

**ISM 6136 – Datamining/Predictive Analytics**

**Dr. Bharti Sharma**

**Class Assignment 8**

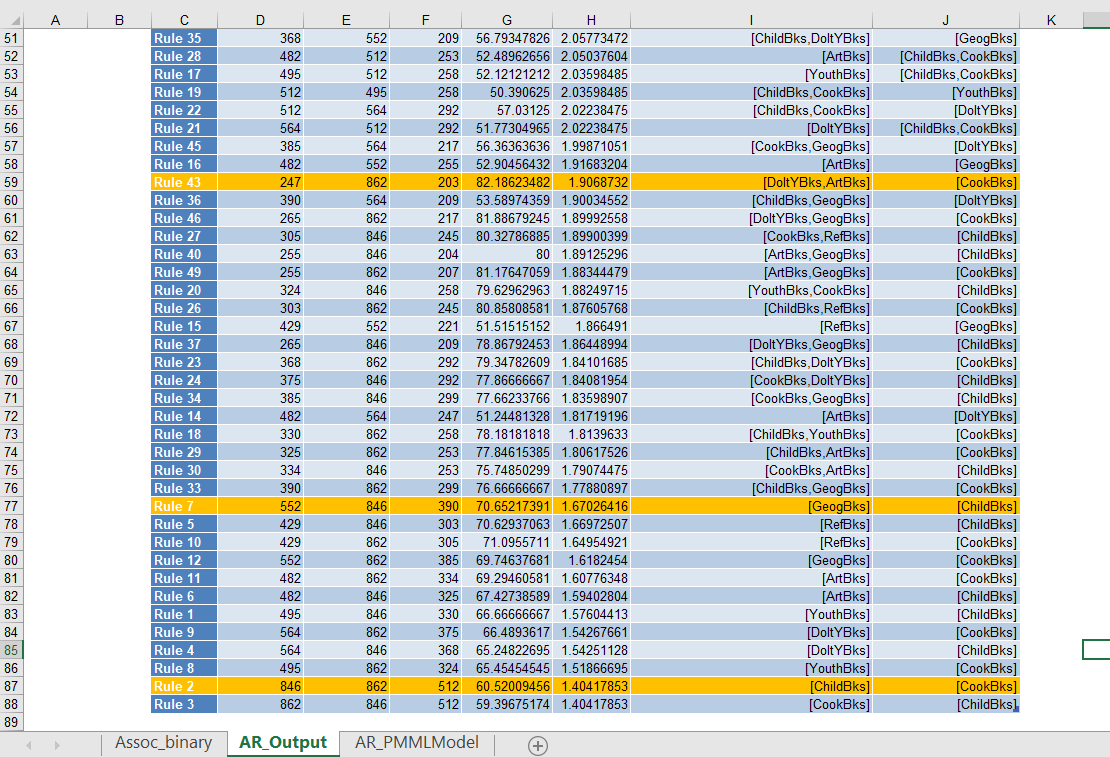
**5 points**

**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.

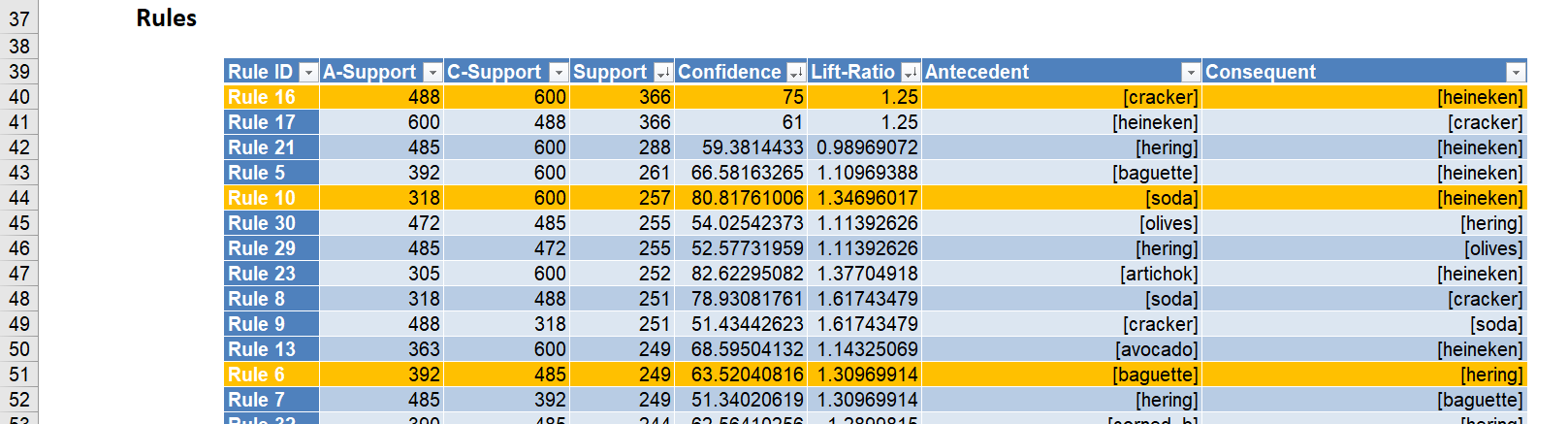


Reason for choosing the rules: These three rules were chosen by sorting the rule table by largest support value (as it measures overall impact and validity of the rule), then looked at the confidence and lift ratio (which shows the rate at which consequents can be found and how effective the rule is 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.

1. 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.

