GOA COLLEGE OF ENGINEERING

"Bhausaheb Bandodkar Technical Education Complex"

Experiment No: 5 **Date:**

Lab Session 5: Comparison of Rule Based Classifier, Decision Tree and KNN Classifier

Aim: Compare the performance of Rule Based Classifier, Decision Tree and KNN Classifier on the following datasets:

- Iris Dataset
- Titanic Dataset
- Placement Dataset

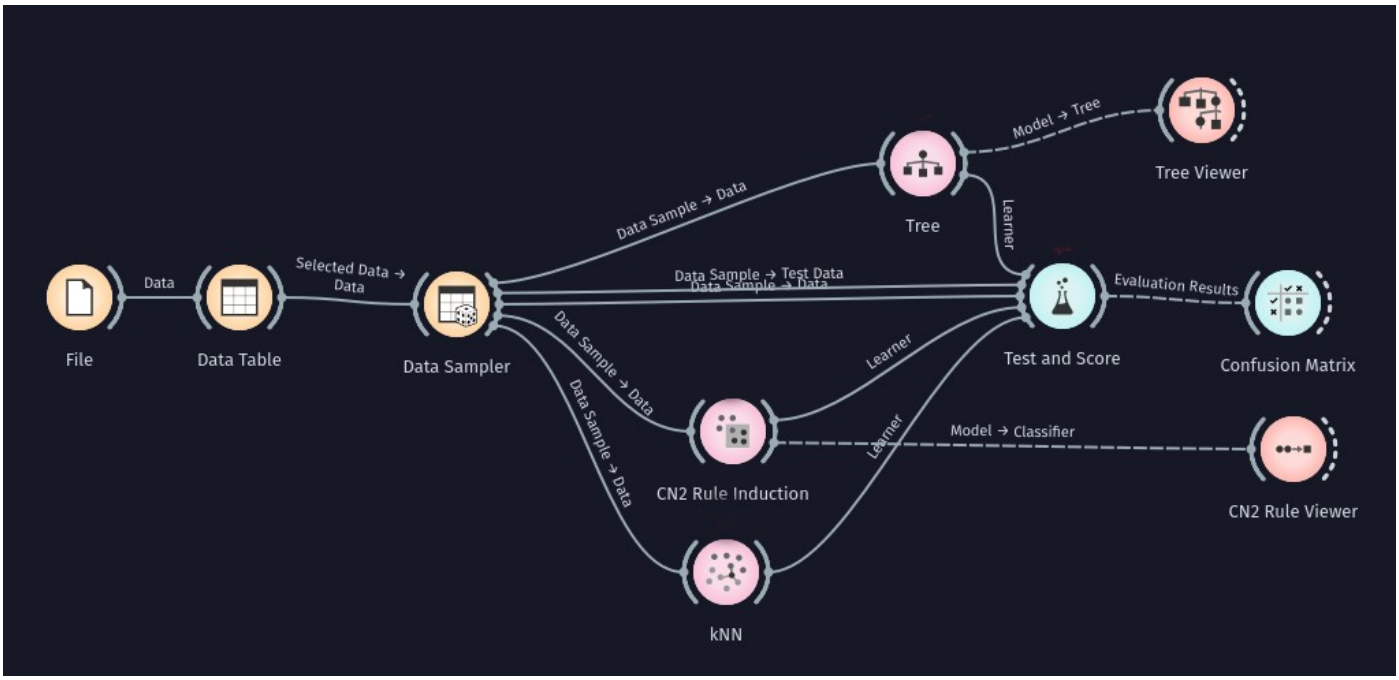
Problem Description: Compare the performance of Rule Based Classifier, Decision Tree and KNN Classifier on the following datasets:

- Iris Dataset
- Titanic Dataset
- Placement Dataset

Widgets Used:

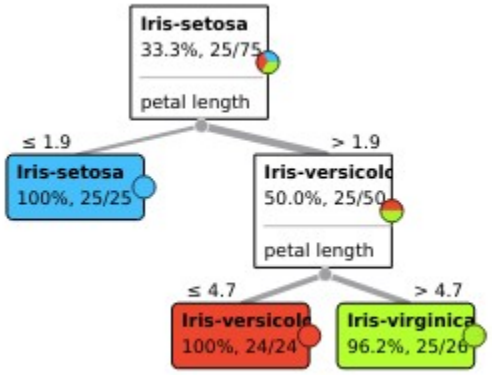
- Data Sampler (Refer Error: Reference source not found)
- Test and Score
- Tree
- Tree Viewer
- CN2 Rules
- KNN
- Confusion Matrix
- CSV File Import

