IE-583 Homework 2

This homework has been divided into four components - Best Association Rules, Best Association Rules with consequent as Democrats ,Best Association Rules with consequent as Republicans and similarity between Decision Trees and Association Rule Mining

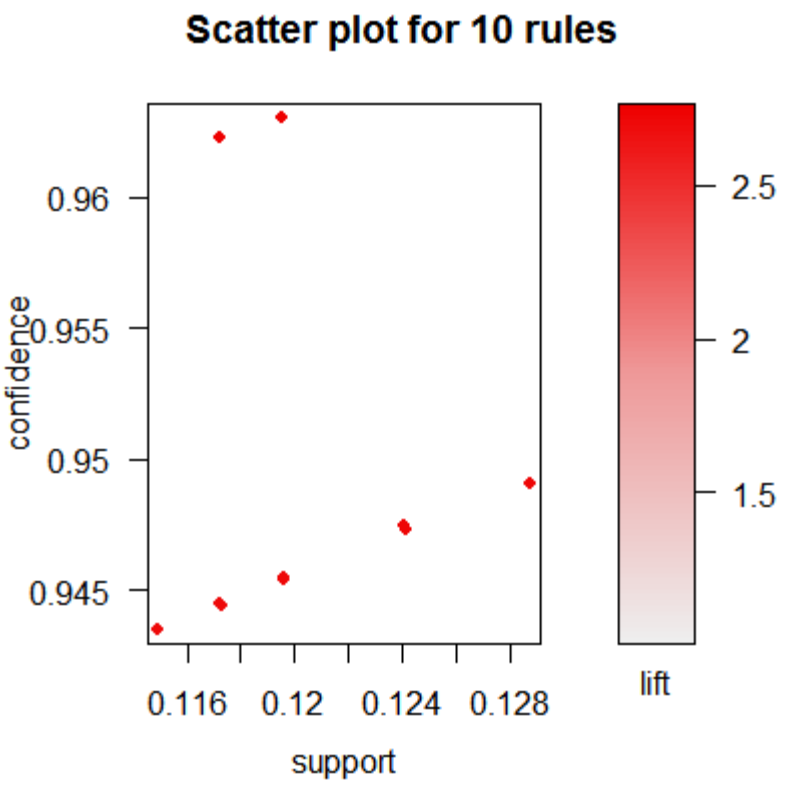
1.

