University of Dhaka

Department of Computer Science and Engineering

CSE-4255:Introduction to Data Mining and Warehousing Lab

Lab Assignment-2

Sihan Tawsik
Roll: 5
Session: 2016-17

Submitted to-

Dr. Chowdhury Farhan Ahmed

Professor
Department of Computer Science and Engineering,
University of Dhaka

This report contains the comparison result of two classifiers: a) Decision tree classifier and b) Naive bayes classifier.

- Language: python
- Libraries: classification report from sklearn

Iris Dataset

- **Length:** 149
- **Trained Data:** 104
- **Tested Data:** 45

Target Values	Decision Tree					Naïve Bayes				
	Avg Precision	Avg Recall	F1-Score	Support	Accuracy	Precision	Recall	F1-Score	Support	Accuracy
Macro Avg	0.96	0.96	0.96	45		0.95	0.95	0.95	45	
Weighted Avg	0.96	0.96	0.96	45	96%	0.96	0.96	0.96	45	97%

lenses Dataset

- **Length:** 24
- **Trained Data:** 17
- **Tested Data:** 7

Target Values	Decision Tree					Naïve Bayes				
	Avg Precision	Avg Recall	F1-Score	Support	Accuracy	Precision	Recall	F1-Score	Support	Accuracy
Macro Avg	1.0	1.0	1.0	7	100%	0.29	0.33	0.31	7	87%
Weighted Avg	1.0	1.0	1.0	7		0.73	0.86	0.79	7	

Adult Dataset

- **Length:** 32560
- **Trained Data:** 22792
- **Tested Data:** 9768

Target Values	Decision Tree					Naïve Bayes				
	Avg Precision	Avg Recall	F1-Score	Support	Accuracy	Precision	Recall	F1-Score	Support	Accuracy
Macro Avg	0.63	0.56	0.68	9768	73%	0.75	0.66	0.88	9768	80%
Weighted Avg	0.68	0.72	0.65	9768		0.79	0.80	0.48	9768	

wine-quality-red Dataset

- **Length:** 1599
- **Trained Data:** 1120
- **Tested Data:** 480

Target Values	Decision Tree					Naïve Bayes				
	Avg Precision	Avg Recall	F1-Score	Support	Accuracy	Precision	Recall	F1-Score	Support	Accuracy
Micro Avg	0.29	0.27	0.27	480	56%	0.30	0.32	0.31	480	57%
Weighted Avg	0.55	0.57	0.55	480		0.56	0.56	0.56	480	

Ballon Dataset

- **Length:** 19
- **Trained Data:** 13
- **Tested Data:** 6

Target Values	Decision Tree					Naïve Bayes				
	Avg Precision	Avg Recall	F1-Score	Support	Accuracy	Precision	Recall	F1-Score	Support	Accuracy
Macro Avg	1.0	1.0	1.0	6	100%	0.25	0.50	0.30	6	52%
Weighted Avg	1	1.0	1.0	6		0.25	0.50	0.33	6	

Wine quality white Dataset

- **Length:** 4898
- **Trained Data:** 490
- **Tested Data:** 1469

Target Values	Decision Tree					Naïve Bayes				
	Avg Precision	Avg Recall	F1-Score	Support	Accuracy	Precision	Recall	F1-Score	Support	Accuracy
Macro Avg	0.34	0.30	0.32	1469	58%	0.36	0.29	0.27	1469	47%
Weighted Avg	0.55	0.56	0.54	1469		0.49	0.45	0.44	1469	

Car Dataset

- **Length:** 1727
- **Trained Data:** 1208
- **Tested Data:** 518

Target Values	Decision Tree					Naïve Bayes				
	Avg Precision	Avg Recall	F1-Score	Support	Accuracy	Precision	Recall	F1-Score	Support	Accuracy
Macro Avg	0.86	0.89	0.88	518	94%	0.35	0.38	0.36	518	79%
Weighted Avg	0.94	0.94	0.94	518		0.74	0.79	0.77	518	

amphibians data set

- **Length:** 189
- **Trained Data:** 132
- **Tested Data:** 57

Target Values	Decision Tree					Naïve Bayes				
	Avg Precision	Avg Recall	F1-Score	Support	Accuracy	Precision	Recall	F1-Score	Support	Accuracy
Macro Avg	0.50	0.59	0.97	57	99%	0.42	0.50	0.84	57	84%
Weighted Avg	0.72	0.99	0.98	57		0.71	0.84	0.84	57	

Conclusion: Decision tree, although takes more time than naive bayes, it gives more accurate results than naive bayes. Specially for small datasets. But for bigger datasets, naive bayes performs better than decision tree.