Journal of Mines, Metals and Fuels, 71(5): 656-660; 2023. DOI: 10.18311/jmmf/2023/34166

Print ISSN : 0022-2755
Journal of Mines, Metals and Fuels
Contents available at: www.informaticsjournals.com/index.php/jmmf

AI Driven Customer Segmentation and Recommendation of Product for Super Mall

Dipanwita Dey and Kallal Banerjee

Department of Management Studies, Swami Vivekananda University Barrackpore, West Bengal 700121.
**E-mail: post.for.phd@gmail.com, kallalb@svu.ac.in*

Abstract:

The prime objective of the paper is to propose customer segmentation for the retail customer for a large retail chain super mall. Along with customer segmentation, product recommendation is integrated with the system. Segmentation is required to identify the customer with specific behaviour. The real-world data comes from the customer support department of a retail chain super mall. For customer segmentation, the k- means algorithm is applied by using RFM data and the Association Rule Mining algorithm is applied to create the mapping rules between customer segments and favourite products. The customers are segmented into "Traditional", "High Spending", "Occasional", "Low Spending", "Disloyal" and "Frequent Buyer". Customer segmentation is necessary for efficient marketing strategy providing discounts, promotions, campaigns, etc. Moreover, the ARM algorithm is used to map the preferred combo for all customer segment. This combined approach is used to create an effective marketing strategy for retail chain management.

Keywords: *Clustering, Segmentation, retail chain, association rules, RFM*

1.0 Introduction

India is one of the largest retail destinations in the world. According to studies, Indian retail market is anticipated to pick up 10.7% by 2024 compared to 4.7% at 2019. Mass marketing is not an effective strategy in present days. Especially in the retail sector, "one-size-fits-all" strategy is ineffective, time consuming and expensive in nature[1]. Each of the segments shares the same traits and behavioural similarity within the cluster. In general customer segmentation can be done in the following ways[2].

- Demographic: personal information e.g. age, income, education, gender, stage of life cycle
- Geographic: location e.g. region, city size, population density
- Behavioural: transactional behaviour e.g. browsing, spend by category, price point
- Psychographic: qualitative traits e.g. social status, life style, habits, attitudes

Nowadays, data mining technique is used to find out the hidden characteristic of the customer . It will also helpful to determine CRM strategy and customer value. The segmentation helps for better campaigning, finding target customer. Moreover, it will serve the following in best ways[3]:

- Customer Retention: To find out the customer who churn out quickly and provide special attention to these customers. Find out the new potential customers.
- Better Brand Strategy: By using segmentation, it is easy to find out prime motivation, needs and wants of customer
- Find out New Market: By customer segmentation, new market can be found out and marketing strategy can be enhanced by finding out different hidden insights of customer behaviour.
- Enhanced Distribution: Segmentation can identify the preferred channel from where customer purchase the products.

**Author for correspondence*

In this paper, the sample product list contain daily FMCG products.

2.0 Literature Review

Due to diversity, customers respond differently for wants and needs. So segmentation separates based on similarities according to different criteria[4]. Segmentation helps to identify new customer, their demand and improving revenue[5,6]. There are different types of parameter used in the customer segmentation. It may be customer purchase behaviour, customer demographic parameter. RFM is first introduced by Huges in the year of 1994[7]. RFM stands for recency: time elapsed after recent purchase, frequency: how many time of purchase during a period, Monetary: total amount spent within the tenure[8]. There are some other variations in RFM. LRFM is another popular variation, which is proposed by Peker. There are parameters L(Length)[9-11]. Parameter G(Group) is used to specify the product category[12]. There is another important parameter P (Periodicity). This techniques differentiate the object based on similarities and differences[13]. There are many techniques for cluster divided into two categories: Hierarchical and non-Hierarchical[14]. Among the techniques *k*-means algorithm is most popular. Before proceeding in *k*-means algorithm, it is required to find the value of *k* i.e, the number of cluster. First of all, random *k* –center of points is chosen arbitrarily and find out the closest distance. Then centre of cluster is chosen again. Same process is repeated again until value of *k* is optimized[15]. Moreover, *k*-means reduces the inter-cluster heterogeneity and increase intra-cluster homogeneity .Association Rule Mining is first defined by R Agrawal, T Imielinski, A Swami[16]. It will find out the pattern from sequential data. *k*-item size item sets will give (*k*-1) number of patterns[17]. ARM has three measure: Lift, Confidence, Support. Support demonstrates how many times an item set appear in a set. Confidence measures how many times rule is true. Lift gives the relation between antecedent and consequent[18]. Some authors treats segmentation[19] and association rule mining[20] separately on retail data. Some of the authors also used the combined approach in the other dataset[21] also. However, there is still a shortage of implementing this combined approach in various types of retail dataset specially with FMCG data. The aim of the paper is to fill up the gap.