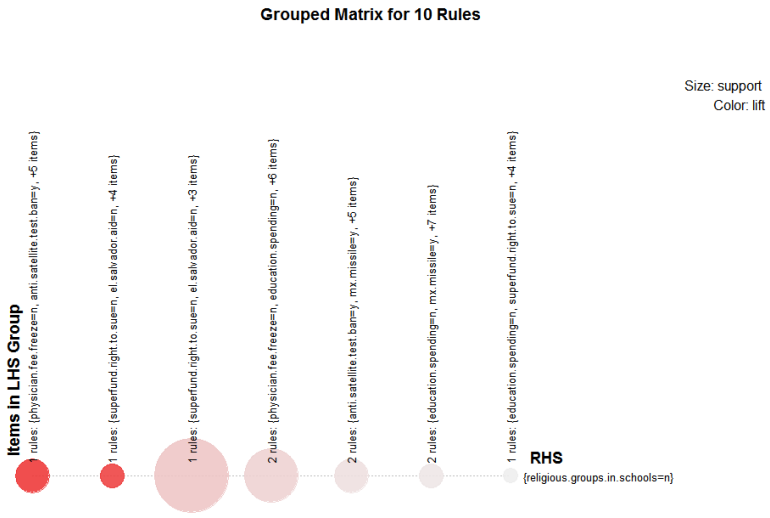
Best Association Rules

lhs		rhs	support	confidence	lift
1.{handicapped.infants=y, mx.missile=y, immigration=n, superfund.right.to.sue=n, crime=n}	=>	{religious.groups.in.schools=n}	0.117241	0.9444444	2.70
2.{handicapped.infants=y, el.salvador.aid=n, immigration=n, superfund.right.to.sue=n, crime=n}	=>	{religious.groups.in.schools=n}	0.128736	0.9491525	2.71
3.{handicapped.infants=y, el.salvador.aid=n, mx.missile=y, immigration=n, superfund.right.to.sue=n, crime=n}	=>	{religious.groups.in.schools=n}	0.117241	0.9622642	2.75
4.{handicapped.infants=y, el.salvador.aid=n, anti.satellite.test.ban=y, mx.missile=y, immigration=n, crime=n}	=>	{religious.groups.in.schools=n}	0.11954	0.9454545	2.70
5.{handicapped.infants=y, physician.fee.freeze=n, el.salvador.aid=n, mx.missile=y, immigration=n, crime=n}	=>	{religious.groups.in.schools=n}	0.124138	0.9473684	2.71

6.{handicapped.infants=y, physician.fee.freeze=n, anti.satellite.test.ban=y, mx.missile=y, immigration=n, crime=n}	=>	{religious.groups.in.schools=n}	0.11954	0.9454545	2.70
7.{handicapped.infants=y, el.salvador.aid=n, mx.missile=y, immigration=n, education.spending=n, superfund.right.to.sue=n}	=>	{religious.groups.in.schools=n}	0.114943	0.9433962	2.69
8.{handicapped.infants=y, physician.fee.freeze=n, el.salvador.aid=n, anti.satellite.test.ban=y, mx.missile=y, immigration=n, crime=n}	=>	{religious.groups.in.schools=n}	0.11954	0.962963	2.75
9.{handicapped.infants=y, physician.fee.freeze=n, el.salvador.aid=n, anti.satellite.test.ban=y, immigration=n, education.spending=n, crime=n}	=>	{religious.groups.in.schools=n}	0.124138	0.9473684	2.71
10.{handicapped.infants=y, physician.fee.freeze=n, el.salvador.aid=n, anti.satellite.test.ban=y, mx.missile=y, immigration=n, education.spending=n}	=>	{religious.groups.in.schools=n}	0.117241	0.9444444	2.70

Visualization of Rules





Discussion of the rules –

- Here, the rule 1 is redundant. This is because rule 3 has more number of items and support as compared to rule 1.
- While using the Apriori algorithm, the support was kept extremely low with a high confidence and lift. The support was slowly increased to reduce the number of rules. Initially, the number of rules obtained was 482817 rules. Placing limits on support, confidence and lift, the number of rules were reduced to 252. Once all redundant rules were removed (better_rules2[!is.redundant(better_rules2)) using code from the packages arules, the rules were brought down to 10. However, this code snippet didn't completely remove all redundant rules. Rule 1 was classified by observing and studying the support and items.

- *From the plots above, it is apparent that rule 2 has highest support. Alongside, it has a competitive confidence (only 2 rules have better confidence) and high lift –*

2.{handicapped.infants=y, el.salvador.aid=n, immigration=n, superfund.right.to.sue=n, crime=n}	=>	{religious.groups.in.schools=n}	0.128736	0.9491525	2.716325	56

- *Rule 8 gives the highest confidence. Also, this rule has the highest lift. Therefore, this rule is 2.7 times more likely to occur as compared to the scenario that the items were independent –*

8.{handicapped.infants=y, physician.fee.freeze=n, el.salvador.aid=n, anti.satellite.test.ban=y, mx.missile=y, immigration=n, crime=n}	=>	{religious.groups.in.schools=n}	0.11954	0.962963	2.755848	52

- The reason why the consequent is filled with religious groups in school = n, is that it is the most dominant rule. However, no effort was taken to group the rules here, as we wanted to prune the rules solely on the basis of support, confidence and lift.

2.

