

PROJECT SUMMARY :

In this project a, performance comparison between different machine learning algorithms: *Linear Regression* ,*Decision tree* and *SVM*, on the **Wisconsin Breast Cancer datasets** is done. We have to predict the Stage of Breast Cancer - M (Malignant) and B (Benign). The main objective is to assess the correctness in classifying data with respect to accuracy and efficiency of each algorithm.

JUPYTER NOTEBOOK : Breastcancer.ipynb

DATASET DESCRIPTION :

Breast Cancer Prediction using Machine Learning

We have to predict the Stage of Breast Cancer - M (Malignant) and B (Benign)

ATTRIBUTE INFORMATION :

- 1) ID number
- 2) Diagnosis (M = malignant, B = benign)
- 3-32.Ten 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 ($\text{perimeter}^2 / \text{area} - 1.0$)
 - 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)

Attributes (3-32) are divided into three parts each containing ten features:

Mean (3-13),

Stranded Error(13-23)

Worst(23-32)

RESULTS :

ALGORITHM	ACCURACY	RUNNING TIME OF ALGORITHM
LOGISTIC REGRESSION	89.2	0.012010
DECISION TREE	100.0	0.010737
SUPPORT-VECTOR MACHINE	93.72	0.031914

CONCLUSION :

From the above results we get that Decision tree is given the correct prediction with highest accuracy. So it is the best algorithm for the given dataset.