

EXPERIMENTS USING WEKA

M.ARBABZ SHERIEF

192124175

1.DATA PREPROCESSING AND PREPARATION FOR KNOWLEDGE ANALYSIS USING WEKA.

OUTPUT:

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) Publisher(s)[a] v

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

11:41:05 - SimpleKMeans

11:41:24 - EM

Clusterer output

EM

==

Number of clusters selected by cross validation: 6

Number of iterations performed: 0

Attribute	Cluster 0 (0.1)	Cluster 1 (0.27)	Cluster 2 (0.14)	Cluster 3 (0.18)	Cluster 4 (0.1)	Cluster 5 (0.2)
Rank						
mean	14.2	32.5385	31.2857	14.2222	16.6	30.1
std. dev.	17.3695	11.8224	11.6149	7.0317	15.8209	12.7056
Title						
Minecraft	2	1	1	1	1	1
Grand Theft Auto V	1	1	1	1	2	1
Tetris?y(EA)	2	1	1	1	1	1
Wii Sports	1	1	1	2	1	1
PUBG: Battlegrounds	1	1	1	1	2	1
Super Mario Bros.	1	1	1	1	1	2
Mario Kart 8?y/?yDeluxe	1	1	1	2	1	1
Pok??mon Red / Green / Blue?y/?yYellow	1	2	1	1	1	1
Red Dead Redemption 2	1	1	1	1	2	1
Terraria	2	1	1	1	1	1
Wii Fit?y/?yPlus	1	1	1	2	1	1
Tetris?y(1989) [j]	1	1	1	2	1	1
Pac-Man	2	1	1	1	1	1
Animal Crossing: New Horizons	1	1	1	2	1	1
The Witcher 3?y/?yHearts of Stone?y/?yBlood and Wine	1	1	2	1	1	1
Mario Kart Wii	1	1	1	2	1	1
Wii Sports Resort	1	1	1	2	1	1
New Super Mario Bros.	1	1	1	1	1	2
New Super Mario Bros. Wii	1	1	1	2	1	1

Clusterer

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

Cluster mode

☒ Use training set

☐ Supplied test set

☐ Percentage split %

☐ Classes to clusters evaluation

(Nom) Publisher(s)[a]

☒ Store clusters for visualization

Result list (right-click for options)

11:41:05 - SimpleKMeans

11:41:24 - EM

Clusterer output

Wii Sports Resort	1	1	1	2	1	1
New Super Mario Bros.	1	1	1	1	1	2
New Super Mario Bros. Wii	1	1	1	2	1	1
Call of Duty: Modern Warfare	1	1	2	1	1	1
Diablo III?y/?yReaper of Souls	1	2	1	1	1	1
Human: Fall Flat	1	2	1	1	1	1
The Elder Scrolls V: Skyrim	1	2	1	1	1	1
Super Smash Bros. Ultimate	1	1	1	1	1	2
The Legend of Zelda: Breath of the Wild	1	1	1	1	1	2
Pok?mon Gold / Silver?y/?yCrystal	1	2	1	1	1	1
Duck Hunt	1	1	1	1	1	2
Wii Play	1	1	1	2	1	1
The Walking Dead	1	1	2	1	1	1
Grand Theft Auto: San Andreas	1	1	1	1	2	1
Super Mario World	1	1	1	1	1	2
Call of Duty: Modern Warfare 3	1	1	2	1	1	1
Call of Duty: Black Ops	1	1	2	1	1	1
Borderlands 2	1	2	1	1	1	1
Pok?mon Sun / Moon?y/?yUltra Sun / Ultra Moon	1	2	1	1	1	1
Pok?mon Sword / Shield	1	2	1	1	1	1
Grand Theft Auto IV	1	1	1	1	2	1
Pok?mon Diamond / Pearl?y/?yPlatinum	1	2	1	1	1	1
Super Mario Bros. 3	1	1	1	1	1	2
Super Mario Odyssey	1	1	1	1	1	2
Call of Duty: Black Ops II	1	1	2	1	1	1
FIFA 18	1	2	1	1	1	1
Kinect Adventures!	1	2	1	1	1	1
Sonic the Hedgehog	2	1	1	1	1	1
Nintendogs	1	1	1	1	1	2
Mario Kart DS	1	1	1	1	1	2
God of War	1	2	1	1	1	1
Red Dead Redemption	1	2	1	1	1	1
Call of Duty: Modern Warfare 2	1	1	2	1	1	1
[total]	54	62	56	58	54	59
Sales						
238,000,000[b]	2	1	1	1	1	1
170,000,000	1	1	1	1	2	1

Clusterer

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

Cluster mode

☒ Use training set

☐ Supplied test set

☐ Percentage split %

☐ Classes to clusters evaluation

(Nom) Publisher(s)[a]

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

11:41:05 - SimpleKMeans

11:41:24 - EM

Clusterer output

100,000,000	2	1	1	1	1	1
82,900,000	1	1	1	2	1	1
75,000,000	1	1	1	1	2	1
58,000,000	1	1	1	1	1	2
56,870,000	1	1	1	2	1	1
47,520,000	1	2	1	1	1	1
46,000,000	1	1	1	1	2	1
44,500,000	2	1	1	1	1	1
43,800,000	1	1	1	2	1	1
43,000,000	1	1	1	2	1	1
42,071,635	2	1	1	1	1	1
40,170,000	1	1	1	2	1	1
40,000,000	1	1	2	1	1	1
37,380,000	1	1	1	2	1	1
33,140,000	1	1	1	2	1	1
30,800,000	1	1	1	1	1	2
30,320,000	1	1	1	2	1	1
30,000,000	1	1	2	1	1	1
29,530,000	1	1	1	1	1	2
29,490,000	1	1	1	1	1	2
28,300,000	1	1	1	1	1	2
28,020,000	1	1	1	2	1	1
28,000,000	1	1	2	1	1	1
27,500,000	1	1	1	1	2	1
26,662,500	1	1	1	1	1	2
26,500,000	1	1	2	1	1	1
26,200,000	1	1	2	1	1	1
27,000,000	1	2	1	1	1	1
25,370,000	1	2	1	1	1	1
25,000,000	1	1	1	1	2	1
24,730,000	1	2	1	1	1	1
24,430,000	1	1	1	1	1	2
24,400,000	1	1	1	1	1	2
24,200,000	1	1	2	1	1	1
24,000,000	1	2	1	1	1	1
23,982,960	2	1	1	1	1	1
23,960,000	1	1	1	1	1	2

Clusterer						
Choose EM -l 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						
<input checked="" type="radio"/> Use training set <input type="radio"/> Supplied test set <input type="text" value="Set..."/> <input type="radio"/> Percentage split % <input type="text" value="66"/> <input type="radio"/> Classes to clusters evaluation (Nom) Publisher(s)[a] <input type="text" value=""/> <input checked="" type="checkbox"/> Store clusters for visualization						
Ignore attributes						
<input type="button" value="Start"/> <input type="button" value="Stop"/>						
Result list (right-click for options)						
11:41:05 - SimpleKMeans						
11:41:24 - EM						
Cluster output						
23,600,000	1	1	1	1	1	2
23,000,000	1	9	1	1	1	1
22,700,000	1	1	2	1	1	1
[total]	47	55	49	51	47	52
Series						
Minecraft	2	1	1	1	1	1
Grand Theft Auto	1	1	1	1	4	1
Tetris	2	1	1	2	1	1
Wii	1	1	1	5	1	1
PUBG Universe	1	1	1	1	2	1
Super Mario	1	1	1	2	1	6
Mario Kart	1	1	1	3	1	2
Pokémon	1	6	1	1	1	1
Red Dead	1	2	1	1	2	1
None	2	3	1	1	1	3
Pac-Man	2	1	1	1	1	1
Animal Crossing	1	1	1	2	1	1
The Witcher	1	1	2	1	1	1
Call of Duty	1	1	6	1	1	1
Diablo	1	2	1	1	1	1
The Elder Scrolls	1	2	1	1	1	1
Super Smash Bros.	1	1	1	1	1	2
The Legend of Zelda	1	1	1	1	1	2
The Walking Dead	1	1	2	1	1	1
Borderlands	1	2	1	1	1	1
FIFA	1	2	1	1	1	1
Sonic the Hedgehog	2	1	1	1	1	1
God of War	1	2	1	1	1	1
[total]	28	36	30	32	28	33
Platform(s)						
Multi-platform[c]	5	1	1	1	1	1
Multi-platform	1	6	8	1	6	3
Multi-platform[e]	2	1	1	1	1	1
Wii	1	1	1	7	1	1
Multi-platform[f]	1	1	1	1	1	2
Wii U?y/?ySwitch	1	1	1	2	1	1
Game Boy?y/?yColor	1	2	1	1	1	1