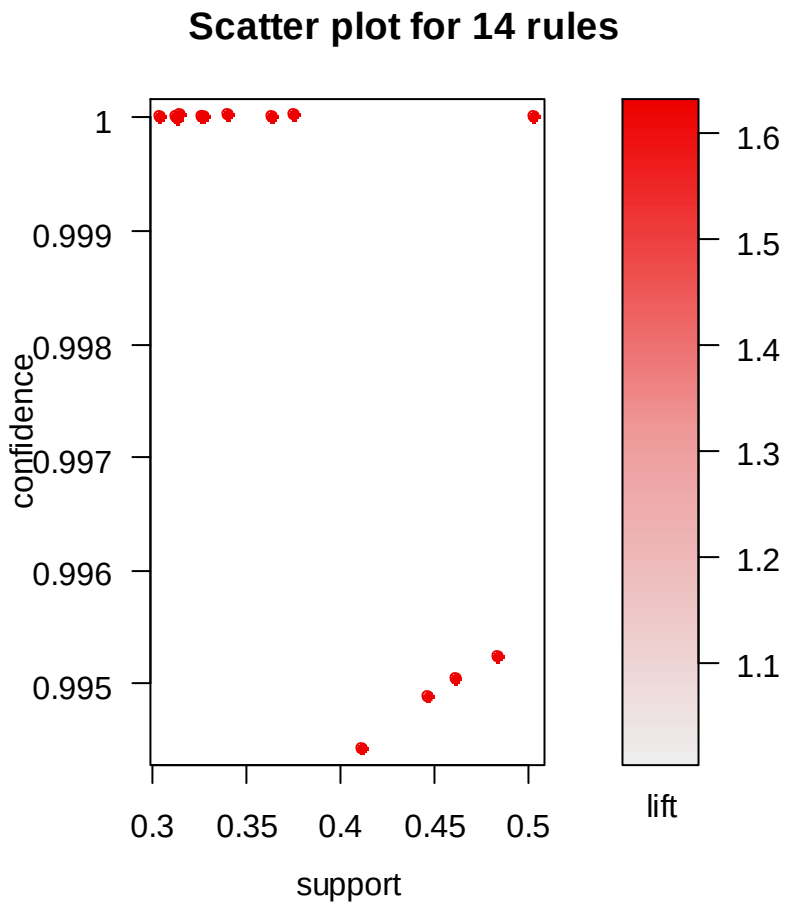
Best Association Rules for a Democrat

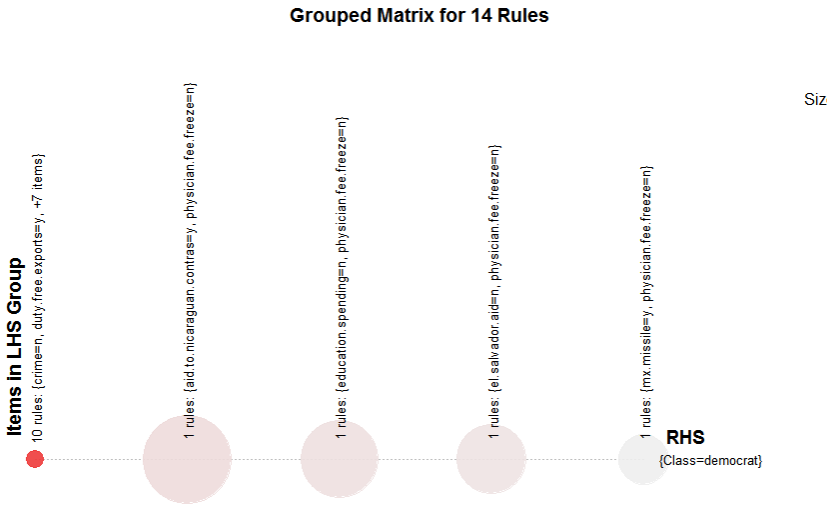
	lhs		rhs	support	confidence	lift	count
[1]	{superfund.right.to.sue=n, crime=n}	=>	{Class=democrat}	0.3126444	1	1.621693	136
[2]	{anti.satellite.test.ban=y, crime=n}	=>	{Class=democrat}	0.340023	1	1.621693	148
[3]	{physician.fee.freeze=n, crime=n}	=>	{Class=democrat}	0.3747133	1	1.621693	163
[4]	{adoption.of.the.budget.resolution=y, crime=n}	=>	{Class=democrat}	0.3632188	1	1.621693	158
[5]	{physician.fee.freeze=n, mx.missile=y}	=>	{Class=democrat}	0.4114943	0.9944444	1.620162	179
[6]	{physician.fee.freeze=n, el.salvador.aid=n}	=>	{Class=democrat}	0.4459777	0.9948718	1.620859	194
[7]	{physician.fee.freeze=n, education.spending=n}	=>	{Class=democrat}	0.4620699	0.9950495	1.621148	201

