

**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.

**Top 3 Rules Highlighted in red (Selection Criteria: High Support, High Confidence, High Lift Ratio) Min Confidence Selected: 70%. Ensured supports are not very low.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rule ID** | **A-Support** | **C-Support** | **Support** | **Confidence** | **Lift-Ratio** | **Antecedent** | **Consequent** | **Selection Criteria** |
| **Rule 2** | 552 | 846 | 390 | 70.65217391 | 1.67026416 | [GeogBks] | [ChildBks] | High Support, >70% confidence and >1 lift ratio |
| **Rule 3** | 429 | 862 | 305 | 71.0955711 | 1.64954921 | [RefBks] | [CookBks] |  |
| **Rule 1** | 429 | 846 | 303 | 70.62937063 | 1.66972507 | [RefBks] | [ChildBks] |  |
| **Rule 14** | 385 | 846 | 299 | 77.66233766 | 1.83598907 | [CookBks,GeogBks] | [ChildBks] |  |
| **Rule 13** | 390 | 862 | 299 | 76.66666667 | 1.77880897 | [ChildBks,GeogBks] | [CookBks] |  |
| **Rule 7** | 368 | 862 | 292 | 79.34782609 | 1.84101685 | [ChildBks,DoItYBks] | [CookBks] |  |
| **Rule 8** | 375 | 846 | 292 | 77.86666667 | 1.84081954 | [CookBks,DoItYBks] | [ChildBks] |  |
| **Rule 6** | 324 | 846 | 258 | 79.62962963 | 1.88249715 | [YouthBks,CookBks] | [ChildBks] | Good Support with high confidence and lift ratio |
| **Rule 5** | 330 | 862 | 258 | 78.18181818 | 1.8139633 | [ChildBks,YouthBks] | [CookBks] |  |
| **Rule 11** | 325 | 862 | 253 | 77.84615385 | 1.80617526 | [ChildBks,ArtBks] | [CookBks] |  |
| **Rule 12** | 334 | 846 | 253 | 75.74850299 | 1.79074475 | [CookBks,ArtBks] | [ChildBks] |  |
| **Rule 9** | 303 | 862 | 245 | 80.85808581 | 1.87605768 | [ChildBks,RefBks] | [CookBks] |  |
| **Rule 10** | 305 | 846 | 245 | 80.32786885 | 1.89900399 | [CookBks,RefBks] | [ChildBks] |  |
| **Rule 4** | 227 | 862 | 227 | 100 | 2.32018561 | [ItalCook] | [CookBks] | 100% confidence level with very good lift ratio and reasonable support |
| **Rule 18** | 265 | 862 | 217 | 81.88679245 | 1.89992558 | [DoItYBks,GeogBks] | [CookBks] |  |
| **Rule 15** | 265 | 846 | 209 | 78.86792453 | 1.86448994 | [DoItYBks,GeogBks] | [ChildBks] |  |
| **Rule 19** | 255 | 862 | 207 | 81.17647059 | 1.88344479 | [ArtBks,GeogBks] | [CookBks] |  |
| **Rule 16** | 255 | 846 | 204 | 80 | 1.89125296 | [ArtBks,GeogBks] | [ChildBks] |  |
| **Rule 17** | 247 | 862 | 203 | 82.18623482 | 1.9068732 | [DoItYBks,ArtBks] | [CookBks] |  |

**Final Statement to Board**

**Generally, it can be said that the following books are often bought together\*\*\*\*:**

**1. If Geography books then there is a 71% chance that a children’s book will be purchased.**

**2. If Italcook is purchased, then there is a 100% chance that a Cook book will be bought.**

**3. If Youth books and cook books are bought together then there is 80% chance that Children books will be bought.**

\*\*\*Based on the data available

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.

**Top 3 Rules Highlighted in red (Selection Criteria: High Support, High Confidence, High Lift Ratio) Min Confidence Selected: 70%. Ensured supports are not very low.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rule ID** | **A-Support** | **C-Support** | **Support** | **Confidence** | **Lift-Ratio** | **Antecedent** | **Consequent** | **Selection Criteria** |
| **Rule 3** | 488 | 600 | 366 | 75 | 1.25 | [cracker] | [heineken] | High Support, >70% confidence and >1 lift-ratio |
| **Rule 2** | 318 | 600 | 257 | 80.81761006 | 1.34696017 | [soda] | [heineken] |  |
| **Rule 4** | 305 | 600 | 252 | 82.62295082 | 1.37704918 | [artichok] | [heineken] |  |
| **Rule 1** | 318 | 488 | 251 | 78.93081761 | 1.61743479 | [soda] | [cracker] |  |
| **Rule 33** | 318 | 366 | 234 | 73.58490566 | 2.01051655 | [soda] | [cracker,heineken] |  |
| **Rule 35** | 257 | 488 | 234 | 91.05058366 | 1.86579065 | [soda,heineken] | [cracker] |  |
| **Rule 34** | 251 | 600 | 234 | 93.22709163 | 1.55378486 | [soda,cracker] | [heineken] | Reasonable supprt with high confidence and >1 lift ratio |
| **Rule 6** | 296 | 312 | 220 | 74.32432432 | 2.38218988 | [coke] | [ice\_crea] |  |
| **Rule 7** | 312 | 296 | 220 | 70.51282051 | 2.38218988 | [ice\_crea] | [coke] |  |
| **Rule 5** | 282 | 472 | 220 | 78.0141844 | 1.65284289 | [turkey] | [olives] |  |
| **Rule 29** | 288 | 392 | 214 | 74.30555556 | 1.89554989 | [heineken,hering] | [baguette] |  |
| **Rule 27** | 261 | 485 | 214 | 81.99233716 | 1.69056365 | [baguette,heineken] | [hering] |  |
| **Rule 28** | 249 | 600 | 214 | 85.9437751 | 1.43239625 | [baguette,hering] | [heineken] |  |
| **Rule 77** | 255 | 390 | 201 | 78.82352941 | 2.02111614 | [hering,olives] | [corned\_b] |  |
| **Rule 79** | 236 | 485 | 201 | 85.16949153 | 1.75607199 | [olives,corned\_b] | [hering] |  |
| **Rule 78** | 244 | 472 | 201 | 82.37704918 | 1.74527647 | [hering,corned\_b] | [olives] |  |
| **Rule 50** | 249 | 305 | 199 | 79.91967871 | 2.62031733 | [avocado,heineken] | [artichok] | Reasonable supprt with high confidence and high lift ratio |

