Predicting Coronary Heart Disease

Heart Disease is one of the leading causes of death for men and women in the United States. It can go undetected for years because of its subtle beginning symptoms. However, despite the genetic factors, the Center of Disease Control and Prevention suggests that there are controllable variables that people can implement themselves to reduce their risk of this disease.

The data I will be using to create a model that predicts the likelihood that someone will experience coronary heart disease during their lifetime comes from the 2022 annual CDC survey data of around 400k adults and which includes questions related to health and whether or not they have already been diagnosed with heart disease. Doctor's offices could use this model when patients fill out their own survey data to calculate their risk of heart disease, or patients could use an app/website using this model to do so themselves.

https://www.kaggle.com/datasets/ambujdevsingh/key-indicators-of-heart-disease

The criteria for success would be that the model can accurately (above 90% accuracy) predict the likelihood of someone experiencing heart disease during their lifetime based on the model that is created by the training of the data provided above. The scope of the solution space would be only using this data alone since we don't have any key columns to tie it to any other tables. This is also true about the constraints since we have to assume that there aren't any duplicates in the data when it comes to having the same person's responses recorded more than once. Another constraint is the data doesn't include diet which can be an important variable as well but would be hard to express in a simple number.

The stakeholders would be whoever would want to implement the model into their software in the future. Currently there aren't any stakeholders who are waiting for this model and it is mainly for educational purposes here at Springboard.