Name - KUMUD KAMBLE | **Div** - TY9 | **Roll no** - 24 | **Subject** - DWM | **Experiment No** - 5

Aim :- Implementation of Association Rule Mining algorithm (Apriori)

Introduction :-

- Association Rule Mining: Discovers relationships between items in a dataset.
- Apriori Algorithm: A popular algorithm for association rule mining based on frequent

itemsets. • Frequent Itemsets: Sets of items frequently occurring together in a dataset.

Two-Step Process:

- Frequent Itemset Generation: Finds itemsets meeting a minimum support threshold.
- Rule Generation: Creates association rules from frequent itemsets (e.g., "if milk, then bread").

Procedure :-

- Import the necessary libraries:
- Define a function to get frequent itemsets
- Define a function to generate candidate itemsets
- Define the Apriori algorithm
- Use the Apriori algorithm to find frequent itemsets

```
from itertools import combinations

# Function to get frequent itemsets based on minimum support
def get_frequent_itemsets(transactions, min_support):
    itemsets = {}
    for transaction in transactions:
        for item in transaction:
            if item in itemsets:
                itemsets[item] += 1
            else:
                itemsets[item] = 1

    # Filter itemsets to only include those that meet or exceed the minimum support
    frequent_itemsets = {item: support for item, support in itemsets.items() if support >= min_support}
    return frequent_itemsets

# Function to generate candidate itemsets of size k
def get_candidate_itemsets(frequent_itemsets, k):
    candidates = []
    frequent_items = list(frequent_itemsets.keys())
    for combination in combinations(frequent_items, k):
        candidates.append(combination)
    return candidates

# Apriori algorithm to find all frequent itemsets
def apriori(transactions, min_support):
    k = 1
    # Initial set of frequent itemsets
    frequent_itemsets = get_frequent_itemsets(transactions, min_support)
    all_frequent_itemsets = [frequent_itemsets]

    # Iterate to find larger itemsets
    while frequent_itemsets:
        k += 1
        # Generate candidate itemsets of size k
        candidates = get_candidate_itemsets(frequent_itemsets, k)
        candidate_supports = {candidate: 0 for candidate in candidates}
        # Calculate support for each candidate itemset
        for transaction in transactions:
            for candidate in candidates:
                if set(candidate).issubset(set(transaction)):
                    candidate_supports[candidate] += 1

        # Filter candidate itemsets to only include those that meet or exceed the minimum support
        frequent_itemsets = {itemset: support for itemset, support in candidate_supports.items() if support >= min_s
        if frequent_itemsets:
            all_frequent_itemsets.append(frequent_itemsets)
    return all_frequent_itemsets
# Example usage
transactions = [
    ['milk', 'bread', 'butter'],
    ['bread', 'butter'],
    ['milk', 'bread'],
    ['milk', 'butter'],
    ['bread', 'butter'],
    ['milk', 'bread', 'butter']
]

min_support = 2
frequent_itemsets = apriori(transactions, min_support)
print(frequent_itemsets)
```

```
[{'milk': 4, 'bread': 5, 'butter': 5}, {('milk', 'bread'): 3, ('milk', 'butter'): 3, ('bread', 'butter'): 4}]
```

Review Questions:

1)What is the Apriori algorithm in Association Rule Mining?

Ans:It is an algorithm that finds frequent itemsets using support and generates association rules based on confidence and lift.

2) What is the significance of support, confidence, and lift in

Apriori? Ans:Support: Measures frequency of an itemset.

Confidence: Indicates the reliability of a rule.

Lift: Evaluates the rule's importance compared to random chance.

Conclusion :- The Apriori algorithm effectively identifies frequent itemsets and generates association rules from transactional data. By using a minimum support threshold, it efficiently finds significant relationships between items. Implementing it in Python allows for a structured approach, uncovering hidden patterns and leading to valuable insights for decision-making in various domains. The Apriori algorithm is a fundamental and practical technique for association rule mining, offering a powerful solution for discovering knowledge from large datasets.