



Easy Engineering Classes – Free YouTube Lectures

EEC Classes GGSIPU, UPTU, Mumbai Univ., Pune Univ., GTU, Anna Univ., PTU and Others EEC Classes

ARM:-

Data Warehouse and Data Mining – Video Lecture Series (For B.Tech, MCA, M.Tech)

Association Rule Mining: (Rules)

"ARM", Also called as Market Basket Analysis (MBA) and Affinity Analysis.

→ Set of items in a transaction is called Market Basket.

→ Mostly used in RETAIL.

→ if 'A' then 'B' { $A \Rightarrow B$ }

Product Consequent
 Antecedent

= Support: (S). Percentage (%) of transactions (T) that contains both 'A' and 'B'.

$(A \Rightarrow B) = P(A \cap B)$ } measures frequency of association.

= Confidence: (C). In a transaction Set 'T' if 'C' is the % of times 'B' is present in all the transactions containing 'A'. (Strength).

$C = P(B|A) = \frac{P(A \cap B)}{P(A)}$ } Strength of association

Conditional Probability.

Parameters:-

- (i) Finding all items that appears frequently in transaction. } min. Support Count.
- (ii) Finding Strong associations among frequent items } Confidence.



Easy Engineering Classes – Free YouTube Lectures

EEC Classes GGSIPU, UPTU, Mumbai Univ., Pune Univ., GTU, Anna Univ., PTU and Others EEC Classes

ARM-2

Data Warehouse and Data Mining – Video Lecture Series (For B.Tech, MCA, M.Tech)

Problems in ARM:-

- i) Levels of frequency of appearance determination.
- ii) Finding strong associations among frequent items.

Functions of ARM:-

- i) Finding set of items - that has significant impact on business.
- ii) Collating infoⁿ from numerous tr^x.
- iii) Generating rules from counts in tr^x.

Strengths of ARM:-

- i) Easy interpretation.
- ii) Easy-to-start
- iii) Flexible data formats
- iv) Simplicity.

(1,2,3,4)

Weakness:-

- i) Exponential Growth in computations
- ii) lumping
- iii) Rule Selection
- iv) Rare items. } frequent items

(1,2), (1,3), (1,2,3) ...

Data Warehouse and Data Mining – Video Lecture Series (For B.Tech, MCA, M.Tech)

Apriori Algorithm: Idea is to generate candidate itemsets of a given size and then scan dataset to check if their counts are really large. The process is iterative.

- (i) All Singleton itemsets are Candidates in the first pass. Any items with less than specified Support Value is eliminated.

- (ii) Two member Candidate itemsets.
 - (iii) Three " " " " "
- eliminated.**

- (iv) Frequent itemsets constitutes set of frequent itemsets.

- (v) Generate Association Rules which have Confidence Values greater than or equal to Specified min. Confidence

50!-

Tid	Items
1	2, 3
2	1, 3, 5
3	1, 2, 4
4	2, 3

min Support = 2

Items Support

1	→	2	
2	→	3	—
3	→	3	
4	→	1	
5	→	1	

eliminated.

Itemsets Support

$\{1, 2\}$	$\rightarrow 1$
$\{1, 3\}$	$\rightarrow 1$
$\{2, 3\}$	$\rightarrow 2$

$$\{2, 3\} \rightarrow 2$$

Easy Engineering Classes – Free YouTube Lectures

For Engineering Students of GGSIPU, UPTU and Other Universities, Colleges of India

Data Warehousing and Data Mining- Lecture Series [Mumbai Univ, GTU, UPTU, GGSIPU, Pune Univ & others]

Ques.) For the following Given Transaction Data-Set, Generate Rules using Apriori Algorithm. Consider the Values as SUPPORT = 50% and CONFIDENCE = 75%.

$$\text{Support}(\text{Bread}) = \frac{n_{\text{Bread}}}{n}$$

Frequent Item Set

Transaction ID.	Items Purchased
1	Bread, cheese, Egg, Juice
2	Bread, cheese, Juice
3	Bread, Milk, Yogurt
4	Bread, Juice, Milk
5	cheese, Juice, Milk

<u>Items</u>	<u>Frequency</u>	<u>Support</u>
1. Bread →	4	→ 4/5 = 80%.
2. Cheese →	3	→ 3/5 = 60%.
Egg →	1	→ 1/5 = 20%.
3. Juice →	4	→ 4/5 = 80%.
4. Milk →	3	→ 3/5 = 60%.
Yogurt →	1	→ 1/5 = 20%.

