UPenn Data Analysis and Visualization - Final Project Proposal

1 Project Title

Breast Cancer Detector

---

2 Team Members

Adesola Fakiyesi
Kimberly Kockenmeister
Kendall Sisk

---

3 Project Description/Outline

The team plans to make a 'Breast Cancer Detector' which would help in the early diagnosis of breast cancer. The objective of this project is to classify whether the user's tumor is benign or malignant. We would be using Machine Learning Classification Methods which would help us fit a function that would predict the discrete diagnosis benign or malignant for every new input.

---

4 Dataset to be Used

Resources:

1. Breast Cancer Wisconsin (Diagnostic) Data Set

    a. https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29

        i. 569   Rows
        ii. 32 Columns

---

6 Rough Breakdown of Tasks

Every team member will work to complete the following:

- We will start our project by choosing a data set from Kaggle called the Breast Cancer Wisconsin (Diagnostic) Data Set. In a study for the data set, ten (10) features were computed from a digitized image of a fine needle aspirate of a breast mass. These features describe the characteristics of the cell nuclei present in the image.

- We will focus on these ten features with mean classification to create an Application Detector that predicts if a tumor is benign or malignant. To do this, we will use Machine Learning Classification Methods which will help us fit a function, helping us predict the discrete diagnosis benign or malignant for every new input. The specific models we will train and test will be the Decision Tree Algorithm, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and the Random Forest Classification (RF).

- We will use the best results for our dataset to aid us in building our application. To connect our model to our API, we will use the pickle operation to serialize our machine learning algorithm and save the serialized format to a file. We will then load this file to de-serialize our model.

- From there, we will create 3 routes: a home route that would render to our index.html, a prediction route that would pull in the inputs from the user and make a prediction based off our random forest model, and a results route that would send results back to the user telling them if the tumor was benign or malignant.