**DG Cardiovascular Health Project**

**Overview**

The goal of this project is to utilize multiple datasets on heart disease and failure to learn important techniques in data science, and educate students on how to provide clear, presentable data. We will be taking a look at how various factors can contribute to this, as well as exploring the relationships between characteristics and traits of individuals in our dataset. You will ultimately create a final deliverable that summarizes the work that you've done, and interesting conclusions you have found over the course of a few weeks.

**Technology Stack**

* NumPy
* Pandas
* Matplotlib
* Seaborn
* Scikit-learn

(Tentative: If time allows)

* Keras
* Tensorflow

**Goals**

**#1**

* Analyze data through the use of common exploratory data analysis techniques, as well as modelling methods in order to get a better understanding of how to create meaningful visualizations that help others extrapolate information from datasets.

**#2**

* Utilize regression tools to predict heart disease or failure based on the provided data.

**#3**

* Create a final deliverable that summarizes the work done and the information found over the course of a few weeks of work.

(If time allows)

**#4**

* Implement a convolutional neural network to analyze x-ray images and create an image classifier to determine certain diseases.

**Spring Semester Timeline (Tentative)**

| **Week** | **Plan** |
| --- | --- |
| Week 1 | * Introduction |
| Week 2 | * EDA + Visualization I |
| Week 3 | * EDA + Visualization II |
| Week 4 | * Modeling + Regression |
| BREAK |  |
| Week 5 (3/28) | * Random Forests + Gradient Boosting |
| Week 6 (4/4) | * PCA + SVM |
| Week 7: (4/11) | * **TBA** |
| Week 8: (4/18) | * YOUR FOCUS GOES HERE |
| Week 9: (4/25) | * Final Deliverable |

**Resources**

* [Heart Disease UCI](https://www.kaggle.com/ronitf/heart-disease-uci)
* [Heart Disease Dataset](https://www.kaggle.com/johnsmith88/heart-disease-dataset) (seems like the same dataset)
* [UCI Machine Learning Repository: Heart Disease Data Set](https://archive.ics.uci.edu/ml/datasets/heart+disease) (original source)
* [Classification models for heart disease prediction using feature selection and PCA](https://www.sciencedirect.com/science/article/pii/S2352914820300125) (Research paper that uses the same dataset with detailed descriptions)
* [Heart Failure Prediction](https://www.kaggle.com/andrewmvd/heart-failure-clinical-data)
* [Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone | BMC Medical Informatics and Decision Making](https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5) (paper from which the dataset comes from)

Both of these datasets will require a bit of prior background knowledge on medical data which will be introduced and discussed in the first few weeks. Note that this project is still primarily focused on the data science aspect and techniques in machine learning rather than the medical side of our data, and it will be up to individuals in terms of how deep you want to learn on your own regarding the surrounding medical knowledge that encompasses these datasets.