

# An Effective Approach of Cross Selling using Balanced Market Basket Analysis

**Sandeep Rathore**

Research Scholar, Rabindranath Tagore University, Raisen

**Dr. Rajendra Gupta**

Associate Professor, Rabindranath Tagore University, Raisen

## Abstract

The Cross selling is an observation of selling an additional item or service to the customer. Indeed the companies define cross-selling in general and cross-selling analytics in many ways and with many names. One common naming used in retail sector is 'market basket analytics'. Most of the people are familiar with cross-selling from the experience as online consumers. The e-commerce websites make product suggestions which are based on a market basket analysis. The number of the possible suggestions is also known as "association rules". In this paper, we have proposed a cross selling strategy using MapReduce algorithm and it is found that the algorithm best perform in Market Basket Analysis.

**Keywords :** Market Basket Analysis, Cross Selling, MapReduce

## 1. Introduction

The modern-day marketers mostly preferred way of generating more revenue from the existing clients. The cross-selling is defined as "an action or practice of selling an additional product or service to an existing customer". At the same time as acquiring new customers have becomes a costliest way to increase the revenues. It has become imperative for the businesses to enhance their growth potential from their presented clients.

## 2. Market Basket Analysis

A modus operandi namely 'Market basket analysis (MBA)' is to discover relationships between pairs of products purchased together by the customer. This technique can be used to un-cover interesting cross-sells and related items. The concept behind the market basket analysis is very simple. Simply observe the customer order for products that have been purchased together. For example, using market basket analysis (MBA), the customer might uncover the fact that he tends to buy hot dogs and buns together. Using this information, the customer may organize the store, so that hot dogs and buns are next to each other. This concept was suggested in [1].

The Market Basket Analysis (MBA) is an investigative technique which identifies the strength of association between pairs of products purchased from an individual retailer. This type of analysis is usually applied to data on purchasing behavior, such as that collected at the point of sale. For example, if applied to grocery shopping, the results of a MBA could inform a supermarket pricing strategy. If the supermarket knows that bread and fruit juice tend to be purchased together it can avoid offering price discounts on both at the same time. Almost all available literature has attempted to address the concept of how part of Market Basket Analysis is done. The author in [3] presented how Market Basket Analysis can be used in managerial research and conducted that the adoption of Market Basket Analysis can help bridge science practice divides. They also argued that the use of Market Basket Analysis can lead to insights in substantive management domains such as human resources, organizational behavior, entrepreneurship and strategic management.

The author in [2] presented, how Market Basket Analysis can be used to estimate potential revenue increases for a little university bookstore and concluded that depending on the customers price sensitivity and on the saturation level of the affinities uncovered revenue can increase by as much as Rs. 75,000/- for the bookstore. Another research was done by Fact-Point Group in on how retailers are using Market Basket Analysis to succeed margin

and market share. Results obtained reveal that retailers are using Market Basket Analysis to develop more profitable advertising and promotions target that can attract into the stores and increase the size and value of the basket of purchases among other things.

### 3 Meaning of Cross-Selling

In spite of its name, cross-selling is a well-known sales strategy in marketing, where the cost of acquiring fresh business is usually significantly higher than selling one additional product to an existing customer. Therefore, implementing a cross-selling strategy can represent a financially attractive investment, if a company can effectively find ample cross-selling opportunities.

### 4. Objective

The objective of the research study is based on the analysis of Market Basket Analysis for Cross-selling and finding efficient method for cross selling using MapReduce.

### 5. Literature Review

The researcher proposed the effectiveness of the MBA algorithm by applying it to the data obtained from a large retailing company. They used attributes for mining qualitative rules for categorical data using association rules. The association rules found application in many research areas such as: Market basket analysis, recommendation systems to Advertising and Strategy formulation [2, 4-5]. In marketing, a literature market basket analysis has been classified into two types: explanatory and exploratory. The main idea behind exploratory models is the discovering of purchasing patterns from Point-of-Sale (POS) data.