Remove these ∵ their Support is less than 50%.

(APRIORI ALGORITHM SOLVED PROBLEM-1) (2)

Data Warehouse and Data Mining [Mumbai Univ, Pune Univ, GTU,]
Lecture Series [UPTU, GGSIPU, DU, PTU and other Universities]

Easy Engineering Classes – Free
YouTube Coaching

For Engineering Students of GGSIPU, UPTU and Other Universities,
Colleges of India

Make 2-Items Candidate Set and
Write their Frequency.

Item Pairs	Frequency	Support
(Bread, Cheese) → 2	→	2/5 = 40%.
(Bread, Juice) → 3	→	3/5 = 60%.
(Bread, Milk) → 2	→	2/5 = 40%.
(Cheese, Juice) → 3	→	3/5 = 60%.
(Cheese, Milk) → 1	→	1/5 = 20%.
(Juice, Milk) → 2	→	2/5 = 40%.

There Support
≥ 50%.

For Rules → (Bread, Juice) (1)

→ (Cheese, Juice) (2)

(1) (Bread, Juice)

(Bread → Juice) (Juice → Bread)

$$\text{Confidence (A} \rightarrow \text{B)} = \frac{\text{Support (A} \cup \text{B)}}{\text{Support (A)}}$$

$$1.) (\text{Bread} \rightarrow \text{Juice}) = \frac{S(\text{B} \cup \text{J})}{S(\text{B})} = \frac{3.5}{5.4} = \frac{3}{4} = 75\%$$

$$2.) (\text{Juice} \rightarrow \text{Bread}) = \frac{3.5}{5.4} = 75\%$$

$$(2) \rightarrow (\text{Cheese} \rightarrow \text{Juice}) = \frac{3.5}{5.3} = 100\%$$

$$\rightarrow (\text{Juice} \rightarrow \text{Cheese}) = \frac{3.5}{5.4} = 75\%$$

All the
Rules
are
Good.

[ASSOCIATION RULE MINING: APRIORI ALGORITHM SOLVED PROBLEM-2]

Easy Engineering Classes – Free YouTube Lectures

For Engineering Students of GGSIPU, UPTU and Other Universities, Colleges of India

Data Warehousing & Data Mining-lecture Series [Mumbai Univ, GTU, UPTU, GGSIPU, Pune Univ and others]

Ques.1) For the following given Transaction Data-Set, Generate rules using Apriori Algorithm. Consider the Values as SUPPORT = 22% and CONFIDENCE = 70%.

Frequent Item Set

Transaction Id.	Items Purchased
1	I ₁ , I ₂ , I ₅
2	I ₂ , I ₄
3	I ₂ , I ₃
4	I ₁ , I ₂ , I ₄
5	I ₁ , I ₃
6	I ₂ , I ₃
7	I ₁ , I ₃
8	I ₁ , I ₂ , I ₃ , I ₅
9	I ₁ , I ₂ , I ₃

All Items
Support \geq
22%

Item	Frequency	Support
I ₁	6	$6/9 = 66\%$
I ₂	7	$7/9 = 80\%$
I ₃	6	$6/9 = 66\%$
I ₄	2	$2/9 = 22.2\%$
I ₅	2	$2/9 = 22.2\%$

Generate Pairs of
itemset

(APRIORI ALGORITHM SOLVED PROBLEM-2) (2)

Data Warehouse and Data Mining [Mumbai Univ, GTU, UPTU]
Lecture Series [GGSIPU, Pune Univ, PTU and other University]

Easy Engineering Classes – Free
YouTube Coaching

For Engineering Students of GGSIPU, UPTU and Other Universities,
Colleges of India

(C2)

<u>Itemset Pairs</u>	<u>Frequency</u>	<u>Support</u>
$(I_1, I_2) \longrightarrow$	4	$4/q = 44.4\%$
$(I_1, I_3) \longrightarrow$	4	$4/q = 44.4\%$
$(I_1, I_4) \longrightarrow$	1	$1/q = 11.1\%$ (R)
$(I_1, I_5) \longrightarrow$	2	$2/q = 22.2\%$
$(I_2, I_3) \longrightarrow$	4	$4/q = 44.4\%$
$(I_2, I_4) \longrightarrow$	2	$2/q = 22.2\%$
$(I_2, I_5) \longrightarrow$	2	$2/q = 22.2\%$
$(I_3, I_4) \longrightarrow$	0	$0/q = 0\%$ (R)
$(I_3, I_5) \longrightarrow$	1	$1/q = 11.1\%$ (R)
$(I_4, I_5) \longrightarrow$	0	$0/q = 0\%$ (R)