3.0 Objective

We tried to develop customer segmentation by *k*-means algorithm to make the cluster of customer according to their financial behaviour by using recency, frequency, monitery. The main objective is to identify the most profitable cluster along with finding the hidden insights of customers with high accuracy of prediction. By using association rule mining, product bundle is created for each of the cluster. Based on the product bundle, product recommendation is given to customer. This model also enhances the chances of cross-sell of the products to customer in future.

4.0 Methodology

Customer Segmentation

In the current approach, segmentation is implemented by applying RFM and clustering methods. For clustering purpose, *k*-means algorithm is used. This is an unsupervised learning technique. Finding the number of cluster has some thumb rules: (a) The smallest cluster should not be too small (not less than 2%-3% of population), (b) The largest cluster should not be too big (35%-40%), (c) The distance between the centroid of cluster should be high as possible, i.e. cluster should be distinct as much as possible, (d) The radius of cluster should be small as possible. (c) and (d) are termed as inter-cluster heterogeneity and intra-cluster homogeneity respectively. Scree plot along with Elbow method is applied to find the optimal cluster.

Association Rule Mining

Let I be an Item set and X and Y be the two subset of I . By “Association Rule” X→Y implies that “If customer purchase item X, he/she will purchase item Y most likely”. Support, Confidence and Lift formula are given below[22]

$$\text{Support } (X \rightarrow Y) = P\ (X \cap Y) \qquad \text{... (1)}$$

$$\text{Lift } (X \rightarrow Y) = P\ (X|Y) \qquad \text{... (2)}$$

$$\text{Confidence } (X \rightarrow Y) = P(X|Y)/\ P(Y) \qquad \text{... (3)}$$

Association Rule X→Y will be useful only if (1) Support (X) and Support (Y) value is greater than threshold value, (2) X→Y has a Confidence value greater than threshold value. This threshold value will be determined by business requirement.

Tools

Python and Azure Machine Learning Studio is used as programming language and development platform respectively for proposed approach. It is a cloud based tool. Azure Machine Learning Tool can perform all end-to-end functionalities for machine learning development life cycle such as data processing, build and train model, deploy and monitoring. The tool is user friendly and cost effective in nature.

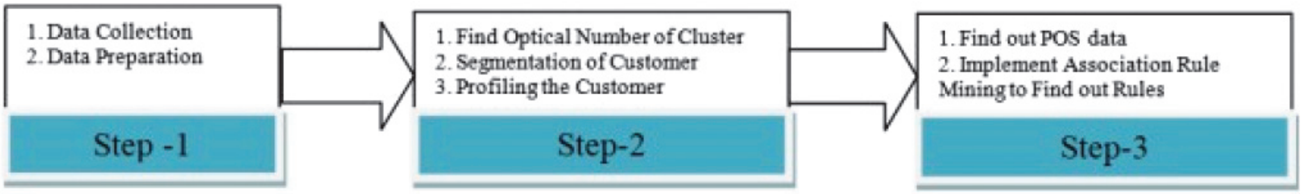

1. Data Collection
2. Data Preparation
Step -1
1. Find Optical Number of Cluster
2. Segmentation of Customer
3. Profiling the Customer
Step-2
1. Find out POS data
2. Implement Association Rule Mining to Find out Rules
Step-3

Fig 1 : Dataflow Diagram of Approach

5.0 Data Analysis and Result

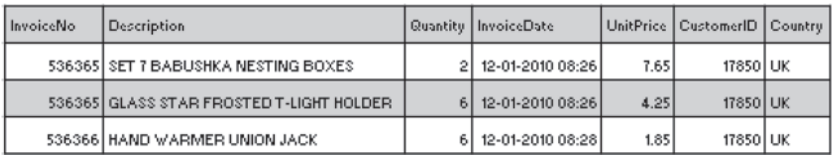
This paper proposes the customer segmentation of retail sector followed by the market basket analysis of products using Aprori algorithm. The flow for the methodology is followed in Fig.1.