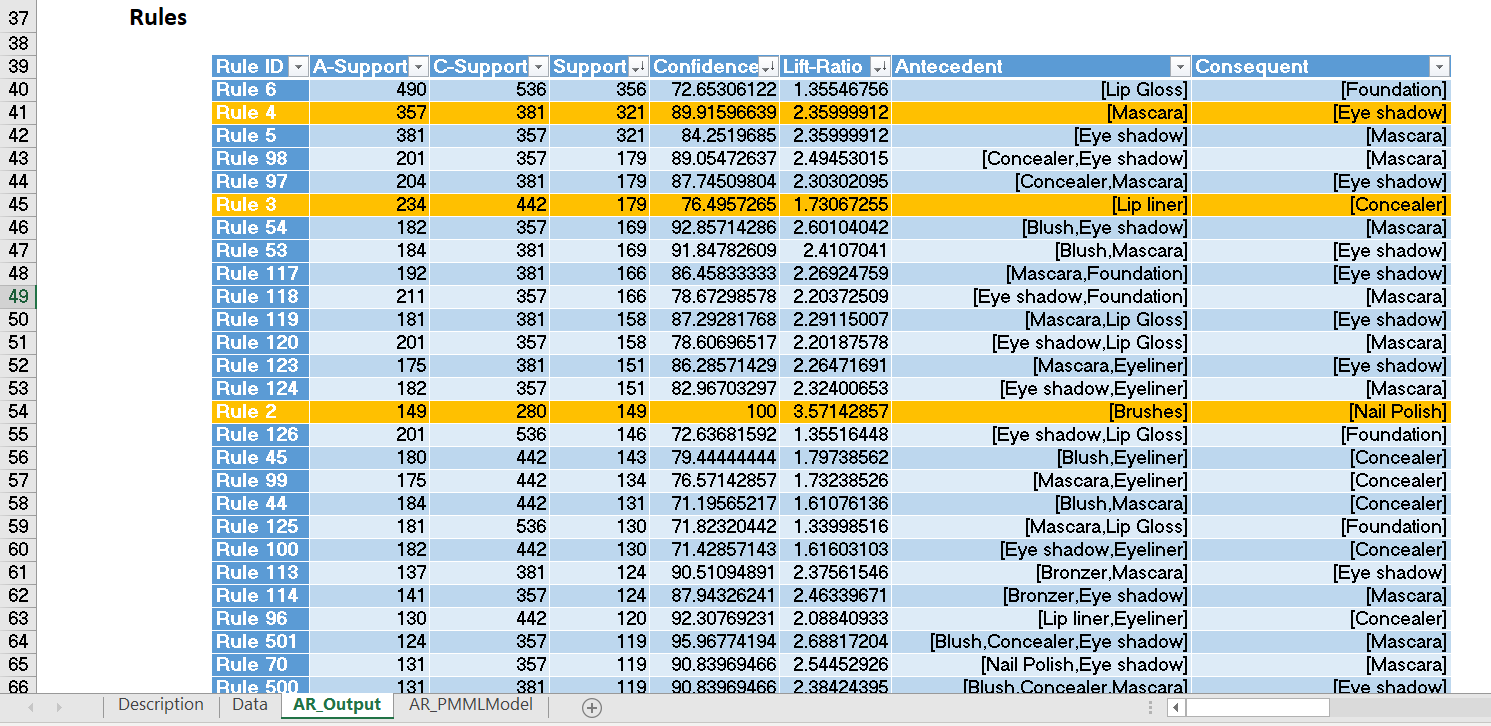


Reason for choosing the rules: These rules were chosen by sorting the rule table by largest support value (as it measures overall impact and validity of the rule), then looked at the confidence and lift ratio (which shows the rate at which consequents can be found and how effective the rule is at finding 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.

1. 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.



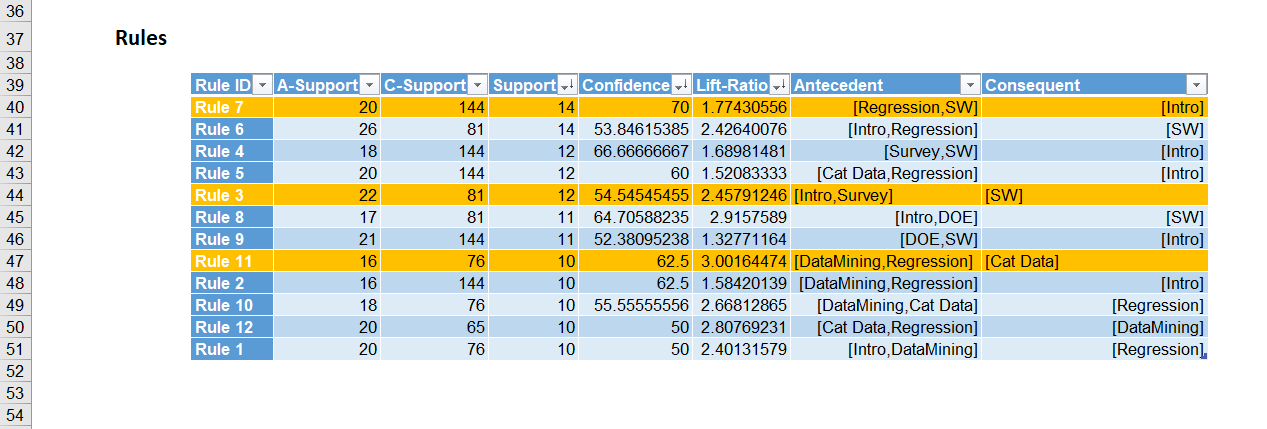
Reason for choosing the rules: These three rules were chosen by sorting the rule table by largest support value (as it measures overall impact and validity of the rule), then looked at the confidence and lift ratio (which shows the rate at which consequents can be found and how effective the rule is 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.

1. 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.



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.