[8]	{physician.fee.freeze=n, aid.to.nicaraguan.contras=y}	=>	{Class=democrat}	0.48275	0.9952607	1.621492	210
[9]	{adoption.of.the.budget.resolution=y, physician.fee.freeze=n}	=>	{Class=democrat}	0.50344	1	1.621923	219
[10]	{el.salvador.aid=n, education.spending=n, crime=n}	=>	{Class=democrat}	0.31494	1	1.621923	137
[11]	{aid.to.nicaraguan.contras=y, education.spending=n, crime=n}	=>	{Class=democrat}	0.32643	1	1.621923	142
[12]	{physician.fee.freeze=n, el.salvador.aid=n, duty.free.exports=y}	=>	{Class=democrat}	0.31264	1	1.621923	136
[13]	{physician.fee.freeze=n, education.spending=n, duty.free.exports=y}	=>	{Class=democrat}	0.30344	1	1.621923	132
[14]	{physician.fee.freeze=n,						

aid.to.nicaraguan.contras=y,						
duty.free.exports=y}	=>	{Class=democrat}	0.32	1	1.62	142

Visualization of rules –



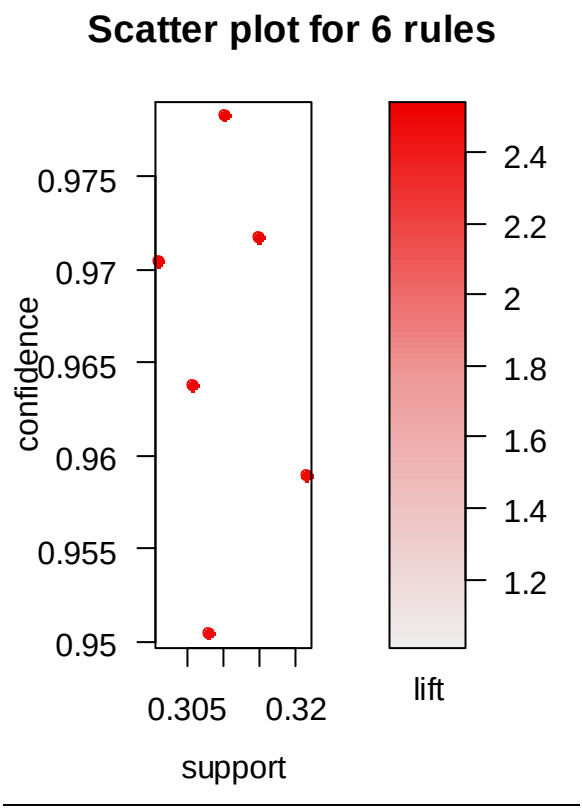


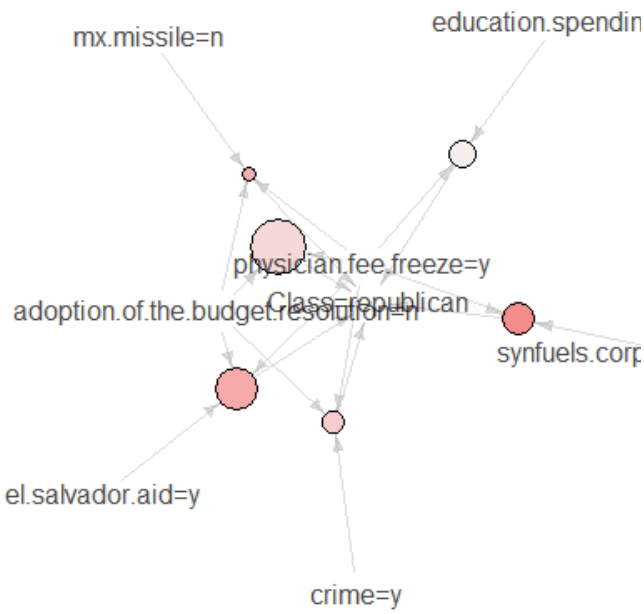
- In order to arrive at the best rules for democrats, the general rules obtained from the apriori algorithm was bound by a support of 0.3, confidence of 0.95 and lift of 1.62. The resulting rules were pruned to remove redundant rules. Next, the rules are pruned by group where rhs was manually fixed to democrats.
- The rules adoption of budget = y and physician fee freeze =n, have an high support, lift as well as confidence.
