# Module 3 – Heart Disease

Andy Peng

Hi my name is Andy Peng. Welcome to my presentation for Flatiron School Module 3 – Heart Disease. For the extent of this project, our stakeholders are a company that focuses on predicting whether a patient has heart disease given patient's medical records.

Overview

- Relationship to heart disease
  - Gender
 - Thallium Stress Result
  - Age
- Modeling

We managed to create a model that can predict heart diseases with great accuracy while also minimizing the chances of classifying a patient as false negative. Before we talk about the results of our model, we will be discussing about a few features and its relationship to heart disease.
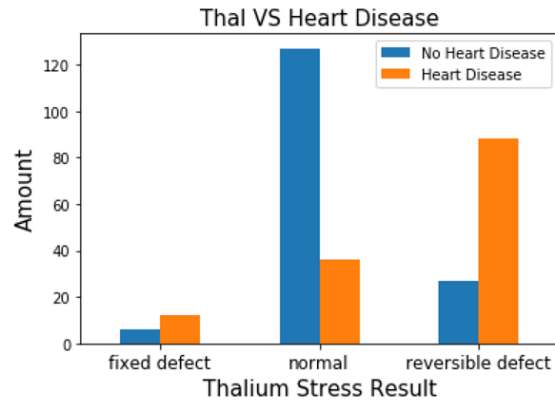
Does Gender affect Heart Diseases?
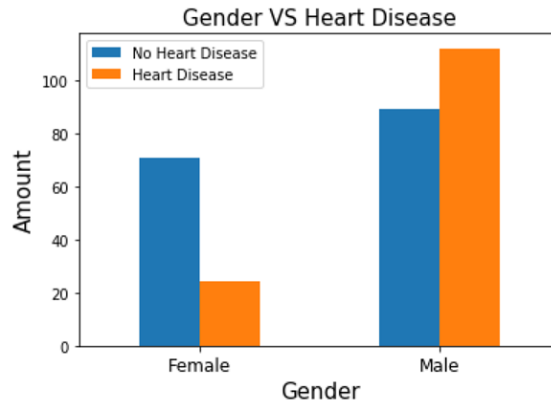
Does gender affect heart disease? Here we have a graph showing the amount of patients with and without heart diseases separated by the sex of the individual. As you can see male patients tend to have more heart diseases than female patients. This led me to question if there is any feature that also appears in individuals with just heart disease.

Chest Pain
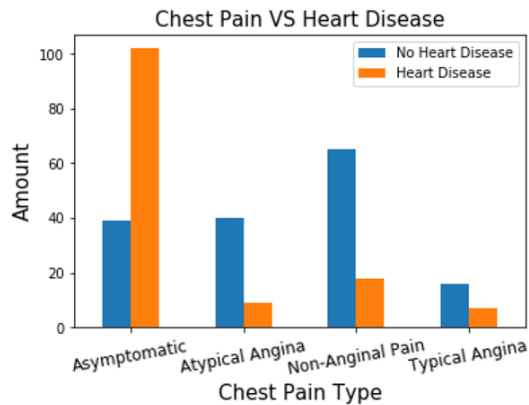
Further investigation shows that chest pain type is a feature that stands out the most. As you can see in this graph, majority of the patients with heart disease have asymptomatic chest pain type. 100 patients that have heart disease have asymptomatic chest pain compare to the second most 18 patients with non-anginal pain.

Thallium Stress Results

Thallium Stress Results. Do they affect heart disease? Patients with Thallium Stress Results as reversible defect tend to have a higher chance to have heart disease compared to patients with normal or fixed defect.

Does age affect Heart Diseases?
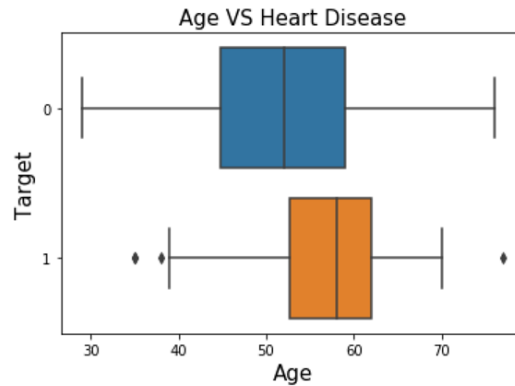
Does age affect Heart Diseases? Based off the boxplot here we can see that individuals with heart disease overall tend to be older than individuals without heart disease. But what other feature are affected when one gets older? Cholesterol, Maximum Heart Rate and Blood Sugar Levels.

