



**Department of Computer Science and Engineering
University of Dhaka**

Course:
CSE-4255: Introduction to Data Mining and Warehousing Lab

Lab Work-02

Submitted By:
Sree Sowmik Kumar Sarker
Roll: 54

Submitted to:
Dr. Chowdhury Farhan Ahmed
Professor
Department of Computer Science and Engineering
University of Dhaka.

Date: 22 February, 2021

Comparison between Decision-tree classification and Naive bayesian classification algorithm.

Naive bayesian classification algorithm	Decision-tree Classification Algorithm																																																																						
<div>Dataset Summary: adult.data Dataset Size: 32561 Number of Attributes: 14</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td><=50K</td><td>0.95</td><td>0.80</td><td>0.87</td><td>4944</td></tr><tr><td>>50K</td><td>0.58</td><td>0.87</td><td>0.70</td><td>1568</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.82</td><td>6512</td></tr><tr><td>macro avg</td><td>0.77</td><td>0.84</td><td>0.78</td><td>6512</td></tr><tr><td>weighted avg</td><td>0.86</td><td>0.82</td><td>0.83</td><td>6512</td></tr></tbody></table>		precision	recall	f1-score	support	<=50K	0.95	0.80	0.87	4944	>50K	0.58	0.87	0.70	1568	accuracy			0.82	6512	macro avg	0.77	0.84	0.78	6512	weighted avg	0.86	0.82	0.83	6512	<div>Dataset Summary: adult.data Dataset Size: 32561 Number of Attributes: 14</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td><=50K</td><td>0.92</td><td>0.80</td><td>0.80</td><td>4944</td></tr><tr><td>>50K</td><td>0.61</td><td>0.87</td><td>0.72</td><td>1568</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.82</td><td>6512</td></tr><tr><td>macro avg</td><td>0.77</td><td>0.74</td><td>0.77</td><td>6512</td></tr><tr><td>weighted avg</td><td>0.81</td><td>0.79</td><td>0.80</td><td>6512</td></tr></tbody></table>		precision	recall	f1-score	support	<=50K	0.92	0.80	0.80	4944	>50K	0.61	0.87	0.72	1568	accuracy			0.82	6512	macro avg	0.77	0.74	0.77	6512	weighted avg	0.81	0.79	0.80	6512										
	precision	recall	f1-score	support																																																																			
<=50K	0.95	0.80	0.87	4944																																																																			
>50K	0.58	0.87	0.70	1568																																																																			
accuracy			0.82	6512																																																																			
macro avg	0.77	0.84	0.78	6512																																																																			
weighted avg	0.86	0.82	0.83	6512																																																																			
	precision	recall	f1-score	support																																																																			
<=50K	0.92	0.80	0.80	4944																																																																			
>50K	0.61	0.87	0.72	1568																																																																			
accuracy			0.82	6512																																																																			
macro avg	0.77	0.74	0.77	6512																																																																			
weighted avg	0.81	0.79	0.80	6512																																																																			
<div>Dataset Summary: connect-4.data Dataset Size: 67557 Number of Attributes: 42</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>draw</td><td>0.26</td><td>0.07</td><td>0.12</td><td>1290</td></tr><tr><td>loss</td><td>0.70</td><td>0.41</td><td>0.52</td><td>3327</td></tr><tr><td>win</td><td>0.74</td><td>0.93</td><td>0.83</td><td>8894</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.72</td><td>13511</td></tr><tr><td>macro avg</td><td>0.57</td><td>0.47</td><td>0.49</td><td>13511</td></tr><tr><td>weighted avg</td><td>0.68</td><td>0.72</td><td>0.68</td><td>13511</td></tr></tbody></table>		precision	recall	f1-score	support	draw	0.26	0.07	0.12	1290	loss	0.70	0.41	0.52	3327	win	0.74	0.93	0.83	8894	accuracy			0.72	13511	macro avg	0.57	0.47	0.49	13511	weighted avg	0.68	0.72	0.68	13511	<div>Dataset Summary: connect-4.data Dataset Size: 67557 Number of Attributes: 42</div> <table><thead><tr><th></th><th>precision</th><th>recall</th><th>f1-score</th><th>support</th></tr></thead><tbody><tr><td>draw</td><td>0.36</td><td>0.27</td><td>0.32</td><td>1290</td></tr><tr><td>loss</td><td>0.80</td><td>0.51</td><td>0.62</td><td>3327</td></tr><tr><td>win</td><td>0.84</td><td>1.00</td><td>0.73</td><td>8894</td></tr><tr><td>accuracy</td><td></td><td></td><td>0.72</td><td>13511</td></tr><tr><td>macro avg</td><td>0.60</td><td>0.67</td><td>0.63</td><td>13511</td></tr><tr><td>weighted avg</td><td>0.88</td><td>0.82</td><td>0.83</td><td>13511</td></tr></tbody></table>		precision	recall	f1-score	support	draw	0.36	0.27	0.32	1290	loss	0.80	0.51	0.62	3327	win	0.84	1.00	0.73	8894	accuracy			0.72	13511	macro avg	0.60	0.67	0.63	13511	weighted avg	0.88	0.82	0.83	13511
	precision	recall	f1-score	support																																																																			
draw	0.26	0.07	0.12	1290																																																																			
loss	0.70	0.41	0.52	3327																																																																			
win	0.74	0.93	0.83	8894																																																																			
accuracy			0.72	13511																																																																			
macro avg	0.57	0.47	0.49	13511																																																																			
weighted avg	0.68	0.72	0.68	13511																																																																			
	precision	recall	f1-score	support																																																																			
draw	0.36	0.27	0.32	1290																																																																			
loss	0.80	0.51	0.62	3327																																																																			
win	0.84	1.00	0.73	8894																																																																			
accuracy			0.72	13511																																																																			
macro avg	0.60	0.67	0.63	13511																																																																			
weighted avg	0.88	0.82	0.83	13511																																																																			

