**Predicting heart disease using machine learning**

* This notebook looks into using various Python-based machine learning and data science libraries in an attempt to build a machine learning model capable of predicting whether or not someone has heart disease based on their medical attributes.

1. **Problem Definition In a statement** –

* Given clinical parameters about a patient, can we predict whether or not they have heart disease?

1. **Data The original data** came from the

* Cleveland data from the UCI Machine Learning Repository …

URL - <https://archive.ics.uci.edu/ml/datasets/heart+Disease>.

* There is also a version of it available on Kaggle.

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

1. **Evaluation**

* If we can reach 95% accuracy at predicting whether or not a patient has heart disease during the proof of concept, we'll pursue the project.

1. **Features**

* This is where you'll get different information about each of the features in your data. You can do this via doing your own research (such as looking at the links above) or by talking to a subject matter expert (someone who knows about the dataset).

**Create data dictionary**

* age -- age in years
* sex -(1 = male; 0 = female)
* chest pain type (4 values)
* resting blood pressure
* serum cholestoral in mg/dl
* fasting blood sugar > 120 mg/dl
* resting electrocardiographic results (values 0,1,2)
* thalach - maximum heart rate achieved
* exercise induced angina
* oldpeak = ST depression induced by exercise relative to rest
* the slope of the peak exercise ST segment
* number of major vessels (0-3) colored by flourosopy
* thal: 3 = normal; 6 = fixed defect; 7 = reversable defect