Capstone Project: Heart Disease Prediction using Machine Learning

By

Rezwanur Rahman (Rez)

Background: Heart disease is very common among different ages. People with certain blood sugar level, cholesterol level etc., may exhibit symptoms for heart disease. It is important to find a smart way to speculate possibility of heart disease given certain physiological status. UCI ML database has a large, collected data for heart disease will be used to develop a ML model.

Problem Statement: This data has several parameters that encompasses a relationship between person’s general health status and potential heart risk. A ML model can be used to extract important parameter (aka features) while analyzing the data. Eventually ML model can be used to predict heart risk.

Dataset and Input: In this project UCI ML dataset on heart disease has been used (<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>). From the dataset has a wide range of input features:

* age: The person’s age in years
* sex: The person’s sex
* trestbps: The person’s resting blood pressure (mm Hg on admission to the hospital)
* chol: The person’s cholesterol measurement in mg/dl
* fbs: The person’s fasting blood sugar
* restecg: resting electrocardiographic results
* thalach: The person’s maximum heart rate achieved
* exang: Exercise induced
* oldpeak: ST depression induced by exercise relative to rest
* slope: the slope of the peak exercise ST segment
* ca: The number of major vessels
* thal: A blood disorder called thalassemia

Solution Statement: To predict heart risk given dataset an exploratory data analysis will be done. Important features will be identified. These will be used to classify possible risk of heart disease or not.

Benchmark Model: Data can be split into train and test parts. Model can be benchmarked based on its performance with test dataset.

Evaluation Metrics: Performance of the ML model can be determined in terms of ROC curve and confusion matrix.

Project Design: Model will be deployed on GitHub repo.