Clusterer						
Choose EM -l 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						
<input checked="" type="radio"/> Use training set <input type="radio"/> Supplied test set <input type="text" value="Set..."/> <input type="radio"/> Percentage split % <input type="text" value="66"/> <input type="radio"/> Classes to clusters evaluation (Nom) Publisher(s)[a] <input type="text" value=""/> <input checked="" type="checkbox"/> Store clusters for visualization						
Ignore attributes						
<input type="button" value="Start"/> <input type="button" value="Stop"/>						
Result list (right-click for options)						
11:41:05 - SimpleKMeans						
11:41:24 - EM						
Cluster output						
Game Boy?y/?yNES	1	1	1	2	1	1
Nintendo Switch	1	2	1	2	1	3
Nintendo DS	1	2	1	1	1	4
Wii U / Switch	1	1	1	1	1	2
Game Boy Color	1	2	1	1	1	1
NES	1	1	1	1	1	2
Nintendo 3DS	1	2	1	1	1	1
Xbox 360	1	2	1	1	1	1
PlayStation 4?y/?yWindows	1	2	1	1	1	1
PS3?y/?yXbox 360	1	2	1	1	1	1
[total]	22	30	24	26	22	27
Initial release date						
November 18, 2011[d]	2	1	1	1	1	1
17-Sep-13	1	1	1	1	2	1
12-Sep-06	2	1	1	1	1	1
19-Nov-06	1	1	1	2	1	1
20-Dec-17	1	1	1	1	2	1
13-Sep-85	1	1	1	1	1	2
29-May-14	1	1	1	2	1	1
27-Feb-96	1	2	1	1	1	1
26-Oct-18	1	1	1	1	2	1
16-May-11	2	1	1	1	1	1
1-Dec-07	1	1	1	2	1	1
14-Jun-89	1	1	1	2	1	1
22-May-80	2	1	1	1	1	1
20-Mar-20	1	1	1	2	1	1
19-May-15	1	1	2	1	1	1
10-Apr-08	1	1	1	2	1	1
25-Jun-09	1	1	1	2	1	1
15-May-06	1	1	1	1	1	2
11-Nov-09	1	1	1	2	1	1
25-Oct-19	1	1	2	1	1	1
16-May-12	1	2	1	1	1	1
22-Jul-16	1	2	1	1	1	1
11-Nov-11	1	2	1	1	1	1
7-Dec-18	1	1	1	1	1	2
3-Mar-17	1	1	1	1	1	2
21-Nov-99	1	2	1	1	1	1

Cluster mode		Clusterer output					
<input checked="" type="radio"/> Use training set		21-Apr-84	1	1	1	1	2
<input type="radio"/> Supplied test set	Set...	2-Dec-06	1	1	1	2	1
<input type="radio"/> Percentage split	% 66	24-Apr-12	1	1	2	1	1
<input type="radio"/> Classes to clusters evaluation		26-Oct-04	1	1	1	1	2
(Nom) Publisher(s)[a]	▼	21-Nov-90	1	1	1	1	2
<input checked="" type="checkbox"/> Store clusters for visualization		8-Nov-11	1	1	2	1	1
		9-Nov-10	1	1	2	1	1
		18-Sep-12	1	2	1	1	1
		18-Nov-16	1	2	1	1	1
		15-Nov-19	1	2	1	1	1
		29-Apr-08	1	1	1	1	2
		28-Sep-06	1	2	1	1	1
		23-Oct-88	1	1	1	1	2
		27-Oct-17	1	1	1	1	2
		12-Nov-12	1	1	2	1	1
		29-Sep-17	1	2	1	1	1
		4-Nov-10	1	2	1	1	1
		23-Jun-91	2	1	1	1	1
		21-Apr-05	1	1	1	1	2
		14-Nov-05	1	1	1	1	2
		20-Apr-18	1	2	1	1	1
		18-May-10	1	2	1	1	1
		10-Nov-09	1	1	2	1	1
		[total]	54	62	56	58	59
		Developer(s)[a]					
		Mojang Studios	2	1	1	1	1
		Rockstar North	1	1	1	1	4
		EA Mobile	2	1	1	1	1
		Nintendo EAD	1	1	1	8	1
		FUBG Corporation	1	1	1	1	2
		Nintendo R&D4	1	1	1	1	2
		Game Freak	1	6	1	1	1
		Rockstar Studios	1	1	1	1	2
		Re-Logic	2	1	1	1	1
		Nintendo R&D1	1	1	1	2	1
		Namco	2	1	1	1	1
		Nintendo EPD	1	1	1	2	1

Cluster mode		Clusterer output					
<input checked="" type="radio"/> Use training set		CD Projekt Red	1	1	2	1	1
<input type="radio"/> Supplied test set	Set...	Infinity Ward	1	1	3	1	1
<input type="radio"/> Percentage split	% 66	Blizzard Entertainment	1	2	1	1	1
<input type="radio"/> Classes to clusters evaluation		No Brakes Games	1	2	1	1	1
(Nom) Publisher(s)[a]	▼	Bethesda Game Studios	1	2	1	1	1
<input checked="" type="checkbox"/> Store clusters for visualization		Bandai Namco Studios?y/?ySora Ltd.	1	1	1	1	2
		Telltale Games	1	1	2	1	1
		Infinity Ward?y/?ySledgehammer	1	1	2	1	1
		Treyarch	1	1	3	1	1
		Gearbox Software	1	2	1	1	1
		EA Vancouver	1	2	1	1	1
		Good Science Studio	1	2	1	1	1
		Sonic Team	2	1	1	1	1
		Santa Monica Studio	1	2	1	1	1
		Rockstar San Diego	1	2	1	1	1
		[total]	32	40	34	36	37
		Publisher(s)[a]					
		Xbox Game Studios	2	2	1	1	1
		Rockstar Games	1	2	1	1	5
		Electronic Arts	2	1	1	1	1
		Nintendo	1	3	1	10	1
		FUBG Corporation	1	1	1	1	2
		Re-Logic?y/?y505 Games	2	1	1	1	1
		Namco	2	1	1	1	1
		CD Projekt	1	1	2	1	1
		Activision	1	1	6	1	1
		Blizzard Entertainment	1	2	1	1	1
		Curve Digital	1	2	1	1	1
		Bethesda Softworks	1	2	1	1	1
		Telltale Games	1	1	2	1	1
		2K Games	1	2	1	1	1
		Nintendo?y/?yThe Pok??mon Company	1	4	1	1	1
		EA Sports	1	2	1	1	1
		Sega	2	1	1	1	1
		Sony Interactive Entertainment	1	2	1	1	1
		[total]	23	31	25	27	28

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
☐ Percentage split
 %
☐ Classes to clusters evaluation
(Nom) Publisher(s)[a]
☒ Store clusters for visualization

Result list (right-click for options)

11:41:05 - SimpleKMeans
11:41:24 - EM

Cluster output

	0	1	2	3	4	5
Atari Game Studios	1	2	1	1	5	1
Rockstar Games	2	1	1	1	1	1
Electronic Arts	1	3	1	10	1	11
Nintendo	1	1	1	1	2	1
PUBG Corporation	2	1	1	1	1	1
Re-Logic?y?y505 Games	2	1	1	1	1	1
Namco	1	1	2	1	1	1
CD Projekt	1	1	6	1	1	1
Activision	1	2	1	1	1	1
Blizzard Entertainment	1	2	1	1	1	1
Curve Digital	1	2	1	1	1	1
Bethesda Softworks	1	2	1	1	1	1
Telltale Games	1	2	1	1	1	1
2K Games	1	2	1	1	1	1
Nintendo?g?yThe Pok??mon Company	1	4	1	1	1	1
EA Sports	2	1	1	1	1	1
Sega	2	1	1	1	1	1
Sony Interactive Entertainment	1	2	1	1	1	1
[total]	23	31	25	27	23	28

Time taken to build model (full training data) : 0.29 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 5 (10%)
1 13 (27%)
2 7 (14%)
3 9 (18%)
4 5 (10%)
5 10 (20%)

Log likelihood: -23.46274

2.K-MEANS CLUSTER ANALYSIS USING WEKA

Clusterer

ChooseSimpleKMeans-init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set

☐ Percentage split

☐ Classes to clusters evaluation

☒ Store clusters for visualization

Set...

%66

(Num) Class

Ignore attributes

Start

Stop

Result list (right-click for options)

12:19:54 - EM

12:20:06 - SimpleKMeans

Clusterer output

kMeans

=====

Number of iterations: 10

Within cluster sum of squared errors: 653.6910342150591

Initial starting points (random):

Cluster 0: 1,1,1,0,0,0,0,1,0,1,1,1,0.049296,0.364198,0,0.019868,0,0.610619,0.56,1,1,1,1,69,200,60,1,1,1,1.39,2159,13.1,90.6,10400,261000,3.6,1.9,42,197,552,335,'

Cluster 1: 1,0,0,0,0,0,0,0,0,1,0,0,0.055944,0.610619,0.56,0.714286,0,0,0,66,0,0,0,1,1,1.29,7,12.1,95.1,3.8,77000,3.8,1.4,37,38,194,161,6.7,0.71,1,10,0.5,

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster#	
		0	1
	(165.0)	(96.0)	(69.0)
1.Gen	0.8061	0.9479	0.6087
2.Sym	0.6395	0.702	0.5524
3.Alc	0.7394	0.9271	0.4783
4.HepB	0.1081	0.112	0.1026
5.HepB	0.0079	0.0123	0.0018
6.HepB	0.2695	0.2733	0.2643
7.HepC	0.2179	0.1676	0.288
8.Cir	0.903	0.9896	0.7826
9.End	0.0794	0.0607	0.1054
10.Smo	0.5081	0.6379	0.3274
11.Dia	0.3457	0.3229	0.3773
12.Obe	0.129	0.1421	0.1108
13.Hem	0.0493	0.0499	0.0485
14.Art	0.3642	0.3267	0.4164
15.CRen	0.1227	0.0938	0.163
16.HIV	0.0199	0.0117	0.0313
17.Nom	0.0559	0.0498	0.0645
18.EVar	0.6106	0.7742	0.3831
19.Spl	0.56	0.7792	0.2551

Clusterer

ChooseSimpleKMeans-init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set

Set...