Dataset Summary: iris.data Dataset Size: 150 Number of Attributes: 4 precision recall f1-score support Iris-setosa 0.91 1.00 0.95 10 Iris-versicolor 0.73 0.80 0.76 10 Iris-virginica 0.88 0.70 0.78 10 accuracy 0.83 30 macro avg 0.84 0.83 0.83 30 weighted avg 0.84 0.83 0.83 30	Dataset Summary: iris.data Dataset Size: 150 Number of Attributes: 4 precision recall f1-score support Iris-setosa 1.00 1.00 1.00 10 Iris-versicolor 1.00 1.00 1.00 10 Iris-virginica 1.00 1.00 1.00 10 accuracy 1.00 30 macro avg 1.00 1.00 1.00 30 weighted avg 1.00 1.00 1.00 30
Dataset Summary: breast-cancer.data Dataset Size: 286 Number of Attributes: 9 precision recall f1-score support no-recurrence-events 0.78 0.80 0.79 40 recurrence-events 0.50 0.47 0.48 17 accuracy 0.70 57 macro avg 0.64 0.64 0.64 57 weighted avg 0.70 0.70 0.70 57	Dataset Summary: breast-cancer.data Dataset Size: 286 Number of Attributes: 9 precision recall f1-score support no-recurrence-events 0.82 0.70 0.79 40 recurrence-events 0.57 0.67 0.48 17 accuracy 0.70 57 macro avg 0.67 0.68 0.67 57 weighted avg 0.67 0.67 0.67 57
Dataset Summary: kr-vs-kp.data Dataset Size: 3196 Number of Attributes: 36 Precision recall f1-score support nowin 0.89 0.85 0.87 306 won 0.87 0.91 0.89 333 accuracy 0.88 639	Dataset Summary: kr-vs-kp.data Dataset Size: 3196 Number of Attributes: 36 Precision recall f1-score support nowin 1.00 1.00 1.00 306 won 1.00 1.00 1.00 333 accuracy 1.00 639

macro avg 0.88 0.88 0.88 639 weighted avg 0.88 0.88 0.88 639	macro avg 1.00 1.00 1.00 639 weighted avg 1.00 1.00 1.00 639
Dataset Summary: krkopt.data Dataset Size: 28056 Number of Attributes: 6 precision recall f1-score support accuracy 0.37 5611 macro avg 0.38 0.33 0.34 5611 weighted avg 0.34 0.37 0.34 5611	Dataset Summary: krkopt.data Dataset Size: 28056 Number of Attributes: 6 precision recall f1-score support accuracy 0.37 5611 macro avg 0.58 0.63 0.54 5611 weighted avg 0.44 0.67 0.54 5611
Dataset Summary: breast-cancer-wisconsin.data Dataset Size: 699 Number of Attributes: 10 precision recall f1-score support accuracy 0.64 320 macro avg 0.66 0.79 0.67 320 weighted avg 0.77 0.64 0.66 320	Dataset Summary: breast-cancer-wisconsin.data Dataset Size: 699 Number of Attributes: 10 precision recall f1-score support accuracy 0.64 320 macro avg 0.76 1.00 0.77 320 weighted avg 1.00 0.94 0.86 320
Dataset Summary: agaricus-lepiota.data Dataset Size: 8124 Number of Attributes: 22 precision recall f1-score support	Dataset Summary: agaricus-lepiota.data Dataset Size: 8124 Number of Attributes: 22 precision recall f1-score support

accuracy 0.66 1624					accuracy 1.00 815				
macro avg 0.67 0.79 0.68 1624					macro avg 1.00 1.00 1.00 815				
weighted avg 0.77 0.66 0.67 1624					weighted avg 1.00 1.00 1.00 815				