

## Lab-2 Association Mining

### Introduction:

Association mining is a data mining technique used to discover interesting relationships, patterns, or associations among items in large datasets. In this lab, association rule mining is performed using the Apriori algorithm on a small transactional dataset. The process includes transforming the data into one-hot encoded format, identifying frequent itemsets with minimum support, and generating strong association rules based on confidence and lift.

### Implementation Code:

```
import pandas as pd
from mlxtend.frequent_patterns import apriori, association_rules
#Step1:Load Datasets
data = {
    'TransactionID': [1, 2, 3, 4, 5],
    'Items': [
        ['Bread', 'Butter'],
        ['Milk', 'Diaper', 'Juice', 'Cookies'],
        ['Butter', 'Diaper', 'Juice', 'Soda'],
        ['Bread', 'Milk', 'Diaper', 'Juice'],
        ['Bread', 'Butter', 'Diaper', 'Soda']
    ]
}
df = pd.DataFrame(data)
print("Kishor Lab-2")
print("Initial Data:\n", df)
# Step 2: Convert dataset into a format suitable for the Apriori algorithm# Convert the list of items into one-hot encoded format
df_items = df['Items'].apply(lambda x: pd.Series(1, index=x).fillna(0))
print("\nOne-Hot Encoded Data:\n", df_items)
# Step 3: Apply the Apriori algorithm to find frequent itemsets# Use a minimum support threshold of 0.6 (at least 60% of transactions)
frequent_itemsets = apriori(df_items, min_support=0.6, use_colnames=True)
print("\nFrequent Itemsets:\n", frequent_itemsets)
# Step 4: Generate association rules from the frequent itemsets# Use a minimum confidence threshold of 0.7 (at least 70% confidence)
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.7)
print("\nAssociation Rules:\n", rules)
# Step 5: Interpret the results# Display the rules in a simple format
for _, row in rules.iterrows():
    print(f"\nRule: {set(row['antecedents'])} -> {set(row['consequents'])}")
    print(f"Support: {row['support']:.2f}")
    print(f"Confidence: {row['confidence']:.2f}")
    print(f"Lift: {row['lift']:.2f}")
```

### Output SnapShot:

```
Kishor Lab-2
Initial Data:
  TransactionID  Items
0             1  [Bread, Butter]
1             2  [Milk, Diaper, Juice, Cookies]
2             3  [Butter, Diaper, Juice, Soda]
3             4  [Bread, Milk, Diaper, Juice]
4             5  [Bread, Butter, Diaper, Soda]

One-Hot Encoded Data:
   Bread  Butter  Milk  Diaper  Juice  Cookies  Soda
0     1.0    1.0    0.0    0.0    0.0    0.0    0.0
1     0.0    0.0    1.0    1.0    1.0    1.0    0.0
2     0.0    1.0    0.0    1.0    1.0    0.0    1.0
3     1.0    0.0    1.0    1.0    1.0    0.0    0.0
4     1.0    1.0    0.0    1.0    0.0    0.0    1.0
```

# Frequent Itemsets:

	support	itemsets
0	0.6	(Bread)
1	0.6	(Butter)
2	0.8	(Diaper)
3	0.6	(Juice)
4	0.6	(Diaper, Juice)

## Association Rules:

	antecedents	consequents	antecedent support	consequent support	support \
0	(Diaper)	(Juice)	0.8	0.6	0.6
1	(Juice)	(Diaper)	0.6	0.8	0.6

	confidence	lift	representativity	leverage	conviction	zhangs_metric	\
0	0.75	1.25	1.0	0.12	1.6	1.0	
1	1.00	1.25	1.0	0.12	inf	0.5	

	jaccard	certainty	kulczynski
0	0.75	0.375	0.875
1	0.75	1.000	0.875

Rule: {'Diaper'} -> {'Juice'}

Support: 0.60

Confidence: 0.75

Lift: 1.25

Rule: {'Juice'} -> {'Diaper'}

Support: 0.60

Confidence: 1.00

Lift: 1.25