☐ Percentage split

%66

☐ Classes to clusters evaluation

(Num) Class

☒ Store clusters for visualization

Ignore attributes

StartStop

Result list (right-click for options)

12:19:54 - EM

12:20:06 - SimpleKMeans

Clusterer output

20.PHypp	0.7143	0.9881	0.3333
21.Thr	0.2222	0.3356	0.0644
22.LMet	0.2236	0.2442	0.1949
23.Rad	0.681	0.7779	0.5461
24.Agedia	64.6909	65.8542	63.0725
25.Alcpd	71.0085	92.0443	41.7414
26.cigpy	20.4643	21.5115	19.0072
27.Sta	1.0182	1.4375	0.4348
28.Enodeg	1.1585	1.25	1.0313
29.Ascdeg	1.4417	1.6771	1.1143
30.IntNorRat	1.4219	1.5616	1.2274
31.Alp	19299.9511	29657.7129	4889.1522
32.Hae	12.879	12.4865	13.4252
33.MCorVol	95.1198	96.276	93.511
34.Leu	1473.9615	1455.2479	1499.9979
35.Flat	113206.4427	98567.5386	133573.6134
36.Alb	3.4455	3.2643	3.6976
37.Bil	3.0879	4.328	1.3626
38.Ala	67.0932	67.2718	66.8446
39.Aspa	96.3827	103.7396	86.1471
40.Gam	268.0265	267.3646	268.9475
41.Alk	212.2116	205.7842	221.1541
42.Prot	8.961	9.4319	8.3059
43.Crea	1.1271	1.0822	1.1896
44.NNod	2.7362	2.8202	2.6194
45.dnod	6.8512	6.284	7.6403
46.Bil	1.93	2.4674	1.1823
47.Iro	85.5988	87.7829	82.5602
48.Oxy	37.0289	39.4884	33.6071
49.Per	438.9976	458.0491	412.4913
Class	0.6182	0.5313	0.7391

Time taken to build model (full training data) : 0 seconds

Clusterer

Choose

SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set

☐ Percentage split

☐ Classes to clusters evaluation

☒ Store clusters for visualization

Set...

% 66

(Num) Class

Ignore attributes

Start

Stop

Result list (right-click for options)

12:19:54 - EM

12:20:06 - SimpleKMeans

Clusterer output

28.Encdeg	1.1585	1.25	1.0313
29.Ascdeg	1.4417	1.6771	1.1143
30.IntNorRat	1.4219	1.5616	1.2274
31.Alp	19299.9511	29657.7129	4889.1522
32.Hae	12.879	12.4865	13.4252
33.MCorVol	95.1198	96.276	93.511
34.Leu	1473.9615	1455.2479	1499.9979
35.Flat	113206.4427	98567.5386	133573.6134
36.Alb	3.4455	3.2643	3.6976
37.Bil	3.0879	4.328	1.3626
38.Ala	67.0932	67.2718	66.8446
39.Asps	96.3827	103.7396	86.1471
40.Gam	268.0265	267.3646	268.9475
41.Alk	212.2116	205.7842	221.1541
42.Prot	8.961	9.4319	8.3059
43.Crea	1.1271	1.0822	1.1896
44.NNod	2.7362	2.8202	2.6194
45.dnod	6.8512	6.284	7.6403
46.Bil	1.93	2.4674	1.1823
47.Iro	85.5988	87.7829	82.5602
48.Oxy	37.0289	39.4884	33.6071
49.Fer	438.9976	458.0491	412.4913
Class	0.6182	0.5313	0.7391

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	96 (58%)
1	69 (42%)

Clusterer

ChooseSimpleKMeans-init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 6 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

Use training set

Supplied test set

Percentage split

Classes to clusters evaluation

Set...

%

66

(Num) Class

Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

12:19:54 - EM

12:20:06 - SimpleKMeans

12:24:46 - SimpleKMeans

Clusterer output

kMeans

=====

Number of iterations: 10

Within cluster sum of squared errors: 554.0663475375386

Initial starting points (random):

Cluster 0: 1,1,1,0,0,0,0,1,0,1,1,0,0.049296,0.364198,0,0,0.019868,0,0.610619,0.56,1,1,1,1,69,200,60,1,1,1,1,39,2159,13,1,90,6,10400,261000,3,6,1,9,42,197,552,335,

Cluster 1: 1,0,0,0,0,0,0,0,0,1,0,0,0.055944,0.610619,0.56,0.714286,0,0,0,66,0,0,0,1,1,1,29,7,12,1,95,1,3,8,77000,3,8,1,4,37,38,194,161,6,7,0,71,1,10,0,5,

Cluster 2: 1,1,1,0,0.007937,0.269504,1,1,0.079365,0.508065,0,0,0,0,0,0,1,1,1,0,0,1,61,100,20.464286,0,1,1,1,92,2269,12,1,119,5,1,80000,3,445535,9,6,204,357,19,

Cluster 3: 1,1,1,0,0,0,0,1,0,0,0,0,0,0,0,0,0,1,1,1,0,1,0,69,100,0,2,1,3,1,41,123,10,1,89,5,2,3,89000,4,4,3,31,60,75,177,6,8,0,7,3,3,5,1,37,11,173,0

Cluster 4: 1,0,1,0,0.007937,0,0,1,0.079365,0.508065,1,1,0.049296,1,0,0,0,1,1,1,0,0,1,75,71.008547,20.464286,0,1,1,1,71,5,2,11,1,106,9,2100,52000,3,3,2,6,15,47,9,

Cluster 5: 1,1,1,0,0,0,1,0,1,0,1,0,0,0,0,0,0,0,1,0,0,1,71,71.008547,0,1,1,2,1,11,5,9,1,90,9,6,36,307000,2,47,3.087938,31,29,339,254,5,5,1,18,3,6,3,1,93,85,598,

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (165.0)	Cluster#					
		0 (22.0)	1 (38.0)	2 (24.0)	3 (25.0)	4 (41.0)	5 (15.0)
1.Gen	0.8061	0.9545	0.6053	0.7917	0.92	0.8293	0.8667
2.Sym	0.6395	1	0.5936	0.7166	0.8912	0.2243	0.8186
3.Alc	0.7394	1	0.2895	0.625	0.88	0.9024	1
4.HepB	0.1081	0.0504	0.1017	0.25	0.1373	0.0053	0.2144
5.HepB	0.0079	0.0007	0.0025	0.001	0.0419	0.0021	0.0026
6.HepB	0.2695	0.1486	0.2217	0.8141	0.2247	0.0197	0.4539
7.HepC	0.2179	0.0099	0.2277	0.3333	0.4887	0.0594	0.2957
8.Cir	0.903	1	0.6842	0.9167	1	1	0.8667
9.End	0.0794	0.0743	0.043	0.2249	0.0222	0.0174	0.2106
10.Smo	0.5081	0.6162	0.2644	0.6885	0.5029	0.5517	0.5672
11.Dia	0.3457	0.3182	0.3775	0	0.32	0.5047	0.4667
12.Obe	0.129	0.2786	0.0823	0.0417	0.0103	0.2352	0.0753
13.Rem	0.0493	0.0522	0.0315	0.0021	0.0538	0.1036	0.0099
14.Art	0.3642	0.5166	0.4043	0.0833	0.12	0.3747	0.8667
15.CRen	0.1227	0.1364	0.1348	0	0.08	0.1249	0.3333

Clusterer

ChooseSimpleKMeans-init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 6 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set

☐ Percentage split

☐ Classes to clusters evaluation

☒ Store clusters for visualization

Set...

% 66

(Num) Class

Ignore attributes

StartStop

Result list (right-click for options)

12:19:54 - EM

12:20:06 - SimpleKMeans

12:24:46 - SimpleKMeans

Clusterer output

16.HIV	0.0199	0.0009	0.0295	0.0833	0.0024	0.0005	0.004
17.Mon	0.0559	0.1036	0.1126	0.0417	0.0067	0.0339	0.0075
18.EVar	0.6106	0.578	0.2455	0.6965	0.8754	0.8942	0.2295
19.Sp1	0.56	0.5255	0.1526	0.8017	0.9424	0.7863	0
20.PByp	0.7143	0.9545	0.0376	0.9107	1	0.9547	0.6286
21.Thr	0.2222	0.6818	0.0643	0.25	0.24	0.1762	0
22.LMet	0.2236	0.5	0.2486	0.0833	0.4089	0.0055	0.2667
23.Rad	0.681	0.8636	0.4653	0.7083	0.6	0.7239	0.9333
24.Agedia	64.6909	70.2727	61.3947	57.5833	62.76	65.7805	76.4667
25.Alcpd	71.0085	125.4565	23.6068	46.5442	82.8014	83.7115	96.0028
26.cigpy	20.4643	24.3295	21.7415	21.2009	17.4443	18.7639	20.0619
27.Sta	1.0182	1.9545	0.5	0.5	1.64	0.8049	1.3333
28.Enodeg	1.1585	1.2727	1.0305	1.125	1.4	1.122	1.0667
29.Aacdeg	1.4417	1.8182	1.0643	1.25	2.08	1.3766	1.2667
30.IntNorRat	1.4219	1.6005	1.2243	1.3589	1.5088	1.5095	1.3127
31.Alp	19299.9511	108153.4359	7433.5606	5177.5438	13930.8761	992.312	628.22
32.Hae	12.879	12.3545	13.2805	14.1116	11.668	13.0829	12.12
33.MCorVol	95.1198	95.1273	92.8642	95.1175	95.184	97.6585	93.78
34.Leu	1473.9615	2638.7182	1961.6151	420.8418	1332.836	1087.8732	1505.772
35.Flat	113206.4427	149100.7727	143736.8654	76135.9768	117313.84	79858.6027	126035.5333
36.Alb	3.4455	3.1312	3.6927	3.7702	3.0386	3.3779	3.624
37.Bil	3.0879	3.8404	1.1286	1.6328	8.092	3.038	1.0725
38.Ala	67.0932	70.7727	62.3733	84.7578	76.88	57.3415	55.7333
39.Aapa	96.3827	106.9545	80.8096	95.7659	155.28	73.6098	85.4
40.Gam	268.0265	397.1818	245.5882	245.8761	281.24	219	282.8667
41.Alk	212.2116	254.5127	251.9059	176.1755	258.84	152.9024	191.6667
42.Prot	8.961	9.4755	8.9453	12.4676	7.0849	8.2874	7.6041
43.Crea	1.1271	0.9525	1.0812	0.8793	1.4395	0.8751	2.064
44.MMod	2.7362	4.2273	2.572	2.7083	2.72	2.2131	2.4667
45.dnod	6.8512	8.4321	8.7915	5.4271	6.1203	5.0635	8
46.Bil	1.93	2.2982	1.3232	0.9867	5.0688	1.2422	1.0853
47.Iro	85.5988	78.1179	77.0915	105.6078	72.3756	95.3993	81.3593
48.Oxy	37.0289	40.5761	32.8837	43.9803	32.7704	39.5482	31.4174
49.Per	438.9976	564.6356	441.0776	443.7487	450.0392	377.1646	325.7986
Class	0.6182	0.2727	0.6316	0.9167	0.12	0.9268	0.6

