**Breast Cancer Detection**

The Breast Cancer in women can be classified as one of the most dangerous disease.Even after the surgery, most often women not able to survive.So it is important to diagnose the disease.

This dataset has been taken from kaggle.

**Objective: -** Given the reports of the patient we should be able to classify whether the patient is having the cancer or not.

**EDA and Preprocessing :** We have 31 columns all are numerical and not containing any big scale values.Since the features are less and correlation matrix also shows there is not much relation between the features.Hence I am directly using raw data for modelling.

**Modelling :**

Since the problem statement is a classification problem, I tried several classification algorithms.

1. Logistic Regression :

Confusion marix = [[67 0]

[47 0]]

Training Score: 63.73626373626373

If you observe for the logistic regression, for the any input, model predicting as class 0.

1. Decision Tree Classifier :

Confusion marix = [[63 4]

[ 2 45]]

Training Score: 99.34065934065934

From the above confusion matrix, the decision tree classifier is working exceptionally well.

1. Random Forest Classifier :

Confusion marix = [[64 3]

[ 1 46]]

Training Score: 100.0

Random Forest model is working even better compared to Decision Tree Classifier

1. KNeighborsClassifier :

Confusion matrix :[[63 4]

[30 17]]

Training Score: 80.21978021978022

1. SVC :

Confusion matrix : [[67 0]

[47 0]]

Training Score: 63.73626373626373

1. AdaBoostClassifier :

Confusion matrix : [[67 0]

[ 2 45]]

Training Score: 100.0

1. Gradient Boosting Classifier:

Confusion Matrix = [[64 3]

[ 2 45]]

Training Score: 100.0

1. XGBClassifier :

Confusion Matrix = [[65 2]

[ 2 45]]

Training Score: 99.34065934065934

1. Naive Bayes :

Confusion Matrix = [[66 1]

[45 2]]

Training Score: 99.34065934065934

**Accuracies :**

Find the accuracies for each Model

\*\*So now we conclude the accuracy of different models:\*\*

\*\*1. AdaBoost Classifier = 98.24 %\*\*

\*\*2. XGB Classifier= 96.49 %\*\*

\*\*3. Random Forest Classifier =96.49 %\*\*

\*\*4. Gradient Boosting Classifier= 95.61%\*\*

\*\*5. Decision Tree Classifier= 94.74 %\*\*

\*\*6. K Neighbours Classifier= 70.18 %\*\*

\*\*7. SVC = 63.73 %\*\*

\*\*8. Naiye Bayes= 59.64 %\*\*

\*\*9. Logistic Regression = 58.77%\*\*

**Conclusion :**

Out of all models ADA Boost has given the best Accuracy.