Market Basket Analysis: A Case Study of a Dehradun Based Bakery Shop

Dr. Smita Ghoshal[1]
Ms. Sukanya Paul[2]

Abstract

Market Basket Analysis is a modeling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items. The main goal of such an analysis is to entice customer to buy more than they usually do. Market basket analysis (MBA) is an example of an analytics technique employed by retailers to understand customer purchase behaviors, which can aid the retailer in correct decision making. It is used to determine what items are frequently bought together or placed in the same basket by customers (Kaur & Kang, 2016). In this research we study the new trends, challenges, and the impact of market basket analysis on consumer buying behavior by using Data Analytics software Rapidminer to create frequent item-sets and associations between the products purchased together in a bakery shop in Dehradun. The study exploits the certain marketing activities which later can be leveraged effectively by the Bakery shop and hence increase their sales and profit margin.

Keywords: Market Basket Analysis, MBA, Predictive Analysis, Rapidminer, Modeling Technique, Data Analytics

Introduction

Market basket analysis (MBA) is an example of an analytics technique employed by retailers to understand customer purchase behaviors, which can help the retailer in correct decision making. It is applied to determine what items are usually bought together or placed in the same basket by customers (Kaur M & Kang S. 2016). MBA is an upcoming technique being used by retailers to manage inventory. Collecting large no of data of the customers makes no sense if that cannot be used to extract important information for gaining competitive advantage. ANN backpropagation is used to predict product's inventory requirement for each product Market basket analysis (also known as association rule mining) is a method of discovering customer purchasing patterns by extracting associations or co-occurrences from stores' transactional databases (Chen Y, Tang K, Shen R, Hu Y. 2004). Market Basket Analysis is done to check whether the purchase of one product raises the likelihood of the purchase of other products. This knowledge is a tool for the marketers to bundle the products or strategize a product cross sell to a customer.

According to Szymkowiak M et.al If the traditional contents of a market basket is replaced with specific sociodemo graphic variables, the technique can also be applied to data from sampling surveys or censuses in order to discover association rules and co-occurrence relationships. MBA analysis is better than multicategory choice model. Explanatory models of multi category choice behavior explains the dependency of purchases. They estimate own and cross category effects of marketing-mix variables on purchase incidences for a predefined set of product categories. Because of analytical restrictions, multi-category choice models can only handle a small number of categories. Hence, for large retail arrangements, the problem arises of how to determine the composition of shopping baskets with a meaningful selection of categories. Traditionally, this is

[1]*Associate Professor, Doon Business School, Dehradun*
[2]*Student, PGDM, Doon Business School, Dehradun*

resolved by managerial intuition (Boztug Y, Reutterer T. 2006). In recent years, the progress of bar code technology has made it possible to collect information containing consumer transaction data efficiently. Information about market baskets is collected in the form of sets of items which are bought together in a transaction. (Aggarwal C & Wolf J, 1999). MBA is the oldest areas of data mining. Mining of databases has attracted a growing amount of attention in database communities due to its wide applicability to improving marketing strategies. Among others, data clustering is an important technique for exploratory data analysis. (Huang C et.al 2002). The typical solution involves the mining and analysis of association rules, which take the form of statements such as people who buy diapers are likely to buy beer." Market basket analysis has been an elementary part of quantitative decision support in retail marketing for many years and it is regularly cited as a prime application area of data mining. (Decker R & Monein K. 2003). Store-chain association rules, an Apriori like algorithm for automatically extracting association rules is proposed specifically for a multi-store environment, where stores may have different product-mix strategies that can be adjusted over time. (Chen Y, Tang K, Shen R, Hu Y. 2004) MBA using K-Apriori algorithm extracts a set of frequent itemsets from the data, and then pulls out the rules with the highest information content for different groups of customers by dividing the customers in different clusters.(Annie M.C, Kumar D).

Research Methodology

Objective:

The main objective of this research is to see:

1. How different products in a retail shop assortment inter-relate.
2. Mining association rules from transactional data will provide us with valuable information about co-occurrences and co - purchases of products so that we can focus on cross-sell and increase the profit margin.

Data Collection:

The methodology of market basket analysis in a retail stores is to find the selling documents with the items for the transactions. In this research the copy bills are the selling documents considered here. Transaction data for about 45-50 days are taken into consideration. 200 bills from the Dehradun bakery shop were collected for the study.

