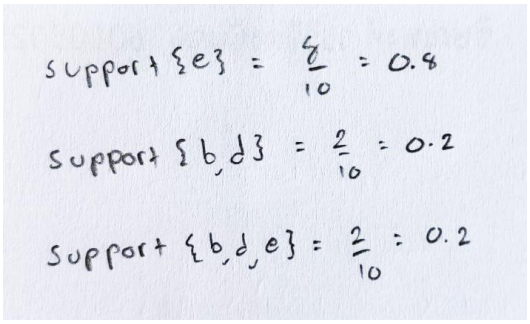
Homework 8

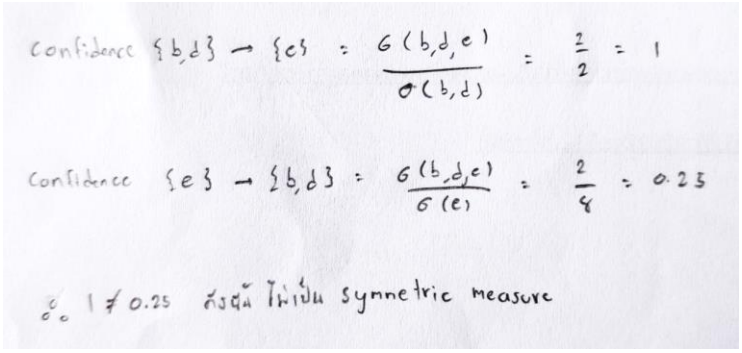
Exercise 1: Consider the data set shown in the following table.

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	$\{a, b, e\}$

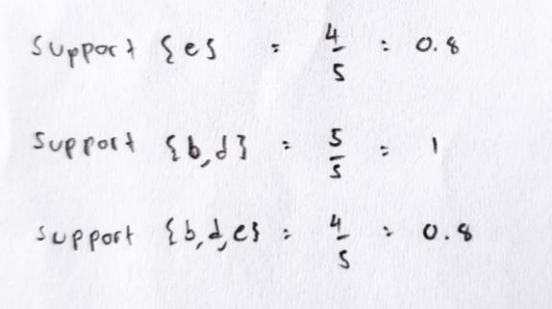
- Compute the support for itemsets {e}, {b, d}, and {b, d, e} by treating each transaction ID as a market basket.



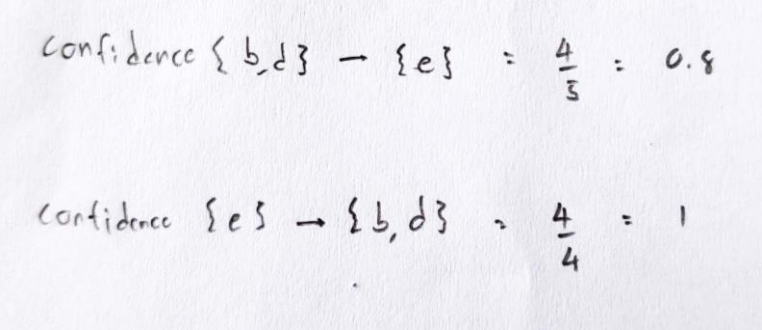
- Use the results in part (a) to compute the confidence for the association rules {b, d} ⟶ {e} and {e} ⟶ {b, d}. Is confidence a symmetric measure?



- 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.)



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



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.

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

Itemset	Support
F	1
G	6
H	4
I	4
J	5

L1

Itemset	Support
G	6
H	4
I	4
J	5

Itemset
{GH}
{GI}
{GJ}
{HI}
{HJ}
{IJ}

Scan D ---------->

C2

Itemset	Support
{GH}	4
{GI}	4
{GJ}	5
{HI}	2
{HJ}	3
{IJ}	3

L2

Itemset	Support
{GH}	4
{GI}	4
{GJ}	5
{HJ}	3
{IJ}	3

Itemset
{GHI}
{GHJ}
{GIJ}
{HIJ}

Scan D ---------->

C3

Itemset	Support
{GHI}	2
{GHJ}	3
{GIJ}	3
{HIJ}	1

L3

Itemset	Support
{GHJ}	3
{GIJ}	3

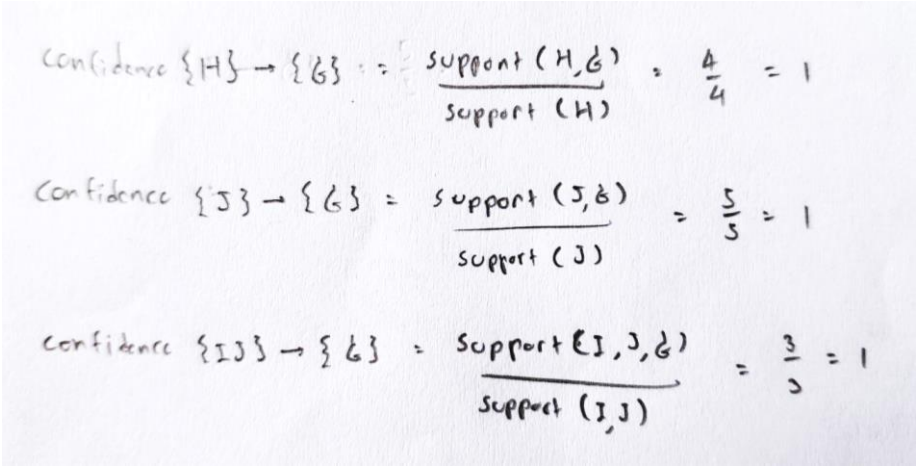
Itemset
{GHIJ}

Scan D ---------->

C4

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.



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