data into training and testing sets and apply feature selection methods suited best to the dataset and train machine learning model.

We will perform model training using Support Vector Machine classifier using the training data. Experiment with different kernel Logistic Regression and Random Forest classifier using the training data and tune hyperparameters.

After that we will evaluate the performance of each model using appropriate evaluation metrics such as accuracy, precision, sensitivity, specificity and area under the ROC curve (AUC-ROC). Compare the performance of the models on the test set to select the best-performing model.

Using techniques like grid search or random search to systematically search through a hyperparameter space and find the optimal hyperparameters for each model. Cross-validation can also be used to assess the generalization performance of different hyperparameter combinations.

Based on the evaluation results and hyperparameter tuning we will select the best-performing model for predicting heart attacks.

Dataset:

https://archive.ics.uci.edu/dataset/45/heart+disease