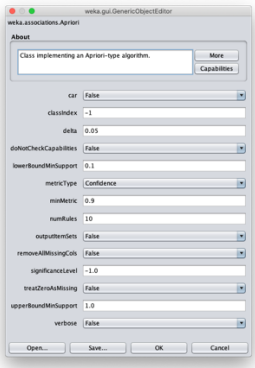
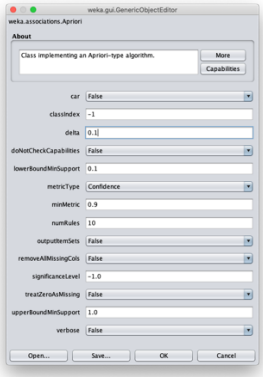
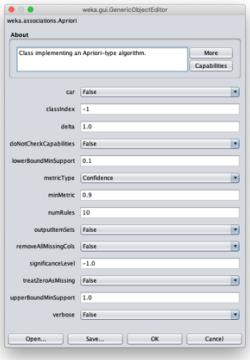
Man Tik Li (ml3546)
INFO 371 – Data Mining Applications
Assignment 3

Task 1

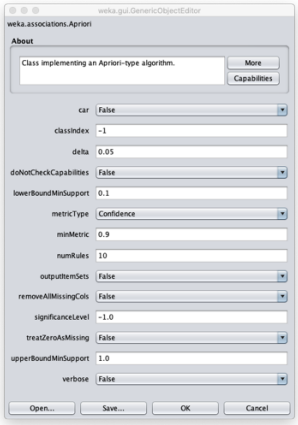
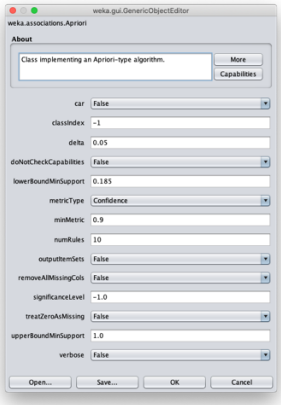
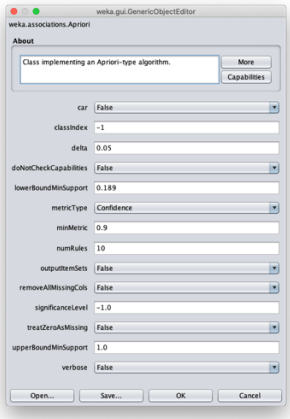
delta

