Experiment 6

Title: Demonstration of APRIORI algorithm along with frequent item sets, non-frequent item sets and strong & weak association rules.

1. What is the APRIORI algorithm in Data Mining?
The APRIORI algorithm is a classic algorithm used for mining frequent item sets and discovering association rules from large transactional datasets. It is particularly useful in market basket analysis, where the goal is to find which items are frequently bought together.

2. What is the frequent item set?
A frequent itemset is a group (or set) of items that appear together frequently in a dataset, such as a list of customer transactions.

3. What is association rule?
An association rule is a rule-based method used to discover interesting relationships, patterns, or correlations between items in a dataset, typically in transactional data like market basket analysis.

An association rule is an implication of the form: $A \Rightarrow B$
Where:
A = antecedent (if part)
B = consequent (then part)
This means: "If A is bought, then B is likely to be bought too."

4. Applications for APRIORI Algorithm & Association Rules
   1. Market Basket Analysis
   2. E-Commerce Recommendation Systems
   3. Healthcare and Medical Diagnosis
   4. Education and E-Learning
   5. Gaming and App Usage Analysis
   6. Fraud Detection
   7. Website Navigation Patterns

Example:

Transaction ID	Items
T1	Pen, Notebook, Eraser
T2	Pen, Pencil
T3	Pen, Notebook
T4	Pencil, Eraser
T5	Pen, Notebook, Pencil

Table.6.1

Minimum Support = 60% ≈2

Step 1: Candidate itemset of size 1 (In this step count of all available items in the given table and then make the table)

Item Sets	Support
{Pen}	4
{Notebook}	3
{Eraser}	2
{Pencil}	3

Table.6.2

Step 2: Frequent itemset of size 1 (In this step select those items which have 2 or more support in previous candidate itemset table)

Item Sets	Support
{Pen}	4
{Notebook}	3
{Eraser}	2
{Pencil}	3

Table.6.3

// In this step all items are selected because of no items there which have below 2 supports

Step 3: Candidate itemset of size 2 (In this step make a combination (2 items in each set) of items)

Item Sets	Support
{Pen, Notebook}	3
{Pen, Eraser}	1
{Pen, Pencil}	2
{Notebook, Eraser}	1
{Notebook, Pencil}	1
{Eraser, Pencil}	1

Table.6.4

//We highlighted rows because we are removing them in the next step

Step 4: Frequent itemset of size 2 (In this step select those items which have 2 or more support in previous candidate itemset table)

Item Sets	Support
{Pen, Notebook}	3
{Pen, Pencil}	2

Table.6.5

We removed the 4 combinations (Highlighted rows in step 3) because of their all-items support is below 2

Step 5: Candidate itemset of size 3 (In this step make a combination (3 items in each set) of items)

Item Sets	Support
{Pen, Notebook, Pencil}	1

Table.6.6

//We highlighted rows because we are removing them in the next step

No item set in size 3 which have 2 or more support, now we consider frequent itemset of size 2 (from step 4)

Largest frequent item sets

Item Sets	Support
{Pen, Notebook}	3
{Pen, Pencil}	2

Table.6.7

1. Association Rules from {Pen, Notebook}

Rule: {Pen} ⇒ {Notebook}

Support = 3/5 = 0.6

Confidence = Support (Pen ∪ Notebook) / Support (Pen)
= 3/4 = 75%

Lift = Confidence / Support (Notebook)
= 0.75 / (3/5) = 0.75 / 0.6 = 1.25

Rule: {Notebook} ⇒ {Pen}

Support = 3/5 = 0.6

Confidence = 3/3 = 100%

Lift = 1.00 / 0.6= 1.25

2. Association Rules from {Pen, Pencil}

Rule: {Pen} ⇒ {Pencil}

Support = 2/5 = 0.4

Confidence = 2/4 = 50%

Lift = 0.50 / (3/5) = 0.50 / 0.6 = 0.83

Rule: {Pencil} ⇒ {Pen}

Support = 2/5 = 0.4

Confidence = 2/3 = 66.66%

Lift = 0.67 / (4/5) = 0.67 / 0.8 = 0.83

Rule	Support	Confidence	Lift
{Pen} ⇒ {Notebook}	0.6	75%	1.25
{Notebook} ⇒ {Pen}	0.6	100%	1.25
{Pen} ⇒ {Pencil}	0.4	50%	0.83
{Pencil} ⇒ {Pen}	0.4	66.66%	0.83

Table.6.8

Rules with Lift > 1 show strong association.
Rules with Lift < 1 show weak association.

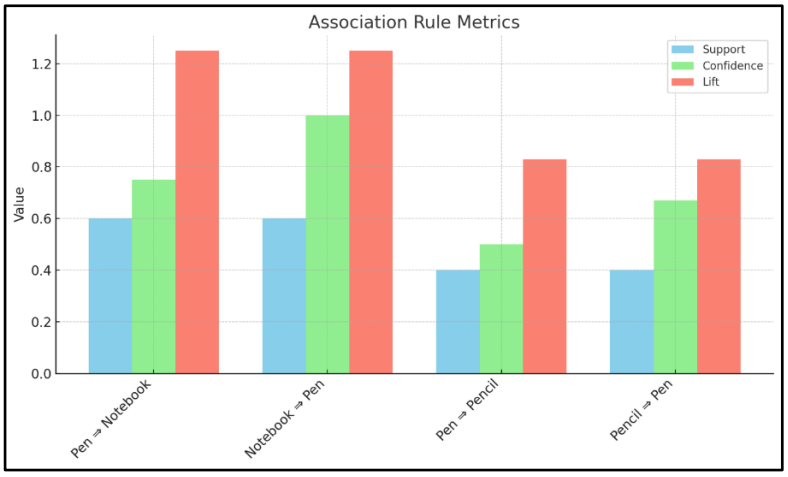

Fig.6.1