Clusterer

Choose
SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 6 -A "weka.core.EuclideanDistance" -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set
☐ Supplied test set
☐ Percentage split
☐ Classes to clusters evaluation

(Num) Class

▼

☒ Store clusters for visualization

Ignore attributes

Start
Stop

Result list (right-click for options)

12:19:54 - EM
12:20:06 - SimpleKMeans

Clusterer output

32.Hae	12.879	12.3545	13.2805	14.1116	11.668	13.0829	12.12
33.MCorVol	95.1198	95.1273	92.8642	95.1175	95.184	97.6585	93.78
34.Leu	1473.9615	2638.7182	1961.6151	420.8418	1332.836	1087.8732	1505.772
35.Plat	113206.4427	149100.7727	149736.8654	76135.9768	117313.84	79858.6027	126835.5333
36.Alb	3.4455	3.1312	3.6927	3.7702	3.0386	3.3779	3.624
37.Bil	3.0879	3.8404	1.1286	1.6328	8.092	3.038	1.0725
38.Ala	67.0932	70.7727	62.3733	84.7578	76.88	57.3415	55.7333
39.Aspe	96.3827	106.9545	80.8096	95.7659	155.28	73.6098	85.4
40.Gam	268.0265	397.1818	245.5882	245.8761	281.24	219	282.8667
41.Alk	212.2116	254.5127	251.9059	176.1755	258.84	152.9024	191.6667
42.Frot	8.961	9.4755	8.9453	12.4676	7.0849	8.2874	7.6041
43.Crea	1.1271	0.9525	1.0812	0.8793	1.4395	0.8751	2.064
44.NMod	2.7362	4.2273	2.572	2.7083	2.72	2.2131	2.4667
45.dmod	6.8512	8.4321	8.7915	5.4271	6.1203	5.0635	8
46.Bil	1.93	2.2982	1.3232	0.9867	5.0688	1.2422	1.0853
47.Iro	85.5988	78.1179	77.0915	105.6078	72.3756	95.3993	81.3593
48.Oxy	37.0289	40.5761	32.8837	43.9803	32.7704	39.5482	31.4174
49.Per	438.9976	564.6356	441.0776	443.7487	490.0392	377.1646	325.7986
Class	0.6182	0.2727	0.6316	0.9167	0.12	0.9268	0.6

Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0

22 (13%)

1

38 (23%)

2

24 (15%)

3

25 (15%)

4

41 (25%)

5

15 (9%)

3.DATA ANALYSIS BY EXPECTATION MAXIMISATION ALGORITHM USING WEKA.

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

(Num) Class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

12:45:20 - EM

Clusterer output

EM

==

Number of clusters selected by cross validation: 2

Number of iterations performed: 1

Attribute	Cluster	
	0	1
	(0.51)	(0.49)
=====		
1.Gen		
mean	0.9239	0.6823
std. dev.	0.2652	0.4656
2.Sym		
mean	0.6701	0.6073
std. dev.	0.4494	0.4549
3.Alc		
mean	0.9126	0.5574
std. dev.	0.2824	0.4967
4.HepB		
mean	0.1142	0.1017
std. dev.	0.307	0.2798
5.HepB		
mean	0.0137	0.0019
std. dev.	0.108	0.0034
6.HepB		
mean	0.258	0.2816
std. dev.	0.4094	0.4106
7.HepC		
mean	0.1947	0.2423
std. dev.	0.3833	0.4183

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

(Num) Class ▾

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

12:45:20 - EM

Clusterer output

8.Cir

mean10.8012

std. dev.00.3991

9.End

mean0.06630.0931

std. dev.0.21180.2587

10.Smo

mean0.62830.3817

std. dev.0.39730.4338

11.Dia

mean0.3260.3664

std. dev.0.46820.4736

12.cbe

mean0.14870.1083

std. dev.0.3460.2998

13.Rem

mean0.0430.0559

std. dev.0.1840.2169

14.Art

mean0.30580.4255

std. dev.0.45780.4886

15.CRen

mean0.06590.1824

std. dev.0.24810.3827

16.HIV

mean0.02470.0148

std. dev.0.15140.1114

17.Non

mean0.03190.0812

Clusterer

Choose **EM** -I 100 -N -1 -X 10 -max -1 -ll-cv 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 (Num) Class

▼

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

12:45:20 - EM

Clusterer output

std. dev.	0.1519	0.2615
18.EVar		
mean	0.774	0.435
std. dev.	0.331	0.4016
19.Sp1		
mean	0.8584	0.2465
std. dev.	0.3351	0.3867
20.FHyp		
mean	0.9966	0.4177
std. dev.	0.0309	0.4668
21.Thr		
mean	0.3811	0.0553
std. dev.	0.4836	0.219
22.LMet		
mean	0.2255	0.2217
std. dev.	0.4129	0.4101
23.Rad		
mean	0.7697	0.5878
std. dev.	0.418	0.4895
24.Agedia		
mean	64.6327	64.7521
std. dev.	10.7887	15.4692
25.Alopd		
mean	51.6537	49.3155
std. dev.	68.1823	50.8489
26.cigpy		
mean	21.2032	19.688
std. dev.	17.1905	57.9338

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

☐ Percentage split

☐ Classes to clusters evaluation

☒ Store clusters for visualization

Set...

% 66

(Num) Class

Ignore attributes

Start

Stop

Result list (right-click for options)

12:45:20 - EM

Clusterer output

27.Sta

mean1.38280.6351

std. dev.1.22010.9974

28.Encdeg

mean1.29561.0145

std. dev.0.55030.1126

29.Ascdeg

mean1.75551.112

std. dev.0.75330.3724

30.IntMorRat

mean1.60851.2258

std. dev.0.56940.1958

31.Alp

mean33664.48644209.0737

std. dev.200435.978321160.066

32.Hae

mean12.673413.0951

std. dev.1.98512.2309

33.MCorVol

mean96.620993.5427

std. dev.8.70267.5471

34.Leu

mean1302.45021654.1454

std. dev.2634.79443094.5896

35.Flat

mean99178.8625127943.3269

std. dev.104014.41105681.8202

Clusterer

ChooseEM -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

(Num) Class

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

124520 - EM

Clusterer output

36.A1b

mean3.25943.6411

std. dev.0.66350.6202

37.B1l

mean4.80841.2805

std. dev.7.07330.9073

38.A1a

mean69.597464.4624

std. dev.62.427849.7547

39.Aspa

mean104.919887.414

std. dev.86.871585.0192

40.Gam

mean262.0466274.3089

std. dev.253.4341257.6957

41.A1k

mean199.4116225.6588

std. dev.128.1886197.0228

42.Prot

mean9.78978.0904

std. dev.13.6757.9767

43.Crea

mean1.05081.2073

std. dev.0.7171.1091

44.MMod

mean2.81062.6581

std. dev.1.77811.7817

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

(Num) Class ▾

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

12:45:20 - EM

Clusterer output

45.dmod

mean

5.8052

7.9501

std. dev.

4.09

5.1503

46.Bil

mean

2.7084

1.1123

std. dev.

4.8339

0.7648

47.Iro

mean

90.5607

80.3861

std. dev.

45.9998

31.6416

48.Oxy

mean

40.5441

33.3361

std. dev.

23.1491

16.9693

49.Per

mean

456.0817

421.0498

std. dev.

382.4285

252.6978

Class

mean

0.526

0.715

std. dev.

0.4993

0.4514

Time taken to build model (full training data) : 0.36 seconds

=== Model and evaluation on training set ===

Clustered Instances

0

89 (54%)

1

76 (46%)

Log likelihood: -114.21025

Clusterer

Choose

SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set

☐ Percentage split

☐ Classes to clusters evaluation

Set...

%

66

(Num) Cla

Train on a percentage of the data and cluster the remainder

☒ Store clusters for visualization

Ignore attributes

Start

Stop

Result list (right-click for options)

12:45:20 - EM

12:50:30 - SimpleKMeans

Cluster output

kMeans

=====

Number of iterations: 10

Within cluster sum of squared errors: 653.6910342150591

=====

Starting points (random):

Cluster 0: 1,1,1,0,0,0,0,1,0,1,1,1,0.049296,0.364198,0,0,0.019868,0,0,0.610619,0.56,1,1,1,1,69,200,60,1,1,1,1.38,2159,13.1,90.6,10400,261000,3.6,1.9,42,197,552,335,'

Cluster 1: 1,0,0,0,0,0,0,0,0,1,0,0,1,0,0,0.055944,0.610619,0.56,0.714286,0,0,0,66,0,0,0,1,1,1.29,7,12.1,95.1,3.8,77000,3.8,1.4,37,38,194,161,6.7,0.71,1,10,0.5,

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster#	
		0	1
	(165.0)	(96.0)	(69.0)
1.Gen	0.8061	0.9479	0.6087
2.Sym	0.6395	0.702	0.5524
3.Alc	0.7394	0.9271	0.4783
4.HepB	0.1081	0.112	0.1026
5.HepB	0.0079	0.0123	0.0018
6.HepB	0.2695	0.2733	0.2643
7.HepC	0.2179	0.1676	0.288
8.Cir	0.903	0.9896	0.7826
9.End	0.0794	0.0607	0.1054
10.Smo	0.5081	0.6379	0.3274
11.Dia	0.3457	0.3229	0.3773
12.Obe	0.129	0.1421	0.1108
13.Hem	0.0493	0.0499	0.0485
14.Art	0.3642	0.3267	0.4164
15.CRen	0.1227	0.0938	0.163
16.HIV	0.0199	0.0117	0.0313
17.Non	0.0559	0.0498	0.0645
18.EVar	0.6106	0.7742	0.3831
19.Sp1	0.56	0.7792	0.2551

Clusterer

Choose

SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set

☐ Percentage split

☐ Classes to clusters evaluation

☒ Store clusters for visualization

Set...