Maximum Heart Rate

Maximum Heart Rate VS Heart Disease
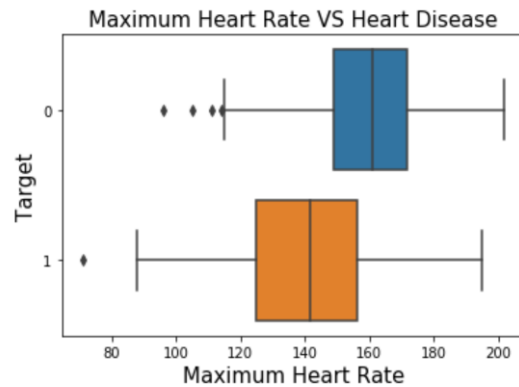
There are many other features that are affected by old age, but we will be focusing on maximum heart rate. As an individual gets older, their heart won't be able to pump at the same maximum heart rate as a 20-year-old. By graphing a box plot of maximum heart rate of patients with and without heart disease we can see that patients with heart disease tend to have a lower maximum heart rate.

Maximum Heart Rate

Age VS Maximum Heart rate
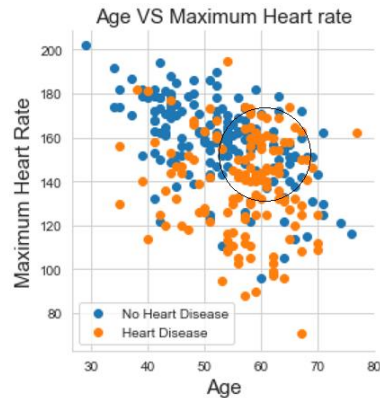
Now let's look at the scatter plot of age versus maximum heart rate. We can see there is a negative correlation between these two features. As you get older you have a lower maximum heart rate. But also notice that majority of patients with heart disease are older and have a lower maximum heart rate.

Modeling

- Logistic Regression
- K Nearest Neighbors
- Decision Tree
- Random Forest
- XGBoost
- GridSearchCV

For modeling we used a combination of Logistic Regression, K Nearest Neighbors, Decision Tree, Random Forest, XGBoost and GridSearchCV to get the best model for our problem.

Results

- AUC – Random Forest with GridSearchCV
- Accuracy/Precision/F1 Score – XGBoost with GridSearchCV
- Recall - XGBoost

And here are the results. For best AUC values we used Random Forest with GridSearchCV. For Accuracy/Precision/F1 Score we used XGBoost with GridSearchCV. Finally for Recall we used XGBoost.

Recommendations

 XGBoost for Recall Score

	precision	recall	f1-score	support
0	0.87	0.87	0.87	31
1	0.86	0.86	0.86	29
accuracy			0.87	60
macro avg	0.87	0.87	0.87	60
weighted avg	0.87	0.87	0.87	60

Since we are working to classify patients on whether they have heart disease or not. Our goal would be to minimize the chances of us telling a patient they don't have heart disease when they indeed do have heart disease. In other words we want to maximize our recall score, which is exactly what we obtained by using XGBoost.

Recommendations

- Gender
- Asymptomatic Chest Pain
- Reversible Defect
- Age & Maximum Heart Rate

Here are some features we should focus on when identifying whether a patient has heart disease or not. Gender – Males tend to have a higher chance at having heart disease than Females. Asymptomatic Chest Pain – When patients have this chest pain type, they tend to have a higher chance at having heart disease than other individuals. Reversible Defect – From the Thallium Stress Results, more patients with this result have heart disease than other results. Age & Maximum Heart Rate – As individuals get older and their maximum heart rate starts to decrease, they will have a higher chance of having heart disease.

Next Steps

Features to Explore

- Genetic Disorders
- Body Weight Percentages
- Patient's Diet

Gather More Data

For our next step we can explore other key features such as whether the patient have genetic disorder, their body weight percentages and their diet. If we want to improve our model, gathering more data would help us do a better job at predicting whether a patient has heart disease or not.

Thank You

Thank You