

## **OBJECTIVES**

### **ASSOCIATION RULE MINING:**

Association rule mining is a data mining approach used to explore and interpret large transactional datasets to identify unique patterns and rules. During transactions, these patterns define interesting relationships and interactions between different items.

### **DOMAIN OF ASSOCIATION RULE MINING:**

The association rules are useful for analysing and predicting customer behaviour. They play an important part in customer analysis, market-based analysis, product clustering, catalogue design and store layout.

### **BENEFITS FROM ASSOCIATION RULE MINING:**

- Applying the algorithms to supermarkets, the scientists were able to discover links between different items purchased, called association rules, and ultimately use that information to predict the likelihood of different products being purchased together.
- For retailers, association rule mining offered a way to better understand customer purchase behaviours. Because of its retail origins, association rule mining is often referred to as market basket analysis.

### **USING PATTERNS IN ASSOCIATION RULE MINING:**

Frequent Pattern Mining (aka Association Rule Mining) is an analytical process that finds frequent patterns, associations, or causal structures from data sets. Given a set of transactions, this process aims to find the rules that enable us to predict the occurrence of a specific item based on the occurrence of other items in the transaction.

## **DATASET**

Groceries dataset downloaded from [Kaggle](#) and The dataset has 38765 rows of the purchase orders of people from the grocery stores.

Dataset has 3 columns:

- Member\_number: a number that represents a person or to uniquely identify a person or buyer
- Date: date at which the particular item was bought
- ItemDescription: Name of the item

([https://www.kaggle.com/heeraldedhia/groceries-dataset?select=Groceries\\_dataset.csv](https://www.kaggle.com/heeraldedhia/groceries-dataset?select=Groceries_dataset.csv))

## **PREPROCESSING**

- The dataset contained 38765 rows of purchases, Member\_number associated for each row
- First had to group by Member\_number and then find a unique set of Member\_number's
- Find the list of items bought by a particular person
- Then store these items as a transaction made by a particular person
- Now we have a list of transactions, where each row represents a transaction

## **REASONING:**

- This dataset will be perfect for association rule mining because association mining is useful for analyzing and predicting customer behaviour and our dataset deals with items bought by customers
- The preprocessing is done to get Member\_number and the items bought by that particular person

## **RULE OF MINING**

The parameters considered for this rule mining process are

1. Support
2. Confidence
3. Lift

## CHOICE OF ALGORITHM

The algorithm chosen for this association mining for the retail dataset is the Apriori algorithm. Apriori is an algorithm for frequent itemset mining and association rule learning over relational databases. Apriori is designed to operate on databases containing transaction data. Apriori uses the following steps.

1. Set a minimum value for support and confidence. This means that we are only interested in finding rules for the items that have certain default existence (e.g. support) and have a minimum value for co-occurrence with other items (e.g. confidence).
2. Extract all the subsets having a higher value of support than the minimum threshold.
3. Select all the rules from the subsets with confidence value higher than the minimum threshold.
4. Order the rules by descending order of Lift.

## TIME REQUIRED

Apriori Algorithm can be slow. The main limitation is the time required to hold a vast number of candidate sets with many frequent itemsets, low minimum support or large itemsets i.e. it is not an efficient approach for a large number of datasets.

## RESULTING RULES

### TEST - 1:

```
min_support=0.05, min_confidence=0.2
```

```
LAST RULE IN THE ANS SET:
```

```
Rule: beef -> other vegetables
```

```
Support: 0.05079527963057978
```

```
Confidence: 0.4248927038626609
```

```
Lift: 1.1282232695208803
```

```
No of item sets = 107
```

## TEST - 2:

`min_support=0.06, min_confidence=0.3`

LAST RULE IN THE ANS SET:

**Rule:** `yogurt -> rolls/buns`

**Support:** `0.0659312467932273`

**Confidence:** `0.36925287356321845`

**Lift:** `1.304938985629579`

**No of item sets = 69**

## TEST - 3:

`min_support=0.07, min_confidence=0.4`

LAST RULE IN THE ANS SET:

**Rule:** `yoghurt -> whole milk`

**Support:** `0.07183170856849666`

**Confidence:** `0.5970149253731343`

**Lift:** `1.303003459744948`

**No of item sets=28**

# RECOMMENDATIONS

Based on the rules presented to the client, the client should use the rules to organise the items present in the rule together which indicates that these two items are bought together more frequently. By keeping these two items together, the number of transactions can be increased thereby increasing profit for the client.

=====

## CODE

[https://github.com/sevenwik/College/blob/master/Data\\_Mining/Worksheet2/solution2.ipynb](https://github.com/sevenwik/College/blob/master/Data_Mining/Worksheet2/solution2.ipynb)

