Market Basket Analysis

Mahendra Gupta

Lecturer(CSE)

Govt. Polytechnic College Anuppur (M.P.)

Contents

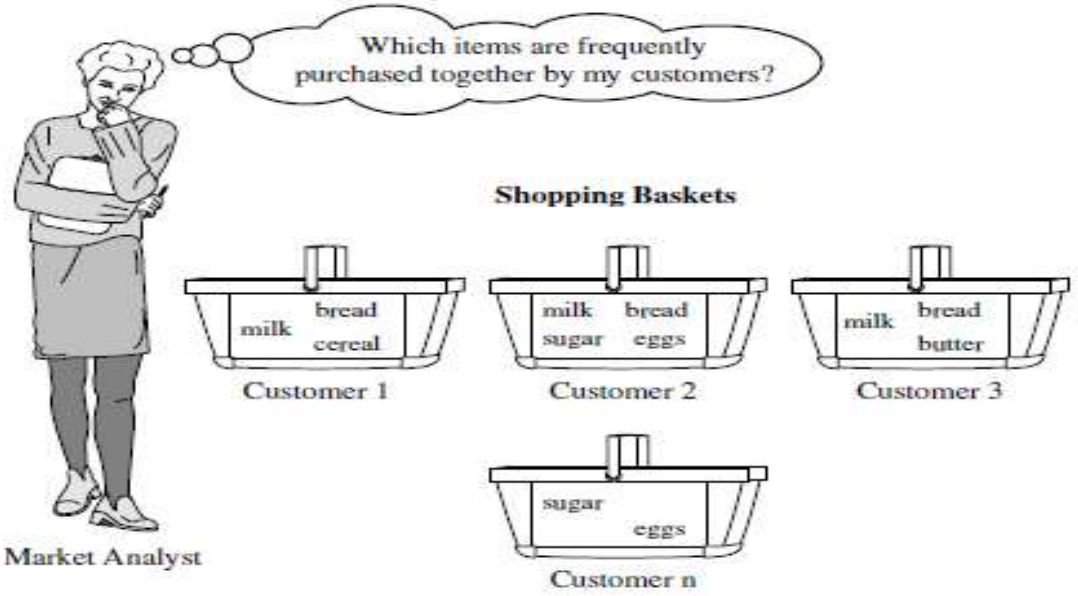
1. Introduction
2. Objective
3. Procedure(Apriori Algorithm)
4. Applications
5. Summarization
6. References

Introduction

- A large number of people use to visit shopping malls and shopping centers[1].
- Massive amounts of data continuously being collected and stored as transactions
- Those collected data can be useful from the business prospective such as

Objective

1. To find frequently purchased item sets from large transactional database[1].
2. Ap



Definition

- The process of discovering frequent item sets in large transactional database is called market basket analysis[1].
- Frequent item set mining leads to the discovery of associations and correlations among items[1].

The Apriori Algorithm[1]
(For finding frequent item sets)

Step:1 This step simply scans all of the transactions in order to count the number of occurrences of each item, named it as C1.

Step:2 Apply minimum support count, and items set that satisfies the condition named as L1. (min support count =2)

Table:1

Transactional data for an *AllElectronics* branch.

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

Scan D for count of each candidate →

C_1

Itemset	Sup. count
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2

Compare candidate support count with minimum support count →

L_1

Itemset	Sup. count
{I1}	6
{I2}	7
{I3}	6
{I4}	2
{I5}	2

Step:3 To discover the set of frequent 2-itemsets, L2, the algorithm uses the join L1XL1 to generate a candidate set of 2-itemsets, and count the occurrences of each set.

Step:4 Apply min support count on C2, result will be 2-frequent item set L2.

