**Final Project**

**Member:** Aparna Srikanti and Helen Wu

1. DataSet: Breast Cancer Related Data. We get the dataset from <http://docs.w3cub.com/scikit_learn/modules/generated/sklearn.datasets.load_breast_cancer/>

The copy of UCI ML Breast Cancer Wisconsin (Diagnostic) dataset is:

<https://goo.gl/U2Uwz2>

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)>

1. The field in the datasets are:

1) ID number

2) Diagnosis (M = malignant, B = benign)

3-32)

Real-valued features are computed for each cell nucleus:

a) radius (mean of distances from center to points on the perimeter)

b) texture (standard deviation of gray-scale values)

c) perimeter

d) area

e) smoothness (local variation in radius lengths)

f) compactness (perimeter^2 / area - 1.0)

g) concavity (severity of concave portions of the contour)

h) concave points (number of concave portions of the contour)

i) symmetry

j) fractal dimension ("coastline approximation" - 1)

1. Project:
   1. Perform a Deep Dive of Existing Data Using Machine Learning
   2. Create an analysis of existing data to make a prediction, classification, or regression
2. Will be using Python Pandas with Machine Learning Libraries and Matplotlib (Web if time permits)
   1. Machine Learning Tree Algorithm