Is it Possible to Predict Heart Disease with Machine Learning?

# **Michael Gonzalez DSC 680 Fall 2020 https://micgonzalez.github.io/**

# Which Domain?

The reason why I selected a heart disease data for my first project, 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 in most Hispanic families. In my first project, I am hoping to use machine learning to help in predicting the presence of heart disease by common traits.

10 References:

Bhanot, K. Predicting presence of Heart Diseases using Machine Learning. February 12, 2019. <https://towardsdatascience.com/predicting-presence-of-heart-diseases-using-machine-learning-36f00f3edb2c> (Article on using machine learning to predict heart disease)

Green, W. Machine Learning with a Heart: Predicting Heart Disease. July 31, 2018. <https://medium.com/@dskswu/machine-learning-with-a-heart-predicting-heart-disease-b2e9f24fee84> (Article on using machine learning to find the presence of heart disease)

Muthuvel, Marimuthu & Sivaraju, Deivarani & Ramamoorthy, Gayathri. (2019). Analysis of Heart Disease Prediction using Various Machine Learning Techniques. <https://www.researchgate.net/publication/330981991_Analysis_of_Heart_Disease_Prediction_using_Various_Machine_Learning_Techniques> (Article on the analysis of different machine learning techniques on heart disease predictions)

Nashif, S., Raihan, Md.R., Islam, Md.R. and Imam, M.H. (2018) Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System. <https://www.scirp.org/html/14-1560633_88650.htm> (Article on detecting heart disease traits with machine learning)

European Society of Cardiology. (2019, May 13). Machine learning overtakes humans in predicting early death or heart attack: Machine algorithm uses 85 variables to calculate risk in individuals. <https://www.sciencedaily.com/releases/2019/05/190513081412.htm> (Article on the benefits of using machine learning to predict heart disease)

S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019, doi: 10.1109/ACCESS.2019.2923707. <https://ieeexplore.ieee.org/document/8740989> (Article on using hybrid machine learning techniques on predicting heart disease)

Sharma, H. (2017). Prediction of Heart Disease using Machine Learning Algorithms: A Survey. <https://www.semanticscholar.org/paper/Prediction-of-Heart-Disease-using-Machine-Learning-Sharma/d0a5d4b8e8da3ee2a6bf8ac5d44196fb0365cf1c?p2df> (Article on the survey results of the effectiveness on using machine learning for predicting heart disease)

Amin Ul Haq, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, Ruinan Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Mobile Information Systems, vol. 2018, Article ID 3860146, 21 pages, 2018. <https://www.hindawi.com/journals/misy/2018/3860146/> (Article on using a hybrid framework for using machine learning techniques to predict heart disease)

Ronit. Heart Disease UCI Dataset. Kaggle. N.D. <https://www.kaggle.com/ronitf/heart-disease-uci> (Dataset of UCI’s heart disease information)

Apurb Rajdhan , Avi Agarwal , Milan Sai , Dundigalla Ravi, Dr. Poonam Ghuli, 2020, Heart Disease Prediction using Machine Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 04 (April 2020). <https://www.ijert.org/heart-disease-prediction-using-machine-learning> (Article on predicting heart disease with machine learning techniques)

# 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 useful data points that include blood pressure, heart rates at different times and if the patient was a smoker? These data points will help in giving insights on the factors that can lead to developing 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 machine learning techniques used to predict the presence of heart disease in a patient. Typically doctors and healthcare companies use similar attributes and factors to predict this type of disease with patients that have similar lifestyles and family history. The focus of this project is developing a model that can be used to make early prediction of people who are a high risk of developing heart disease.

# What Method?

I am planning 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 am planning 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 of the results. This will lead me to have 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 from 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 to 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. There is another issue on results that are based on ethnicity since it could be used in a negative way.

# Concluding Remarks

The selected dataset has many useful data points that have detailed information from the profiles of different patients. The data points correspond the attributes of blood pressure, heart rate and others have been correctly connected to related patient. Having this useful information will help in predicting the presence of heart disease in high risk people. At the moment, humans are the only ones to make that type of prediction. By 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. This model could be used to prevent people from developing heart disease in the future. The model will have gone through a trail of testing and experimenting of using different algorithms that can predict the presence of heart disease in all people and genders.