Generate C_2 candidates from L_1 →

C_2

Itemset
{I1, I2}
{I1, I3}
{I1, I4}
{I1, I5}
{I2, I3}
{I2, I4}
{I2, I5}
{I3, I4}
{I3, I5}
{I4, I5}

Scan D for count of each candidate →

C_2

Itemset	Sup. count
{I1, I2}	4
{I1, I3}	4
{I1, I4}	1
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I2, I5}	2
{I3, I4}	0
{I3, I5}	1
{I4, I5}	0

Compare candidate support count with minimum support count →

L_2

Itemset	Sup. count
{I1, I2}	4
{I1, I3}	4
{I1, I5}	2
{I2, I3}	4
{I2, I4}	2
{I2, I5}	2

Step:5 To generate 2-frequent item set, generate 3-item using L2XL2 and apply min support count which will be 3-frequent item set.

Generate C_3 candidates from L_2 → C_3

Itemset
{I1, I2, I3}
{I1, I2, I5}

Scan D for count of each candidate → C_3

Itemset	Sup. count
{I1, I2, I3}	2
{I1, I2, I5}	2

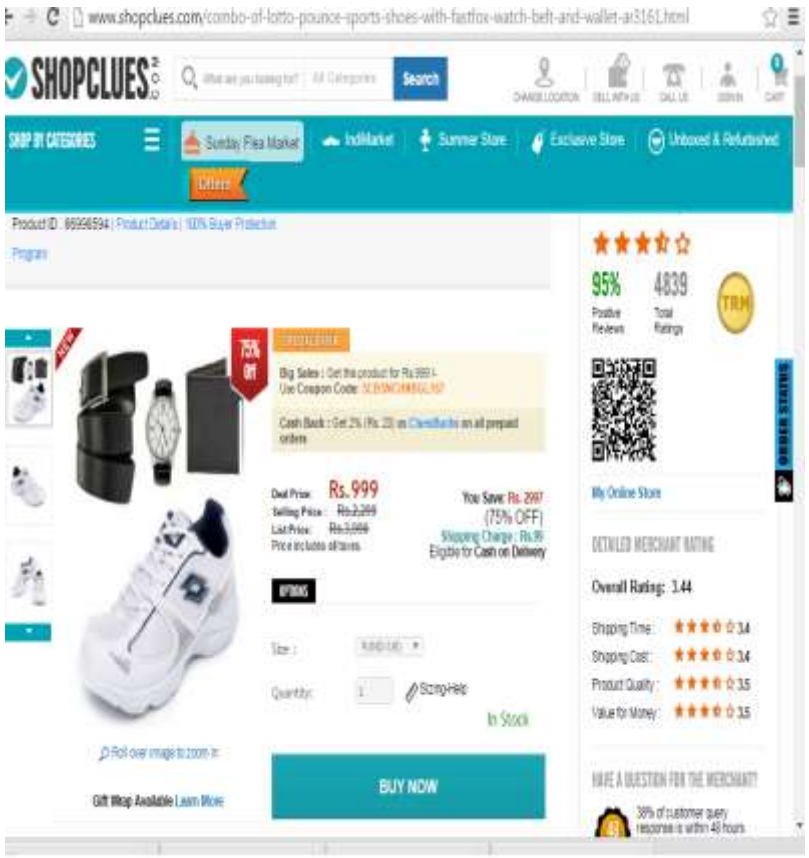
Compare candidate support count with minimum support count → L_3

Itemset	Sup. count
{I1, I2, I3}	2
{I1, I2, I5}	2

Step:6 Continue until you get empty sets.

Application

1. Cross Selling of the Items[3][4]



Application

2. Proper Placement of Items[5].

3. Fraud detection.
4. Customer Behavior.
5. Affinity promotion[6].

Summary

- Market basket analysis.
- Discovering Frequent item sets.
- Application of frequent item sets.

References

[1] *Jiwawi Han and Micheline Kamber, Concepts and Techniques of Data Mining, 2nd ed., Jim Gray, Ed.: Morgon Kaufmann,Elsevier, 2012*

[2] http://www.alamy.com/stock-photo/shopping-trolley.html

[3]http://www.amazon.in/

[4] http://search.shopclues.com/

[5] http://www.salemmarafi.com/code/market-basket-analysis-with-r/

[6]http://people.revoledu.com/kardi/tutorial/MarketBasket/Applications.htm

[7] http://youtube.com.


THANK YOU