

Homework 8

Exercise 1: Consider the data set shown in the following table.

Customer ID	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- Compute the support for itemsets {e}, {b, d}, and {b, d, e} by treating each transaction ID as a market basket.

$$\begin{aligned} \text{support } \{e\} &= \frac{4}{10} = 0.4 \\ \text{support } \{b, d\} &= \frac{2}{10} = 0.2 \\ \text{support } \{b, d, e\} &= \frac{2}{10} = 0.2 \end{aligned}$$

- Use the results in part (a) to compute the confidence for the association rules {b, d} → {e} and {e} → {b, d}. Is confidence a symmetric measure?

$$\begin{aligned} \text{confidence } \{b, d\} \rightarrow \{e\} &= \frac{G(b, d, e)}{G(b, d)} = \frac{2}{2} = 1 \\ \text{confidence } \{e\} \rightarrow \{b, d\} &= \frac{G(b, d, e)}{G(e)} = \frac{2}{4} = 0.25 \end{aligned}$$

∵ 1 ≠ 0.25 ∴ confidence is not a symmetric measure

- Repeat part (a) by treating each customer ID as a market basket. Each item should be treated as a binary variable (1 if an item appears in at least one transaction bought by the customer, and 0 otherwise.)

$$\begin{aligned}\text{Support } \{e\} &= \frac{4}{5} = 0.8 \\ \text{Support } \{b, d\} &= \frac{5}{5} = 1 \\ \text{Support } \{b, d, e\} &= \frac{4}{5} = 0.8\end{aligned}$$

- Use the results in part (c) to compute the confidence for the association rules $\{b, d\} \rightarrow \{e\}$ and $\{e\} \rightarrow \{b, d\}$.

$$\begin{aligned}\text{Confidence } \{b, d\} \rightarrow \{e\} &= \frac{4}{5} = 0.8 \\ \text{Confidence } \{e\} \rightarrow \{b, d\} &= \frac{4}{4} = 1\end{aligned}$$

Exercise 2. A database contains 6 transactions. Let minimum support = 50% and minimum confidence = 75%.

- Find all the frequent itemsets using Apriori. Please show every step Apriori (). For each frequent itemset, please indicate its support.

