Florida Atlantic University
COLLEGE OF BUSINESS

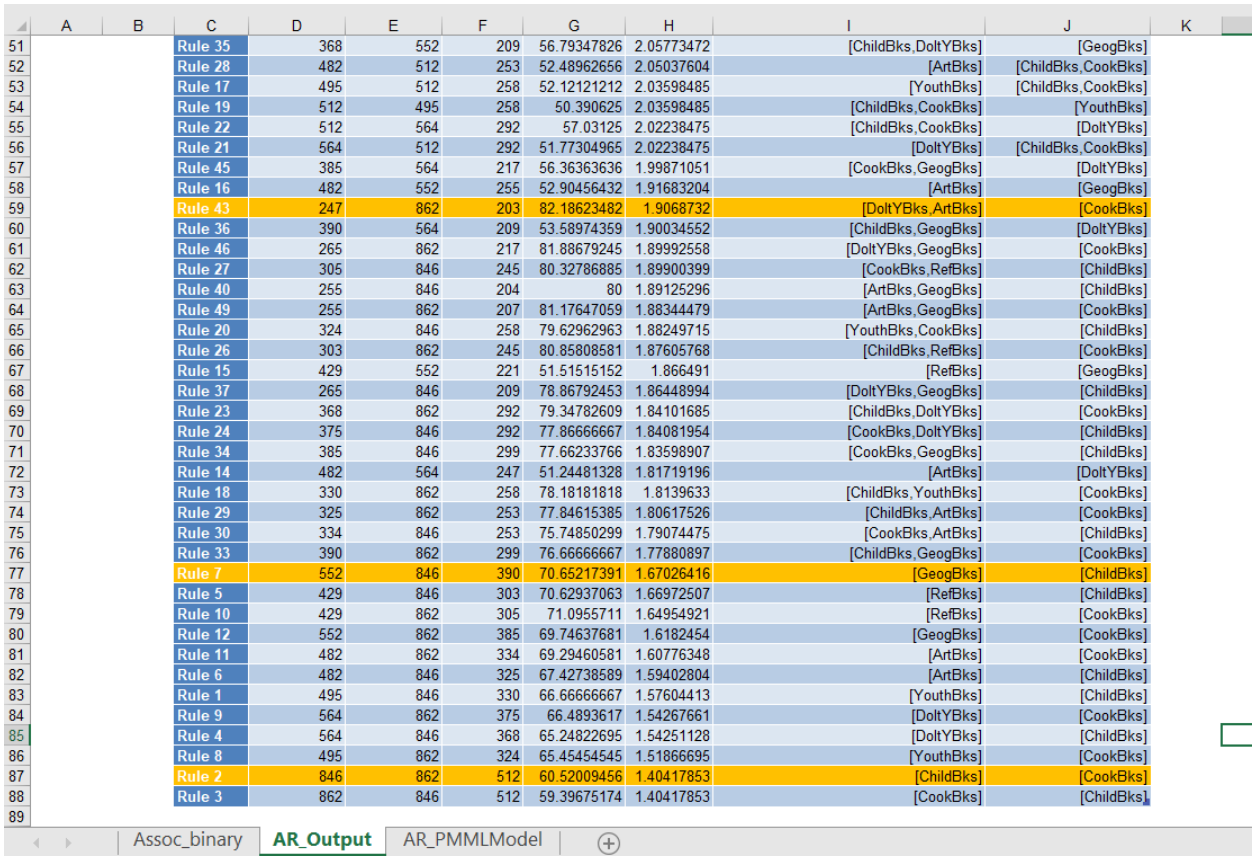
ISM 6136 – Datamining/Predictive Analytics
Dr. Bharti Sharma

Class Assignment 8
5 points

TASK: Performing Association – Data Mining Task using XLMiner

(Note: Check Support A & C together and not individual support).

1. Charles_Book_Club – Determine which group of books will likely to be purchased based on given data. Determine at least 3 strongest association rules. Remember the rules you select should not have the exactly same combination of Antecedents and Consequents.
   Paste the screen shot of your rules and explain your rule selection criteria for each one. (Note: Check Support A & C together and not individual support). Write a final statement to present these rules to the Charles Book Club.