(C3)

<u>(Itemset)</u>	<u>(Frequency)</u>	<u>(Support)</u>
$(I_1, I_2, I_3) \longrightarrow$	2	$2/q = 22.2\%$
$(I_1, I_2, I_5) \longrightarrow$	2	$2/q = 22.2\%$

Confidence.


- $\hookrightarrow (I_1, I_2) \rightarrow (I_5) = 2/4 = 50\% \rightarrow \times$
- $\hookrightarrow (I_1, I_5) \rightarrow (I_2) = 2/2 = 100\% \checkmark$
- $\hookrightarrow (I_2, I_5) \rightarrow (I_1) = 2/2 = 100\% \checkmark$
- $\hookrightarrow (I_1) \rightarrow (I_2, I_5) = 2/6 = 33\% \rightarrow \times$
- $\hookrightarrow (I_2) \rightarrow (I_1, I_5) = 2/7 = 29\% \rightarrow \times$
- $\hookrightarrow I_5 \rightarrow (I_2, I_1) = 2/2 = 100\% \checkmark$

Association Rules Exercise

- Here are a dozen sales transactions.
- The objective is to use this transaction data to find affinities between products, that is, which products sell together often.
- The support level will be set at 33 percent; the confidence level will be set at 50 percent.

Association Rules Exercise

Rule : $X \Rightarrow Y$


$$\text{Support} = \frac{\text{freq}(X, Y)}{N}$$
$$\text{Confidence} = \frac{\text{freq}(X, Y)}{\text{freq}(X)}$$

Transactions List

1	Milk	Egg	Bread	Butter
2	Milk	Butter	Egg	Ketchup
3	Bread	Butter	Ketchup	
4	Milk	Bread	Butter	
5	Bread	Butter	Cookies	
6	Milk	Bread	Butter	Cookies
7	Milk	Cookies		
8	Milk	Bread	Butter	
9	Bread	Butter	Egg	Cookies
10	Milk	Butter	Bread	
11	Milk	Bread	Butter	
12	Milk	Bread	Cookies	Ketchup

1-item Sets	Frequency
Milk	9
Bread	10
Butter	10
Egg	3
Ketchup	3
Cookies	5

Frequent 1-item Sets	Frequency
Milk	9
Bread	10
Butter	10
Cookies	5

Transactions List

1	Milk	Egg	Bread	Butter
2	Milk	Butter	Egg	Ketchup
3	Bread	Butter	Ketchup	
4	Milk	Bread	Butter	
5	Bread	Butter	Cookies	
6	Milk	Bread	Butter	Cookies
7	Milk	Cookies		
8	Milk	Bread	Butter	
9	Bread	Butter	Egg	Cookies
10	Milk	Butter	Bread	
11	Milk	Bread	Butter	
12	Milk	Bread	Cookies	Ketchup

2-item Sets	Frequency
Milk, Bread	7
Milk, Butter	7
Milk, Cookies	3
Bread, Butter	9
Butter, Cookies	3
Bread, Cookies	4

Frequent 2-item Sets	Frequency
Milk, Bread	7
Milk, Butter	7
Bread, Butter	9
Bread, Cookies	4

Transactions List

1	Milk	Egg	Bread	Butter
2	Milk	Butter	Egg	Ketchup
3	Bread	Butter	Ketchup	
4	Milk	Bread	Butter	
5	Bread	Butter	Cookies	
6	Milk	Bread	Butter	Cookies
7	Milk	Cookies		
8	Milk	Bread	Butter	
9	Bread	Butter	Egg	Cookies
10	Milk	Butter	Bread	
11	Milk	Bread	Butter	
12	Milk	Bread	Cookies	Ketchup

Milk, Bread, Butter, Cookies

3-item Sets	Frequency
Milk, Bread, Butter	6
Milk, Bread, Cookies	1
Bread, Butter, Cookies	3
Milk, Butter, Cookies	2