% 66

(Num) Class

Ignore attributes

Start

Stop

Result list (right-click for options)

12:45:20 - EM

12:50:30 - SimpleKMeans

Clusterer output

20.PHyp	0.7143	0.9881	0.3333
21.Thr	0.2222	0.3356	0.0644
22.IMet	0.2236	0.2442	0.1949
23.Rad	0.681	0.7779	0.5461
24.Agedia	64.6909	65.8542	63.0725
25.Alcpd	71.0085	92.0443	41.7414
26.cigpy	20.4643	21.5115	19.0072
27.Sta	1.0182	1.4375	0.4348
28.Encdeg	1.1585	1.25	1.0313
29.Ascdeg	1.4417	1.6771	1.1143
30.IntNorRat	1.4219	1.5616	1.2274
31.Alp	19299.9511	29657.7129	4889.1522
32.Hae	12.879	12.4865	13.4252
33.MCorVol	95.1198	96.276	93.511
34.Leu	1473.9615	1455.2479	1499.9979
35.Plat	113206.4427	98567.5386	133573.6134
36.Alb	3.4455	3.2643	3.6976
37.Bil	3.0879	4.328	1.3626
38.Ala	67.0532	67.2718	66.8446
39.Asps	96.3827	103.7396	86.1471
40.Gam	268.0265	267.3646	268.9475
41.Alk	212.2116	205.7842	221.1541
42.Prot	8.961	9.4319	8.3059
43.Crea	1.1271	1.0822	1.1896
44.NNod	2.7362	2.8202	2.6194
45.dnod	6.8512	6.284	7.6403
46.Bil	1.93	2.4674	1.1823
47.Iro	85.5588	87.7829	82.5602
48.Oxy	37.0289	39.4884	33.6071
49.Per	438.9976	458.0491	412.4913
Class	0.6182	0.5313	0.7391

Time taken to build model (full training data) : 0 seconds

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer

Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set Set...

☐ Percentage split % 66

☐ Classes to clusters evaluation
(Num) Class v

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

12:45:20 - EM

12:50:30 - SimpleKMeans

Clusterer output

28.Encdeg	1.1585	1.25	1.0313
29.Ascdeg	1.4417	1.6771	1.1143
30.IntNorRat	1.4219	1.5616	1.2274
31.Alp	19299.9511	29657.7129	4889.1522
32.Hae	12.879	12.4865	13.4252
33.MCorVol	95.1198	96.276	93.511
34.Leu	1473.9615	1455.2479	1499.9979
35.Plat	113206.4427	98567.5386	133573.6134
36.Alb	3.4455	3.2643	3.6976
37.Bil	3.0879	4.328	1.3626
38.Ala	67.0932	67.2718	66.8446
39.Asps	96.3827	103.7396	86.1471
40.Gam	268.0265	267.3646	268.9475
41.Alk	212.2116	205.7842	221.1541
42.Prot	8.961	9.4319	8.3059
43.Crea	1.1271	1.0822	1.1896
44.NNod	2.7362	2.8202	2.6194
45.dnod	6.8512	6.284	7.6403
46.Bil	1.93	2.4674	1.1823
47.Iro	85.5988	87.7829	82.5602
48.Oxy	37.0289	39.4884	33.6071
49.Fer	438.9976	458.0491	412.4913
Class	0.6182	0.5313	0.7391

Time taken to build model (full training data) : 0 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	96 (58%)
1	69 (42%)

4.DATA ANALYSIS BY COBWEB-HIERARCHICAL CLUSTERING ALGORITHM USING WEKA.

Associator

Choose **Apriori** -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

StartStop

Associator output

Result list (right-click for ...)

130437 - Apriori

Apriori

=====

Minimum support: 0.15 (694 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 44

Size of set of large itemsets L(2): 380

Size of set of large itemsets L(3): 910

Size of set of large itemsets L(4): 633

Size of set of large itemsets L(5): 105

Size of set of large itemsets L(6): 1

Best rules found:

1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 <conf:(0.92)> lift:(1.27) lev:(0.03) [155] conv:(3.35)

2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf:(0.92)> lift:(1.27) lev:(0.03) [149] conv:(3.28)

3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 <conf:(0.92)> lift:(1.27) lev:(0.03) [150] conv:(3.27)

4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746 <conf:(0.92)> lift:(1.27) lev:(0.03) [159] conv:(3.26)

5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779 <conf:(0.91)> lift:(1.27) lev:(0.04) [164] conv:(3.15)

6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 <conf:(0.91)> lift:(1.26) lev:(0.03) [151] conv:(3.06)

7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701 <conf:(0.91)> lift:(1.26) lev:(0.03) [145] conv:(3.01)

8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(3)

9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757 <conf:(0.91)> lift:(1.26) lev:(0.03) [156] conv:(3)

10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(2.92)

Associator

Choose **Apriori** -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -< -1

Start

Stop

Result list (right-click for ...)

130437 - Apriori

Apriori

=====

Minimum support: 0.15 (694 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 44

Size of set of large itemsets L(2): 380

Size of set of large itemsets L(3): 910

Size of set of large itemsets L(4): 633

Size of set of large itemsets L(5): 105

Size of set of large itemsets L(6): 1

Best rules found:

1. biscuits=t frozen foods=t fruit=t total=high

2. baking needs=t biscuits=t fruit=t total=high

3. baking needs=t frozen foods=t fruit=t total=high

4. biscuits=t fruit=t vegetables=t total=high 81

5. party snack foods=t fruit=t total=high 854 ==

6. biscuits=t frozen foods=t vegetables=t total=

7. baking needs=t biscuits=t vegetables=t total=

8. biscuits=t fruit=t total=high 954 ==> bread a

9. frozen foods=t fruit=t vegetables=t total=high

10. frozen foods=t fruit=t total=high 969 ==> bre

weka.gui.GenericObjectEditor

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm.

More

Capabilities

car False

classIndex -1

delta 0.05

doNotCheckCapabilities False

lowerBoundMinSupport 0.1

metricType Confidence

minMetric 0.9

numRules 10

outputItemSets False

removeAllMissingCols False

significanceLevel -1.0

treatZeroAsMissing False

upperBoundMinSupport 1.0

verbose False

Open...

Save...

OK

Cancel

[155] conv: (3.35)

[149] conv: (3.28)

,03] [150] conv: (3.27)

[159] conv: (3.26)

conv: (3.15)

(0.03) [151] conv: (3.06)

(0.03) [145] conv: (3.01)

[3] [156] conv: (3)

, (2.92)

Associator

Close April 8, 10 10:08:10 AM EDT 3/20/17

Start Stop

Results (Right-click for ...)

Times/ Apr 8

10:05:55 - April 8

Associator output

=====

Minimum support: 0.10 (699 instances)

Minimum metric: Count(Support) * 5

Number of cycles performed: 72

Generated sets of large itemsets:

Size of set of large itemsets L(1): 16

Size of set of large itemsets L(2): 380

Size of set of large itemsets L(3): 910

Size of set of large itemsets L(4): 600

Size of set of large itemsets L(5): 180

Size of set of large itemsets L(6): 1

Item rules found:

1. item1=item2 frozen breads= fruit= total=high 700 ==> bread and cake= 923 Count:(0.92)> lift:(11.20) lev:(0.03) [158] covr:(0.03)
2. baking needs= frozen food= fruit= total=high 760 ==> bread and cake= 886 Count:(0.82)> lift:(11.27) lev:(0.03) [149] covr:(3.28)
3. baking needs= frozen food= fruit= total=high 770 ==> bread and cake= 703 Count:(0.82)> lift:(11.27) lev:(0.03) [126] covr:(3.27)
4. item1=item2 frozen breads= total=high 875 ==> bread and cake= 740 Count:(0.92)> lift:(11.20) lev:(0.03) [158] covr:(0.03)
5. cake= bread food= fruit= total=high 825 ==> bread and cake= 779 Count:(0.81)> lift:(11.27) lev:(0.04) [164] covr:(3.15)
6. item1=item2 frozen breads= vegetable= total=high 920 ==> bread and cake= 628 Count:(0.67)> lift:(11.26) lev:(0.03) [151] covr:(0.04)
7. baking needs= item1=item2 vegetable= total=high 720 ==> bread and cake= 701 Count:(0.67)> lift:(11.26) lev:(0.03) [146] covr:(0.03)
8. baking needs= fruit= total=high 851 ==> bread and cake= 866 Count:(0.81)> lift:(11.26) lev:(0.04) [179] covr:(3)
9. frozen breads= fruit= vegetable= total=high 808 ==> bread and cake= 787 Count:(0.91)> lift:(11.26) lev:(0.03) [156] covr:(0)
10. frozen food= fruit= total=high 869 ==> bread and cake= 877 Count:(0.81)> lift:(11.26) lev:(0.04) [179] covr:(0.52)

6.FP GROWTH ALGORITHM USING WEKA.

Associator

Choose **FPGrowth** -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1

Start Stop

Result list (right-click for ...)

