PART III -

CLASSIFICATION AND CLUSTERING

Section I: Classification

Link to Dataset :

Breast Cancer Wisconsin (Original) Data Set

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original)>

Two classification algorithms used:

* Decision Tree
* Random Forest

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| S.No | Number of Instances | Total  number  of  attribute  s | Number of  attributes  used  (in case of  dimensionalit  y reduction) | Classification  Method  e.g. Logistic  Regression | Parameters | Train/Test  Ratio | Test  Accuracy/  Error |
| 1 | 699 | 10 |  | Decision Tree | numClasses =2  Impurity = Gini  maxDepth = 5  maxBins = 32 | 80/20 | 0.9428 |
| 2 | 699 | 10 | 5 | Decision Tree | numClasses =2  Impurity = Gini  maxDepth = 5  maxBins = 32 | 80/20 | 0.9391 |
| 3 | 699 | 10 |  | Random Forest | Impurity = Gini  maxDepth = 3  treeCount = 20  featureSubsetStrategy = ‘auto’ | 80/20 | 0.9841 |
| 4 | 699 | 10 | 5 | Random Forest | Impurity = Gini  maxDepth = 3  treeCount = 20  featureSubsetStrategy = ‘auto’ | 80/20 | 0.9760 |

|  |  |
| --- | --- |
| S.No | Other Metrics |
| 1 | Recall : 0.9428, F-Measure:0.9428, FPR(0): 0.1086, FPR(1): 0.0319, TPR(0): 0.9680, TPR(1): 0.8913  Confusion Matrix:  91 3  5 41 |
| 2 | Recall : 0.9391, F-Measure:0.939, FPR(0): 0.06 , FPR(1): 0.0612, TPR(0): 0.9387, TPR(1): 0.94  Confusion Matrix:  92 6  3 47 |
| 3 | Recall : 0.9841 F-Measure:0.9841 FPR(0): 0.0476, FPR(1): 0, TPR(0): 1, TPR(1): 0.9523  Confusion Matrix:  84 0  2 40 |
| 4 | Recall : 0.9760 F-Measure:0.9760 FPR(0): 0.0655, FPR(1):0, TPR(0): 1, TPR(1): 0.9344  Confusion Matrix:  106 0  4 57 |