51			Rule 35	368	552	209	56.79347826	2.05773472	[ChildBks,DoltYBks]	[GeogBks]
52			Rule 28	482	512	253	52.48962656	2.05037604	[ArtBks]	[ChildBks,CookBks]
53			Rule 17	495	512	258	52.12121212	2.03598485	[YouthBks]	[ChildBks,CookBks]
54			Rule 19	512	495	258	50.390625	2.03598485	[ChildBks,CookBks]	[YouthBks]
55			Rule 22	512	564	292	57.03125	2.02238475	[ChildBks,CookBks]	[DoltYBks]
56			Rule 21	564	512	292	51.77304965	2.02238475	[DoltYBks]	[ChildBks,CookBks]
57			Rule 45	385	564	217	56.36363636	1.99871051	[CookBks,GeogBks]	[DoltYBks]
58			Rule 16	482	552	255	52.90456432	1.91683204	[ArtBks]	[GeogBks]
59			Rule 43	247	862	203	82.18623482	1.9068732	[DoltYBks,ArtBks]	[CookBks]
60			Rule 36	390	564	209	53.58974359	1.90034552	[ChildBks,GeogBks]	[DoltYBks]
61			Rule 46	265	862	217	81.88679245	1.89992558	[DoltYBks,GeogBks]	[CookBks]
62			Rule 27	305	846	245	80.32786885	1.89900399	[CookBks,RefBks]	[ChildBks]
63			Rule 40	255	846	204	80	1.89125296	[ArtBks,GeogBks]	[ChildBks]
64			Rule 49	255	862	207	81.17647059	1.88344479	[ArtBks,GeogBks]	[CookBks]
65			Rule 20	324	846	258	79.62962963	1.88249715	[YouthBks,CookBks]	[ChildBks]
66			Rule 26	303	862	245	80.85808581	1.87605768	[ChildBks,RefBks]	[CookBks]
67			Rule 15	429	552	221	51.51515152	1.866491	[RefBks]	[GeogBks]
68			Rule 37	265	846	209	78.86792453	1.86448994	[DoltYBks,GeogBks]	[ChildBks]
69			Rule 23	368	862	292	79.34782609	1.84101685	[ChildBks,DoltYBks]	[CookBks]
70			Rule 24	375	846	292	77.86666667	1.84081954	[CookBks,DoltYBks]	[ChildBks]
71			Rule 34	385	846	299	77.66233766	1.83598907	[CookBks,GeogBks]	[ChildBks]
72			Rule 14	482	564	247	51.24481328	1.81719196	[ArtBks]	[DoltYBks]
73			Rule 18	330	862	258	78.18181818	1.8139633	[ChildBks,YouthBks]	[CookBks]
74			Rule 29	325	862	253	77.84615385	1.80617526	[ChildBks,ArtBks]	[CookBks]
75			Rule 30	334	846	253	75.74850299	1.79074475	[CookBks,ArtBks]	[ChildBks]
76			Rule 33	390	862	299	76.66666667	1.77880897	[ChildBks,GeogBks]	[CookBks]
77			Rule 7	552	846	390	70.65217391	1.67026416	[GeogBks]	[ChildBks]
78			Rule 5	429	846	303	70.62937063	1.66972507	[RefBks]	[ChildBks]
79			Rule 10	429	862	305	71.0955711	1.64954921	[RefBks]	[CookBks]
80			Rule 12	552	862	385	69.74637681	1.6182454	[GeogBks]	[CookBks]
81			Rule 11	482	862	334	69.29460581	1.60776348	[ArtBks]	[CookBks]
82			Rule 6	482	846	325	67.42738589	1.59402804	[ArtBks]	[ChildBks]
83			Rule 1	495	846	330	66.66666667	1.57604413	[YouthBks]	[ChildBks]
84			Rule 9	564	862	375	66.4893617	1.54267661	[DoltYBks]	[CookBks]
85			Rule 4	564	846	368	65.24822695	1.54251128	[DoltYBks]	[ChildBks]
86			Rule 8	495	862	324	65.45454545	1.51866695	[YouthBks]	[CookBks]
87			Rule 2	846	862	512	60.52009456	1.40417853	[ChildBks]	[CookBks]
88			Rule 3	862	846	512	59.39675174	1.40417853	[CookBks]	[ChildBks]
89										

