**K22-8729 Sajad Ali BAI-5C**

**LAB04\_Assignment Task:**

UCI ML Repository contains many datasets for classification. You need to find 2 datasets

with at least 10 attributes

<https://archive.ics.uci.edu/ml/datasets.php>

Complete the following tables and calculate Training and Testing Accuracy using the accuracy Score

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Dataset1**  **Training Accuracy** | **Dataset1**  **Testing Accuracy** | **Dataset2**  **Training Accuracy** | **Dataset2**  **Testing Accuracy** |
| **DT using gini**  **(without pruning)** | **1.0** | **1.0** | **1.0** | **0.83** |
| **DT using gini**  **(with pruning)** | **0.826** | **0.823** | **0.985** | **0.916** |
| **DT using entropy**  **(without pruning)** | **1.0** | **1.0** | **1.0** | **0.97** |
| **DT using entropy**  **(with pruning)** | **0.84** | **0.835** | **1.0** | **0.97** |

* Hint: Check ccp\_alpha parameter for pruning. Use ccp\_alpha = 0.015 for pruning
* Combine the Actual VS. Predicted output of Testing Data and find out the total number of correct classifications and wrong predictions of DT using **Entropy with Pruning on Dataset2.**

|  |  |  |
| --- | --- | --- |
| **Correct Classifications** | **Wrong Predictions** | **Total** |
| **35** | **01** | **36** |

* Display its results as (Correctly Predict: 20 Out 30)
* Plot a Graph for training and testing accuracy for Data1 and Data2 for all 4 methods (Total 8 Graphs) that will show your model is overfit, underfit or Generalized.