Heart Disease Binary Classifier

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

# Any surprises from your domain from these data?

I choose healthcare as a domain. I worked with healthcare for 2 years. The most difficult part I have to remember is the dataset I have chosen has not considered all the parameters. Only 14 attributes are there. I am confident there could be multiple factors which could contribute the heart disease. I found multiple dataset on heart decease related.

# The dataset is what you thought it was?

I would say there is not perfect dataset and there is not 100% guaranty of prediction model which I am working on. I feel there should be some additional attributes would have been there that could have made model much more reliable and give correct results.

1. Stress level for each individual. That too we cant or there is not way or I don’t know how to calculate the stress level of individual.
2. Physically active or not.
3. Daily exercise with hours
4. The most important part for this analysis the type of food each individual is taking.
   1. Proteins
   2. Carbs
   3. Fibers
   4. Vitamins
   5. etc.
5. All vitamin levels and what is impact of each on heart.

# Have you had to adjust your approach or research questions?

For now I don’t think there will be any problems in current approach.

Except As I stated above. 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.

# Is your method working?

Yes as per train and test data the model is performing very well.

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?

# What challenges are you having?

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.