CLASSMATE
Date :
Page :

Name : Tejas Redkar
PRN : 10322 10937
Panel-C, Batch-C2
Roll No : PC-44

DEC Lab Assignment-6

* Aim : Apply a-priori algorithm to find frequency of occuring items from given data & generate strong association rules.

* Objective :
  - Learn frequent itemsets, closed itemsets, Market Basket Analysis
  - Calculate support, confidence & lift
  - To generate the association roles.

* Inputs : Dataset

* Outputs : Generating association rules.

* Platform : Windows - Version

* FAQ's

Q1) What is the a-priori principle?

Ans The a-priori algorithm is known as a priori knowledge. If an interest is frequent, then all of its subsets must be frequent.

Q2) What are the different steps involved in a-priori algorithm?

Ans 1) Initialize
2) Generate Frequent 1 - itemset
3) Generate frequent K - itemset (K>1)
4) Generate association rules
5) Calculate confidence
6) Prune weak rules.

Q3) What is a minimum support threshold? What is a minimum confidences threshold?

Ans 1) It is minimum frequency of occurence that an itemset most have in ~~common~~ a dataset to be considered "Frequent".
2) Itemset with support below this threshold are considered infrequent & are pruned during the mining process.
3) Minimum confidence threshold:
It is the minimum acceptable probability that a rule must have to be considered strong.

Rules below confidence ~~below~~ below this threshold are considered weak & may not provide meaningful insights.

Q4) How are the frequent item sets identified using the a-priori algorithm?

Ans 1) Initialize

2) Generate frequent 1-itemsets
3) Generate frequent k-itemsets (K>1)

The identified frequent itemsets are those sets of items that meet or exceed specified minimum support threshold.

The frequent itemsets capture patterns of co-occurences in dataset.

Q5) How are association rules generated using a-priori algorithm?

Ans 1) Generate Frequent itemsets
2) Generate candidate rules
3) Calculate confidence
4) Prune weak rules.

The remaining rules, after pruning are considered strong association rules.

Q6) How can the a-priori algorithm be used to detect credit card fraud in real time?

Ans 1) Transaction Representation
2) Set minimum support threshold to capture common patterns
3) Generate association rules with relevant confidence levels
4) Monitor Real-time transactions:

5) Evaluate the new transaction against the generates association rules.

Q7) How can star schema & concept hierarchies be used to analyse real-time sales data to identify trending products by product category, product subcategory & product brand in real time?

Ans Star Schema:

- Utilize a star schema with a central fact table containing real time sales data (measures like sales quantity)
- Connect this fact table to dimension table for product category, product subcategory, product brand, & time.
- Enables efficient & quick analysis by providing a clear structure for querying & aggregating sales data.

Concept Hierarchies:

- Create concept hierarchies within the product, dimension
- Product Category
- Product Subcategory
- Product Brand
- Allows for more easy drill down analysis from higher-level categories to more detailed brand insights in real time.