Frequent 3-item Sets	Frequency
Milk, Bread, Butter	6

Association Rule Mining - Subset Creation

- Frequent 3-Item Set = $I \Rightarrow \{\text{Milk, Bread, Butter}\}$
- Non-Empty subset are
 - $\{\{\text{Milk}\}, \{\text{Bread}\}, \{\text{Butter}\}, \{\text{Milk, Bread}\}, \{\text{Milk, Butter}\}, \{\text{Bread, Butter}\}\}$
- How to form Association Rule...?
 - For every non-empty subset S of I , the association rule is,
 - $S \rightarrow (I-S)$
 - If $\text{support}(I) / \text{support}(S) \geq \text{min_confidence}$

Association Rule Mining - Subset Creation

- Non-Empty subset are
 - $\{\{\text{Milk}\}, \{\text{Bread}\}, \{\text{Butter}\}, \{\text{Milk, Bread}\}, \{\text{Milk, Butter}\}, \{\text{Bread, Butter}\}\}$
 - $\text{Min_Support} = 30\%$ and $\text{Min_Confidence} = 60\%$
- Rule 1: $\{\text{Milk}\} \rightarrow \{\text{Bread, Butter}\}$ $\{S=50\%, C=66.67\%\}$
 - $\text{Support} = 6/12 = 50\%$
 - $\text{Confidence} = \text{Support}(\text{Milk, Bread, Butter}) / \text{Support}(\text{Milk}) = \frac{6/12}{9/12} = 6/9 = 66.67\% > 60\%$
 - Valid
- Rule 2: $\{\text{Bread}\} \rightarrow \{\text{Milk, Butter}\}$ $\{S=50\%, C=60\%\}$
 - $\text{Support} = 6/12 = 50\%$
 - $\text{Confidence} = \text{Support}(\text{Milk, Bread, Butter}) / \text{Support}(\text{Bread}) = 6/10 = 60\% \geq 60\%$
 - Valid

Association Rule Mining - Subset Creation

- Non-Empty subset are
 - $\{\{\text{Milk}\}, \{\text{Bread}\}, \{\text{Butter}\}, \{\text{Milk, Bread}\}, \{\text{Milk, Butter}\}, \{\text{Bread, Butter}\}\}$
 - $\text{Min_Support} = 30\%$ and $\text{Min_Confidence} = 60\%$
- Rule 3: $\{\text{Butter}\} \rightarrow \{\text{Milk, Bread}\}$ $\{S=50\%, C=60\%\}$
 - $\text{Support} = 6/12 = 50\%$
 - $\text{Confidence} = \text{Support}(\text{Milk, Bread, Butter}) / \text{Support}(\text{Butter}) = 6/10 = 60\% \geq 60\%$
 - Valid
- Rule 4: $\{\text{Milk, Bread}\} \rightarrow \{\text{Butter}\}$ $\{S=50\%, C=85.7\%\}$
 - $\text{Support} = 6/12 = 50\%$
 - $\text{Confidence} = \text{Support}(\text{Milk, Bread, Butter}) / \text{Support}(\text{Milk, Bread}) = 6/7 = 85.7\% > 60\%$
 - Valid

Association Rule Mining - Subset Creation

- Non-Empty subset are
 - $\{\{\text{Milk}\}, \{\text{Bread}\}, \{\text{Butter}\}, \{\text{Milk, Bread}\}, \{\text{Milk, Butter}\}, \{\text{Bread, Butter}\}\}$
 - $\text{Min_Support} = 30\%$ and $\text{Min_Confidence} = 60\%$
- Rule 5: $\{\text{Milk, Butter}\} \rightarrow \{\text{Bread}\}$ $\{S=50\%, C=85.7\%\}$
 - $\text{Support} = 6/12 = 50\%$
 - $\text{Confidence} = \text{Support}(\text{Milk, Bread, Butter}) / \text{Support}(\text{Milk, Butter}) = 6/7 = 85.7\% \geq 60\%$
 - Valid
- Rule 6: $\{\text{Bread, Butter}\} \rightarrow \{\text{Milk}\}$ $\{S=50\%, C=66.67\%\}$
 - $\text{Support} = 6/12 = 50\%$
 - $\text{Confidence} = \text{Support}(\text{Milk, Bread, Butter}) / \text{Support}(\text{Bread, Butter}) = 6/9 = 66.67\% \geq 60\%$
 - Valid