Pattern Recognition project proposal

-Prathamesh Patil (025910428)

1. **Comparison of performance of different models for breast cancer prediction**
    a. The early diagnosis of Breast cancer can improve the chances of survival significantly. If we can compare some machine learning models for their performance and get the most accurate model, then this model's accurate classification can prevent patients undergoing unnecessary treatments.
    b. Three different classification models will compete for maximum performance
        i. Nearest neighbor
        ii. Decision tree
        iii. Random forest
    c. Data preprocessing:
        i. Using PCA, we can reduce dimensionality of the data set
        ii. Handling missing data points
        iii. Encode categorical data
        iv. Feature scaling
    d. Applying Machine learning classification models
2. **Dataset** : From UC Irvine Machine Learning repository: **Breast Cancer Wisconsin (Diagnostic) Dataset** : http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29
3. **Methods Used for classification**:
    a. Nearest neighbour
    b. Decision tree
    c. Random forest