Assoc_binary | AR_Output | AR_PMMLModel

Reason for choosing the rules: These three rules were chosen by sorting the rule table by largest support value (as it measures overall impact and validity of the rule), then looked at the confidence and lift ratio (which shows the rate at which consequents can be found and how effective the rule is at finding consequents respectively). Also taken into account are the actual antecedents and consequents as for these rules there should be no redundancy. To note, in the screenshot the sorting shown was only for fitting the rules on the same page to get all rules in a single screenshot, not for analysis purposes. Rule 2 was chosen as it has the highest support value as well as a balanced confidence and a lift ratio greater than 1 which shows a stronger association between antecedent and consequent. Rule 7 was then chosen similarly to rule 2, except that rule 7 had a slightly lower support but offered higher confidence and lift ratio and a different antecedent and consequent pair. Rule 43 was chosen based on the antecedent-consequen pair as the other rules had similar ones. Rule 43 had a mid range support value while having the highest confidence and lift ratio out of the three rules chosen.

Final statement to Charles Book Club: Based on my findings, there seems to be a correlation that if a customer buys a child's book then a cookbook is also bought, similarly if a customer buys a geography book then a child's book is also bought, and if a do-it-yourself book and an art book are bought then a cookbook is also bought. Based on these associations, appropriate sales or discounts can be applied or reshelving of books for ease of customers can be done, such as example moving the geography books next to childrens books and children's books next to cookbooks which could be next to do-it-yourself and art books, or a bundle deal containing the correlating books.

2. Perform Market-Basket analysis on the following list of groceries and determine which ones will be very likely to be bought together. Determine at least 3 strongest association rules. Remember the rules you select should not have the exactly same combination of Antecedents and Consequents.
Paste the screen shot of your rules and explain your rule selection criteria for each one. (Note: Check Support A & C). Write a final statement to present these rules to the Grocery store.

Rules

Rule ID	A-Support	C-Support	Support	Confidence	Lift-Ratio	Antecedent	Consequent
Rule 16	488	600	366	75	1.25	[cracker]	[heineken]
Rule 17	600	488	366	61	1.25	[heineken]	[cracker]
Rule 21	485	600	288	59.3814433	0.98969072	[hering]	[heineken]
Rule 5	392	600	261	66.58163265	1.10969388	[baguette]	[heineken]
Rule 10	318	600	257	80.81761006	1.34696017	[soda]	[heineken]
Rule 30	472	485	255	54.02542373	1.11392626	[olives]	[hering]
Rule 29	485	472	255	52.57731959	1.11392626	[hering]	[olives]
Rule 23	305	600	252	82.62295082	1.37704918	[artichok]	[heineken]
Rule 8	318	488	251	78.93081761	1.61743479	[soda]	[cracker]
Rule 9	488	318	251	51.43442623	1.61743479	[cracker]	[soda]
Rule 13	363	600	249	68.59504132	1.14325069	[avocado]	[heineken]
Rule 6	392	485	249	63.52040816	1.30969914	[baguette]	[hering]
Rule 7	485	392	249	51.34020619	1.30969914	[hering]	[baguette]

Reason for choosing the rules: These rules were chosen by sorting the rule table by largest support value (as it measures overall impact and validity of the rule), then looked at the confidence and lift ratio (which shows the rate at which consequents can be found and how effective the rule is at finding consequents respectively). Also taken into account are the actual antecedents and consequents as for these rules there should be no redundancy. Rule 16 was chosen since it has the highest support and a high confidence and a

greater than 1 lift ratio. Rule 19 was chosen for similar reasons, a slightly lowered support value but higher confidence and lift ratio values as well as differing antecedent-consequent pair. Rule 6 was chosen as it had a different antecedent-consequent pair with a high range support value which gives it a higher impact than similar pairs.
Final statement to Market Basket store: Based on these findings, if a shopper buys crackers then they might also buy Heineken, also if shoppers buy soda, then they could also buy Heineken, finally, if shoppers buy a baguette then they might purchase hering. These correlations show what shoppers frequent buy together so weekly deals can be structured around these findings.