The exploratory approach does not include information on consumer demographics or marketing mix variables summaries a vast amount of data into a fewer meaningful rules. This type of method is quite useful for discovering unknown relationships between the items in the data [6, 8].

Moreover, the exploratory approaches are not appropriate for forecasting and finding the cause-roots of complex problems. This is just used to uncover distinguished cross-category interdependencies based on some frequency patterns for product or items categories purchased together [7, 9]. A representative application of this exploratory approach is identifying product category relationships by simple association measures. The pair-wise associations are used to compare entities in pairs and judge which entity is preferred or has greater amount of some quantitative property. While examining the leading products for consumers is extremely important since a large number of shoppers come into contact with this specific product type day by days [10-12].

As the departments with leading products generate much in-store traffic, it is crucial to use this information for placing other specific products nearby. Another significant stream of research in the field of exploratory analysis is the process of generating association rules. According to the research study in 2014, there are few algorithms developed that are not based on the Apriori [20, 23-26]. The researchers have developed a framework for knowledge discovery from market basket data. Combining Apriori and Attribute Oriented Generalization [21-23].

A typical and widely-used example of association rule mining is market basket analysis. It is a technique that discovers relationships between pairs of products purchased together [28-30, 32]. The idea behind market basket analysis is simple, simply examine the order of products have been purchased together. For example in market basket analysis the fact might be uncover that if “an customer buy milk also tends to buy breads” [31].

The Market Basket Analysis report shows better result about given items, for example if a person need to find relationship of Milk with others Bread, Eggs and Cheeses, then market basket analysis display a reports This report consists of the products name, frequency, Support and Confidence.

An author [38] proposed a distributed Apriori association rule and classical Apriori mining algorithms for grid based knowledge discovery. The intention of the paper is to obtain knowledge with the help of predictive Apriori algorithm and distributed grid dependent Apriori algorithms for Association rule mining.

A market-basket analysis with principal component in proposed in [28]. Market-basket analysis is a well-known business crisis that can be solved computationally with the help of association rules, mined from transaction data to reduce the cross-selling results.

## 6 MapReduce Algorithm for Cross Selling

In the above example, there are only four transactions and four products. This limited number of products and deals is usually not the case in big marketing, where manufacturers or distributors handle with hundreds or thousands of products and customers. In this type of circumstances, the computation of an MapReduce algorithm uses additional techniques and can be automated.

A system is automatized, feeding it regularly with increasing numbers of ERP transactions. Moreover, once the algorithm has learned the first rules, the minimum support and minimum confidence is automatically adjusted, enabling an early case of machine learning. The following table shows the value of percentage support and percentage confidence of the grouped products. In the table 'C' represents 'Components', 'T' represent 'Tools', 'S' represent 'Services' and 'M' represent 'Machine'.

**Table 1 :** Percentage Support and Confidence for a set of four items

S.No	X Y	n(X U Y)	N	% Support	n(X)	Confidence
1	C T	1	4	20	3	40
2	C S	3	4	75	3	67
3	C M	3	4	75	3	67
4	C T S	1	4	50	3	40
5	C S M	1	4	50	3	0
6	C T M	0	4	0	3	0
7	C T S M	0	4	0	3	75
8	T C	1	4	50	2	75
9	T S	1	4	50	2	75
10	T M	1	4	50	2	75
11	T C S	1	4	50	2	0
12	T S M	0	4	0	2	0
13	T C M	0	4	0	2	0
14	T C S M	0	4	0	2	75
15	S C	2	4	75	2	100
16	S T	1	4	50	2	75
17	S M	1	4	50	2	75
18	S C T	1	4	50	2	75

In the above table, the confidence level is higher when two item 'S' and 'T' are combined purchase as shown in point number 15.

## 7. Balanced Market Basket Analysis using MapReduce

The algorithm MapReduce was developed to process enormous datasets in a disseminated parallel computation. It is one of the key algorithms that enabled Un-structured Data analytics.

The data requirement for the algorithm is :

1. Baskets
  - This column identifies the individual baskets.
  - Values can be numeric to identify the baskets.
2. Products
  - This column has all the items that are included in each basket.
  - Values of items can be numeric.

