

# Class comparisons association rule mining: Market basket analysis - basic concepts.

Submitted by:

Jay kakdiya: 160410116046

Kaushiki kansara: 160410116048

Devangini kathad: 160410116049

Kaustubh wade: 160410116050

1

# Concept description

- Data mining can be classified into two categories: descriptive data mining and predictive data mining.
- Descriptive data mining describes the data set in a concise and summative manner and presents interesting general properties of the data.
- Predictive data mining analyzes the data in order to construct one or a set of models, and attempts to predict the behavior of new data sets.
- Database is usually storing the large amounts of data in great detail. However users often like to view sets of summarized data in concise, descriptive terms.



## Market basket analysis

- Market Basket Analysis is a modelling technique.
- It is based on, if you buy a certain group of items, you are more (or less) likely to buy another group of items.
- For example, if you are in a store and you buy a car then you are more likely to buy insurance at the same time than somebody who didn't buy insurance.
- The set of items a customer buys is referred to as an itemset.
- Market basket analysis seeks to find relationships between purchases (Items).
  - E.g. IF {Car, Accessories} THEN {Insurance}

{Car, Accessories} → {Insurance}

3



## Market basket analysis (Cont..)

 $\{Car, Accessories\} \rightarrow \{Insurance\}$ 

- The probability that a customer will buy car without an accessories is referred as the support for rule.
- The conditional probability that a customer will purchase Insurance is referred to as the confidence.



## Association rule mining

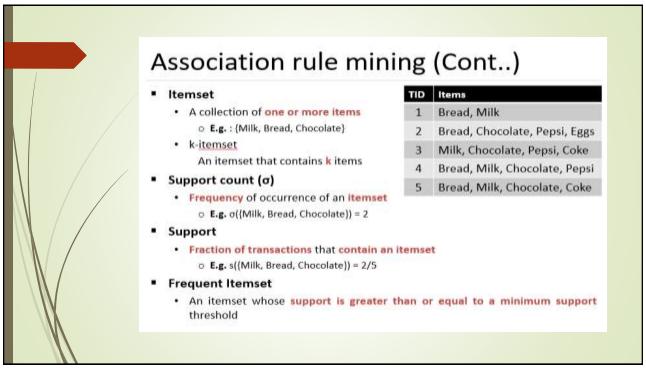
- Given a set of transactions, we need rules that will predict the occurrence of an item based on the occurrences of other items in the transaction.
- Market-Basket transactions

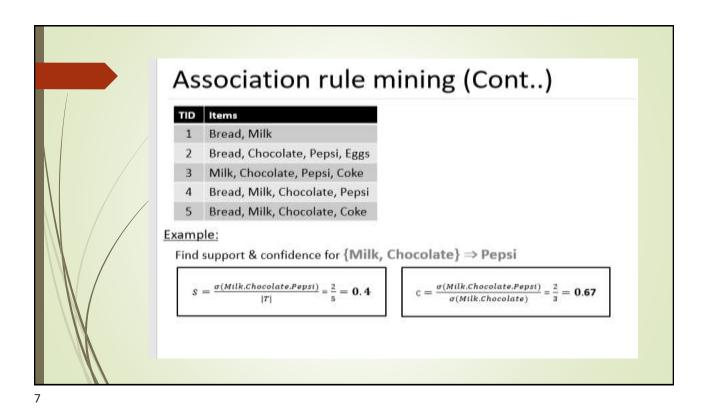
TID	Items
1	Bread, Milk
2	Bread, Chocolate, Pepsi, Eggs
3	Milk, Chocolate, Pepsi, Coke
4	Bread, Milk, Chocolate, Pepsi
5	Bread, Milk, Chocolate, Coke

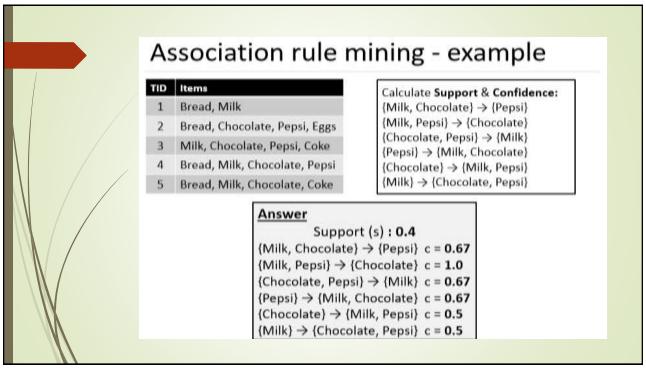
## **Example of Association Rules**

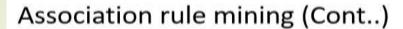
{Chocolate} → {Pepsi}, {Milk, Bread} → {Eggs, Coke}, {Pepsi, Bread} → {Milk}

5









 A common strategy adopted by many association rule mining algorithms is to decompose the problem into two major subtasks:

### 1. Frequent Itemset Generation

- The objective is to find all the item-sets that satisfy the minimum support threshold.
- These itemsets are called frequent itemsets.

#### 2. Rule Generation

- The objective is to extract all the high-confidence rules from the frequent itemsets found in the previous step.
- These rules are called strong rules.

q

## Apriori algorithm

- Purpose: The Apriori Algorithm is an influential algorithm for mining frequent itemsets for Boolean association rules.
- Key Concepts:
  - Frequent Itemsets:

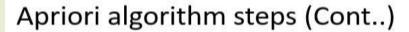
The sets of item which has minimum support (denoted by L<sub>I</sub> for ith-Itemset).

Apriori Property:

Any subset of frequent itemset must be frequent.

• Join Operation:

To find  $L_k$ , a set of candidate k-itemsets is generated by joining  $L_{k-1}$  itself.



- Step 1:
  - · Start with itemsets containing just a single item (Individual items).
- Step 2:
  - Determine the support for itemsets.
  - Keep the itemsets that meet your minimum support threshold and remove itemsets that do not support minimum support.
- Step 3:
  - Using the itemsets you have kept from Step 1, generate all the possible itemset combinations.
- Step 4:
  - Repeat steps 1 & 2 until there are no more new itemsets.

11

## Apriori algorithm - Pseudo code (Cont..)

```
C<sub>k</sub>: Candidate itemset of size k
L_{k}: \text{Frequent itemset of size k}
L_{1}=\{\text{frequent items}\};
\text{for } (k=1; L_{k} \mid = \emptyset; k++) \text{ do begin}
C_{k+1} = \text{candidates generated from } L_{k};
\text{for each transaction } t \text{ in database do}
\text{Increment the count of all candidates in } C_{k+1}
\text{That are contained in } t
L_{k+1} = \text{candidates in } C_{k+1} \text{ with } \underline{\text{min support}}
\text{end}
\text{return } U_{k} L_{k};
```