Data Workflow:



Output:

- Iris Dataset



	IF conditions		THEN class	Distribution	Probabilities [%]	Quality	Length
0	petal width≥1.6	→	iris=Iris-virginica	[0, 0, 23]	4 : 4 : 92	-0.00	1
1	petal length≥5.0	→	iris=Iris-virginica	[0, 0, 2]	20 : 20 : 60	-0.00	1
2	petal length≥3.3	→	iris=Iris-versicolor	[0, 25, 0]	4 : 93 : 4	-0.00	1
3	sepal length≤4.5	→	iris=Iris-setosa	[2, 0, 0]	60 : 20 : 20	-0.00	1
4	sepal length≤4.7	→	iris=Iris-setosa	[3, 0, 0]	67 : 17 : 17	-0.00	1
5	sepal length≤4.9	→	iris=Iris-setosa	[4, 0, 0]	71 : 14 : 14	-0.00	1
6	sepal length≤5.1	→	iris=Iris-setosa	[10, 0, 0]	85 : 8 : 8	-0.00	1
7	sepal length≤5.3	→	iris=Iris-setosa	[2, 0, 0]	60 : 20 : 20	-0.00	1
8	sepal length≤5.7	→	iris=Iris-setosa	[3, 0, 0]	67 : 17 : 17	-0.00	1
9	sepal length≤5.8	→	iris=Iris-setosa	[1, 0, 0]	50 : 25 : 25	-0.00	1
10	TRUE	→	iris=Iris-virginica	[25, 25, 25]	33 : 33 : 33	-1.585	0

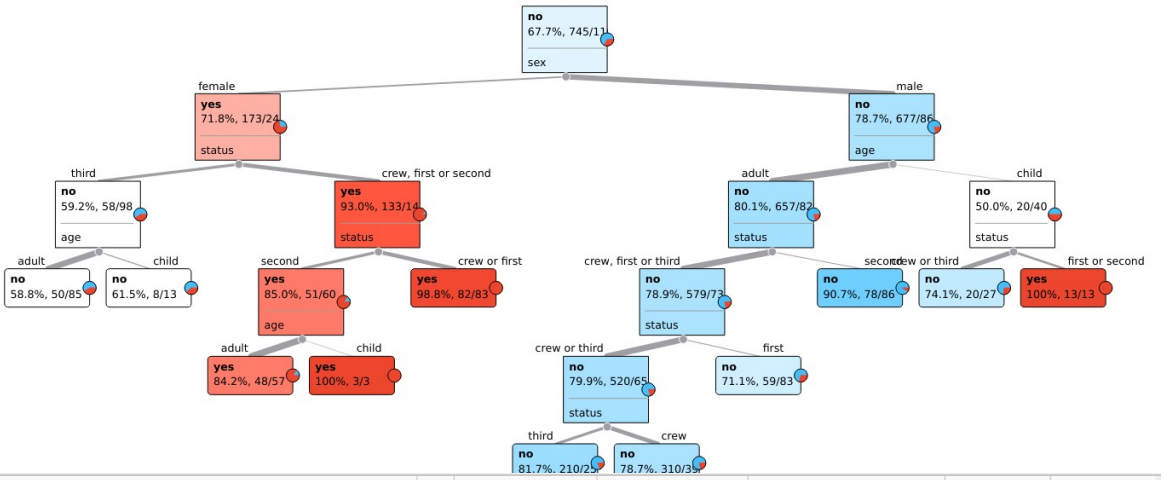
Predicted

Actual	Iris-setosa	Iris-versicolor	Iris-virginica	Σ
Iris-setosa	25	0	0	25
Iris-versicolor	0	20	5	25
Iris-virginica	0	1	24	25
Σ	25	21	29	75

GOA COLLEGE OF ENGINEERING

"Bhausaheb Bandodkar Technical Education Complex"

