DWM
Assignment-3

1812401160 Vishwas Acharya
Page No.:
Date: / /

1) Explain Apriori algorithm with example

Ans - It is given by R. Agrawal and R. Srikant in 1994 for finding frequent itemsets in a dataset for boolean association rule.

- Name of the algorithm is Apriori because it uses prior knowledge of frequent itemset properties.
- We apply an iterative approach or level-wise search where k-frequent itemsets are used to find k+1 itemsets.
- To improve the efficiency of level-wise generation of frequent itemsets, an important property is called Apriori property which helps by reducing the search space.

- **Apriori Property**

↳ All non-empty subset of frequent itemset must be frequent.

↳ The key concept of Apriori algorithm is its anti-monotonicity of support measures.

- **Example:**

↳ Suppose we have the following dataset that has various transactions, and from this dataset, we need to find

the frequent itemsets and generate the association rules using the apriori algo.

TID	ITEMSETS
T_1	A, B
T_2	B, D
T_3	B, C
T_4	A, B, D
T_5	A, C
T_6	B, C
T_7	A, C
T_8	A, B, C, E
T_9	A, B, C

Given: Minimum Support = 2

Minimum Confidence = 50%

Ans:- $A \wedge B \rightarrow C$, $B \wedge C \rightarrow A$, $A \wedge C \rightarrow B$

- **Advantages**

↳ This is easy to understand algorithm
↳ The join and prune steps of the algorithm can be easily implemented on large datasets.

- **Disadvantages**

↳ This apriori algorithm works slow compared to other algorithms.
↳ The overall performance can be reduced as it scans the database for multiple times.

Page No.:
Date: / /

2) Applications of Market Basket Analysis

Ans Market basket analysis is applied to various fields of the retail sector in order to boost sales and generate revenue by identifying the needs of the customers and make purchase suggestion to them.

- **Cross Selling**:- It is a basically a sales technique in which seller suggests some related product to a customer after he buys a product.

- **Product Placement**:- It refers to placing the complimentary (pen and papers) and substitute goods (tea & coffee) together so that the customer addresses the goods and will buy both the goods together.

- **Affinity Promotion**:- It is a method of promotion that design promotional events based on associated products.

- **Fraud Detection**:- It may be possible to identify purchase behaviour that can associate with fraud on the basis of market basket analysis data that contain credit usage.

- **Customer Behaviour** :- Market basket analysis helps to understand customer behaviour. It understants the customer behaviour under different conditions. It provides an insight into customer behaviour.

3) Explain support and confidence for an data set with example.

Ans
- **Support** :- It is the frequency of A or how frequently an then appears in the dataset.
- It is defined as the fraction of the transaction T that contains the itemset X.
- If these are X datasets, then for transaction T, it can be written as

$$\text{support}(x) = \frac{\text{Freq}(x)}{T}$$

- **Confidence** :- It indicates how often the rule has been found to be true.
- Or how often the item x & y occur together in the dataset when the occurence of x is already given.
- It is the ratio of the transaction on that contains x & y to the number of records that contain t

Page No. :
Date : / /

$$\text{confidence} = \frac{\text{Freq}(X, Y)}{\text{Freq}(X)}$$

- Example:- Given a set of transactions, (dataset)

TID	Items
1	Bread, Milk
2	Bread, chocolate, Pepsi, eggs
3	Milk, chocolate, pepsi, coke
4	Bread, milk, chocolate, pepsi
5	Bread, milk, chocolate, coke

- Find support & confidence for {Milk, chocolate} => Pepsi

$$S = \frac{\sigma(\text{Milk, chocolate, Pepsi})}{|T|} = \frac{2}{3}$$

$S = 0.4$

$$C = \frac{\sigma(\text{Milk, chocolate, Pepsi})}{\sigma(\text{Milk, chocolate})} = \frac{2}{3}$$

$C = 0.67$