Section-1

During first phase of approach, (a) Data collection and (b) Data preparation take place as per Fig.1.

Data Collection: Customer data is collected from the supermall customer care department of Taiwan based Super Mall. The number of records extracted is 51904 and the range of time period is 3 years. The data is transferred to Oracle database. There are some raw hypothetical data of customer care looks like in Table 1.

Table 1: Format of raw customer data sample



InvoiceNo	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	SET 7 BABUSHKA NESTING BOXES	2	12-01-2010 08:26	7.65	17850	UK
536365	GLASS STAR FROSTED T-LIGHT HOLDER	6	12-01-2010 08:26	4.25	17850	UK
536366	HAND WARMER UNION JACK	6	12-01-2010 08:28	1.85	17850	UK

Source: Customer Care Department of Taiwan-based Supermall

Data Preparation: At first, missing value treatment is done on raw data. Customer records with any one blank field will be dropped. Then data will be aggregated to create three features: Recency, Frequency and Monetary. Recency is defined as number of days elapsed last purchased. Frequency is the total number of transaction. Monetary is defined as the total money spent by customer. After that, outlier treatment will be done. Here Inter Quartile Range method is applied to cap both lower and upper outliers.

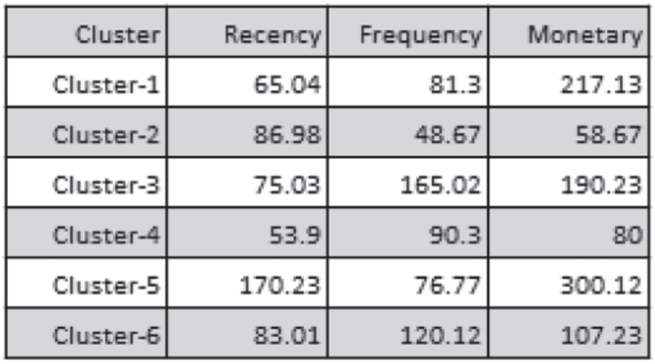
Section-2

During the second phase of approach, (a) Finding optical number of cluster, (b) Segmentation of customer based on RFM, (c) Profiling of customer takes place as per Fig.1.

Finding Optical Number of Customer: In the present case, *k*-means algorithm is used with RFM data. The main objective of *k*-means algorithm is to minimize total – with in sum cluster square WSS. The variables are normalized using the value of 0 and 1. The algorithm reaches an optimal point between inter-cluster heterogeneity and intra-cluster homogeneity. By using Elbow method, optimal cluster number can be implemented. In Fig.2, it is observed that intra-cluster homogeneity is increasing and inter-cluster heterogeneity is decreasing sharply. After *k*=6, even increasing number of cluster intra-cluster homogeneity and intra-cluster homogeneity does not increase or decrease much. So *k*=6 is chosen as optimal solution. Segmentation of Customer: Table 2 represents the mean for each of the feature for each cluster with respect to overall population in percentage format.

Profiling Customer: Table 3 represents the different range of value using different colour. Different value ranges are used to profiling the customer with respect of marketing analyst. Referring Table 3, for Cluster-5 customers Recency is low to medium and Monetary is higher than population mean. These “High Spending” customers are really asset for any business. But Recency is in higher side. So these customers can be give more personalized discounts, offer for maximization of revenue. Cluster-3 customer has higher Monetary and Frequency than population mean. Recency is also low to medium. These type of “Frequent Buyer” maintains a long term relationship with the business. Loyalty points, spot discount will retain these people in long run. Cluster-3 customer has lower Frequency and Monetary and medium Recency than population mean. These “Disloyal” can be offered better pricing than competitors mall.

