ASSIGNMENT-3

In this report we will be using apriori algorithm to calculate association rules for the sample dataset that we have taken

Transaction ID	ITEMS
1	Pizza,
2	Pizza, Garlic bread
3	Pizza, Coke, Wings
4	Wings, Coke, Chocolava
5	Garlic bread, Wings
6	Wings, Coke
7	Chocolava
8	Pizza, Garlic bread, Wings, Coke, Chocolava

Consider Pizza as item1, Garlic bread as item2, Coke as item3, Wings as item4, Chocolava as item5.

CASE1:

Let us consider minimum support count as 50%
minimum_support =50% => 0.5*len(dataset)=0.5*8=4. So, the
minimum_support=4 and the **minimum_confidence** =60%.

Step1: The first step is Create a table containing support count for each

Item present in dataset.

C1:

Itemset	Sup_Count
Item1	4
Item2	3
Item3	4
Item4	5
Item5	3

Step 2: Compare Candidate support count with minimum supportcount (In here **minimum_support**=4 if sup_count of candidate set items is less than **minimum_support** then remove those items).

L1:

Itemset	Sup_Count
Item1	4
Item3	4
Item4	5

Generate candidate set C2 using L1

C2:

Itemset	Sup_Count
Item1, Item3	2
Item1, Item4	2
Item3, Item4	4

L2:

Compare **C2** with minimum support and we generate L2.

Itemset	Sup_Count
Item3, Item4	4

Since, we only have one item left. So, we stop the process and determine the association rules for the L2.

Thus, we have discovered all the frequent item-sets. Now we are generating association rules for these items. For that we need to calculate confidence of each rule.

So, the subsets will be:

{Item3}
{Item4}

{Item3, Item4}.

Itemset {Item3, Item4} //from L2

SO, rules can be

Confidence for Item3->Item4=sup_count{Item3,Item4}/sup_count({Item3}).

I.E confidence=4/4=1

Here the confidence is greater than minimum confidence 1>0.6

Confidence for Item4->Item3=sup_count{Item3,Item4}/sup_count({Item4}).

I.E confidence=4/5=0.8

Here the confidence is greater than minimum confidence 0.8>0.6

Since the minimum confidence is 60%, and both rules confidence is greater than minimum confidence and they can be considered as strong association rules.

RULES Generated:
{Item3}->{Item4} which means {Coke}->{Wings}.

So, people who buy Coke has high chances of purchasing Wings.

{Item4}->{Item3} which means {Wings}->{Coke}.

So, people who buy Wings has high chances of purchasing Coke.

CASE2:

Let us consider minimum support count as 25%
minimum_support =25% => 0.25*len(dataset)=0.25*8=2. So, the
minimum_support=2 and the **minimum_confidence** = 70%.
Step1: The first step is Create a table containing support count for each

Item present in dataset.

C1:

Itemset	Sup_Count
Item1	4
Item2	3
Item3	4
Item4	5
Item5	3

Step 2: Compare Candidate support count with minimum supportcount (In here **minimum_support**=2 if sup_count of candidate set items is less than **minimum_support** then remove those items).

L1:

Itemset	Sup_Count
Item1	4
Item2	3
Item3	4
Item4	5
Item5	3

Generate candidate set C2 using L1

C2:

Itemset	Sup_Count
Item1, Item2	2
Item1, Item3	2
Item1, Item4	2
Item1, Item5	1
Item2, Item3	1
Item2, Item4	2
Item2, Item5	1
Item3, Item4	4
Item3, Item5	2
Item4, Item5	2

L2:

Compare **C2** with minimum support and we generate L2.

Itemset	Sup_Count
Item1, Item2	2
Item1, Item3	2
Item1, Item4	2
Item2, Item4	2
Item3, Item4	4
Item3, Item5	2
Item4, Item5	2

Generate candidate set C3 using L2

C3:

Itemset	Sup_Count
Item1, Item2, item3	1
Item1, Item2, item4	1
Item1, Item2, item5	1
Item1, item3, item4	2
Item1, Item3, item5	1
Item1, Item4, item5	1
Item2, item3, item4	1
Item2, Item3, item5	1
Item2, Item4, item5	1
Item3, Item4, item5	2

L3:

Compare **C3** with minimum support and we generate L3

Itemset	Sup_Count
Item1, Item3, Item5	**2**
Item3, Item4, Item5	**2**

Generate candidate set C4 using L3

C4:

Itemset	Sup_Count
Item1, Item3, Item4, Item5	1

Since the sup_count is less than the minimum support count we consider L3 to

generate association rules

Association Rules:
Itemset {I1, I3, I5} //from L3
SO rules can be

{Item1}-> {Item3, Item5}

confidence=support {Item1, Item3, Item5}/support{item1} =2/4=0.5

{Item3}-> {Item1, Item5}

confidence=support {Item1, Item3, Item5}/support{item3} =2/4=0.5

{Item5}-> {Item1, Item3}

confidence=support {Item1, Item3, Item5}/support{item5} =2/3=0.66

{Item3, Item5}->{Item1}

confidence=support {Item1, Item3, Item5}/support {item3, Item5} =2/2=1

{Item1, Item3}->{Item5}

confidence=support {Item1, Item3, Item5}/support {item1, Item3} =2/2=1

{Item1, Item5}->{Item3}
confidence=support {Item1, Item3, Item5}/support {item1, Item5} =2/1=2

Since the minimum confidence is 70%, then last 3 rules can be considered as

strong association rules.

Itemset {I3, I4, I5} //from L3
SO, rules can be

{Item3}-> {Item4, Item5}

confidence=support {Item3, Item4, Item5}/support{item3} =2/4=0.5

{Item4}-> {Item3, Item5}

confidence=support {Item3, Item4, Item5}/support{item4} =2/5=0.4

{Item5}-> {Item3, Item5}

confidence=support {Item3, Item4, Item5}/support{item5} =2/3=0.66

{Item4, Item5}->{Item3}

confidence=support {Item3, Item4, Item5}/support {item4, Item5} =2/2=1

{Item3, Item5}->{Item4}

confidence=support {Item3, Item4, Item5}/support {item3, Item5} =2/2=1

{Item3, Item4}->{Item5}

confidence=support {Item3, Item4, Item5}/support {item3, Item4} =2/4=0.5

RULES Generated:

{Item3, Item5}->{Item1} which means {Coke, Chocolava}->{Pizza}

So, people who buy Coke, Chocolava has high chances of purchasing Pizza

{Item1, Item3}->{Item5} which means {Pizza, Coke}->{Chocolava}

So, people who buy Pizza, Coke has high chances of purchasing Chocolava

{Item1, Item5}->{Item3} which means {Pizza, Chocolava}-> {Coke}

So, people who buy Pizza, Chocolava has high chances of purchasing Coke

{Item4, Item5}->{Item3} which means {Wings, Chocolava}-> {Coke}

So, people who buy Wings, Chocolava has high chances of purchasing Coke

{Item3, Item5}->{Item4} which means {Coke, Chocolava}->{Wings}

So, people who buy Coke, Chocolava has high chances of purchasing Wings

CASE3:

Let us consider minimum support count as 40%
minimum_support = 40% => 0.40*len(dataset)=0.40*8=3.2. So, the
minimum_support=3.2 and the **minimum_confidence** = 50%.

Step1: The first step is Create a table containing support count for each

Item present in dataset.

C1:

Itemset	Sup_Count
Item1	4
Item2	3
Item3	4
Item4	5
Item5	3

Step 2: Compare Candidate support count with minimum support count (In here **minimum_support**=3.2 if sup_count of candidate set items is less than **minimum_support** then remove those items).

L1:

Itemset	Sup_Count
Item1	4
Item3	4
Item4	5

Generate candidate set C2 using L1

C2:

Itemset	Sup_Count
Item1, Item3	2
Item1, Item4	2
Item3, Item4	4

L2:

Compare **C2** with minimum support and we generate L2.

Itemset	Sup_Count
Item3, Item4	4

Since, we only have one item left. So, we stop the process and determine the association rules for the L2.

Thus, we have discovered all the frequent item-sets. Now we are generating association rules for these items. For that we need to calculate confidence of each rule.

So, the subsets will be:

{Item3}
{Item4}
{Item3, Item4}.

Itemset {Item3, Item4} //from L2

SO, rules can be

Confidence for Item3->Item4=sup_count{Item3,Item4}/sup_count({Item3}).

I.E confidence=4/4=1

Here the confidence is greater than minimum confidence 1>0.5

Confidence for Item4->Item3=sup_count{Item3,Item4}/sup_count({Item4}).

I.E confidence=4/5=0.8

Here the confidence is greater than minimum confidence 0.8>0.5

Since the minimum confidence is 50%, and both rules confidence is greater than minimum confidence and they can be considered as strong association rules.

RULES Generated:
{Item3}->{Item4} which means {Coke}->{Wings}.

So, people who buy Coke has high chances of purchasing Wings.

{Item4}->{Item3} which means {Wings}->{Coke}.

So, people who buy Wings has high chances of purchasing Coke.

INFERENCE:

From the above all three cases, we can see that

People who buy Wings has high probability of buying Coke and similarly those who buy Coke has high chances of buying Wings.

References:

https://www.geeksforgeeks.org/apriori-algorithm/

https://www.youtube.com/watch?v=wpCeFC8-z-k&t=14s