- Titanic Dataset



	IF conditions		THEN class	Distribution	Probabilities [%]	Quality	Length
0	status=crew AND sex≠male	→	survived=yes	[0, 14]	6 : 94	-0.00	2
1	sex≠female AND status=second AND age≠adult	→	survived=yes	[0, 9]	9 : 91	-0.00	3
2	sex≠female AND status=second	→	survived=no	[78, 8]	90 : 10	-0.446	2
3	status=second AND age≠adult	→	survived=yes	[0, 3]	20 : 80	-0.00	2
4	status=second	→	survived=yes	[9, 48]	17 : 83	-0.629	1
5	sex≠female AND status=third AND age≠child	→	survived=no	[210, 47]	81 : 19	-0.686	3
6	status≠third AND age≠adult	→	survived=yes	[0, 5]	14 : 86	-0.00	2
7	status=crew	→	survived=no	[310, 84]	79 : 21	-0.748	1
8	sex=female AND status≠third	→	survived=yes	[1, 67]	3 : 97	-0.111	2
9	sex≠female AND status≠first	→	survived=no	[20, 7]	72 : 28	-0.826	2
10	status=first	→	survived=no	[59, 24]	71 : 29	-0.868	1
11	age≠adult	→	survived=no	[8, 5]	60 : 40	-0.961	1
12	status=third	→	survived=no	[50, 35]	59 : 41	-0.977	1
			d=no	[745, 356]	68 : 32	-0.908	0

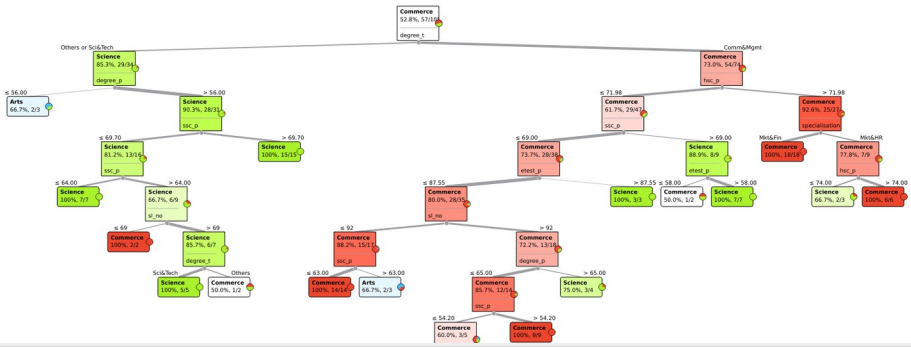
Predicted

Actual	no	yes	Σ
no	735	10	745
yes	231	124	355
Σ	966	134	1100

GOA COLLEGE OF ENGINEERING

"Bhausaheb Bandodkar Technical Education Complex"