Table 2: Mean of six cluster for each features



Cluster	Recency	Frequency	Monetary
Cluster-1	65.04	81.3	217.13
Cluster-2	86.98	48.67	58.67
Cluster-3	75.03	165.02	190.23
Cluster-4	53.9	90.3	80
Cluster-5	170.23	76.77	300.12
Cluster-6	83.01	120.12	107.23

Source: Using Author’s Calculation

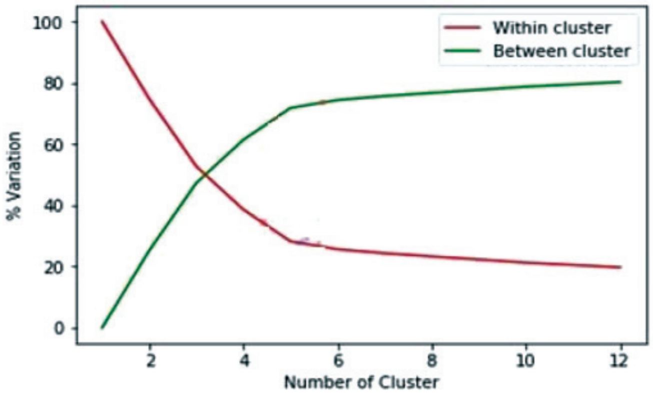
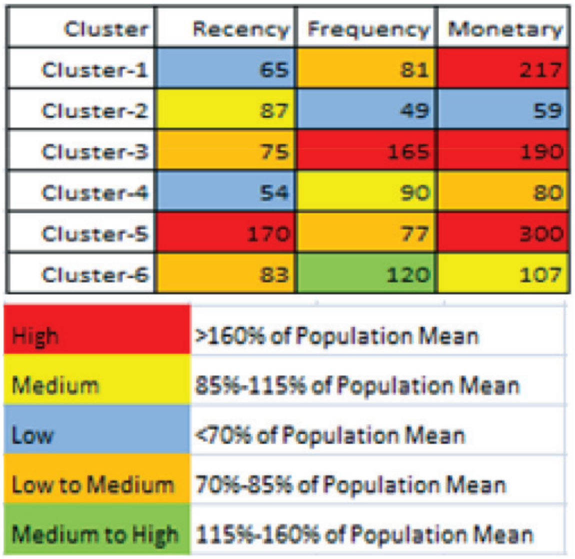


Figure 2 : Finding cluster by elbow method

Table 3: Customer profiling chart for six clusters



Cluster	Recency	Frequency	Monetary
Cluster-1	65	81	217
Cluster-2	87	49	59
Cluster-3	75	165	190
Cluster-4	54	90	80
Cluster-5	170	77	300
Cluster-6	83	120	107

High	>160% of Population Mean
Medium	85%-115% of Population Mean
Low	<70% of Population Mean
Low to Medium	70%-85% of Population Mean
Medium to High	115%-160% of Population Mean

Source: Using Author's Calculation

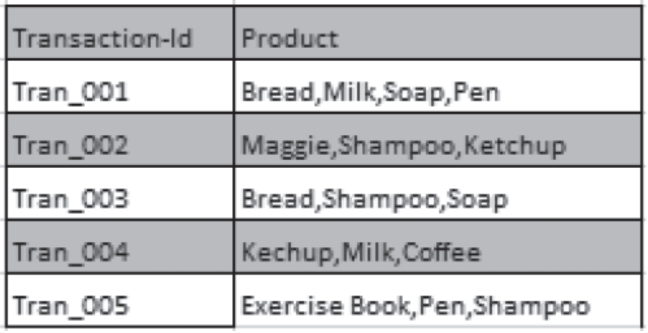
Section 3: During the third phase of approach, (a) Find out POS data, (b) Implement Association Rule Mining to Find out Rule as per Fig.1.

Find out POS data: In present case, association rule mining is applied on point of sale data where billing and transactions are done. The total number of records is 79876 and range of time period is 3 years. There are some raw hypothetical taken from POS are as given in Table 4.