The algorithm is made up of mappers & reducers. The Mappers read the data from file storage one chunk at a time and parse data to generate the key value pairs. After that reducer receive those key-value pairs by key and process these values associated with earlier values. So the following procedure is needs to be apply:

1. The use of datastore to designate data sources
2. To define mappers and reducers function
3. The use of mapreduce with datastore for the mapper and reducer to process data
4. To store the processed data for more analysis

Since the mappers and reducers perform quite simple operations, we need to chain them together to handle further complex operations. In the data mining algorithm known as Apriori algorithm, the most time consuming procedure are the generation of transactions. So the MapReduce algorithm can be used to to address this holdup.

The process of MapReduce is start with setting up of datastore. Now we use a fairly small CSV file on a local drive, but data store can -

- read a data which is too large to fit in the computer memory or
- read files on multiple locations using clusters

It is necessary to start the work with a small subset of the data to prototype and testing the algorithm before we use it on big data applications. The software MATLAB makes it easy to prototype the MapReduce algorithm on local machine and then scale it up to the cluster or cloud later.

The procedure of the MapReduce algorithm for the Balanced Market Basket Analysis is as follows :

### Step 1: Grouping of items by transaction

In the first step, we are treating visitor as a shopper, in which the pages visited as items in the shopping cart or say 'transaction'. A visitor can visit the same page a number of times, but such repeated visits are not factored in item-set counting.

Now the focus is on reducing or minimizing the keys by balancing a MapReduce algorithm which are generated. For this purpose, a good starting point needs to be make group items through transactions by using VisitorCookieID as the key, because we have a finite set of visitors but they can visit a larger number of pages.

#### VisitorCookieID

#### Page

'3821337fdad7a619323b10b602c4616'	'/matlabcentral/answer/'
'3821314fdad7a6132253b10b602c4616'	'/matlabcentral/answer/152931-how-to-translate-the-following-code-from...'
'3821332fdad7a6132253b10b602c4616'	'/matlabcentral/answer/153109-number-greater-than-the-largest-positive...'
'3821319fdad7a6132253b10b602c4616'	'/help/matlab/ref/real.html'
'3821319fdad7a6132253b10b602c4616'	'/matlabcentral/answer/153201-how-to-evluate-large-factorials'
'38213319fdad7a6132253b10b602c4616'	'/help/matlab/matlab_pro/floating-point-numbers.html'
'3821320fdad7a6132253b10b602c4616'	'/matlabcentral/answer/contributors/5560484-tigo/questions'
'3821320fdad7a6132253b10b602c4616'	'/matlabcentral/newreader/view_thread/297445'

Now the reducer receive the key-value pairs by using key, and merges multiple cell arrays with the same key into a single cell array and removes any duplicates. We can then store the result in a new datastore.

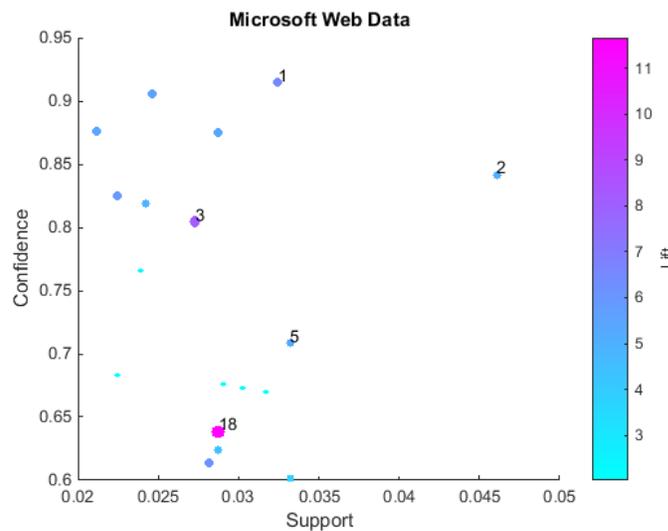
Key	Value
'009279b5566e6327e55463dc9c6273'	{1x12 cell}
'01c5f3d2e475d787ehu711fb43f8'	{1x14 cell}
'07176157c000c8855c21b958b7766d'	{1x10 cell}
'0f295bff2a611053a4333f027b1f1a'	{1x14 cell}
'130f74a9c7402bf1b8f666b557315f'	{1x10 cell}

**Step 2: Generate 1-item set**

Since all items in a transaction are unique. So we need to count the number of times an item comes in the transactions to check how many transactions contained that item. The webpages were stored as a value in the earlier step, therefore it is simply needed to retrieve those and count their contents.