- It is apparent from the the second plot that aid to nicarugan contras =y and physician freeze = n has the highest support.
- More than half of the rules have almost perfect confidence as apparent from the first plot.
- Many of the above rules have a confidence of 1, this implies that all democrats voted for the items in the lhs.
- The lift in most of the above cases is ~1.6. This implies that the democrats voting according to the items in the lhs is 1.6 times more likely than other members of the senate doing the same.

Best Association Rules for a Republican

	lhs		rhs	support	confidence	lift	count
[1]	{physician.fee.freeze=y, education.spending=y}	=>	{Class=republican}	0.308046	0.950355	2.407474	134
[2]	{adoption.of.the.budget.resolution=n, physician.fee.freeze=y}	=>	{Class=republican}	0.321839	0.958904	2.428277	140
[3]	{physician.fee.freeze=y, synfuels.corporation.cutback=n}	=>	{Class=republican}	0.310345	0.971826	2.462997	135
[4]	{adoption.of.the.budget.resolution=n, physician.fee.freeze=y, mx.missile=n}	=>	{Class=republican}	0.301149	0.970370	2.458125	131
[5]	{adoption.of.the.budget.resolution=n, physician.fee.freeze=y, el.salvador.aid=y}	=>	{Class=republican}	0.314943	0.971631	2.461583	137
[6]	{adoption.of.the.budget.resolution=n,						

physician.fee. freeze=y, crime=y}	=>	{Class=republican}	0.30057	0.96376	2.4954	1323