- 130437 - Apriori
- 130655 - Apriori
- 130840 - FPGrowth

Associator output

```

=== Run information ===

Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Relation:     vote
Instances:    435
Attributes:   17
    handicapped-infants
    water-project-cost-sharing
    adoption-of-the-budget-resolution
    physician-fee-freeze
    el-salvador-aid
    religious-groups-in-schools
    anti-satellite-test-ban
    aid-to-nicaraguan-contras
    mx-missile
    immigration
    synfuels-corporation-cutback
    education-spending
    superfund-right-to-sue
    crime
    duty-free-exports
    export-administration-act-south-africa
    Class

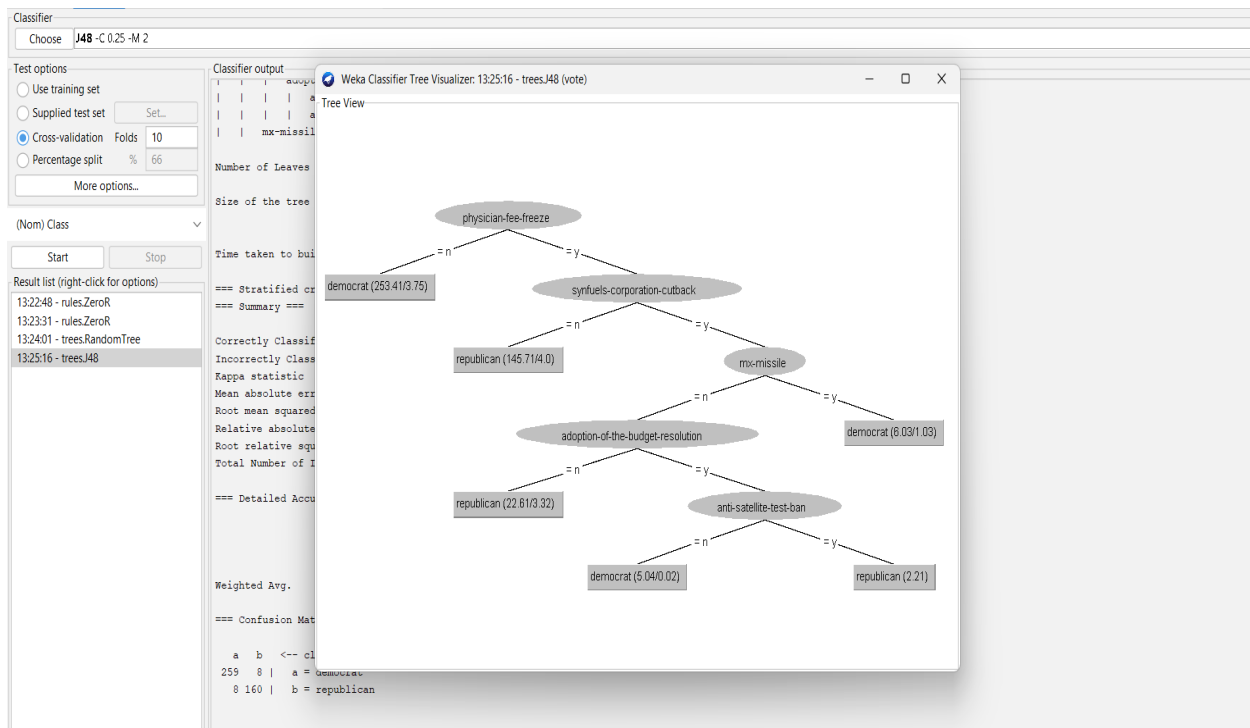
=== Associator model (full training set) ===

FPGrowth found 41 rules (displaying top 10)

1. [el-salvador-aid=y, Class=republican]: 157 ==> [physician-fee-freeze=y]: 156   <conf:(0.99)> lift:(2.44) lev:(0.21) conv:(46.56)
2. [crime=y, Class=republican]: 158 ==> [physician-fee-freeze=y]: 155   <conf:(0.98)> lift:(2.41) lev:(0.21) conv:(23.43)
3. [religious-groups-in-schools=y, physician-fee-freeze=y]: 160 ==> [el-salvador-aid=y]: 156   <conf:(0.97)> lift:(2) lev:(0.18) conv:(16.4)
4. [Class=republican]: 168 ==> [physician-fee-freeze=y]: 163   <conf:(0.97)> lift:(2.38) lev:(0.22) conv:(16.61)
5. [adoption-of-the-budget-resolution=y, anti-satellite-test-ban=y, mx-missile=y]: 161 ==> [aid-to-nicaraguan-contras=y]: 155   <conf:(0.96)> lift:(1.73) lev:(0.15) conv:(10.2)
6. [physician-fee-freeze=y, Class=republican]: 163 ==> [el-salvador-aid=y]: 156   <conf:(0.96)> lift:(1.96) lev:(0.18) conv:(10.45)
7. [religious-groups-in-schools=y, el-salvador-aid=y, superfund-right-to-sue=y]: 160 ==> [crime=y]: 153   <conf:(0.96)> lift:(1.68) lev:(0.14) conv:(8.6)
8. [el-salvador-aid=y, superfund-right-to-sue=y]: 170 ==> [crime=y]: 162   <conf:(0.95)> lift:(1.67) lev:(0.15) conv:(8.12)
9. [crime=y, physician-fee-freeze=y]: 168 ==> [el-salvador-aid=y]: 160   <conf:(0.95)> lift:(1.95) lev:(0.18) conv:(9.57)
10. [el-salvador-aid=y, physician-fee-freeze=y]: 168 ==> [crime=y]: 160   <conf:(0.95)> lift:(1.67) lev:(0.15) conv:(8.02)

```

7.PREDICTION OF CATEGORICAL DATA USING DECISION TREE ALGORITHM USING WEKA.



Classifier

Choose J48 -C 0.25 -M 2

Test options

☐ Use training set

☐ Supplied test set

Set...

☒ Cross-validation

Folds 10

☐ Percentage split

% 66

More options...

(Nom) crime

Start Stop

Result list (right-click for options)

13:22:48 - rules.ZeroR

13:23:31 - rules.ZeroR

13:24:01 - trees.RandomTree

13:25:16 - trees.J48

13:27:03 - trees.J48

Classifier output

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: vote

Instances: 435

Attributes: 17

handicapped-infants

water-project-cost-sharing

adoption-of-the-budget-resolution

physician-fee-freeze

el-salvador-aid

religious-groups-in-schools

anti-satellite-test-ban

aid-to-nicaraguan-contras

mx-missile

immigration

synfuels-corporation-cutback

education-spending

superfund-right-to-sue

crime

duty-free-exports

export-administration-act-south-africa

Class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

Class = democrat

| aid-to-nicaraguan-contras = n

| | el-salvador-aid = n: n (2.34/0.06)

| | el-salvador-aid = y: y (43.01/4.9)

| aid-to-nicaraguan-contras = y

| | religious-groups-in-schools = n: n (133.15/14.63)

| | religious-groups-in-schools = y

Classifier

ChooseJ48 - C 0.25 - M 2

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation

☐ Percentage split

Set...

Folds10

%66

More options...

(Nom) crime

StartStop

Result list (right-click for options)

132248 - rules.ZeroR

132331 - rules.ZeroR

132401 - trees.RandomTree

132516 - trees.J48

132703 - trees.J48

Classifier output

| | el-salvador-aid = y: y (43.01/4.9)

| aid-to-nicaraguan-contras = y

| | religious-groups-in-schools = n: n (133.15/14.63)

| | religious-groups-in-schools = y

| | adoption-of-the-budget-resolution = n: y (5.13/1.07)

| | adoption-of-the-budget-resolution = y

| | | duty-free-exports = n

| | | superfund-right-to-sue = n

| | | | water-project-cost-sharing = n

| | | | synfuels-corporation-ouback = n: y (6.45/1.0)

| | | | synfuels-corporation-ouback = y: n (2.97/0.45)

| | | | water-project-cost-sharing = y

| | | | handicapped-infants = n: n (3.44)

| | | | handicapped-infants = y

| | | | immigration = n: y (2.0)

| | | | immigration = y: n (2.0)

| | | | superfund-right-to-sue = y: y (16.15/4.94)

| | | duty-free-exports = y

| | | education-spending = n

| | | | superfund-right-to-sue = n: n (23.35/5.35)

| | | | superfund-right-to-sue = y

| | | | immigration = n: n (4.55)

| | | | immigration = y: y (7.0/2.33)

| | | | education-spending = y: y (5.46/1.46)

Class = republican: y (161.0/3.0)

Number of Leaves : 15

Size of the tree : 29

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances35484.689 %

8. PREDICTION OF CATEGORICAL DATA USING SMO ALGORITHM USING WEKA.

Classifier

Choose **SMO** -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) class

Result list (right-click for options)

134427 - trees.J48

134646 - functions.SMO

Classifier output

SMO

Kernel used:

Linear Kernel: $K(x,y) = \langle x,y \rangle$

Classifier for classes: bad, good

BinarySMO

Machine linear: showing attribute weights, not support vectors.

```
0.0754 * (normalized) duration
+ 0.7894 * (normalized) wage-increase-first-year
+ 0.8109 * (normalized) wage-increase-second-year
+ 0.339 * (normalized) wage-increase-third-year
+ -0.0216 * (normalized) cost-of-living-adjustment=none
+ 0.2843 * (normalized) cost-of-living-adjustment=rtcf
+ -0.2628 * (normalized) cost-of-living-adjustment=tc
+ -0.5644 * (normalized) working-hours
+ -0.8 * (normalized) pension=none
+ 0.2033 * (normalized) pension=Pret_allw
+ 0.5968 * (normalized) pension=empl_contr
+ 0.3396 * (normalized) standby-pay
+ -0.0055 * (normalized) shift-differential
+ -0.5502 * (normalized) education-allowance=no
+ 0.6464 * (normalized) statutory-holidays
+ -0.2443 * (normalized) vacation=below_average
+ -0.0503 * (normalized) vacation=average
+ 0.2946 * (normalized) vacation=generous
+ -1.2183 * (normalized) longterm-disability-assistance=no
+ -0.2628 * (normalized) contribution-to-dental-plan=none
+ -0.0485 * (normalized) contribution-to-dental-plan=half
+ 0.3113 * (normalized) contribution-to-dental-plan=full
+ -0.6222 * (normalized) contribution-to-health-plan=none
+ 0.2698 * (normalized) contribution-to-health-plan=half
+ 0.3534 * (normalized) contribution-to-health-plan=full
- 0.2873
```

Classifier

Choose SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds 10
☐ Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

13:44:27 - treesJ48
13:46:46 - functions.SMO
13:47:49 - functions.SMO

Classifier output

```

+ -0.6222 * (normalized) contribution-to-health-plan=none
+ 0.2688 * (normalized) contribution-to-health-plan=half
+ 0.3534 * (normalized) contribution-to-health-plan=full
- 0.2873

Number of kernel evaluations: 1055 (93.756% cached)

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      51           89.4737 %
Incorrectly Classified Instances     6           10.5263 %
Kappa statistic                     0.7635
Mean absolute error                  0.1053
Root mean squared error              0.3244
Relative absolute error              23.0111 %
Root relative squared error          67.9505 %
Total Number of Instances           57

=== Detailed Accuracy By Class ===

              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              0.800    0.054    0.889     0.800    0.842     0.766    0.873    0.781    bad
              0.946    0.200    0.897     0.946    0.921     0.766    0.873    0.884    good
Weighted Avg.   0.895    0.149    0.894     0.895    0.893     0.766    0.873    0.848

=== Confusion Matrix ===

  a  b  <-- classified as
16  4  |  a = bad
 2 35  |  b = good

```

9.EVALUATING THE ACCURACY OF THE CLASSIFIERS USING WEKA LOGISTIC:

Classifier

Choose **Logistic -R 1.0E-8 -M -1 -num-decimal-places 4**

Test options

☐ Use training set

☐ Supplied test set

Set...