Implement Association Rule Mining to Find out Rule: First, these raw data are segregated into 6 customer segments. Below is the data for the customer segment "Frequent Buyer". The data will be represented by 0 and 1. Item set for this segment customer is I={Coffee, Milk, Bread, Soap, Shampoo, Exercise Book, Pen}

By using ARM, different product bundle will be generated for each of the customer segment. In present case, confidence threshold=75% and support threshold=20%. From Table 6, "Shampoo→Soap" is not useful as Confidence is low. "{Coffee, Biscuit}→Milk" is not useful as support is low, though confidence is high. For this customer segment, {Bread→Milk} and {Coffee→Milk} are useful association rule mapping. Moreover, high Lift value indicates the high possibility of togetherness. These are offered as product bundling to the customer segment "Frequent Buyer".

Table 4: Format of raw point of sale data sample



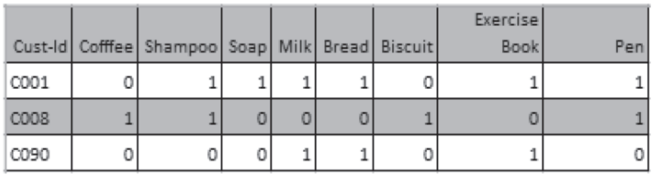
Transaction-Id	Product
Tran_001	Bread,Milk,Soap,Pen
Tran_002	Maggie,Shampoo,Ketchup
Tran_003	Bread,Shampoo,Soap
Tran_004	Kechup,Milk,Coffee
Tran_005	Exercise Book,Pen,Shampoo

Source: Point of sale for Taiwan-based Supermall

Limitation: Though customer segmentation is a very important tool for maximization the profit, it has some limitation which is as follows[23]

1. Data Quality: Since segmentation is based on the historical data, it will give improper result if data quality is not good. Due to improper maintenance of data warehouse, the attributes used in segmentation can give wrong segment.
2. Gap in business translation: Quite often, business people cannot understand the business definition properly and used in wrong way.

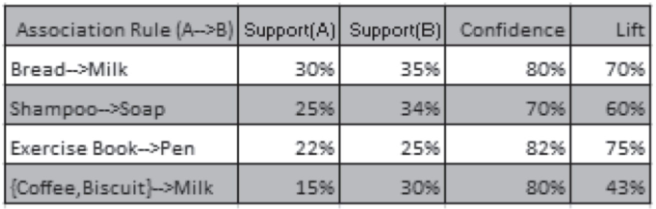
Table 5: Transaction data for "Frequent Buyer" customer sample data



Cust-Id	Cofffee	Shampoo	Soap	Milk	Bread	Biscuit	Exercise Book	Pen
C001	0	1	1	1	1	0	1	1
C008	1	1	0	0	0	1	0	1
C090	0	0	0	1	1	0	1	0

Source: Using author's calculation

Table 6: All possible association rules for "Frequent Buyer" Customer Segment



Association Rule (A-->B)	Support(A)	Support(B)	Confidence	Lift
Bread-->Milk	30%	35%	80%	70%
Shampoo-->Soap	25%	34%	70%	60%
Exercise Book-->Pen	22%	25%	82%	75%
{Coffee,Biscuit}-->Milk	15%	30%	80%	43%

Source: Using Author's Calculation

6.0 Conclusion

Due to rapid change in customer behaviour, personalize customer care is key success factor now a day. Customer also feels themselves privileged and maintains a long term relation with business. This paper basically focuses into super mall customer segmentation and recommendation of customized product bundle. By using the segmentation, customer base is segmented into six cluster namely "Traditional", "High Spending", "Occasional", "Low Spending", "Disloyal", "Frequent Buyer". Each of this segments is dealt by marketing team differently. Association Rule Mining provides the product bundle for each of cluster for future provision of cross-selling. This study can be utilized theoretically and practically for hypermarket, e-commerce store for customer segmentation. Not only retail sector, this study can be implemented in tourism, financial and other sector as well based on their spending behaviour.

