

Project Proposal

Prediction of Probability of Heart Diseases Through Different Parameters With Machine Learning

Hua Mo

Background

Heart disease is the leading cause of death for men, women, and people of most racial and ethnic groups in the United States. One person dies every 37 seconds in the United States from cardiovascular disease. About 647,000 Americans die from heart disease each year, which is 1 in every 4 deaths. Heart disease costs the United States about \$219 billion each year from 2014 to 2015. This includes the cost of health care services, medicines, and lost productivity due to death.

There are different root causes of the heart disease. There are a number of lifestyle choices that can increase the risk of heart disease. These include high blood pressure and cholesterol, smoking, overweight and obesity, diabetes, family history, diet of junk food, age, a history of preeclampsia during pregnancy, staying in a stationary position for extended periods of time. Having any of these risk factors greatly increases the risk of heart disease.

In this project, we will use the heart data from Kaggle and combined with Supervised Machine Learning method to set up a model to predict the possibility of heart diseases. The link of Kaggle data is shown below:

(<https://www.kaggle.com/ronitf/heart-disease-uci>)

The model will be used to in the personal health care business as key to help people adjust the behavior to decrease the probability of heart disease. The personal in the personal health care will be the stakeholders to provide key insights.

Criteria of Success

The criterial of success is to setup a model, which could predict heart disease with supervised machine learning technology, such as RainForest Regression Model, etc..

Scope of Solution Space

To achieve the goal, we will divide the question into several steps:

- Data Wrangling: Understand the features of data and identify which column will be output
- Exploratory Data Analysis: Understand the relationship of the data of the data
- Preprocess and Setup Training Data: pickup the model for the training/test data for prediction of heart disease
- Modeling: Establish Machine Learning Model for the test
- Prediction: Use model to predict the probability of heart disease

Constraints within Solution Space

The constraints within solution space could be the model could not correctly predict the probability of heart disease. The way to resolve this issue is to use several different models and optimize the operation parameters