Number of Attributes that can be taken:

Attribute Information:

- Invoice No: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.
- Stock Code: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.
- Description: Product (item) name. Nominal.
- Quantity: The numbers of each product (item) per transaction. Numeric.
- Invoice Date: Invoice Date and time. Numeric, the day and time when each transaction was generated.
- Unit Price: Unit price. Numeric, Product price per unit in sterling.
- Customer ID: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.

Market Basket Analysis Using Rapid Miner:

A market basket database typically consists of many transaction records. Each record lists all items purchased during a single customer transaction. The objective of this data mining exercise is to identify if certain groups of items are usually purchased together. The result is a set of rules, called association rules which summarize item associations as follows:

If [A] is purchased --> then [B] is also purchased, [x%] of time.

These association rules can be applied in an old-fashioned brick and mortar setting as well as in an online setting for real-time cross-selling or ad placement.

Two essential concepts - Support and Confidence:

A key idea to get comfortable with is that of frequent item sets. An item set can consist of one item or more. An example of typical data says consist of customer transactions involving purchases of typical cosmetics items, one frequent item set example could be [brushes, nail polish].

Frequent item sets are quantified by support which is the ratio of the number of instances where [brushes, nail polish] appeared together in a single transaction to the total number of transactions.

Support = occurrences of [brushes, nail polish]/total # of transactions

The next important metric that you will need to run a market basket analysis is confidence. Extending the above example, the confidence of finding [brushes, nail polish] together is defined as confidence [brushes, nail polish] = occurrences of [brushes, nail polish]/total # of [brushes]

Setting up a market basket analysis using Rapid Miner

In Rapid Miner, association rules are extracted using two operators in a sequence. The first operator, called FP Growth, is required to generate frequent item sets. The second operator, Create Association Rules, then produces the IF-THEN rules based on the confidence requirement. But before that you may need some pre-processing steps for selecting the attributes you want and more importantly, to convert the input data to binomial (true/false) format which is required by the FP Growth operator
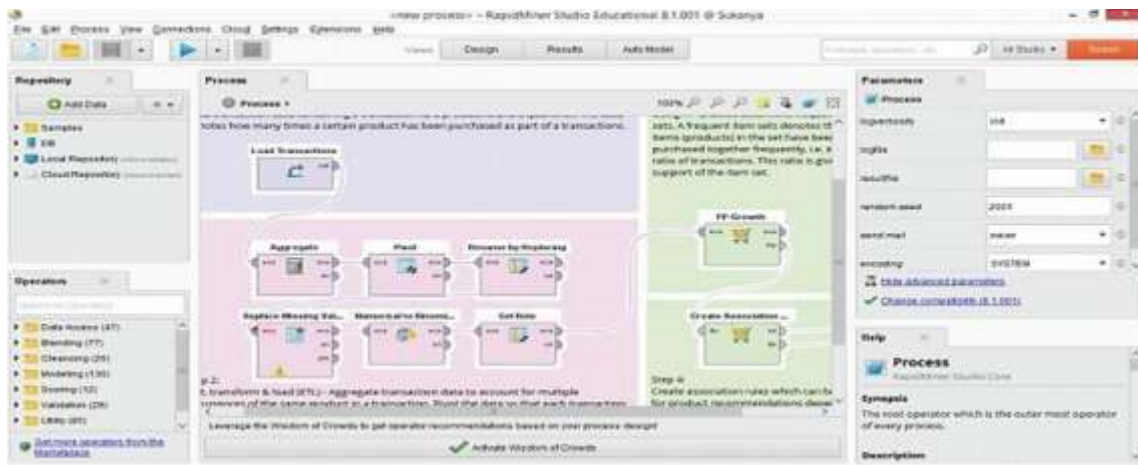
Model associations between products by determining sets of items frequently purchased together and building association rule to drive recommendation.

Step 1: load transaction data containing a transaction id, a product id and a quantifier. The data denotes how many times a certain product has been purchased as part of a transactions.

Step 2: Edit transform & load (ETL) - aggregate transaction data to account for multiple occurrences of the same product in a transaction. Pivot the data so each transaction is represented by row. Transform purchase amounts to binary "product purchased yes/no" indicators.

Step 3: Using FP-Growth, determine frequent item sets. A frequent item sets. A frequent item sets denotes that the items (products) in the set have been purchased together frequently, that is, in a certain ratio of transactions. This ratio is given by the support of the item set.

Step 4: Create association rules which can be used for product recommendations depending on the confidences of the rules.