☒ Cross-validation

Folds 10

☐ Percentage split

% 66

More options...

(Nom) contact-lenses

▼

Start

Stop

Result list (right-click for options)

13:44:27 - trees.J48

13:46:46 - functions.SMO

13:47:49 - functions.SMO

13:51:25 - functions.Logistic

Classifier output

Logistic Regression with ridge parameter of 1.0E-8

Coefficients...

Variable	Class soft	hard
age=young	9.3413	20.7606
age=pre-presbyopic	16.3153	-10.1411
age=presbyopic	-25.6566	-10.6195
spectacle-prescrip=hypermetrope	44.5377	-36.4096
astigmatism=yes	-87.2302	73.4651
tear-prod-rate=normal	77.4902	64.4602
Intercept	-75.3586	-108.1806

Odds Ratios...

Variable	Class soft	hard
age=young	11399.0493	1038062286.0301
age=pre-presbyopic	12179623.85	0
age=presbyopic	0	0
spectacle-prescrip=hypermetrope	2.200240946477201E19	0
astigmatism=yes	0	8.044615217744127E31
tear-prod-rate=normal	4.50352724620167E33	9.879004839403287E27

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	16	66.6667 %
Incorrectly Classified Instances	8	33.3333 %
Kappa statistic	0.3826	
Mean absolute error	0.2241	
Root mean squared error	0.4568	
Relative absolute error	59.3095 %	

Classifier

Choose **Logistic** -R 1.0E-8 -M -1 -num-decimal-places 4

Test options

☐ Use training set
 ☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

More options...

(Nom) contact-lenses

Start

Stop

Result list (right-click for options)

13:44:27 - trees.J48

13:46:46 - functions.SMO

13:47:49 - functions.SMO

13:51:25 - functions.Logistic

Classifier output

```

age=young          1155570755  1000000000000
age=pre-presbyopic 12179623.85      0
age=presbyopic     0              0
spectacle=prescrip=hypermetropes 2.200240946477201E19 0
astigmatism=yes    0              8.044615217744127E31
tear-prod-rate=normal 4.50352724620167E33 9.879004839403287E27

```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances

16

66.6667 %

Incorrectly Classified Instances

8

33.3333 %

Kappa statistic

0.3826

Mean absolute error

0.2241

Root mean squared error

0.4568

Relative absolute error

59.3095 %

Root relative squared error

104.5911 %

Total Number of Instances

24

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.400	0.105	0.500	0.400	0.444	0.321	0.505	0.753	soft
	0.750	0.100	0.600	0.750	0.667	0.596	0.888	0.636	hard
	0.733	0.444	0.733	0.733	0.733	0.289	0.711	0.778	none
Weighted Avg.	0.667	0.316	0.663	0.667	0.662	0.347	0.781	0.749	

=== Confusion Matrix ===

a b c

-- classified as

2 0 3 | a = soft

0 3 1 | b = hard

2 2 11 | c = none

NAÏVE BAYES ALGORITHM:

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set
 ☐ Supplied test set Set...
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

More options...

(Nom) contact-lenses

Start

Stop

Result list (right-click for options)

13:44:27 - trees.J48

13:46:46 - functions.SMO

13:47:49 - functions.SMO

13:51:25 - functions.Logistic

13:54:10 - bayes.NaiveBayes

Classifier output

Naive Bayes Classifier

Attribute	Class	soft	hard	none
		(0.22)	(0.19)	(0.59)

age

young

3.0

3.0

5.0

pre-presbyopic

3.0

2.0

6.0

presbyopic

2.0

2.0

7.0

[total]

8.0

7.0

18.0

spectacle-prescrip

myope

3.0

4.0

8.0

hypermetropes

4.0

2.0

9.0

[total]

7.0

6.0

17.0

astigmatism

no

6.0

1.0

8.0

yes

1.0

5.0

9.0

[total]

7.0

6.0

17.0

tear-prod-rate

reduced

1.0

1.0

13.0

normal

6.0

5.0

4.0

[total]

7.0

6.0

17.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances

17

70.8333 %

Incorrectly Classified Instances

7

29.1667 %

Kappa statistic

0.4381

Classifier

Choose NaiveBayes

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) contact-lenses

Start Stop

Result list (right-click for options)

13:44:27 - trees.J48
13:46:46 - functions.SMO
13:47:49 - functions.SMO
13:51:25 - functions.Logistic
13:54:10 - bayes.NaiveBayes

Classifier output

tear-prod-rate
reduced 1.0 1.0 13.0
normal 6.0 5.0 4.0
[total] 7.0 6.0 17.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 17 70.8333 %
Incorrectly Classified Instances 7 29.1667 %
Kappa statistic 0.4381
Mean absolute error 0.2545
Root mean squared error 0.3326
Relative absolute error 67.3578 %
Root relative squared error 76.1544 %
Total Number of Instances 24

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.800	0.053	0.800	0.800	0.800	0.747	0.947	0.710	soft
	0.250	0.100	0.333	0.250	0.286	0.169	0.925	0.692	hard
	0.800	0.444	0.750	0.800	0.774	0.365	0.830	0.930	none
Weighted Avg.	0.708	0.305	0.691	0.708	0.698	0.412	0.870	0.844	

=== Confusion Matrix ===

a b c <-- classified as
4 0 1 | a = soft
0 1 3 | b = hard
1 2 12 | c = none

J48 ALGORITHM:

Classifier

Choose J48 C4.5 Rules

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) contact-lenses

Start Stop

Result list (right-click for options)

13:44:27 - trees.J48
13:46:46 - functions.SMO
13:47:49 - functions.SMO
13:51:25 - functions.Logistic
13:54:10 - bayes.NaiveBayes
13:55:55 - trees.J48

Classifier output

=== Stratified cross-validation ===

J48 pruned tree

tear-prod-rate = reduced: none (13.0)
Contact-lenses = normal
| soft-prod-rate = not soft (6.0/1.0)
| | soft-prod-rate = yes
| | | soft-prod-rate = yes: hard (3.0)
| | | soft-prod-rate = hyperbolic: none (0.0/1.0)

Number of nodes : 4

Size of the tree : 7

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 20 83.3333 %
Incorrectly Classified Instances 4 16.6667 %
Kappa statistic 0.71
Mean absolute error 0.15
Root mean squared error 0.3849
Relative absolute error 39.7028 %
Root relative squared error 71.3889 %
Total Number of Instances 24

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	0.989	0.947	0.933	soft
	0.750	0.100	0.800	0.750	0.767	0.596	0.813	0.752	hard
	0.800	0.111	0.823	0.800	0.827	0.688	0.811	0.840	none
Weighted Avg.	0.833	0.092	0.857	0.833	0.835	0.703	0.910	0.913	

Classifier

Choose **IBk - K 1 - W 0 - A "\weka.core.neighboursearch.LinearNNSearch - A "\weka.core.EuclideanDistance - R first-last"**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds: 10

☐ Percentage split %: 66

More options...

(Nom) contact-lenses

Start Stop

Result list (right-click for options)

134427 - trees.J48

134646 - functions.SMO

134749 - functions.SMO

135125 - functions.Logistic

135410 - bayes.NaiveBayes

135555 - trees.J48

135651 - lazy.IBk

Classifier output

==== Classifier model (full training set) ====

IB1 instance-based classifier
using 1 nearest neighbour(s) for classification

Time taken to build model: 0 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	19	79.1667 %
Incorrectly Classified Instances	5	20.8333 %
Kappa statistic	0.6262	
Mean absolute error	0.2262	
Root mean squared error	0.3165	
Relative absolute error	59.8856 %	
Root relative squared error	72.4707 %	
Total Number of Instances	24	

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.800	0.053	0.800	0.800	0.800	0.747	0.958	0.833		soft
0.750	0.100	0.600	0.750	0.667	0.596	0.925	0.750		hard
0.800	0.222	0.857	0.800	0.828	0.567	0.896	0.953		none
Weighted Avg.	0.752	0.167	0.802	0.752	0.795	0.610	0.914	0.894	

==== Confusion Matrix ====

```

a b c <-- classified as
4 0 1 | a = soft
0 3 1 | b = hard
1 2 12 | c = none