**Final Statement to Board**

**Generally, it can be said that the following grocery items are often bought together\*\*\*:**

**1. If Cracker is bought then there is 75% chance that Heineken will be bought.**

**2. If soda and cracker are bought together then there is a 93% chance that Heineken will be bought.**

**3. If avocado and Heineken are bought together then there is 80% chance that Artichok will be bought.**

\*\*\*Based on the data available

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.

**Top 3 Rules Highlighted in red (Selection Criteria: High Support, High Confidence, High Lift Ratio) Min Confidence Selected: 80%. Ensured supports are not very low.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rule ID** | **A-Support** | **C-Support** | **Support** | **Confidence** | **Lift-Ratio** | **Antecedent** | **Consequent** | **Selection Criteria** |
| **Rule 2** | 357 | 381 | 321 | 89.915966 | 2.3599991 | [Mascara] | [Eye shadow] | High Support, Good confidence level and good lift-ratio |
| **Rule 3** | 381 | 357 | 321 | 84.251969 | 2.3599991 | [Eye shadow] | [Mascara] |  |
| **Rule 47** | 201 | 357 | 179 | 89.054726 | 2.4945302 | [Concealer,Eye shadow] | [Mascara] |  |
| **Rule 46** | 204 | 381 | 179 | 87.745098 | 2.3030209 | [Concealer,Mascara] | [Eye shadow] |  |
| **Rule 20** | 182 | 357 | 169 | 92.857143 | 2.6010404 | [Blush,Eye shadow] | [Mascara] | Good confidence level and good lift-ratio with reasonable support |
| **Rule 19** | 184 | 381 | 169 | 91.847826 | 2.4107041 | [Blush,Mascara] | [Eye shadow] |  |
| **Rule 57** | 192 | 381 | 166 | 86.458333 | 2.2692476 | [Mascara,Foundation] | [Eye shadow] |  |
| **Rule 58** | 181 | 381 | 158 | 87.292818 | 2.2911501 | [Mascara,Lip Gloss] | [Eye shadow] |  |
| **Rule 61** | 175 | 381 | 151 | 86.285714 | 2.2647169 | [Mascara,Eyeliner] | [Eye shadow] |  |
| **Rule 62** | 182 | 357 | 151 | 82.967033 | 2.3240065 | [Eye shadow,Eyeliner] | [Mascara] |  |
| **Rule 1** | 149 | 280 | 149 | 100 | 3.5714286 | [Brushes] | [Nail Polish] | Very good confidence level and good lift-ratio with reasonable support |

**Final Statement to Board**

**Generally, it can be said that the following cosmetics are often bought together\*\*\*:**

**1. If Mascara is bought then there is a 90% chance that Eyeshadow will be bought.**

**2. If Brushes are bought then there is 100% chance that nailpolish will be bought.**

**3. If blush and eye shadow are bought together then there is 93% chance that Mascara will be bought.**

\*\*\*Based on the data available

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.

**Top 3 Rules Highlighted in red (Selection Criteria: High Support, High Confidence, High Lift Ratio) Min Confidence Selected: 70%. Ensured supports are not very low.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Rule ID** | **A-Support** | **C-Support** | **Support** | **Confidence** | **Lift-Ratio** | **Antecedent** | **Consequent** | **Selection Criteria** |
| **Rule 2** | 20 | 144 | 14 | 70 | 1.77430556 | [Regression,SW] | [Intro] | High support with 70% confidence and >1 lift-ratio |
| **Rule 1** | 11 | 144 | 9 | 81.81818182 | 2.07386364 | [Regression,DOE] | [Intro] | High support with reasonable confidence and >1 lift-ratio |
| **Rule 18** | 9 | 81 | 7 | 77.77777778 | 3.5048011 | [Intro,Regression,DOE] | [SW] |  |
| **Rule 19** | 7 | 144 | 7 | 100 | 2.53472222 | [Regression,DOE,SW] | [Intro] | 100% confidence and >1 lift-ratio with reasonable support |
| **Rule 4** | 8 | 76 | 6 | 75 | 3.60197368 | [Intro,DataMining,Cat Data] | [Regression] |  |
| **Rule 7** | 7 | 65 | 5 | 71.42857143 | 4.01098901 | [Intro,Regression,Forecast] | [DataMining] |  |

**Generally, it can be said that the following courses are often bought together\*\*\*:**

**1. If regression and SW are bought together then there is a 70% chance that the Intro course will be bought.**

**2. If regression and DOE are bought together then there is a 82% chance that the Intro course will be bought.**

**3. If regression, SW and DOE are bought together then there is 100% chance that the Intro course will be bought.**

\*\*\*Based on the data available

**Submit your answers for all the 4 datasets a word document along with 4 Excel spreadsheets on Canvas.**