Predicting Heart Disease

# **Tyler Anderson Fall https://github.com/TylerYinAnderson**

# Which Domain?

The domain where I got my dataset is the BMC Medical Informatics and Decision Making site. This dataset was used in a research paper called “Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone”

# Which Data?

I will be examining the heart failure dataset provided in the research paper. This dataset is about cardiovascular diseases (CVDs) and the impact of CVDs globally as the number one cause of death. Heart failure is a common ailment caused by CVDs and the dataset has 12 features that can be used to predict death from heart failure. CVDs can also be prevented by looking at behavior and risk factors like tobacco use, obesity, physical inactivity and use of alchohol.  
<https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5>  
<https://plos.figshare.com/articles/Survival_analysis_of_heart_failure_patients_A_case_study/5227684/1>  
<https://github.com/davidechicco/cardiovascular_heart_disease>

# Research Questions? Benefits? Why analyze these data?

I plan to analyze this dataset by predicting patients who will have heart failure using one of the many models we have learned in this program. I have not decided what I will use yet but possible models I can use are Logistic Regression, KNN, Decision Tree, Random Forest and Neural Network. I would like to know how big a role each risk factor like smoking or alcohol use can contribute to the development of heart disease. I would also like to know how large an impact age plays in the development of heart disease.

# What Method?

I plan to use a model that gives the best result. After the creation of my train and test set, I can pass it through any of these possible models: Logistic Regression, KNN, Decision Tree, Random Forest and Neural Network.

# Potential Issues?

One of the most glaring problems is the types of methods I want to use for prediction of heart disease. I can either pick one method but I don’t know if it’s the best method for my data and if I used multiple methods and pick the model that gives the best performance it could take me much longer to do taking time away from the quality of the analysis. Another issue is the data feature for sex is binary and does not specify specific sex just 1 or 0 so I have to use assumption to pick the sex.

# Concluding Remarks

Overall I think this is an interesting dataset and People with cardiovascular disease or who are at high cardiovascular risk, due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidemia or already established disease need early detection and management, a machine learning model can be of help for early detection.

# References

Galarnyk, M. (2019, Jul 31). *Understanding Decision Trees for Classification (Python).* Retrieved from towardsdatascience: https://towardsdatascience.com/understanding-decision-trees-for-classification-python-9663d683c952#:~:text=Understanding%20Decision%20Trees%20for%20Classification%20%28Python%29%201%20Classification,that%20they%20are%20relatively%20easy%20to%20interpret.%

Jaroli, H. (2019, April 8). *K-Nearest Neighbors (KNN) with Python.* Retrieved from datascienceplus: https://datascienceplus.com/k-nearest-neighbors-knn-with-python/#:~:text=%20K-Nearest%20Neighbors%20%28KNN%29%20with%20Python%20%201,whether%20someone%20will%20TARGET%20CLASS%20or...%20More%20

Li, S. (2017, Sep 28). *Building A Logistic Regression in Python, Step by Step.* Retrieved from towardsdatascience: https://towardsdatascience.com/building-a-logistic-regression-in-python-step-by-step-becd4d56c9c8#:~:text=Building%20A%20Logistic%20Regression%20in%20Python%2C%20Step%20by,%28yes%2C%20success%2C%20etc.%29%20or%200%20%28no%2C%20failure%2C%20etc.%29.

Chicco, D., Jurman, G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med Inform Decis Mak 20, 16 (2020). <https://doi.org/10.1186/s12911-020-1023-5>

Ahmad T, Munir A, Bhatti SH, Aftab M, Ali Raza M. Survival analysis of heart failure patients: a case study. Dataset. [https://plos.figshare.com/articles/Survival\_analysis\_of\_heart\_failure\_patients\_A\_case\_study/5227684/1. Accessed 25 Jan 2019](https://plos.figshare.com/articles/Survival_analysis_of_heart_failure_patients_A_case_study/5227684/1.%20Accessed%2025%20Jan%202019).

National Heart Lung and Blood Institute (NHLBI). Heart failure. <https://www.nhlbi.nih.gov/health-topics/heart-failure>.

The Guardian. UK heart disease fatalities on the rise for first time in 50 years. <https://www.theguardian.com/society/2019/may/13/heart-circulatory-disease-fatalities-on-rise-in-uk>.

World Health Organization, World Heart Day. <https://www.who.int/cardiovascular_diseases/world-heart-day/en/>.

<https://github.com/davidechicco/cardiovascular_heart_disease>