Key	Value
'/loren/'	[2]
'/loren/2016/07/05/when-is-a-numeric-not-a-number/'	[1]
'/loren/2019/10/02/using-parfor-getting-up-and-running/'	[1]
'/loren/2001/11/14/generating-from-your-matlab-algorithms/'	[1]
'/loren/2014/02/06/using-gpus-matlab/'	[1]

Here we have taken Microsoft Website Data code using publicly available dataset such as Anonymous Microsoft Web Data Set from UCI Machine Learning Repository source. When Balanced Market Basket Analysis is applied on this dataset, we get the following outcome :



**Figure 1 :** Microsoft Web Data Analysis using Balanced Market Basket Analysis

The above figure is made up by generating the following rules :

### Step 3: Generate the rules

```

Rule #1
  {Windows95} => {Windows Family of OSs}
  Conf: 0.91, Lift: 6.46
Rule #2
  {Windows95 Support} => {isapi}
  Conf: 0.84, Lift: 5.16
Rule #3
  {SiteBuilder Network Membership} => {Internet Site Construction for
  Developers}
  Conf: 0.80, Lift: 8.17
Rule #5
  {Knowledge Base, isapi} => {Support Desktop}
  Conf: 0.71, Lift: 5.21
Rule #18
  {Windows Family of OSs, isapi} => {Windows95 Support}
  Conf: 0.64, Lift: 11.66

```

In the above given rules, MapReduce was used only for the initial data processing, and the rest was still done in computer memory. However, if the processed data gets larger, then it is needed to make more use of MapReduce. Instead, we subdivide the dataset into several chunks and complete the entire process through rule generation in each. We need to adjust the minimum support to account for the reduction of transaction counts in subsets if we do so. Then we can combine the final output. This may not provide the same result as the single thread process, but it should be fairly close.

## 6. Result and Discussion

The association rules are generated for the given input which is taken from UCI depository. A very well-organized rule satisfies the following conditions:

1. The item set exceed minimum support determined based on the business need.
2. To exceed the minimum confidence
3. Having greater Lift Ratio

After the model has been processed, we can use the rules and itemsets to make predictions. In an association model, a prediction tells us what item is likely to occur given the presence of the specified item, and the prediction can include such information as the probability, the support, or the importance. A Marketer would consider rules with high Lift ratio, high Confidence and good support. The confidence of 100% tells us that this rule appears to be a very promising rule for the business.

*The graph in the earlier section shows that the components which are having higher Lift value are the most proper impact. So according to the graph, the components in between 0.025 to 0.03 are having appropriate outcome.*

The process of creating itemsets and counting correlations can be time-consuming. Although the Microsoft Association Rules algorithm uses optimization techniques to save space and make processing faster, we need to know that the performance issues can occur under conditions such as the following:

- Data set is large with many individual items.
- Minimum itemset size is set too low.

To minimize processing time and reduce the complexity of the itemsets, we might try grouping related items by categories before we analyze the data.

## 7. Conclusion

It is important to have solutions for each step of the data analysis. Market Basket Analysis using MapReduce algorithm is a perfect solution for the treatment and deployment of data for Cross selling. In this research study, we have demonstrated the potential of combining MapReduce algorithm and Balanced Market Basket Algorithm to generate applications such as recommendation systems.

The market basket problem can be found as the best example of mining association rules. The building up on previous researches views by using Market Based methods for mining association rules allowed for discovering useful information for this research work. After aggregating the data and finding product affinities, the Balanced Market Basket Analysis extends the analysis by adding up some probabilities of a consumer purchasing certain products in different conditions and in certain times.