3. A drug store chain wants to learn more about cosmetics buyers purchase patterns. Specifically, they want to know what items are purchased in conjunction with each other, for purposes of display, point of sale special offers and eventually implement a real time recommender system to cross-sell items at time of purchase. Paste the screen shot of your rules and determine at least 3 strongest association rules. Remember the rules you select should not have the exactly same combination of Antecedents and Consequents. Explain your rule selection criteria for each one. (Note: Check Support A & C). Write a final statement to present these rules to the Drug store.

Rules

Rule ID	A-Support	C-Support	Support	Confidence	Lift-Ratio	Antecedent	Consequent
Rule 6	490	536	356	72.65306122	1.35546756	[Lip Gloss]	[Foundation]
Rule 4	357	381	321	89.91596639	2.35999912	[Mascara]	[Eye shadow]
Rule 5	381	357	321	84.2519685	2.35999912	[Eye shadow]	[Mascara]
Rule 98	201	357	179	89.05472637	2.49453015	[Concealer,Eye shadow]	[Mascara]
Rule 97	204	381	179	87.74509804	2.30302095	[Concealer,Mascara]	[Eye shadow]
Rule 3	234	442	179	76.4957265	1.73067255	[Lip liner]	[Concealer]
Rule 54	182	357	169	92.85714286	2.60104042	[Blush,Eye shadow]	[Mascara]
Rule 53	184	381	169	91.84782609	2.4107041	[Blush,Mascara]	[Eye shadow]
Rule 117	192	381	166	86.45833333	2.26924759	[Mascara,Foundation]	[Eye shadow]
Rule 118	211	357	166	78.67298578	2.20372509	[Eye shadow,Foundation]	[Mascara]
Rule 119	181	381	158	87.29281768	2.29115007	[Mascara,Lip Gloss]	[Eye shadow]
Rule 120	201	357	158	78.60696517	2.20187578	[Eye shadow,Lip Gloss]	[Mascara]
Rule 123	175	381	151	86.28571429	2.26471691	[Mascara,Eyeliner]	[Eye shadow]
Rule 124	182	357	151	82.96703297	2.32400653	[Eye shadow,Eyeliner]	[Mascara]
Rule 2	149	280	149	100	3.57142857	[Brushes]	[Nail Polish]
Rule 126	201	536	146	72.63681592	1.35516448	[Eye shadow,Lip Gloss]	[Foundation]
Rule 45	180	442	143	79.44444444	1.79738562	[Blush,Eyeliner]	[Concealer]
Rule 99	175	442	134	76.57142857	1.73238526	[Mascara,Eyeliner]	[Concealer]
Rule 44	184	442	131	71.19565217	1.61076136	[Blush,Mascara]	[Concealer]
Rule 125	181	536	130	71.82320442	1.33998516	[Mascara,Lip Gloss]	[Foundation]
Rule 100	182	442	130	71.42857143	1.61603103	[Eye shadow,Eyeliner]	[Concealer]
Rule 113	137	381	124	90.51094891	2.37561546	[Bronzer,Mascara]	[Eye shadow]
Rule 114	141	357	124	87.94326241	2.46339671	[Bronzer,Eye shadow]	[Mascara]
Rule 96	130	442	120	92.30769231	2.08840933	[Lip liner,Eyeliner]	[Concealer]
Rule 501	124	357	119	95.96774194	2.68817204	[Blush,Concealer,Eye shadow]	[Mascara]
Rule 70	131	357	119	90.83969466	2.54452926	[Nail Polish,Eye shadow]	[Mascara]
Rule 500	131	381	119	90.83969466	2.38424395	[Blush,Concealer,Mascara]	[Eye shadow]