```

10. PREDICTION OF CATEGORICAL DATA USING BAYESIAN ALGORITHM USING WEKA.

Classifier

Choose **NaiveBayes**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds: 10

☐ Percentage split %: 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

134200 - bayes.NaiveBayes

134201 - bayes.NaiveBayes

Classifier output

Naive Bayes Classifier

Attribute	Class	
	bad	good
	(0.36)	(0.64)
distribution		
age	2	2.25
std. dev.	0.7071	0.6031
weight sum	20	36
prediction	1	1
Naive-Bayes-First-Year		
mean	2.6667	4.0000
std. dev.	0.6643	1.7713
weight sum	20	36
prediction	0.0128	0.0128
Naive-Bayes-Second-Year		
mean	2.9634	4.4444
std. dev.	0.6183	0.9809
weight sum	14	31
prediction	0.0541	0.0541
Naive-Bayes-Third-Year		
age	2.0314	4.0789
std. dev.	0.1478	0.7882
weight sum	4	11
prediction	0.3875	0.3875
Cost of living adjustment		
home	10.0	14.0
rent	2.0	8.0
lv	4.0	3.0
[total]	16.0	25.0
working=house		

Classifier	
Choose	NaiveBayes
Test options	
<input type="radio"/> Use training set <input type="radio"/> Supplied test set Set... <input checked="" type="radio"/> Cross-validation Folds: 10 <input type="radio"/> Percentage split %: 66 More options...	
(Nom) class	
Start Stop Result list (right-click for options) 144204 - BayesNaiveBayes 111221 - BayesNaiveBayes	
Classifier output	
NaiveBayes Mean 19.41617 19.4493 Std. dev. 1.11900 1.19366 N of valid cases 15 32 Precision 1.00000 1.00000 Confusion none 12.0 1.0 half 3.0 3.0 full 6.0 8.0 [total] 21.0 12.0 Standing pay none 2.0 11.0 std. dev. 0.866 2.0396 weight sum 4 5 precision 2 2 Active parameters none 2.1691 5.6818 Std. dev. 1.27208 2.18154 N of valid cases 5 22 precision 2.7778 2.7778 Education-allowance yes 4.0 8.0 no 10.0 4.0 [total] 14.0 12.0 Statutory-holidays none 10.0 11.1802 Std. dev. 0.4046 1.20204 N of valid cases 20 33 precision 1.0 1.0	

Classifier	
Choose	NaiveBayes
Test options	
<input type="radio"/> Use training set <input type="radio"/> Supplied test set Set... <input checked="" type="radio"/> Cross-validation Folds: 10 <input type="radio"/> Percentage split %: 66 More options...	
(Nom) class	
Start Stop Result list (right-click for options) 141203 - BayesNaiveBayes 144204 - BayesNaiveBayes	
Classifier output	
NaiveBayes Mean average 10.0 11.0 accuracy 8.0 11.0 precision 10.0 15.0 [total] 20.0 26.0 Long-term disability assistance yes 6.0 16.0 no 9.0 1.0 [total] 15.0 17.0 Contribution to dental plan none 0.0 0.0 half 8.0 9.0 full 1.0 14.0 [total] 17.0 23.0 Maternity leave assistance yes 10.0 19.0 no 4.0 1.0 [total] 14.0 20.0 Contribution to health plan none 9.0 1.0 half 3.0 8.0 full 1.0 15.0 [total] 13.0 24.0 Time taken to build model: 0 seconds --- Classified cases-validation --- --- Summary --- Correctly classified instances 31 89.4012 % Incorrectly classified instances 4 10.5988 % Error statistic 0.7761	

Classifier

Choose NaiveBayes

Test options

☐ Use training set
 ☐ Supplied test set

Set...

☒ Cross-validation

Folds 10

☐ Percentage split

% 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

14:42:03 - bayes.NaiveBayes

14:43:31 - bayes.NaiveBayes

Classifier output

```

[total] 19.0 24.0

contribution-to-health-plan
none 9.0 1.0
half 3.0 8.0
full 7.0 15.0
[total] 19.0 24.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 51 89.4737 %
Incorrectly Classified Instances 6 10.5263 %
Kappa statistic 0.7741
Mean absolute error 0.1042
Root mean squared error 0.2637
Relative absolute error 22.7763 %
Root relative squared error 55.2266 %
Total Number of Instances 57

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0.900 0.108 0.818 0.900 0.857 0.776 0.965 0.926 bad
0.892 0.100 0.943 0.892 0.917 0.776 0.965 0.983 good
Weighted Avg. 0.895 0.103 0.899 0.895 0.896 0.776 0.965 0.963

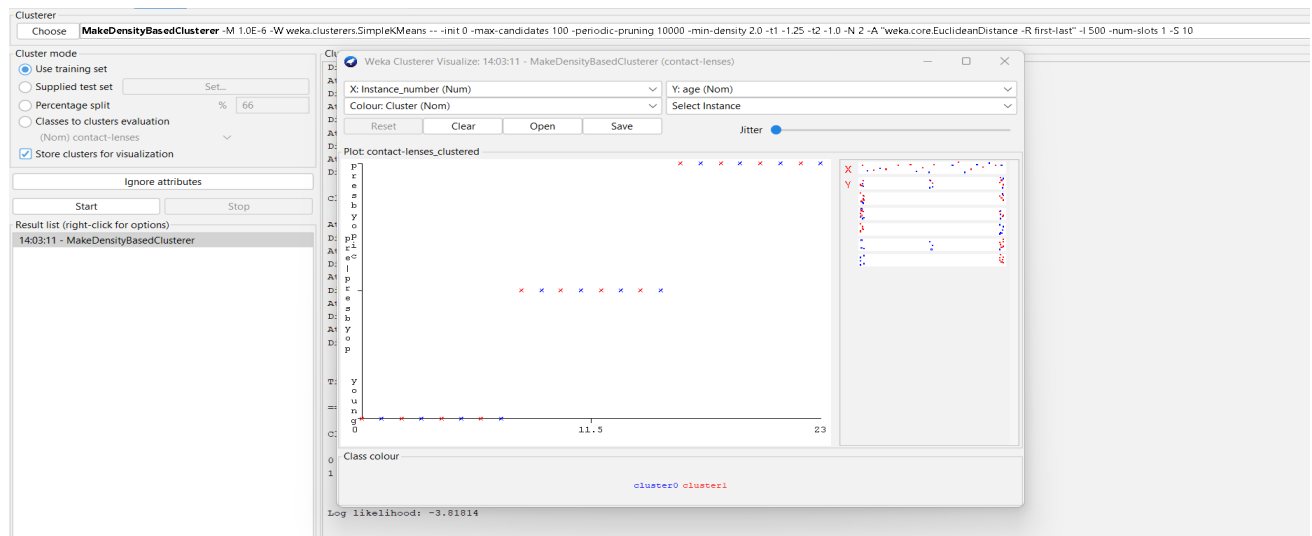
=== Confusion Matrix ===

 a b <-- classified as
18 2 | a = bad
 4 33 | b = good

```

11.DATA ANALYSIS BY DENSITY BASED CLUSTERING ALGORITHM USING WEKA.

[illegible]



12.GIVING THE FOLLOWING DATABASE WITH 5 TRANSACTIONS AND A MINIMUM SUPPORT THRESHOLD OF 60% AND A MINIMUM CONFIDENCE THRESHOLD OF 80%, FIND ALL FREQUENT ITEMSETS USING (A) APRIORI AND (B) FP-GROWTH.

TID	Transaction
T1	{A, B, C, D, E, F}
T2	{B, C, D, E, F, G}
T3	{A, D, E, H}
T4	{A, D, F, I, J}
T5	{B, D, E, K}

OUTPUT:

Associator output

```

=== Run information ===

Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    EXP-20
Instances:   5
Attributes:  2
             TID
             TRANSACTION

=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.3 (1 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Size of set of large itemsets L(2): 5

Best rules found:

1. TRANSACTION=A B C D E F 1 ==> TID=T1 1    <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
2. TID=T1 1 ==> TRANSACTION=A B C D E F 1    <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
3. TRANSACTION=B C D E F G 1 ==> TID=T2 1    <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
4. TID=T2 1 ==> TRANSACTION=B C D E F G 1    <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
5. TRANSACTION=A D E H 1 ==> TID=T3 1        <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
6. TID=T3 1 ==> TRANSACTION=A D E H 1        <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
7. TRANSACTION=A D F I J 1 ==> TID=T4 1      <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
8. TID=T4 1 ==> TRANSACTION=A D F I J 1      <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
9. TRANSACTION=B D E K 1 ==> TID=T5 1        <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)
10. TID=T5 1 ==> TRANSACTION=B D E K 1       <conf:(1)> lift:(5) lev:(0.16) [0] conv:(0.8)

```

13.THE 'DATABASE' BELOW HAS NINE TRANSACTIONS. WHAT ASSOCIATION RULES CAN BE FOUND IN THIS SET, IF THE MINIMUM SUPPORT (I.E COVERAGE) IS 60% AND THE MINIMUM CONFIDENCE (I.E. ACCURACY) IS 80% ?

TRANS_ID ITEMLIST

TID	List of Items
T100	I1, I2, I5
T100	I2, I4
T100	I2, I3
T100	I1, I2, I4
T100	I1, I3
T100	I2, I3
T100	I1, I3
T100	I1, I2, I3, I5
T100	I1, I2, I3

OUTPUT:

Associator output

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: EX21
Instances: 9
Attributes: 2
TID
LIST OF ITEMS

=== Associator model (full training set) ===

Apriori
=====

Minimum support: 0.11 (1 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8

Size of set of large itemsets L(2): 7

Best rules found:

1. LIST OF ITEMS=I2 I3 2 ==> TID=T100 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. LIST OF ITEMS=I1 I3 2 ==> TID=T100 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. LIST OF ITEMS=I1 I2 I5 1 ==> TID=T100 1 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. LIST OF ITEMS=I2 I4 1 ==> TID=T100 1 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. LIST OF ITEMS=I1 I2 I4 1 ==> TID=T100 1 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. LIST OF ITEMS=I1 I2 I3 I5 1 ==> TID=T100 1 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. LIST OF ITEMS=I1 I2 I3 1 ==> TID=T100 1 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