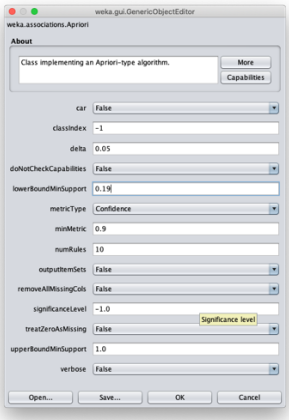
No.	Parameter	Associator Output
1		=== Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === Apriori ======= Minimum support: 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)
2		=== Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.1 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === Apriori ======= Minimum support: 0.2 (925 instances) Minimum metric <confidence>: 0.9 Number of cycles performed: 9 Generated sets of large itemsets: Size of set of large itemsets L(1): 52 Size of set of large itemsets L(2): 634 Size of set of large itemsets L(3): 2598 Size of set of large itemsets L(4): 3950 Size of set of large itemsets L(5): 2470 Size of set of large itemsets L(6): 558 Size of set of large itemsets L(7): 20 Best rules found: 1. biscuits=t frozen foods=t party snack foods=t fruit=t vegetables=t total=high 510 ==> bread and cake=t 478 <conf:(0.94)> lift:(1.3) lev:(0.02) [110] conv:(4.33) 2. biscuits=t frozen foods=t cheese=t fruit=t total=high 495 ==> bread and cake=t 463 <conf:(0.94)> lift:(1.3) lev:(0.02) [106] conv:(4.2) 3. biscuits=t cheese=t fruit=t vegetables=t total=high 513 ==> bread and cake=t 479 <conf:(0.93)> lift:(1.3) lev:(0.02) [109] conv:(4.11) 4. baking needs=t biscuits=t party snack foods=t fruit=t total=high 557 ==> bread and cake=t 520 <conf:(0.93)> lift:(1.3) lev:(0.03) [119] conv:(4.11) 5. baking needs=t cheese=t fruit=t vegetables=t total=high 519 ==> bread and cake=t 483 <conf:(0.93)> lift:(1.29) lev:(0.02) [109] conv:(3.93) 6. frozen foods=t party snack foods=t tissues-paper prd=t fruit=t total=high 518 ==> bread and cake=t 482 <conf:(0.93)> lift:(1.29) lev:(0.02) [109] conv:(3.92) 7. juice-sat-cord-ms=t biscuits=t party snack foods=t fruit=t total=high 529 ==> bread and cake=t 492 <conf:(0.93)> lift:(1.29) lev:(0.02) [111] conv:(3.9) 8. biscuits=t cheese=t fruit=t total=high 584 ==> bread and cake=t 543 <conf:(0.93)> lift:(1.29) lev:(0.03) [122] conv:(3.9) 9. biscuits=t party snack foods=t fruit=t vegetables=t total=high 596 ==> bread and cake=t 554 <conf:(0.93)> lift:(1.29) lev:(0.03) [125] conv:(3.89) 10. baking needs=t biscuits=t frozen foods=t fruit=t vegetables=t total=high 561 ==> bread and cake=t 521 <conf:(0.93)> lift:(1.29) lev:(0.03) [117] conv:(3.84)
3		=== Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 1.0 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === Apriori ======= Minimum support: 0.1 (463 instances) Minimum metric <confidence>: 0.9 Number of cycles performed: 1 Generated sets of large itemsets: Size of set of large itemsets L(1): 52 Size of set of large itemsets L(2): 634 Size of set of large itemsets L(3): 2598 Size of set of large itemsets L(4): 3950 Size of set of large itemsets L(5): 2470 Size of set of large itemsets L(6): 558 Size of set of large itemsets L(7): 20 Best rules found: 1. biscuits=t frozen foods=t party snack foods=t fruit=t vegetables=t total=high 510 ==> bread and cake=t 478 <conf:(0.94)> lift:(1.3) lev:(0.02) [110] conv:(4.33) 2. biscuits=t frozen foods=t cheese=t fruit=t total=high 495 ==> bread and cake=t 463 <conf:(0.94)> lift:(1.3) lev:(0.02) [106] conv:(4.2) 3. biscuits=t cheese=t fruit=t vegetables=t total=high 513 ==> bread and cake=t 479 <conf:(0.93)> lift:(1.3) lev:(0.02) [109] conv:(4.11) 4. baking needs=t biscuits=t party snack foods=t fruit=t total=high 557 ==> bread and cake=t 520 <conf:(0.93)> lift:(1.3) lev:(0.03) [119] conv:(4.11) 5. baking needs=t cheese=t fruit=t vegetables=t total=high 519 ==> bread and cake=t 483 <conf:(0.93)> lift:(1.29) lev:(0.02) [109] conv:(3.93) 6. frozen foods=t party snack foods=t tissues-paper prd=t fruit=t total=high 518 ==> bread and cake=t 482 <conf:(0.93)> lift:(1.29) lev:(0.02) [109] conv:(3.92) 7. juice-sat-cord-ms=t biscuits=t party snack foods=t fruit=t total=high 529 ==> bread and cake=t 492 <conf:(0.93)> lift:(1.29) lev:(0.02) [111] conv:(3.9) 8. biscuits=t cheese=t fruit=t total=high 584 ==> bread and cake=t 543 <conf:(0.93)> lift:(1.29) lev:(0.03) [122] conv:(3.9) 9. biscuits=t party snack foods=t fruit=t vegetables=t total=high 596 ==> bread and cake=t 554 <conf:(0.93)> lift:(1.29) lev:(0.03) [125] conv:(3.89) 10. baking needs=t biscuits=t frozen foods=t fruit=t vegetables=t total=high 561 ==> bread and cake=t 521 <conf:(0.93)> lift:(1.29) lev:(0.03) [117] conv:(3.84)

