**502571-3 2nd Trimester 2022/2023 HW#4**

# Topics: ARM Sections: 2233 & 3827

# Due Date: Monday 13/02/2023 – 2:00 pm

# \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Name: ………………………………………

ID number: …………………………………

# \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Exercise 1:**

We want to use the ARM on the following Database representing seven transactions of many items: F, M B, E, O, A, H, S, D

|  |  |
| --- | --- |
| **Transaction number** | **items** |
| 1 | {F, M, B, E} |
| 2 | {O, F, E} |
| 3 | {O, A, H, S} |
| 4 | {D, B, F, E} |
| 5 | {D, A, H, S} |
| 6 | {O, M, E} |
| 7 | {O, A, D, H, S} |

1. Without calculation, could you list an association rule with a confidence equal to 100%. Justify your answer.

An example of ARM with a confidence of 100% is: F 🡪E

Justification: each transaction having the item F will also have the item E

1. Apply the A-Priori algorithm to determine the important association rules with an assumption that the minsup = 0,4 and the minconf = 0.9

Solution:

Step 1: generating the 1-itemset frequent pattern

|  |  |
| --- | --- |
| Itemset | Support |
| {F} | 3/7=42,8% |
| ~~{M}~~ | ~~2/7=28,6%~~ |
| ~~{B}~~ | ~~2/7=28,6%~~ |
| {E} | 4/7=57,1% |
| {0} | 4/7=57,1% |
| {A} | 3/7=42,8% |
| {H} | 3/7=42,8% |
| {S} | 3/7=42,8% |
| {D} | 3/7=42,8% |

C1

|  |
| --- |
| Itemset |
| {F} |
| {E} |
| {0} |
| {A} |
| {H} |
| {S} |
| {D} |

L1

Step 2: generating the 2-itemset frequent pattern

|  |  |
| --- | --- |
| Itemset | Support |
| {F, E} | 3/7=42,8% |
| ~~{F, O}~~ | ~~1/7=14,3%~~ |
| ~~{F, A}~~ | ~~0%~~ |
| ~~{F, H}~~ | ~~0%~~ |
| ~~{F, S}~~ | ~~0%~~ |
| ~~{F, D}~~ | ~~1/7=14,3%~~ |
| ~~{E, O}~~ | ~~2/7=28,6%~~ |
| ~~{E, A}~~ | ~~0%~~ |
| ~~{E, H}~~ | ~~0%~~ |
| ~~{E, S}~~ | ~~0%~~ |
| ~~{E, D}~~ | ~~1/7=14,3%~~ |
| ~~{O, A}~~ | ~~2/7=28,6%~~ |
| ~~{O, H}~~ | ~~2/7=28,6%~~ |
| ~~{O, S}~~ | ~~2/7=28,6%~~ |
| ~~{O, D}~~ | ~~1/7=14,3%~~ |
| {A, H} | 3/7=42,8% |
| {A, S} | 3/7=42,8% |
| ~~{A, D}~~ | ~~2/7=28,6%~~ |
| {H,S} | 3/7=42,8% |
| ~~{H, D}~~ | ~~2/7=28,6,3%~~ |
| ~~{S, D}~~ | ~~2/7=28,6%~~ |

C2



|  |  |
| --- | --- |
| Itemset | Support |
| {F, E} | 3/7=42,8% |
| {A, H} | 3/7=42,8% |
| {A, S} | 3/7=42,8% |
| {H,S} | 3/7=42,8% |

L2



Step 3: generating the 3-itemset frequent pattern

3.1. Join step:



C3 = L2 Join L2 = {{F, E, A}, {F, E, H}, {E, A, H}, {F, A, H}, {F, E, S}, {F, A, S}, {E, A, S}, {E, H, S}, {A, H, S}, {F, H, S}}



3.2. Prune step:

* {F, E, A} 🡪 2-itemsets: {F, A} does not occur in L2 so this itemset is removed



* {F, E, H} 🡪 2-itemsets: {F, H} does not occur in L2 so this itemset is removed



* {E, A, H} 🡪 2-itemsets: {E, A} does not occur in L2 so this itemset is removed
* {F, A, H} 🡪 2-itemsets: {F, H} does not occur in L2 so this itemset is removed
* {F, E, S} 🡪 2-itemsets: {E, S} does not occur in L2 so this itemset is removed
* {F, A, S} 🡪 2-itemsets: {F, A} does not occur in L2 so this itemset is removed
* {E, A, S} 🡪 2-itemsets: {E, A} does not occur in L2 so this itemset is removed
* {E, H, S} 🡪 2-itemsets: {E, H} does not occur in L2 so this itemset is removed



* {A, H, S} 🡪 2-itemsets: {A, H}, {H, S}, {A, S} 🡪 all occurs in L2 🡪 OK



|  |  |
| --- | --- |
| Itemset | Support |
| {A, H, S} | 3/7=42,8% |

C3



|  |  |
| --- | --- |
| Itemset | Support |
| {A, H, S} | 3/7=42,8% |

L3



Step 4: generating the 4-itemset frequent pattern



We cannot generate 4-itemset since we have only one 3-itemset in L3. Therefore, we stop the algorithm.



Step 5: Generating Association Rules for Supported (Frequent) Itemset

|  |  |  |
| --- | --- | --- |
| Rule | Confidence | Satisfy minconf (0,9) ? |
| A 🡪H,S | 3/3 = 100% | YES |
| H 🡪 A,S | 3/3 = 100% | YES |
| S 🡪 A,H | 3/3 = 100% | YES |
| A, H 🡪S | 3/3 = 100% | YES |
| A, S 🡪H | 3/3 = 100% | YES |
| H, S 🡪 A | 3/3 = 100% | YES |