

DATE:15/09/22

COURSE NAME:DATA WAREHOUSING AND DATA MINING FOR MEDICAL APPLICATIONS

COURSE CODE: CSA1654

NAME OF THE STUDENT: M.OOHA

REG NO:192111142

BRANCH:CSE

EXPERIMENT:23

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"'. The test options are set to 'Cross-validation' with 'Folds' set to 10. The classifier output displays the following metrics:

Time taken to build model: 13.89 seconds

=== Stratified cross-validation ===

=== Summary ===

Metric	Value
Correctly Classified Instances	3432
Incorrectly Classified Instances	1195
Kappa statistic	0.4046
Mean absolute error	0.35
Root mean squared error	0.4218
Relative absolute error	75.6846 %
Root relative squared error	87.7203 %
Total Number of Instances	4627

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.879	0.500	0.755	0.879	0.813	0.417	0.778	0.835	low
	0.500	0.121	0.702	0.500	0.584	0.417	0.778	0.680	high

=== Confusion Matrix ===

	a	b	<-- classified as
2592	356		a = low
839	840		b = high

EXPERIMENT:24

The screenshot shows the Weka Explorer interface with the 'Preprocess' tab selected. The filter chosen is 'Normalize -S 1.0 -T 0.0'. The current relation is 'german\_credit-weka.filters.unsupervised.attribute.Normalize-S1.0-T0.0' with 21 attributes and a sum of weights of 1000. The selected attribute is 'checking\_status' with 4 distinct values and a nominal type. The filter output shows the following distribution:

No.	Label	Count	Weight
1	<0	274	274
2	0<=X<200	269	269
3	>=200	63	63
4	no checking	394	394

The 'Class: class (Nom)' dropdown is set to 'Visualize All', showing a bar chart with four bars representing the counts for each category: 274, 269, 63, and 394.

## EXPERIMENT:25

**Weka Explorer**

Preprocess   **Classify**   **Cluster**   Associate   Select attributes   Visualize

Clusterer  
Choose **EM** -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100

Cluster mode  
☒ Use training set  
☐ Supplied test set Set...  
☐ Percentage split % 66  
☐ Classes to clusters evaluation (Nom) class  
☒ Store clusters for visualization

Ignore attributes  
Start Stop

Result list (right-click for options)  
17:53:34 - MakeDensityBasedClusterer  
18:11:04 - EM

Cluster output

=== Clustering model (full training set) ===

EM  
==

Number of clusters selected by cross validation: 16  
Number of iterations performed: 5

Attribute	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
	(0.09)	(0.03)	(0.03)	(0.06)	(0.06)	(0.03)	(0.03)	(0.07)	(0.12)	(0.06)	(0.07)
region-centroid-col											
mean	71.5227	87.808	122.0319	124.75	126.4601	118.2615	123.235	107.491	126.3789	115.8083	149.4251
std. dev.	42.8881	61.9287	91.9883	68.2498	72.5099	81.9168	79.1897	69.0148	68.895	66.8399	72.142
region-centroid-row											
mean	84.6219	106.0211	102.9603	128.9615	111.1966	134.4826	208.7378	120.8464	204.8754	85.5932	125.9052
std. dev.	23.8605	22.0862	28.7779	9.3951	46.9799	11.4221	25.1011	16.4719	28.4701	37.743	50.3096
region-pixel-count											
mean	9	9	9	9	9	9	9	9	9	9	9
std. dev.	0	0	0	0	0	0	0	0	0	0	0
short-line-density-5											
mean	0.0363	0.0235	0.0001	0	0.0098	0	0.1157	0.0019	0.0008	0.0207	0.0037

Status  
OK

Log x 0

91°F Mostly cloudy   18:11 15-09-2022