Darshan UNIVERSITY
योग: कर्मसु कौशलम्

Data Mining

Lab - 7 (Part 2)

Name: Smit Maru

Enrollment No: 23010101161

Step 1: Load the Dataset

Load the `Tdata.csv` file and display the first few rows.

In [1]:
```
import pandas as pd

df = pd.read_csv("Tdata.csv")
print(df.head())
```

```
  Transaction  bread  butter  coffee  eggs  jam  milk
0          T1      1       1       0     0    0     1
1          T2      1       1       0     0    1     0
2          T3      1       0       0     1    0     1
3          T4      1       1       0     0    0     1
4          T5      1       0       1     0    0     0
```

Step 2: Drop the 'Transaction' Column

We're only interested in the items (not the transaction IDs).

In [2]:
```
df = df.drop(columns=['Transaction'])
```

Step 3: Count Single Items

See how many transactions include each item.

In [3]:
```
item_counts = df.sum()
print(item_counts)
```

```
bread     5
butter    3
coffee    2
eggs      2
jam       2
milk      3
dtype: int64
```

Step 4: Define Apriori Function

This function finds frequent itemsets of size 1, 2, and 3 with minimum support.

In [4]:
```
from itertools import combinations
import pandas as pd

def apriori(df, min_support=0.6):
    num_txns = len(df)
    items = df.columns
    freq_itemsets = []

    L1 = []
    for item in items:
        support = df[item].sum() / num_txns
        if support >= min_support:
            L1.append((frozenset([item]), support))
    freq_itemsets.extend(L1)

    L2 = []
    for itemset in combinations([i[0] for i in L1], 2):
        union_items = itemset[0].union(itemset[1])
        support = (df[list(union_items)].sum(axis=1) == len(union_items)).mean()
        if support >= min_support:
            L2.append((union_items, support))
    freq_itemsets.extend(L2)

    L3 = []
    for itemset in combinations([i[0] for i in L1], 3):
        union_items = itemset[0].union(itemset[1]).union(itemset[2])
        support = (df[list(union_items)].sum(axis=1) == len(union_items)).mean()
        if support >= min_support:
            L3.append((union_items, support))
    freq_itemsets.extend(L3)

    return freq_itemsets

frequent_itemsets = apriori(df, min_support=0.6)

result_df = pd.DataFrame(frequent_itemsets, columns=["Itemset", "Support"])
print(result_df)
```

```
   Itemset   Support
0  (bread)  0.833333
```

Step 5: Run Apriori

Set `min_support = 0.6` and display the frequent itemsets.

In [5]:
```
frequent_itemsets = apriori(df, min_support=0.6)
print(frequent_itemsets)
```

```
[(frozenset({'bread'}), 0.8333333333333334)]
```

Step 6 Display as a DataFrame

In [6]:
```
import pandas as pd

frequent_itemsets = apriori(df, min_support=0.6)

result_df = pd.DataFrame(frequent_itemsets, columns=["Itemset", "Support"])
print(result_df)
```

```
   Itemset   Support
0  (bread)  0.833333
```

Orange Tool : - >Generate Same Frequent Patterns in Orange tools

Extra : - > Define Apriori Function without itertools