Experiment No: 4 **Date:**

Lab Session 4: Decision Tree Classifier

Aim: Implement Decision Tree classifier on the following datasets:
a. Iris Dataset
b. Titanic Dataset
c. Placement Dataset

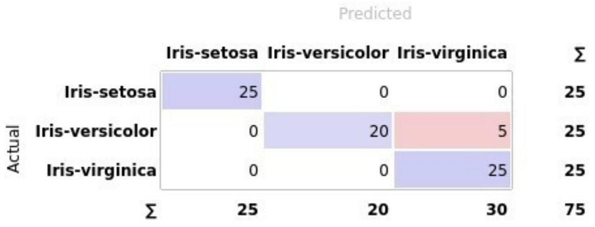
Problem Description: Implement the Decision Tree Classifier andnote the following observations for each of the datasets:
a. The need for Data Sampler
b. The impact of the splitting ratio on the performance
Explore the Predictions widget to understand how we can classify unseen instances.

Widgets Used:
Data Sampler (Refer Error: Reference source not found)

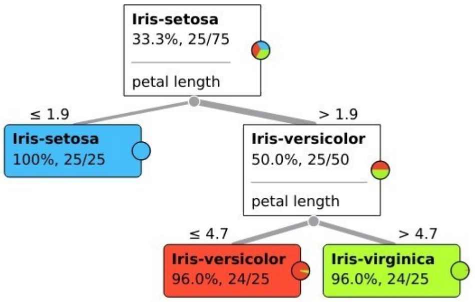
- Data Table
- Data Sampler
- Test and Score
- Tree
- Tree Viewer
- Confusion Matrix
- CSV File Import

Output:

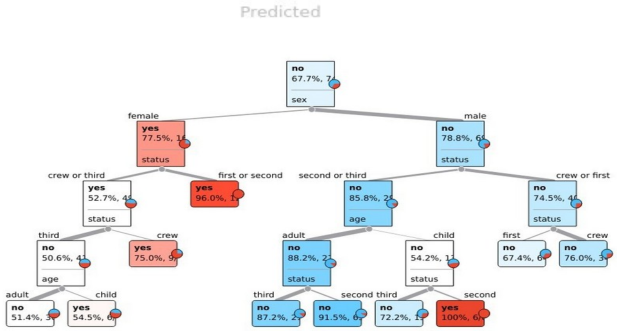
- Iris Dataset 50-50 split



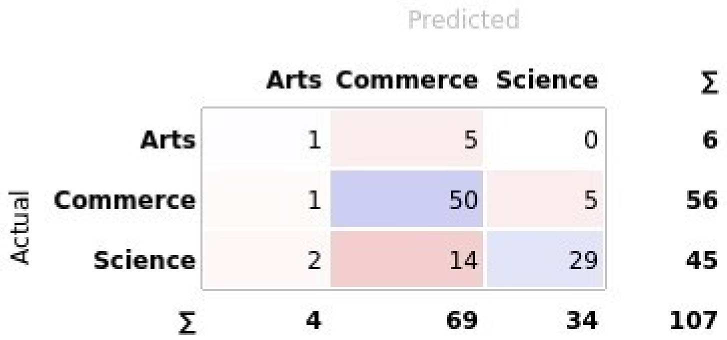
		Predicted			
		Iris-setosa	Iris-versicolor	Iris-virginica	Σ
Actual	Iris-setosa	25	0	0	25
	Iris-versicolor	0	20	5	25
	Iris-virginica	0	0	25	25
	Σ	25	20	30	75



- Titanic Dataset 50-50 split

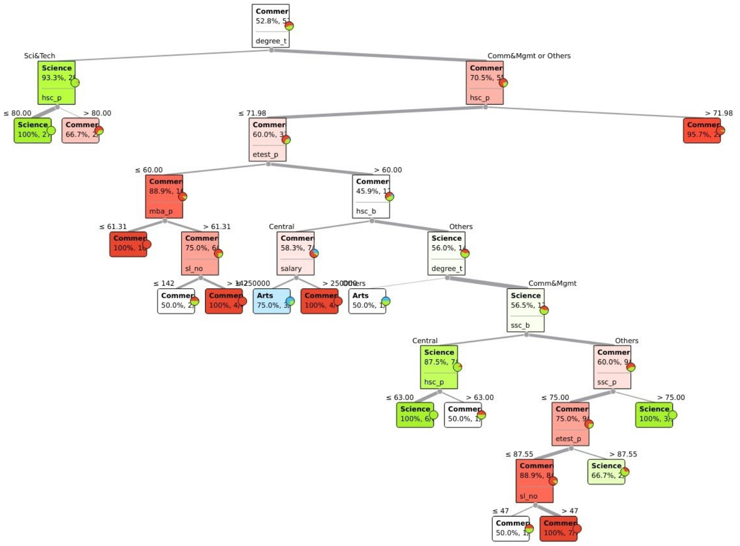


- Placement Dataset 50-50 split



Predicted

Actual	Arts	Commerce	Science	∑
Arts	1	5	0	6
Commerce	1	50	5	56
Science	2	14	29	45
∑	4	69	34	107



Observations and Conclusion:

- The impact of split ratio on performance
    - Iris Dataset

Split Ratio	Accuracy	Precision	Recall	F score
50-50	0.933	0.944	0.933	0.933
60-40	0.967	0.967	0.967	0.967
70-30	0.956	0.956	0.956	0.956
90-10	0.933	0.944	0.933	0.933

- Titanic Dataset

Split Ratio	Accuracy	Precision	Recall	F score
50-50	0.781	0.796	0781	0.753
60-40	0.781	0.789	0.781	0.751
70-30	0.795	0.828	0.795	0.766
90-10	0.795	0.808	0.795	0.733

- Placement Dataset

Split Ratio	Accuracy	Precision	Recall	F score
50-50	0.748	0.752	0.748	0.739
60-40	0.709	0.687	0.709	0.687
70-30	0.719	0.685	0.719	0.701
90-10	0.714	0.752	0.714	0.732

Conclusion:

1. The Decision tree performed best on the Iris dataset
2. The Decision tree performed worst on the placement dataset
3. The proportion of false positives was highest in the titanic dataset
4. The proportion of false positives was lowest in the iris dataset
5. The performance of the classifier appears to go up as the sample size increases