Note:

1. When using the FP Growth operator, the important parameter is "min support". Rapid Miner will find only those item sets which exceed this minimum support value. However, if you check the box for "find min number of item sets", then the priority is given to "Min Number of item sets", in which case it will continue to reduce the support threshold until it finds at least that many item sets indicated in the "Min Number of item sets" field.

2. After finding the frequent item sets, the next step in the process is to extract rules which meet the confidence requirement. You can provide this in the "min confidence" field under the parameter options for Create Association Rules operator.

3. When the above process is run, Rapid Miner will generate outputs for both FP Growth and Create Association Rules operators. The FP Growth output is a table with support values for the minimum number of item sets requested in Tip 1. The association rules output consists of a text view, table view and graphical views of the extracted rules. The easiest and most insightful view is surprisingly the text view which will show rules such as these below:

Output: 1. Association Rules

[Blush] --> [Concealer] (confidence: 0.738)

[Brushes] --> [Nail Polish] (confidence: 1.000)

Data Analysis & Interpretation

Association-rule mining is a useful method of discovering customer purchasing patterns by extracting association or co-occurrences from stores' transactional databases. In Rapid miner:

Frequent item Set using FP growth:

Size	Support	Item 1	Item 2	Item 3
1	0.475	BREAD/BUN		
1	0.45	RUSK		
1	0.41	LOOSE BUISCUIT		
1	0.235	SNACKS		
1	0.2	CAKE/PASTRY		
1	0.145	NAMKEENS		
1	0.135	BUTTER/CHEESE		
1	0.13	DRY CAKES		
2	0.215	BREAD/BUNS	RUSK	
2	0.19	BREAD/BUNS	LOOSE BISCUIT	
2	0.125	BREAD/BUNS	BUTTER/ CHEESE	
2	0.29	RUSK	LOOSE BISCUIT	
3	0.125	BREAD/BUN	RUSK	LOOSE BISCUIT

Association Rules using Create Association:

No	Premises	Conclusion	Support	Confidence	La Place	Gain	p-s	Lift	Conviction
5	BREAD/BUN	LOOSE BISCUIT	0.19	0.4	0.80677966	-0.76	-0.00475	0.97560976	0.98333333
6	RUSK, LOOSE BISCUIT	BREAD/BUN	0.125	0.43103448	0.87209302	-0.45	-0.01275	0.90744102	0.92272727
7	BREAD/BUN	RUSK	0.215	0.45263158	0.82372881	- 0.74	0.00125	1.00584795	1.00480769
8	LOOSE BISCUIT	BREAD/BUN	0.19	0.46341463	0.84397163	- 0.63	-0.00475	0.97560976	0.97840909
9	RUSK	BREAD/BUN	0.215	0.47777778	0.83793103	- 0.69	0.00125	1.00584795	1.00531915
10	BREAD/BUN, RUSK	LOOSE BISCUIT	0.125	0.58139535	0.92592593	- 0.31	0.03685	1.41803744	1.40944444
11	RUSK	LOOSE BISCUIT	0.29	0.64444444	0.88965517	- 0.61	0.1055	1.57181572	1.659375
12	BREAD/BUN LOOSE BISCUIT	RUSK	0.125	0.65789474	0.94537815	- 0.26	0.0395	1.4619883	1.60769231
13	LOOSE BISCUIT	RUSK	0.29	0.70731707	0.91489362	- 0.53	0.1055	1.57181572	1.87916667
14	BUTTER/ CHEESE	BREAD/BUN	0.125	0.92592593	0.99118943	- 0.15	0.06088	1.94931774	7.0875

After processing the transaction data of the bakery shop using Rapid Miner researchers have conclude that Bread and Buns are the item mostly sold in the bakery shop followed by Rusk and loose biscuits. Focusing on the Marketing 4Ps (place , price, product and promotion) to increase more sales in the bakery shop the chef or owner needs to add more variety to the products that are the basic revenue generators (e.g. Bread, buns, cookies,