7.0 Reference

1. Indian Retail Industry Analysis, available at: https://www.ibef.org/ industry/indian-retail-industry-analysis-presentation
2. Emily Smith, "Five Types of Market Segmentation & How To Use Them in 2021, available at: https://blog.remesh. ai/5-types-of-market-segmentation-how-to-use-them.
3. The Importance of Customer Segmentation, available at: https://uplandsoftware.com/bluevenn/r esources/blog/the-importance-of-customer-segmentation/
4. M. McDonald, and I. Dunbar, Market Segmentation: How to Do It, Howto Do It, How to Profit From It. Oxford, a. U.K.: Butterworth-Heinemann, 2004.
5. C. Chan, "Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer," *Expert Syst.Appl.*, vol. 34, no. 4, pp. 2754– 2762, May 2008.
6. T. Chen, A. Lu, and S.-M. Hu, (2012): "Visual storylines: Semantic visualization of movie sequence," *Comput. Graph.,* vol.36, no.4, pp. 241–249, Jun.
7. B. Izadi and A. Sabaghinia, (2014): "RFM- based e-markets segmentation using self-organizing maps," *J. Econ. Manage.*, vol.3, no.12, pp.86-96.
8. A. M. Hughes, (1996): "Boosting response with RFM," *Marketing Tools*, vol.3, no.3,pp. 4–8.
9. S. M. S. Hosseini, A. Maleki, and M. a. R. Gholamian, "Cluster analysis using datamining approach to develop CRM methodology to assess the customer loyalty," *Expert Syst. Appl.*, vol.37, no.7, pp.5259–5264, Jul. 2010.
10. Y. T. Kao, H. H. Wu, H. K. Chen, and E. C. Chang, (2011): "A case study of applying LRFM model and clustering techniques to evaluate customer values," *J. Statist. Manage. Syst.*, vol.14, no.2, pp.267– 276, March.
11. D.C. Li, W.L. Dai, and W.T. Tseng, (2011): "A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business," *Expert Syst. Appl.*,vol. 38, no. 6, pp. 7186–7191, Jun.
12. C. Chang and H.P. Tsai, (2011): "Group RFM analysis as a novel framework to discover better customer consumption behaviour," *Expert Syst. Appl.*,vol. 38, no. 12, pp. 14499–14513, Nov.
13. H. Jiawei, and M. Kamber, (2001): Data Mining: Concepts and Techniques, vol.5. San Francisco, CA, USA: Morgan Kaufmann.
14. Comparison of Hierarchical and Non- Hierarchical Clustering Algorithms, available at:https://www.proquest.com/ openview/f 509f82f8d5184afd7b935efcadc0c3e/1?p q-origsite=gscholar &cbl=2044551
15. P. Michaud, (1997): "Clustering techniques," Future Gener. *Comput. Syst.,* vol.13, nos. 2–3, pp.135–147, Nov.
16. Y. Liu, H. Li, G. Peng, B. Lv, and C. Zhang, (2015): "Online purchaser segmentation and promotion strategy selection: Evidence from Chinese E- commerce market," *Ann. Oper. Res.*, vol.233, no.1, pp. 263–279, Oct.
17. Mining Association Rules Between Sets of Items in Large Databases, available at:https://dl.acm.org/doi/ abs/10.1145/17 0035.170072
18. U. Gürsoy, Ö. A. Kasapo§lu, and K. Atalay, (2019): "Association rules analysis with R programming: Analyzing customer shopping data with Apriori and eclatalgorithms," *Alphanumer. J.*, vol. 7, no. 2, pp. 357–368, Dec.
19. Chen, D., Sain, S. & Guo, K. (2012): Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *J Database Mark Cust Strategy Manag* 19, 197–208.
20. Customer profiling and segmentation based on association rule mining technique available at: https://www.researchgate.net/publication/283654015_Customer_profiling_and_segmentati on_based_on_association_ rule_mining_technique
21. S. Guney, S. Peker and C. Turhan, (2020): "A Combined Approach for Customer Profiling in Video on Demand Services Using Clustering and Association Rule Mining," in *IEEE Access*, vol.8, pp. 84326-84335.
22. Z. Kou, "Association rule mining using chaotic gravitational search algorithm for discovering relations between manufacturing system capabilities and product features," *Concurrent Eng.*, vol.27, no.3, pp.213–232, Sep. 2019.
23. Challenges of Customer Segmentation, available at:https://www.ironsidegroup.com/201 9/02/06/challenges-customer-segmentation.