Heart Disease Binary Classifier

Dhiraj Bankar  
DSC680 - Summer 2021  
<https://github.com/dhirajbankar/DSC680>

# Which Domain?

Healthcare domain- I used to be a part of BCBS healthcare for 2 years and that I determined to travel with a dataset that corresponds to the healthcare domain. Preponderantly, I'm focusing in the datasets wherever I will apply the machine learning algorithms to realize a lot of insight in terms their application in disease prediction.

# Which Data?

Below is the URL address for the heart disease dataset.

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

The dataset has fourteen attributes that are principally the health profile attributes of various patients. As an example, for every of the patients listed within the dataset, the info points embody their heart rates, pressure level within the resting versus active time heart care, whether or not they are smokers or non-smokers etc. Given such data concerning the patients, I'm planning to see if these algorithms are capable of predicting cardiovascular disease in patients.

# Research Questions? Benefits? Why analyze these data?

The analysis question within the current dataset is all concerning characteristic the attributes that are a lot of possible to predict the presence of cardiovascular disease during a given patient. The end result are going to be very useful and includes a heap of sensible applications within the globe. As an example, healthcare suppliers and hospitals, medical care physicians will use such data to predict the disease within the patients and therefore are going to be able to give preventive measures to their patients.

Even though the question within the dataset is dealing concerning predicting the guts disease within the patients- however the top goal of the project is to create a generic model which will be wont to apply and obtain the ends up in just about all alternative disease areas.

# What Method?

I am planning to take a unique approach in terms of selecting the strategy. In alternative words- rather than simply selecting one classification algorithmic rule to predict a disease, I will be able to be employing a combination of various algorithms in predicting the end result. As an example, since this is often a classification drawback, I’m planning to apply completely different classification algorithms like logistic regression, Naïve Bayes algorithm, decision trees, random forests and SVM etc.

The main reason for taking this approach is just to get a global picture of how different algorithms analyze the same dataset and go further to understand on how to handle the results when the outcome from different algorithms is vastly different. Do we need to pick and choose or do we perform a logical elimination of one algorithm versus the other algorithm?

# Potential Issues?

As mentioned on top of, the potential issue may well be the very fact that completely different algorithms could provide altogether different output for a given dataset given their limitation related to the individual machine learning models. this is often one space wherever I would run into problems. Hence, I would like to return up with arrange of action to affect such ambiguity as once it arises. A way to select one best algorithmic rule would be in all probability by gazing every of those algorithms intimately to know the logic related to the classification and will it match well with the dataset in hand.

# Concluding Remarks

Summarizing everything we have a tendency to mentioned on top of – Given my expertise within the biotechnology field, I’m planning to work on the dataset associated with the health care domain.  
As a locality of it I select to go with a dataset wherever the info enclosed the small print concerning the health profile of various patients. As an example, every row corresponds data of one patient that has details like their pulse rate, pressure level, glucose levels, cholesterol levels etc. By using these details concerning the patient, we've got to predict whether or not he or she includes a cardiovascular disease. If a model that's able to predict the guts disease outcome supported this dataset, it'd have a large real-world application wherever folks within the aid domain will use it to use on completely different patients and supported the prediction the patients may be supplied with the mandatory preventive care. coming back to the particular model building facet, I’m planning to preponderantly keep on with the thought of ensemble modeling wherever I’d

 like check and mix completely different algorithmic rule as a goal to induce the most effective model; which will predict the end result with tight accuracy.

**References:**

<https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c>

<https://medium.com/@dskswu/machine-learning-with-a-heart-predicting-heart-disease-b2e9f24fee84>

<https://www.researchgate.net/publication/330981991_Analysis_of_Heart_Disease_Prediction_using_Various_Machine_Learning_Techniques>

<https://www.scirp.org/html/14-1560633_88650.htm>

<https://www.sciencedaily.com/releases/2019/05/190513081412.htm>

<https://ieeexplore.ieee.org/document/8740989>

<https://pdfs.semanticscholar.org/d0a5/d4b8e8da3ee2a6bf8ac5d44196fb0365cf1c.pdf>

<https://www.hindawi.com/journals/misy/2018/3860146/>