rusk) of his business like adding more number of flavor of loose biscuit types or of rusks or they can start making customized and more decorative cakes to increase the sales of cakes and Pastry, they need to add more number of pastry to the display self so that it attracts the customer to buy them as these product also come under impulsive buying. They should introduce more items of different price ranges so that more customers are attracted to buy many items at the same time. They should display or place their varieties of product on the self by specifically labelling the sections so that customers can easily pick up items of their choice. To increase sales of cheese and butter which is a very slow-moving inventory item the refrigerator containing these items should be kept just next to the bread and buns counter. So that customers purchasing the breadsor bun get an idea of buying bread or cheese along with it. The biscuits and namkeens self or counter should be right next to the café area so that people who are purchasing coffee gets and impulse of buying coolies or rusk along with it. Now to increase more sales they should focus on promotion by creating useful and attractive combinations, first the chocolates and the toffees which has a very low support should be made into a big and small chocolate hampers consisting of wide variety chocolate product that can be served for gifting purpose and should be displayed in near about the bill counter so that the customers are able to get a proper view while billing and end up making the last moment purchase decision. Hot beverages are very less sold therefore they can be combined with the certain snacks and at a low rate to drive more traffic to the shop like for people who like to hangout over coffee, which will automatically generate more sales of other items as people will end up purchasing for a pastry or say a dry cake or bread for their next day breakfast or tiffin while leaving. Snacks and coffee can be very well focused on to drive more customers to enter the shop for chitchat over coffee. Misc. goods like the birthday poppers, hats, candles should be well displayed and made attractive combo along with the birthday cake so that people feel benefited and lured to purchase the combo. This will make the sales of birthday cake increase and this might become as USP of the shop that they serve a great birthday party combination which is cost-effective and worth buying.

Conclusion:

Market basket analysis (MBA) is an example of an analytics technique employed by retailers to understand customer purchase behaviors. It is used to determine what items are frequently bought together or placed in the same basket by customers. It uses this purchase information to leverage effectiveness of sales and marketing. MBA looks for combinations of products that frequently occur in purchases and has been prolifically used since the introduction of electronic point of sale systems that have allowed the collection of immense amounts of data. MBA using association rule mining can be used to find relation between various phenomena. Various techniques can be used in the method of producing suitable outcome with the obtained market basket data.

In this case the copy bills of a bakery shop have been collected and analyzed using the software Rapid Miner and the suitable outcome was produced regarding the store arrangements, creating combinations, introducing more varieties to the ones highly sold and maintain products of various price ranges. These strategies will surely help increase the sales of the bakery shop and hence increase its profit margins.

Scope of Further Research:

Market basket analysis has gone one step forward with the combination of proximity marketing using low energy Bluetooth Beacons technology and other internet of things (IOT) with market basket analysis (Shinde P. et.al, 2017). This advancement can make the shopping experience of acustomer better and cost-effective and very easy. An attractive campaign can be organized to ensure customer engagement and sales conversions.

Bibliography

- Aggarwal C, Wolf J. A New Method for Similarity Indexing of Market Basket Data. New York. (1999)
- Boztug Y, Reutterer T. 2006. A combined approach for segment-specific market basket analysis. Austria: Elsevier, 294-312
- Chen Y, Tang K, Shen R, Hu Y. 2004. Market basket analysis in a multiple store environment. USA: Elsevier,339-354
- Chib S, Seetharaman P, Strijnev A. 2002. Analysis of multi-category Purchase incidence decisions Using IRI market basket data
- Decker R, Monein K. 2003. Market basket analysis with neural gas networks and self- organising maps. Germany: Henry Stewart Publications, 337-386.
- Huang C, Chaung K, Chen M. 2002 An Efficient Clustering Algorithm for Market Basket Data Based on Small Large Ratios. Taiwan.
- Kaur M, Kang S. 2016. Market Basket Analysis: Identify the changing trends of market data using association rule mining. Sanguar: Elsevier, 78-85
- Lee J, Jun C, Lee J, Kim S. 2005. Classification-based collaborative filtering using market basket data. South korea. Elsevier, 700-704
- Mild A, Reutterer T. 2002. An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data. Austria.
- Muller, Henning, Pun, Thierry, Squire, David. Learning from user behavior in image retrieval : application of market basket analysis. International Journal of Computer Vision,2004, vol. 56, no. 1- 2, p. 65-77
- Nanopoulos A, Manopoulos Y. 2002. Efficient similarity search for market basket data. Greece: Springer-verlag.
- Szymkowiak M, Klimanek T, Józefowski T. 2018. Applying market basket analysis to official Statistical data. Poland. DE Gruyter. Vol-22.
- Samaraweera WJ, Waduge CP, and Meththananda R. Market Basket Analysis: A Profit Based Approach to Apriori Algorithm. Sri lanka.2017
- Shende P; Mehendarge S; Chougule S; Kulkarni P ; Hatwar U. 2017 .Innovative ideas to improve shopping mall experience over E-commerce websites using beacon technology and data mining algorithms. IEEE.
- Yun C, Chaung K, Chen M. Using Category-Based Adherence to Cluster Market-Basket Data. Taiwan.