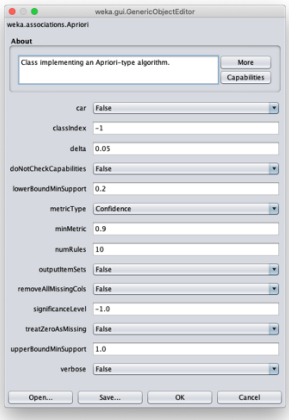
Analysis

The delta parameter does not affect the number of rule.

lowerBoundMinSupport

No.	Parameter	Associator Output
1		=== 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: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === 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)
2		=== Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.185 -S -1.0 -c -1 Relation: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === Apriori ======= Minimum support: 0.18 (856 instances) Minimum metric <confidence>: 0.9 Number of cycles performed: 17 Generated sets of large itemsets: Size of set of large itemsets L(1): 39 Size of set of large itemsets L(2): 266 Size of set of large itemsets L(3): 419 Size of set of large itemsets L(4): 149 Size of set of large itemsets L(5): 6 Best rules found: 1. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(3) 2. 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) 3. biscuits=t vegetables=t total=high 950 ==> bread and cake=t 858 <conf:(0.9)> lift:(1.25) lev:(0.04) [174] conv:(2.86) 4. baking needs=t fruit=t total=high 963 ==> bread and cake=t 869 <conf:(0.9)> lift:(1.25) lev:(0.04) [175] conv:(2.84)
3		=== Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.189 -S -1.0 -c -1 Relation: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === Apriori ======= Minimum support: 0.19 (875 instances) Minimum metric <confidence>: 0.9 Number of cycles performed: 17 Generated sets of large itemsets: Size of set of large itemsets L(1): 39 Size of set of large itemsets L(2): 255 Size of set of large itemsets L(3): 393 Size of set of large itemsets L(4): 129 Size of set of large itemsets L(5): 6 Best rules found: 1. 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)

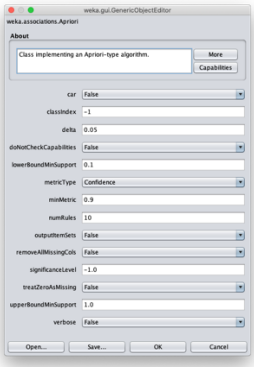
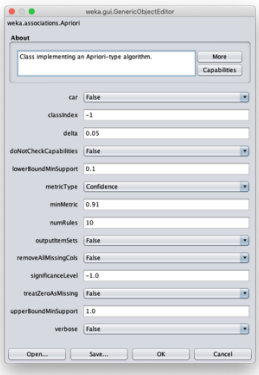
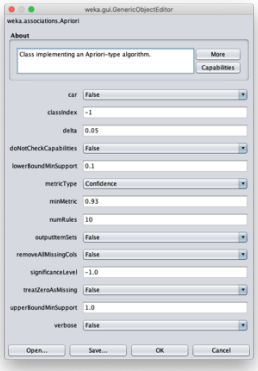
4		`=== Run information ===`
		
		`Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.19 -S -1.0 -c -1`
		`Relation: supermarket`
		`Instances: 4627`
		`Attributes: 217`
		`[list of attributes omitted]`
		`=== Associator model (full training set) ===`
		
		`Apriori`
		`=======`
		
		`Minimum support: 0.19 (879 instances)`
		`Minimum metric <confidence>: 0.9`
		`Number of cycles performed: 17`
		
		`Generated sets of large itemsets:`
		
		`Size of set of large itemsets L(1): 39`
		
		`Size of set of large itemsets L(2): 253`
		
		`Size of set of large itemsets L(3): 387`
		
		`Size of set of large itemsets L(4): 126`
		
		`Size of set of large itemsets L(5): 6`
		
		`Best rules found:`

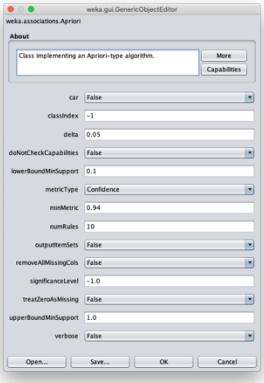
5		`=== Run information ===`
		
		`Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1`
		`Relation: supermarket`
		`Instances: 4627`
		`Attributes: 217`
		`[list of attributes omitted]`
		`=== Associator model (full training set) ===`
		
