WEEK 9 LAB

Apriori and Association Rules

In this lab, I apply the Apriori algorithm to a dataset of five transactions. With a minimum support threshold of 60% and a minimum confidence threshold of 80%, the task required identifying frequent itemsets and deriving strong association rules.

1. A database has 5 transactions. Let's say min_sup = 60% and min_conf = 80%
   a) Find all frequent itemsets using Apriori
   b) List all of the strong association rules (with support s and confidence and lift)

   Answer => support calculation:
   Total transaction (T) =5
   Minimum support = 60%

   Step 1: in first step I separated each items appeared in the transaction and counting how many times each items are appearing in the five transaction and also calculating their support count as shown below in the table:

   apply apriori

   Counting the frequency of each item :

Items	Frequency(f)	Support count(f/t)*100%
Key-chain	5	100%
Eggs	4	80%
Onion	3	60%
Mango	3	60%
Yo-yo	3	60%
Nintendo	2	40%
Corn	2	40%
Doll	1	20%
Apple	1	20%
Umbrella	1	20%
Icecream	1	20%

Removing all the items which support count is below 60% and taking all the items above or equals to 60%

Support count>=60%:
Key-chain, eggs, onions, mango, yo-yo

Step 2: in this steps, taking all possible 2-items sets from the frequent single item sets from above table. Counting the total number of times appearing in the all five transactions and calculating their support count as shown in the table below:

candidate 2-itemsets (C2):

2-itemsets	frequency	Support count(f/t)*100%
Key-chain, Eggs	4	80%
Key-chain, Onion	3	60%
Key-chain, mango	3	60%
Key-chain, Yo-yo	3	60%
Eggs, onions	3	60%
Eggs, mangoes	2	40%
Eggs, yo-yo	2	40%
Onions, mango	1	20%
Onion, yo-yo	2	40%
Mango, yo-yo	2	40%

Taking only 2-itemsets which support count is more than equall to 60% as below:
Frequent pairs = suppoet count>=60%
Frequent candidate 2-itemsets :

Key-chain, Eggs = 4

Key-chain, Onion = 3

Key-chain, Mango = 3

Key-chain, Yo-yo = 3

Eggs, Onion = 3

Step 3 : in this step, taking all possible 3-itemssets from the all five transaction and calculating their total number of repetation and calculating support count for each 3-itemsets as shown in below table:

candidate 3-itemsets (C3)

items	frequency	Support count(f/t)*100%
Key-chain, Eggs, Onion	3	60%
Key-chain, Eggs, Mango	2	40%
Key-chain, Eggs, Yo-yo	2	40%
Key-chain, Mango, Yo-yo	2	40%

From the table only taking 3-itemsets which support count is more than 60% :

Frequent 3-itemsets (C3)>=60% :

Key-chain, Eggs, Onion =60%

Calculations and Rules: using 3-itemsets, I have converted to different subsets and after that calculated the confidence using formula confidence = (frequency of set B / frequency of set A) and lift (confidence/ support B) as shown in the table below:

Rules	Support	confidence	Lift(confidence/support B)
{key-chain, eggs} → {onions}	60%	¾ = 75%	75/60 = 1.25
{key-chain, onion} → {eggs}	60%	3/3 = 100%	100/80 = 1.25
{onion, eggs} → {key-chain}	60%	3/3 = 100%	100/100 = 1.0
{key-chain} → {eggs, onion}	60%	3/5 = 60%	60/60 = 1.0
{onion} → {key-chain, eggs}	60%	3/3 = 100%	100/80 =1.25
{eggs} → {onion, key-chain}	60%	¾ = 75%	75/60 = 1.25

Strong Rules : based on the 80% confidence threshold, the strong rule identifies these items are most sold and probability of purchasing these items are high:

Strong rules are those with Confidence ≥ 80%:

1. {Onion, Key-chain} → {Eggs}
   Support = 60%, Confidence = 100%, Lift = 1.25

2. {Onion, Eggs} → {Key-chain}
   Support = 60%, Confidence = 100%, Lift = 1.00

3. {Onion} → {Key-chain, Eggs}
   Support = 60%, Confidence = 100%, Lift = 1.25

Reflection:

This lab provides hands-on experience with the Apriori method, expanding my grasp of how association rules are mined from a dataset. The effectiveness of Apriori's pruning strategy in eliminating superfluous candidate sets was demonstrated by going through each stage, beginning with basic frequency counts and progressing to larger combinations.

I also witnessed how support, confidence, and lift each play a part in judging rules. Support assures that patterns are common enough to matter, confidence assesses reliability, and lift demonstrates whether a link is actually important or just due to chance.

The most useful insight was recognising that even with a tiny dataset, Apriori's systematic methodology can uncover associations that might not be immediately visible. This highlights how crucial data-driven decision-making is in fields like customer behaviour analysis, retail, and recommendation systems.

Overall, this exercise boosted my comprehension of data-mining topics and improved my ability to articulate and comprehend association rules in a straightforward and orderly fashion.