- Placement Dataset



	IF conditions		THEN class	Distribution	Probabilities [%]	Quality	Length
0	etest_p≥95.5	→	hsc_s=Science	[0, 0, 3]	17 : 17 : 67	-0.00	1
1	degree_t=Sci&Tech AND sl_no≥78.0	→	hsc_s=Science	[0, 0, 21]	4 : 4 : 92	-0.00	2
2	hsc_p≥74.66	→	hsc_s=Commerce	[0, 18, 0]	5 : 90 : 5	-0.00	1
3	mba_p≤55.01 AND gender=F	→	hsc_s=Commerce	[0, 1, 0]	25 : 50 : 25	-0.00	2
4	degree_t=Sci&Tech AND gender≠F	→	hsc_s=Science	[0, 0, 2]	20 : 20 : 60	-0.00	2
5	mba_p≤55.01 AND ssc_b≠Central	→	hsc_s=Commerce	[0, 2, 0]	20 : 60 : 20	-0.00	2
6	ssc_p≥70.89 AND ssc_p≥83.0	→	hsc_s=Science	[0, 0, 4]	14 : 14 : 71	-0.00	2
7	degree_t=Sci&Tech	→	hsc_s=Commerce	[0, 1, 0]	25 : 50 : 25	-0.00	1
8	etest_p≥89.0 AND sl_no≥66.0	→	hsc_s=Science	[0, 0, 4]	14 : 14 : 71	-0.00	2
9	etest_p≥82.66	→	hsc_s=Commerce	[0, 5, 0]	12 : 75 : 12	-0.00	1
10	mba_p≤55.01 AND hsc_b=Central	→	hsc_s=Commerce	[0, 4, 0]	14 : 71 : 14	-0.00	2
11	etest_p≤53.7 AND sl_no≥165.0	→	hsc_s=Commerce	[0, 2, 0]	20 : 60 : 20	-0.00	2
12	ssc_p≥67.0 AND ssc_p≥80.0	→	hsc_s=Commerce	[0, 1, 0]	25 : 50 : 25	-0.00	2
13	ssc_p≥67.0 AND degree_t≠Comm&Mgmt	→	hsc_s=Science	[0, 0, 1]	25 : 25 : 50	-0.00	2
14	ssc_p≥70.89 AND gender≠F	→	hsc_s=Science	[0, 0, 1]	25 : 25 : 50	-0.00	2
15	ssc_p≥70.89 AND ssc_b=Central	→	hsc_s=Science	[0, 0, 3]	17 : 17 : 67	-0.00	2
16	workex≠No AND gender≠F	→	hsc_s=Commerce	[0, 3, 0]	17 : 67 : 17	-0.00	2
17	workex≠No AND ssc_b=Central	→	hsc_s=Arts	[1, 0, 0]	50 : 25 : 25	-0.00	2
18	workex≠No	→	hsc_s=Commerce	[0, 4, 0]	14 : 71 : 14	-0.00	1
19	ssc_p≥70.89	→	hsc_s=Science	[0, 0, 2]	20 : 20 : 60	-0.00	1
20	sl_no≤18.0 AND gender=F	→	hsc_s=Commerce	[0, 1, 0]	25 : 50 : 25	-0.00	2
21	degree_p≥73.0	→	hsc_s=Commerce	[0, 2, 0]	20 : 60 : 20	-0.00	1
22	mba_p≥64.27	→	hsc_s=Arts	[1, 0, 0]	50 : 25 : 25	-0.00	1
23	salary≤230000.0 AND sl_no≥99.0	→	hsc_s=Commerce	[0, 4, 0]	14 : 71 : 14	-0.00	2
24	degree_p≥69.0	→	hsc_s=Science	[0, 0, 2]	20 : 20 : 60	-0.00	1
25	mba_p≥59.47 AND degree_t≠Comm&Mgmt	→	hsc_s=Science	[0, 0, 1]	25 : 25 : 50	-0.00	2
26	degree_t≠Comm&Mgmt	→	hsc_s=Arts	[1, 0, 0]	50 : 25 : 25	-0.00	1
27	ssc_p≥65.0	→	hsc_s=Arts	[1, 0, 0]	50 : 25 : 25	-0.00	1
28	mba_p≥59.5	→	hsc_s=Commerce	[0, 6, 0]	11 : 78 : 11	-0.00	1
29	gender=F	→	hsc_s=Science	[0, 0, 1]	25 : 25 : 50	-0.00	1
30	ssc_p≥55.68	→	hsc_s=Commerce	[0, 2, 0]	20 : 60 : 20	-0.00	1
					25 : 25 : 50	-0.00	1
					50 : 25 : 25	-0.00	1
					25 : 50 : 25	-0.00	1
					5 : 52 : 42	-1.216	0

Predicted

Actual	Arts	Commerce	Science	Σ
Arts	3	1	2	6
Commerce	4	36	16	56
Science	4	8	33	45
Σ	11	45	51	107

Observations and Conclusion:

- Iris Dataset

Algorithm	Accuracy	Precision	Recall	F Score
Decision Tree	0.947	0.954	0.947	0.947
KNN	0.960	0.964	0.960	0.960
Rule Based Classifier	0.933	0.937	0.933	0.933

- Titanic Dataset

Algorithm	Accuracy	Precision	Recall	F Score
Decision Tree	0.777	0.808	0.777	0.741
KNN	0.777	0.808	0.777	0.741
Rule Based Classifier	0.781	0.811	0.781	0.747

- Placement Dataset

Algorithm	Accuracy	Precision	Recall	F Score
Decision Tree	0.710	0.729	0.710	0.719
KNN	0.467	0.436	0.467	0.450
Rule Based Classifier	0.785	0.768	0.785	0.774