		`Apriori`
		`=======`
		
		`Minimum support: 0.2 (925 instances)`
		`Minimum metric <confidence>: 0.9`
		`Number of cycles performed: 16`
		
		`Generated sets of large itemsets:`
		
		`Size of set of large itemsets L(1): 38`
		
		`Size of set of large itemsets L(2): 225`
		
		`Size of set of large itemsets L(3): 302`
		
		`Size of set of large itemsets L(4): 80`
		
		`Size of set of large itemsets L(5): 2`
		
		`Best rules found:`

Analysis

The algorithm stops when lowerBoundMinSupport is reached. I tested the lowerBoundMinSupport with 5 different parameters. At the default value of 0.1, the 10 best rules were generated with the bin numbers. The best rules decreasing when the value of lowerBoundMinSupport is increasing. The lowerBoundMinSupport only generated one of best rules when it set to 0.189 and it stop generated at 0.19.

minMetricType

No.	Parameter	Associator Output
1		=== 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: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === 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)
2		=== Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.91 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === Apriori ======= Minimum support: 0.1 (463 instances) Minimum metric <confidence>: 0.91 Number of cycles performed: 18 Generated sets of large itemsets: Size of set of large itemsets L(1): 52 Size of set of large itemsets L(2): 634 Size of set of large itemsets L(3): 2598 Size of set of large itemsets L(4): 3950 Size of set of large itemsets L(5): 2470 Size of set of large itemsets L(6): 558 Size of set of large itemsets L(7): 20 Best rules found: 1. biscuits=t frozen foods=t party snack foods=t fruit=t vegetables=t total=high 510 ==> bread and cake=t 478 <conf:(0.94)> lift:(1.3) lev:(0.02) [110] conv:(4.33) 2. biscuits=t frozen foods=t cheese=t fruit=t total=high 495 ==> bread and cake=t 463 <conf:(0.94)> lift:(1.3) lev:(0.02) [106] conv:(4.2) 3. biscuits=t cheese=t fruit=t vegetables=t total=high 513 ==> bread and cake=t 479 <conf:(0.93)> lift:(1.3) lev:(0.02) [109] conv:(4.11) 4. baking needs=t biscuits=t party snack foods=t fruit=t total=high 557 ==> bread and cake=t 520 <conf:(0.93)> lift:(1.3) lev:(0.03) [119] conv:(4.11) 5. baking needs=t cheese=t fruit=t vegetables=t total=high 519 ==> bread and cake=t 483 <conf:(0.93)> lift:(1.29) lev:(0.02) [109] conv:(3.93) 6. frozen foods=t party snack foods=t tissues-paper prd=t fruit=t total=high 518 ==> bread and cake=t 482 <conf:(0.93)> lift:(1.29) lev:(0.02) [109] conv:(3.92) 7. juice-sat-cord-ms=t biscuits=t party snack foods=t fruit=t total=high 529 ==> bread and cake=t 492 <conf:(0.93)> lift:(1.29) lev:(0.02) [111] conv:(3.9) 8. biscuits=t cheese=t fruit=t total=high 584 ==> bread and cake=t 543 <conf:(0.93)> lift:(1.29) lev:(0.03) [122] conv:(3.9) 9. biscuits=t party snack foods=t fruit=t vegetables=t total=high 596 ==> bread and cake=t 554 <conf:(0.93)> lift:(1.29) lev:(0.03) [125] conv:(3.89) 10. baking needs=t biscuits=t frozen foods=t fruit=t vegetables=t total=high 561 ==> bread and cake=t 521 <conf:(0.93)> lift:(1.29) lev:(0.03) [117] conv:(3.84)
3		=== Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.93 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === Apriori ======= Minimum support: 0.1 (463 instances) Minimum metric <confidence>: 0.93 Number of cycles performed: 18 Generated sets of large itemsets: Size of set of large itemsets L(1): 52 Size of set of large itemsets L(2): 634 Size of set of large itemsets L(3): 2598 Size of set of large itemsets L(4): 3950 Size of set of large itemsets L(5): 2470 Size of set of large itemsets L(6): 558 Size of set of large itemsets L(7): 20 Best rules found: 1. biscuits=t frozen foods=t party snack foods=t fruit=t vegetables=t total=high 510 ==> bread and cake=t 478 <conf:(0.94)> lift:(1.3) lev:(0.02) [110] conv:(4.33) 2. biscuits=t frozen foods=t cheese=t fruit=t total=high 495 ==> bread and cake=t 463 <conf:(0.94)> lift:(1.3) lev:(0.02) [106] conv:(4.2) 3. biscuits=t cheese=t fruit=t vegetables=t total=high 513 ==> bread and cake=t 479 <conf:(0.93)> lift:(1.3) lev:(0.02) [109] conv:(4.11) 4. baking needs=t biscuits=t party snack foods=t fruit=t total=high 557 ==> bread and cake=t 520 <conf:(0.93)> lift:(1.3) lev:(0.03) [119] conv:(4.11) 5. baking needs=t cheese=t fruit=t vegetables=t total=high 519 ==> bread and cake=t 483 <conf:(0.93)> lift:(1.29) lev:(0.02) [109] conv:(3.93) 6. frozen foods=t party snack foods=t tissues-paper prd=t fruit=t total=high 518 ==> bread and cake=t 482 <conf:(0.93)> lift:(1.29) lev:(0.02) [109] conv:(3.92) 7. juice-sat-cord-ms=t biscuits=t party snack foods=t fruit=t total=high 529 ==> bread and cake=t 492 <conf:(0.93)> lift:(1.29) lev:(0.02) [111] conv:(3.9)

