Devanshu Surana
PC-23, 1032210755
Panel C, Batch C1

DEC Lab Assignment-6

Aim- Apply a-priori algorithm to find frequently occuring items from given data and genrate strong association rules.

Objective:
- learn frequent itemsets, closed itemsets, Market basket analysis.
- Calculate support, confidence and lift
- To generate the association rules.

Theory:-
- A - priori Algorithm.

F_k - Frequent k itemsets
L_k - Candidate k itemsets.

Let k = 1
Generate F_1 = { Frequent 1-itemsets }
Repeat until F_k is empty

- Candidate Generation (Join Step): Generate L_{k+1} From F_k
- Support Counting Count the support of each Candidate in L_{k+1} by scanning the DB

- Candidate pruning: Prune candidate itemsets in L_{k+1} containing subsets of length k that are infrequent Eliminate candidates in L_{k+1} that are infrequent leaving only those that are frequent => F_{k+1}

Rule. Merge two Frequent itemsets with sizen if their first n-1 items are identical

- F_3 = [ABC, ABD, ABE, ACD, BCD, BDE, CDE)
  Merge (ABC, ABD) = ABCD
  Merge (ABC, ABE) = ABCE
  Merge (ABD, ABE) = ABDE
  We cant merge (ABD, ACD) because they share only prefix of length 1 instead of length 2

-- Association Rule Generation
- Given a Frequent itenset L, Find all non-empty subsets f⊂L such that f → L
- f satisfies the minimum confidence requirement.

If {A, B, C, D} is a frequent itemset, that condidate rules:

ABCD → D	ABD → C	ACD → B
A → BCD	B → ACD	C → ABD
AB → CD	AC → BD	AD → BC
BD → AC	CD → AB	

If (L) = k then there are $2^k - 2$ Candidate association rules (ignoring L → ∅ and ∅ → L)

Page No.
Date

Input: Datasets

Output: Generating association Rules

Plat form: Windows.

Conclusion: Thus, we have learn to perform a-priori Algorithm in python Pandas

FAQs.

1. What is the a-priori principle?

Ans. "If an itemset is frequent, then all of its subsets must also be frequent."
Apriori principle holds due to the following property of the support measure:
$\forall X, Y: (X \subseteq Y) \Rightarrow s(X) \geq s(Y)$

2. What are the different steps involved in the a priori algorithm?

Ans. Itemset generation Create candidate itemsets of L=1 by scanning the database for individual items.
- Frequent itemset generation count the support of these candidate itemsets in the db and retain only those itemsets that meet the min. support
- Join: Generate candidate itemsets of length (k+1) by joining pairs of Frequent itemsets of l=k
- Prune: Eliminate candidate itemsets that contain subsets of length k with low support as they are not frequent.

Repeat: Repeat the process by incrementing the length of itemsets untill no more frequent itemsets can be found

3. What is a minimum Support threshold? What is a minimum confidence threhold?

Ans. Minimum Support Threshold: Sets the minimum occurance Frequency for itemsets to be considered significant.
- Minimum Confidence Threshold: Estabilises the minimum confidence level for association rules to be Considered intersting.

4. How are frequent itemsets identified using a-prori algorithim?

Ans. Frequent itemsets are identified using A-prori algorithim by iteratively generating and checking candidate itemsets of increasing lengths, keeping only those that meet the minimum Support threshold This process continues untill no more frequent itemsets can be found.

5. How are association rules generated using a-priori algorithim?

Ans. To generate association rules A-priori Algo —
1. Find Frequent itemsets in the database.
2. Create candidate rules, from these itemsets.
3. Calculate the confidence for each-rules.
4. Keep rules that meet a minimum confidece threshold.
5. The retained rules are the generated acciation

rules, Showing item associations in th database.

6. How can be a-priori algorithm be used to ~~deb~~ detect credit card fraud in real time?

Ans. Detecting credit card froud in reel time involves

1- Data collection Gather transaction data in real time, including transaction amounts, user details.

2- Feature Extraction: Extract revlvant features from the transaction data.

3- Model Building:- utilize machine learning models for rules based systems and current transaction data.

4- Real time monitoing continuosly monitoring incoming transactions and compare them to the established models and rules

5- Anomaly Detection:- Identify transactions that devial significantly from established pattens.

6- Alert Generation: Flag Suspecious transa- cations for review.

7. Froud Preventions:- Take actions like blocking, notifying or encharing security.