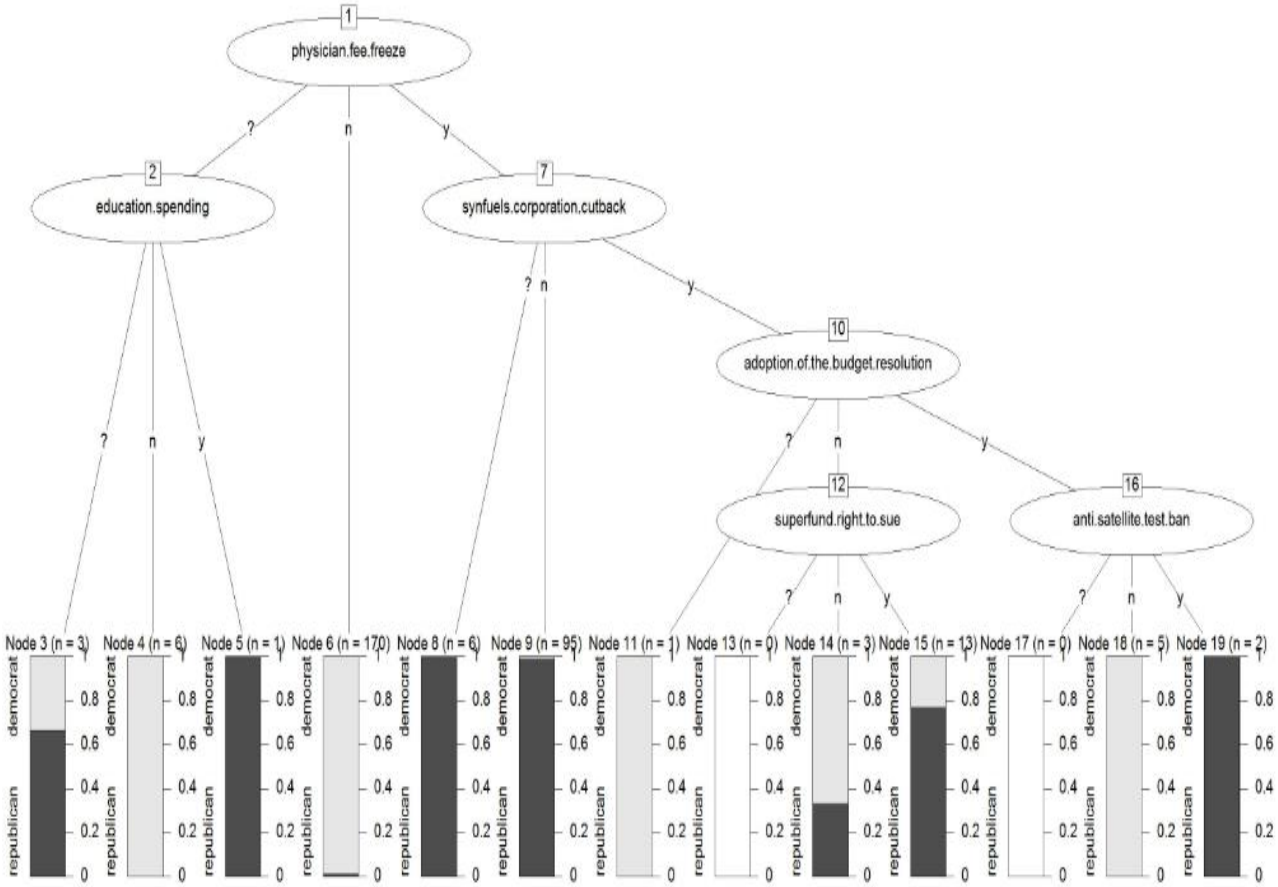


- A confidence of 0.95, support of 0.3 and a lift of 1 was applied to the general set of rules obtained by the apriori algorithm.
- The rules obtained were then pruned to remove redundancies and then grouped together. The consequent of the rules was fixed at republican. This led to attaining 6 best rules.
- The rule # 2, physician fee freeze = y and adoption of budget =n, obtained the highest support.
- The rule #3, physician fee freeze =y & syncfuels corporation cutback = n have the highest lift. This implies that the republicans are 2.53 times more likely to vote according this rule than any other member of the senate. This rule also has the highest confidence. This implies that 97.8% of the republicans voted for the same.
- It is apparent from the plot that the rule with the highest support has a comparatively lower confidence.
- The second plot shows the connection between the items with the consequent (class = republican) in the centre.

- Here we observe that most of all the rules contain physician free f

4.

Similarity between decision trees and Assoication Rule Mining Results



- From the decision it is apparent that an overwhelming majority of democrats are against physician fee freeze. This is si
- A majority of the the votes that have been placed against synfuels cutback are republican votes. The same has bee organizations.

- A (unpruned) decision tree gives us complete and top down information regarding the data set. However, in such cas certain attributes become important due to random noise. Finally, we obtain a whole bunch of information in the last l

- In Association Rule Mining, we can't study the entire data. However, we can always gain information from a small subs