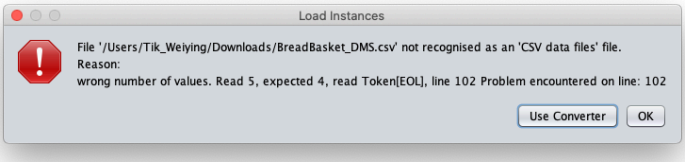
4		=== Run information === Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.94 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1 Relation: supermarket Instances: 4627 Attributes: 217 [list of attributes omitted] === Associator model (full training set) === Apriori ======= Minimum support: 0.1 (463 instances) Minimum metric <confidence>: 0.94 Number of cycles performed: 18 Generated sets of large itemsets: Size of set of large itemsets L(1): 52 Size of set of large itemsets L(2): 634 Size of set of large itemsets L(3): 2598 Size of set of large itemsets L(4): 3950 Size of set of large itemsets L(5): 2470 Size of set of large itemsets L(6): 558 Size of set of large itemsets L(7): 20 Best rules found:

Analysis

The value of minMetric found 10 rules between 0.9 to 0.91. It starts decreasing from 0.93 to 7 rules. When the value sets to 0.94, no rules found.

Task 2

I occurred an error message, when open the CSV file from Kaggle in Weka.

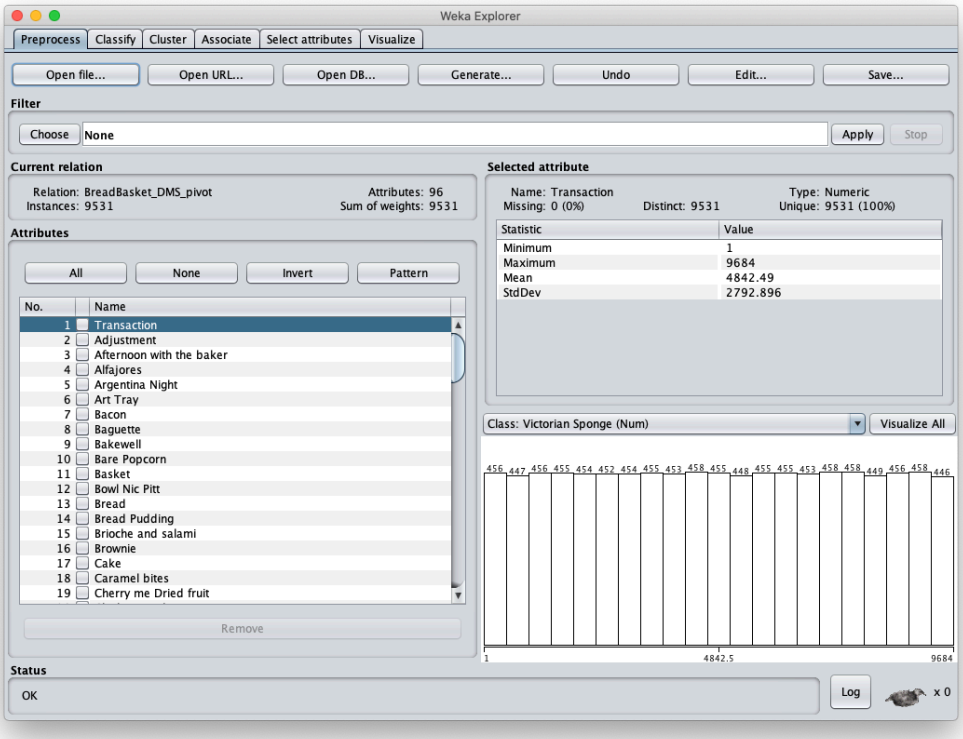


Line 102: `102  2016-10-30,12:15:29,47,Ella's Kitchen Pouches`

The error was caused by the apostrophes from Ella. In order to fix this issue, I replaced all the apostrophes to empty string. For instance, “Ella Kitchen Pouches”. In the CSV file, I detected there’s 30 attributes that contained the apostrophes.

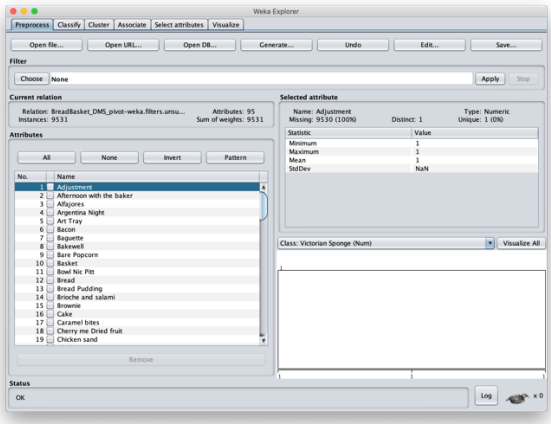
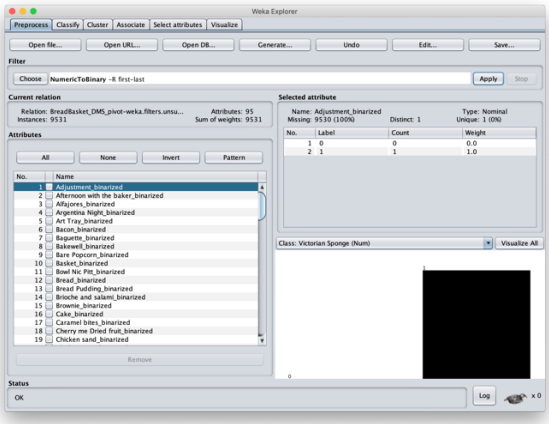
I used the Microsoft excel to create the pivot table for the transaction and item columns. After the pivot table created, I deleted the last two rows and columns in excel. I also clean up the first row and keep the second row. After the cleaning completed, I save the file as BreadBasket_DMS_pviot.csv

The .csv file will look like the screenshot below when opened by Weka:

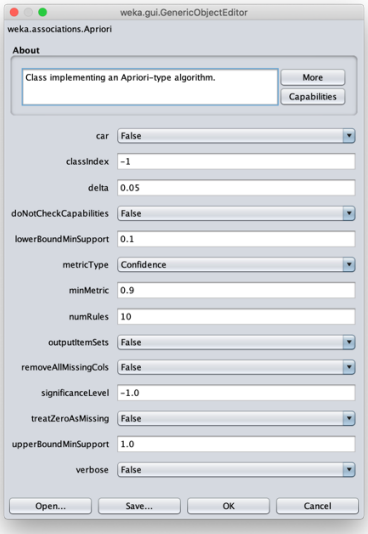
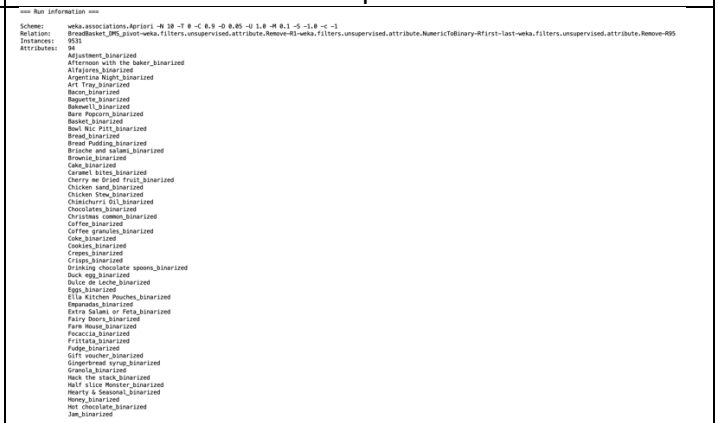


