Project 3

Requirements:

- Create a GitHub account
- Revisit the codes on classification provided in LMS

Dataset:

The breast cancer dataset is a classic and very easy binary classification dataset. This dataset will be used for this project.

Classes	2
Samples per class	212(M),357(B)
Samples total	569
Dimensionality	30
Features	real, positive

Import the dataset using the following lines of code:

```
>>> from sklearn.datasets import load_breast_cancer
>>> df = pd.DataFrame(load_breast_cancer()['data'],
                      columns=load_breast_cancer()['feature_names'])
>>> df['y'] = load_breast_cancer()['target']
```

Project description:

This project is set to test the knowledge about classifier models that we learnt in week 3.

1. Download the dataset from the above lines of codes
2. Do the necessary data pre-processing if required
3. Implement the following classification models (You can use default hyper-parameters for this)
    a. Decision Tree
    b. Random Forest
    c. Gradient Boosting Method
    d. K-Nearest Neighbour
4. Use the following parameter for tuning the hyper-parameters of the Decision Tree model
    a. >>> {'criterion': ["gini", "entropy"], 'max_depth': [10, 20, 50, 100], 'min_samples_leaf': [10, 20, 50]}
5. Compare the results of various models

Submission Process:

- Upload the Project 3 in GitHub
- Provide the repository link as submission