Description | Data | AR_Output | AR_PMMLModel

Reason for choosing the rules: These three rules were chosen by sorting the rule table by largest support value (as it measures overall impact and validity of the rule), then looked at the confidence and lift ratio (which shows the rate at which consequents can be found and how effective the rule is at finding consequents respectively). Also considered are the actual antecedents and consequents as for these rules there should be no redundancy. Rule 4 was chosen because of the high support value and higher confidence and lift ratio than rule 6 (the top rule in screenshot). Rule 8 was chosen as it had another high support value and high confidence and lift ratio greater than 1. Rule 2 was chosen because it had a high confidence and a very high lift ratio, and the support was mid-range in the data which when looked at with the confidence and lift ratio seems very good. All antecedent-consequent pairs are different in this case.
Final statement to Drug store: Based on my findings, if a customer buys mascara then they will most likely buy eye shadow as well, also if a customer buys lip liner then they will most likely buy concealer as well, and finally if a customer buys brushes then they will buy nail polish as well. From this putting the mascara and eye shadow close to each other as well as the lip liner and the concealer, and the brushes and nail polish would be ideal. These item pairs can be bundled for deals or added to the recommendation system for online customer purchase.

4. The institute for Statistics Education at Statistics.com offers online course and is seeking information that will help in packaging and sequencing the courses. Consider the data in the file CourseTopics.xls. This data is for purchases of online statistics courses at Statistics.com. Each row represents the courses attended by a single student. The firm wishes to assess alternative sequencings and bundling of courses. Use association rules to analyze this data and interpret the resulting rules. Determine at least 3 strongest association rules. Remember the rules you select should not have the exactly same combination of Antecedents and Consequents.
   Paste the screen shot of your rules and explain your rule selection criteria for each one.
   (Note: Check for Support A & C). Write a final statement to present these rules to Statistics.com.

Rules

Rule ID	A-Support	C-Support	Support	Confidence	Lift-Ratio	Antecedent	Consequent
Rule 7	20	144	14	70	1.77430556	[Regression,SW]	[Intro]
Rule 6	26	81	14	53.84615385	2.42640076	[Intro,Regression]	[SW]
Rule 4	18	144	12	66.66666667	1.68981481	[Survey,SW]	[Intro]
Rule 5	20	144	12	60	1.52083333	[Cat Data,Regression]	[Intro]
Rule 3	22	81	12	54.54545455	2.45791246	[Intro,Survey]	[SW]
Rule 8	17	81	11	64.70588235	2.9157589	[Intro,DOE]	[SW]
Rule 9	21	144	11	52.38095238	1.32771164	[DOE,SW]	[Intro]
Rule 11	16	76	10	62.5	3.00164474	[DataMining,Regression]	[Cat Data]
Rule 2	16	144	10	62.5	1.58420139	[DataMining,Regression]	[Intro]
Rule 10	18	76	10	55.55555556	2.66812865	[DataMining,Cat Data]	[Regression]
Rule 12	20	65	10	50	2.80769231	[Cat Data,Regression]	[DataMining]
Rule 1	20	76	10	50	2.40131579	[Intro,DataMining]	[Regression]

Reason for choosing the rules: These three rules were chosen by sorting the rule table by largest support value (as it measures overall impact and validity of the rule), then looked at the confidence and lift ratio (which shows the rate at which consequents can be found and how effective the rule is at finding consequents respectively). Also taken into account are the actual antecedents and consequents as for these rules there should be no redundancy. Rule 7 and Rule 3 were chosen because they are on the higher side of the support value for this data as well as having high balanced confidence and lift ratios and differing antecedent-consequent pairs. Rule 11 was chosen for having a unique antecedent-consequent pair and a balanced support-confidence-lift ratio values.

Final statement to statistics.com : Based on my findings, if a student buys the regression analysis and the software only courses then they will also buy the introductory statistics course, and if a student buys the introductory statistics course and the survey design and sampling procedure course then they will also buy the software only courses, and if a student buys the data mining course and the regression analysis course then they will also buy the categorical data analysis course as well. From this course bundles can be made ad data mining, regression analysis and categorical analysis correlate well and also could bundle introductory statistics course and software only courses with either regression analysis or survey design and sampling procedures. Sequencing the courses could be like having the introductory statistics taken with data mining and then taking the regression and software only courses based on those course correlations.