

Executive Summary of the Project

It is believed from information that one person dies every 36 seconds in USA. Close to 700,000 Americans die every year. This is a huge number considering only USA, but we have patients from all over the world. Identifying a heart disease early would go a long way in preventing the disease and saving more lives in the future. Having said that, it is important for doctors to identify heart disease or symptoms of heart disease based on the parameters of patients to prevent invasive procedures. This would save a lot of money for patients and time for healthcare workers.

According to the data we have, RestBP=130, Age=60, Chol=250, MaxHR=165 are the peaks of parameters. This means that patients having values ranging in these areas have a higher risk of a heart disease.

RestBP and Chol have pretty much similar density for heart disease and no heart disease. MaxHR has higher density for cases with no heart disease whereas Age has higher density for cases with heart disease.

There is a high negative correlation between Age and MaxHR. As the age of the patient increases, MaxHR reduces significantly.

Patients who have higher RestBP and Age are more likely to get a heart disease. Patients who have higher Chol and MaxHR are more likely not to get a heart disease, but this is not always necessarily true.

MaxHR is a more decisive parameter because a maximum heart rate achieved is better attributed compared to the age of an individual. Nowadays, we see people in their early 30s and 40s also suffering from a heart disease. So, MaxHR might be a more crucial factor. Given the Age of the patient, we can predict the MaxHR, which helps us in predicting the heart disease.

Since the linear model is statistically significant, MaxHR is a good predictor variable compared to other variables.

In our Logistic Regression Model, For every one-unit change in RestBP, the log odds of AHD (versus No AHD) increases by 0.016. For every one-unit change in Age, the log odds of AHD (versus No AHD) increases by 0.004. For every one-unit change in Chol, the log odds of AHD (versus No AHD) increases by 0.0003. For every one-unit change in MaxHR, the log odds of AHD (versus No AHD) decreases by 0.0043.

We need to understand that a computerized model does not always ensure higher accuracy. Any false positives or misleading analysis could be dangerous. But, from our analysis, Linear Regression is a better fit to the model compared to Logistic Regression. Taking into account other parameters and building a more complex model would be a better idea in the future.