<table><tr><th>Trans_Id</th><th>Items</th></tr><tr><td>1</td><td>F, G, I, J</td></tr><tr><td>2</td><td>G, H, J</td></tr><tr><td>3</td><td>G, I, J</td></tr><tr><td>4</td><td>G, H, J</td></tr><tr><td>5</td><td>G, H, I, J</td></tr><tr><td>6</td><td>G, H, I</td></tr></table>	Trans_Id	Items	1	F, G, I, J	2	G, H, J	3	G, I, J	4	G, H, J	5	G, H, I, J	6	G, H, I	Scan D ----->	C1 <table><tr><th>Itemset</th><th>Support</th></tr><tr><td>F</td><td>1</td></tr><tr><td>G</td><td>6</td></tr><tr><td>H</td><td>4</td></tr><tr><td>I</td><td>4</td></tr><tr><td>J</td><td>5</td></tr></table>	Itemset	Support	F	1	G	6	H	4	I	4	J	5	L1 <table><tr><th>Itemset</th><th>Support</th></tr><tr><td>G</td><td>6</td></tr><tr><td>H</td><td>4</td></tr><tr><td>I</td><td>4</td></tr><tr><td>J</td><td>5</td></tr></table>	Itemset	Support	G	6	H	4	I	4	J	5
Trans_Id	Items																																						
1	F, G, I, J																																						
2	G, H, J																																						
3	G, I, J																																						
4	G, H, J																																						
5	G, H, I, J																																						
6	G, H, I																																						
Itemset	Support																																						
F	1																																						
G	6																																						
H	4																																						
I	4																																						
J	5																																						
Itemset	Support																																						
G	6																																						
H	4																																						
I	4																																						
J	5																																						
<table><tr><th>Itemset</th></tr><tr><td>{GH}</td></tr><tr><td>{GI}</td></tr><tr><td>{GJ}</td></tr><tr><td>{HI}</td></tr><tr><td>{HJ}</td></tr><tr><td>{IJ}</td></tr></table>	Itemset	{GH}	{GI}	{GJ}	{HI}	{HJ}	{IJ}	Scan D ----->	C2 <table><tr><th>Itemset</th><th>Support</th></tr><tr><td>{GH}</td><td>4</td></tr><tr><td>{GI}</td><td>4</td></tr><tr><td>{GJ}</td><td>5</td></tr><tr><td>{HI}</td><td>2</td></tr><tr><td>{HJ}</td><td>3</td></tr><tr><td>{IJ}</td><td>3</td></tr></table>	Itemset	Support	{GH}	4	{GI}	4	{GJ}	5	{HI}	2	{HJ}	3	{IJ}	3	L2 <table><tr><th>Itemset</th><th>Support</th></tr><tr><td>{GH}</td><td>4</td></tr><tr><td>{GI}</td><td>4</td></tr><tr><td>{GJ}</td><td>5</td></tr><tr><td>{HJ}</td><td>3</td></tr><tr><td>{IJ}</td><td>3</td></tr></table>	Itemset	Support	{GH}	4	{GI}	4	{GJ}	5	{HJ}	3	{IJ}	3			
Itemset																																							
{GH}																																							
{GI}																																							
{GJ}																																							
{HI}																																							
{HJ}																																							
{IJ}																																							
Itemset	Support																																						
{GH}	4																																						
{GI}	4																																						
{GJ}	5																																						
{HI}	2																																						
{HJ}	3																																						
{IJ}	3																																						
Itemset	Support																																						
{GH}	4																																						
{GI}	4																																						
{GJ}	5																																						
{HJ}	3																																						
{IJ}	3																																						
<table><tr><th>Itemset</th></tr><tr><td>{GHI}</td></tr><tr><td>{GHJ}</td></tr><tr><td>{GIJ}</td></tr><tr><td>{HIJ}</td></tr></table>	Itemset	{GHI}	{GHJ}	{GIJ}	{HIJ}	Scan D ----->	C3 <table><tr><th>Itemset</th><th>Support</th></tr><tr><td>{GHI}</td><td>2</td></tr><tr><td>{GHJ}</td><td>3</td></tr><tr><td>{GIJ}</td><td>3</td></tr><tr><td>{HIJ}</td><td>1</td></tr></table>	Itemset	Support	{GHI}	2	{GHJ}	3	{GIJ}	3	{HIJ}	1	L3 <table><tr><th>Itemset</th><th>Support</th></tr><tr><td>{GHJ}</td><td>3</td></tr><tr><td>{GIJ}</td><td>3</td></tr></table>	Itemset	Support	{GHJ}	3	{GIJ}	3															
Itemset																																							
{GHI}																																							
{GHJ}																																							
{GIJ}																																							
{HIJ}																																							
Itemset	Support																																						
{GHI}	2																																						
{GHJ}	3																																						
{GIJ}	3																																						
{HIJ}	1																																						
Itemset	Support																																						
{GHJ}	3																																						
{GIJ}	3																																						
<table><tr><th>Itemset</th></tr><tr><td>{GHIJ}</td></tr></table>	Itemset	{GHIJ}	Scan D ----->	C4 <table><tr><th>Itemset</th><th>Support</th></tr><tr><td>{GHIJ}</td><td>1</td></tr></table>	Itemset	Support	{GHIJ}	1																															
Itemset																																							
{GHIJ}																																							
Itemset	Support																																						
{GHIJ}	1																																						

- Give 3 examples of strong association rules that are discovered from the database. For each association rule example, indicate its support & confidence values.

$$\text{Confidence } \{H\} \rightarrow \{G\} = \frac{\text{support}(H, G)}{\text{support}(H)} = \frac{4}{4} = 1$$

$$\text{Confidence } \{J\} \rightarrow \{G\} = \frac{\text{support}(J, G)}{\text{support}(J)} = \frac{5}{5} = 1$$

$$\text{Confidence } \{IJ\} \rightarrow \{G\} = \frac{\text{support}(I, J, G)}{\text{support}(I, J)} = \frac{3}{3} = 1$$

Exercise 3. Repeat the Exercise 2, using RapidMiner

- Find all the frequent itemsets using FP-Growth operator. For each frequent itemset, please indicate its support. => Capture Screen

Size	Support	Item 1	Item 2	Item 3
1	1.000	G		
1	0.833	J		
1	0.667	H		
1	0.667	I		
2	0.833	G	J	
2	0.667	G	H	
2	0.667	G	I	
2	0.500	J	H	
2	0.500	J	I	
3	0.500	G	J	H
3	0.500	G	J	I

- Give 3 examples of strong association rules that are discovered from the database. For each association rule example, indicate its support & confidence values. => Capture Screen

No.	Premises	Conclusion	Support	Confiden... ↓
8	J	G	0.833	1
9	H	G	0.667	1
10	I	G	0.667	1