

Is it Possible to Predict Heart Disease with Machine Learning?

Which Domain?

I selected heart disease data for my first project because it relates to the healthcare industry that is currently benefiting from data science. I also wanted to focus on heart disease since it is a common disease in my father's family and most Hispanic families. Therefore, in my first project, I hope to use machine learning to help predict the presence of heart disease by common traits.

Which Data?

The selected dataset comes from the Kaggle competition section that used UCI's data of heart disease information. The dataset has fourteen attributes that are based on the profiles of different patients who are dealing with heart disease. This dataset has many helpful data points, including blood pressure, heart rates at other times, and if the patient was a smoker? These data points will help give insights into the factors that can lead to heart disease. Below is a link to the dataset:

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

Research Questions? Benefits? Why analyze these data?

The research question is the same as the title of this proposal. Is it possible to predict heart disease with machine learning? With the help of the dataset's attributes and applying machine learning techniques, we can identify the indication of the presence of heart disease in a patient. Typically doctors and healthcare companies use similar features and factors to predict this type of disease with patients with similar lifestyles and family histories. This project focuses on developing a model that can be used to make early predictions about people at high risk of developing heart disease.

What Method?

I plan to use different approaches to predicting the presence of heart disease. Boiling down this project, I see this being a classification problem in predicting heart disease in men and women of different age groups and lifestyles. With this in mind, I plan on using classification algorithms like random forests, naïve Bayes, decision trees, and logistic regression. Using different approaches will give an insightful image of which one works and the differences in the results. This will give me a better understanding of the results and how to handle them.

Potential Issues?

One obvious issue that comes to mind is how to deal with wildly different results using different approaches that might have many limitations. I will need to develop a plan to deal with such irregularities that might lead me away from the focus of this project. Another issue would be the results that are based on gender. This could lead to misleading factors that might not be helpful to all people. Finally, there is another issue with results based on ethnicity since it could be used negatively.

Concluding Remarks

The selected dataset has many helpful data points with detailed information from different patient profiles. The data points correspond to the attributes of blood pressure, heart rate, and others that have been correctly connected to a related patient. This helpful information will help predict the presence of heart disease in high-risk people. At the moment, humans are the only ones to make that type of prediction. Developing a model that can predict the presence of heart disease in high-risk people, I believe

it would be a big help in the healthcare industry. In addition, this model could be used to prevent people from developing heart disease. The model will have gone through a trail of testing and experimenting using different algorithms that can predict the presence of heart disease in all people and genders.