Assignment No. 3

* Problem Statement:
  Apply a-Priori algorithm to find freq. occuring items from given data and generate strong association rules using support & confidence thresholds.

* Objective: Model associations between Products by determining sets of item freq. Purchased together & building association rules to drive recommendations.

* Outcome: Demonstrated market basket analysis using a-priori algorithm to find freq. occuring items from given data & generate strong association rules using support & confidence algo.

* Theory:
  Association rule mining finds interesting associations and relationships between large sets of data items. this rule shows how frequently an item occurs in a transaction. It is defined as an implementation expression of the form $X \rightarrow Y$ where X & Y are any 2 item sets.

Market Basket analysis is one of the key techniques used by large retailers to show association between items. It allows retailers to identify relationships between the items that people buy together freq, So that items can be strategically placed next to each other to boost sales.

Some definations :

↳ Support count : freq. of itemset occurence

↳ freq. Itemset : Itemset whose support >= minup threshold

The association rule evaluation metrix are

① Support (s) : The number of transactions that include items in the {x} & {y} parts of the rules as percentage of total number of transaction.

② confidence (c) : It is the ratio of number of transactions that includes all items in an itemset {A} as well as the transaction that include all items in {A} to the number of transaction that includ all items in {A}

③ Lift (l) : The lift of the rule $X \rightarrow Y$ is the confidence of the rule divided by

the expected confidence, assuming {X} & {Y} are indipendent of each other. expected confidence is confidence divided by freq. {Y}. Itemset occur together (i) as expected if lift = 1 (ii) more than expected if lift > 1 and (iii) less than expected if lift < 1

(A) Apriori Algorithm

→ It is an algorithm for freq. itemset mining & association rule learning over relational databases, first Possesed by Agrawal & Srikant in 1994.

It is designed to operate on databases containing transactions, with each transaction being a set of items (itemsets).

Given a threshold C, the algo identifies the itemsets, which are subsets of atleast C transactions in database

It uses a bottom-up approach, where frequent subsets are extended one at a time (called candidate generation), and groups of candidate are tasked against the data. the algo terminates when no other successful extension are found.

A-Priori used BFS & a hash tree to count candidate item sets efficiently. It generates K-itemsets of length k-1; then it prunes the candidate, which have

in-frequent sub pattern. After that it scans the transaction database, to determine freq. item sets among the candidate.

* Conclusion:

Thus frequently occuring items from given market basket dataset and strong association rules using support & confidence thresholds found using a-priori algorithm.