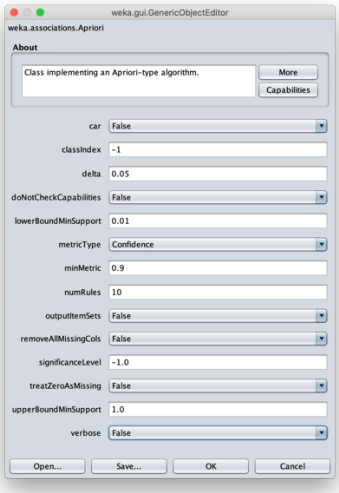
Task 3

I opens the BreadBasket_DMS_pivot.csv and I removed the transaction attribute. Then, apply filter to it. After that, I removed the “Victorian Sponge” attribute to enable the aprior start button.

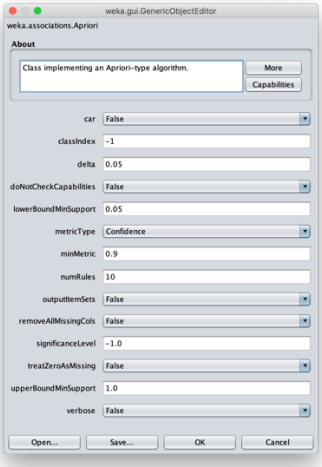
Before Filter	After filtered
	

Apriori

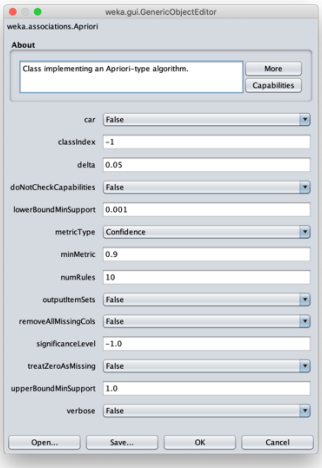
Apriori Parameters	Output
	 … (Bottom) === Associator model (full training set) === No large itemsets and rules found!



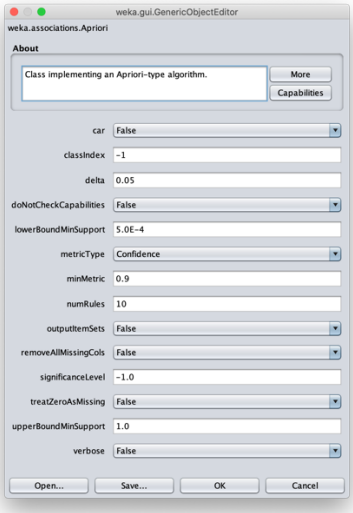

```
Apriori
=======

Minimum support: 0.01 (95 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 31

Size of set of large itemsets L(2): 31

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

Best rules found:
```