## References

- [1] Karthiyayini R. and Balasubramanian R. "Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis", IJARCSSE, Volume 6, Issue 10, pp 241 (2017)
- [2] Svatina M. and Zupencic "How to increase Sales in Retail with Market Analysis, Systems Integrations" (2017), pp 418
- [3] Agrawal R. and Srikant R. Fast Algorithm for Mining Association Rules. Proc. of the Int. Conf on Very Large Database, pp. 487 (2016)
- [4] Julander. Basket Analysis: A New Way of Analyzing Scanner Data. International Journal of Retail and Distribution Management, V (7), pp 10-18 (2016)
- [5] Berry and Linoff. Data Mining Techniques for Marketing, Sales and Customer Relationship Management (second edition), Hungry Minds Inc., (2016)
- [6] Berry, M.J.A., Linoff, G.S.: Data Mining Techniques: for Marketing, Sales and Customer Relationship Management (second edition), Hungry Minds Inc., (2014)
- [7] Agrawal R, Srikant R, Fast algorithms for mining association rules. In: Proceedings of the 20th VLDB conference, pp 487–499 (2016)
- [8] J. Han, H. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In: Proc. Conf. on the Management of Data SIGMOD'00, ACM Press, New York, NY, USA (2015)
- [9] J. Han and M. Kamber. Data Mining: Concepts and Techniuqes, Morgan Kaufmann Publishers, San Francisco, CA, (2015)
- [10] Loraine Charlet, Julander and Ashok Kumar D, "Frequent Item set mining for Market Basket Data using K-Apriori algorithm" , International Journal of Computational Intelligence and Informatics, Volume 1, No. 1, 2011, pp.14-18 (2016)
- [11] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence (2016)
- [12] Julander, C.-R. "Basket Analysis: A New Way of Analyzing Scanner Data" International Journal of Retail and Distribution Management, Volume 20 (7), 10-18 (2016)
- [13] Gregory Piatetsky-Shapiro, William Frawley "Knowledge Discovery in Databases" AAI/ MIT Press. 6. Garry J.Russel, Wagner A. Kamakura. (1997). Modeling Multiple Category Brand Preference with Household Basket Data. Journal of Retailing, Volume 73(4), 439-461. (2015)
- [14] Qiankun Zhao, Sourav S. Bhowmick. "Association Rule Mining: A Survey" Singapore: CAIS, Nanyang Technological University, No. 2003116 (2015)
- [15] RakeshAgrawal, Tomasz Imielinski, Arun Swami. "Mining Association Rules between Sets of Items in Large Databases". ACM SIGMOND Int'l Conference on Management Data, pp. 207- 216 (2015)
- [16] Andreas Mild, Thomas Reutterer. "An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data" Journal of Retailing and Consumer Services vol.10, 123-133 (2015)
- [17] Xiaohui Yu, Yang Liu, Jimmy Xiangji Huang, Aijun An. "Mining Online Reviews for Predicting Sales Performance: A Case Studyin the Movie Domain". IEEE TransactionsOn Knowledge and Data Engineering, VOL. 24, NO. 4, APRIL 2016.

- [18] Julie Marcoux, and Sid-Ahmed Selouani. University de Moncton, Campus de Shippagan, Canada. "A Hybrid Subspace-Connectionist Data Mining Approach for Sales Forecasting in the Video Game Industry World Congress on Computer Science and Information Engineering (2014)
- [19] Han, J., Kamber, M.: "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, (2014)
- [20] M.J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li, "New algorithms for fast discovery of association rules," in Third International Conference on Knowledge Discovery and Data Mining, (2014)
- [21] Robert J. Hilderman, Colin L. Carter, Howard J. Hamilton, and Nick Cercone "Comparative Analysis of Various Approaches Used in Frequent Pattern Mining" (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence (2014)
- [22] Kanwal Garg, Deepak Kumar. Kurukshetra University, Haryana, India. "Comparing the performance of frequent pattern mining algorithm", International Journal of Computer Applications, Vol 69- No. 25, (2013)