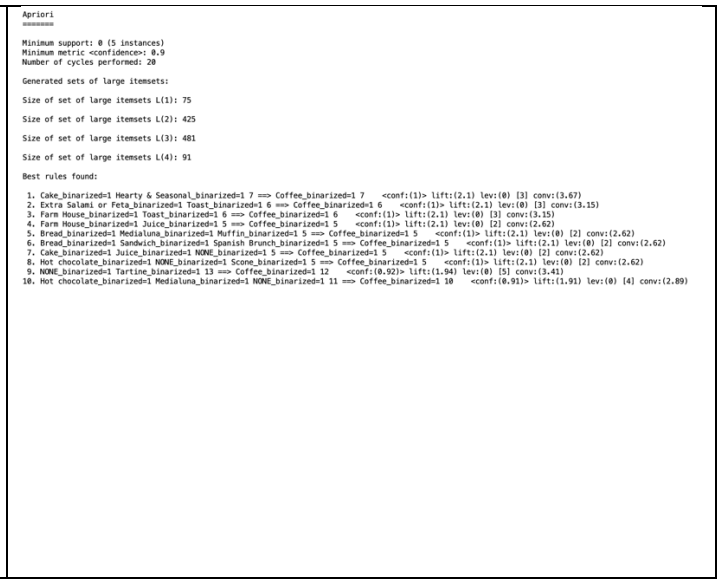



```
Apriori
=======

Minimum support: 0.05 (477 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 19

Generated sets of large itemsets:

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

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

Best rules found:
```




```
Apriori
=======

Minimum support: 0 (10 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 58

Size of set of large itemsets L(2): 271

Size of set of large itemsets L(3): 201

Size of set of large itemsets L(4): 7

Best rules found:

 1. NONE_binarized=1 Tartine_binarized=1 13 ==> Coffee_binarized=1 12    <conf:(0.92)> lift:(1.94) lev:(0) [5] conv:(3.41)
 2. Hot chocolate_binarized=1 Medialuna_binarized=1 NONE_binarized=1 11 ==> Coffee_binarized=1 10    <conf:(0.91)> lift:(1.91) lev:(0) [4] conv:(2.89)
```





```
Apriori
=======

Minimum support: 0 (5 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 20

Generated sets of large itemsets:

Size of set of large itemsets L(1): 75

Size of set of large itemsets L(2): 425

Size of set of large itemsets L(3): 481

Size of set of large itemsets L(4): 91

Best rules found:

 1. Cake_binarized=1 Hearty & Seasonal_binarized=1 7 ==> Coffee_binarized=1 7    <conf:(1)> lift:(2.1) lev:(0) [3] conv:(3.67)
 2. Extra Salami or Feta_binarized=1 Toast_binarized=1 6 ==> Coffee_binarized=1 6    <conf:(1)> lift:(2.1) lev:(0) [3] conv:(3.15)
 3. Farm House_binarized=1 Toast_binarized=1 6 ==> Coffee_binarized=1 6    <conf:(1)> lift:(2.1) lev:(0) [3] conv:(3.15)
 4. Farm House_binarized=1 Juice_binarized=1 5 ==> Coffee_binarized=1 5    <conf:(1)> lift:(2.1) lev:(0) [2] conv:(2.62)
 5. Bread_binarized=1 Medialuna_binarized=1 Muffin_binarized=1 5 ==> Coffee_binarized=1 5    <conf:(1)> lift:(2.1) lev:(0) [2] conv:(2.62)
 6. Bread_binarized=1 Sandwich_binarized=1 Spanish Brunch_binarized=1 5 ==> Coffee_binarized=1 5    <conf:(1)> lift:(2.1) lev:(0) [2] conv:(2.62)
 7. Cake_binarized=1 Juice_binarized=1 NONE_binarized=1 5 ==> Coffee_binarized=1 5    <conf:(1)> lift:(2.1) lev:(0) [2] conv:(2.62)
 8. Hot chocolate_binarized=1 NONE_binarized=1 Scone_binarized=1 5 ==> Coffee_binarized=1 5    <conf:(1)> lift:(2.1) lev:(0) [2] conv:(2.62)
 9. NONE_binarized=1 Tartine_binarized=1 13 ==> Coffee_binarized=1 12    <conf:(0.92)> lift:(1.94) lev:(0) [5] conv:(3.41)
10. Hot chocolate_binarized=1 Medialuna_binarized=1 NONE_binarized=1 11 ==> Coffee_binarized=1 10    <conf:(0.91)> lift:(1.91) lev:(0) [4] conv:(2.89)
```

Analysis

The default parameter of lowerBoundMinSupport didn't find any rules at all and same as 0.05 value. When I set the value to 0.001, weka found 2 rules and it's a good sign. While the value getting smaller you may able to find the rules. Therefore, I found the 10 best rules when the value set to 0.0005. Overall, when the value of lowerBoundMinSupport lower than 0.0005, weka can generated 10 best rules.