Discovering adjacency relationships of goods by using association rules to optimizing hypermarket layout

Mahdi Shakeri[1]
Mahmood Mirzaei Hesari[2]

1. PhD Student of System Management, Semnan University, Semnan, Iran
Email: mahdi.shakeri@semnan.ac.ir
2. PhD International Marketing, Tarbiat Modares University of Tehran, Tehran, Iran.
Email: mahmod.mirzaei@gmail.com

Abstract: These days understanding of customer preferences and orientation of industrial activities and services to recognize customer needs is especially important. One of the most useful areas of data mining is customer relationship management. The importance of issues such as understanding customer needs, maintain and increase the profitability of organization, the need for using data mining techniques have been highlighted. Among this, in the retail industry is also due to a variety of databases, data mining can be used to identify customer needs and improve sales. The research implements data mining processes to understanding customer behavior in relation to mean products purchased on the basket. Knowledge of this relationship will appropriate layout for the sale of goods in the vicinity of them. By using the technique of association rules as a data mining tools, discovering the relationships between groups of goods purchased by customers has been done in one of the best-selling hypermarket in Mashhad that achieving the proper layout within determining the importance of commodities of groups together. **Keywords:** data mining, association rules mining, Adjacent of commodity Groups, hypermarket layout.

1. INTRODUCTION

Today's in the world of technology, there is a large amount of raw data that alone is not applicable. But we can use data mining techniques to make the best utilization of these raw data. Data mining is a good tool to extract best pattern and information from the raw data. With increasing competition around the world, companies must use raw data and information technology to forecast market condition in the coming months (Nafari & Shahrabi, 2010). This process leading to companies take important decisions in their working environment that is very effective on their performance (Ciprian & Alexandra, 2013). Also, the issues of customer orientation and attention to customer needs have importance in marketing industry and even the potential needs of customers greater importance (Massara et al, 2014). In this study, conducted transactions of customers who already had completed their shopping, have been examined and by using of data mining techniques, items have been purchased by the customer are analyzed to present good offers for customers potential purchases. In this way any customer with any selection of products, can recieve best offers for other products within the vicinty of goods.

Marketing research shows that there are many incentives factors in the hypermarket such as allocation of shelf space and different product views that can have a significant impact on consumer buying behavior and create a considerable demand for other products (Valenzuela et al, 2013). There are few empirical researches has been made on the effect of proximity of goods within the layout on the shelves on changing their sales. Detecting knowledge for the identification and classification of these relationships can be very important for executive managers in this industry (Castelli, 2018). The lack of a clear pattern in the arrangement of goods on hypermarket shelves lead to customers don't remind their potential needs and then these hypermarkets lose part of their potential sales (Liang Chen et.al, 2006). Knowing what groups of goods can be brought together to improve their sales and what goods if that are close to each other decrease their sales, can help in the proper ordering of the goods to boost sales (Moncer et al, 2015). Predicting customer behavior can be very useful to increase his purchasing and proper layout of store help customers to select products with least searches and most remindings. Visual effect of proximity can increase stimulation of purchasing decisions in hypermarkets up to 70% (Goswami & Gupta, 2013). Due to these points, in this study the implicit and important relationship between the distance of the displayed products and sale of items in a retail store using data mining techniques will be explored.

2. LITERATURE REVIEW

Searching in data mining literature review and its applications, it reflects the growing expansion of this tool in all disciplines including practical and academic. Applicability of this tool caused, in addition to researchers, increased focus of business owners and investors on this issue. So that a substantial number of studies that have been made in recent years in the field of data mining, devoted to applications of this field in industry and business. Ngai (2009) stated that between 2004 and 2009, in 24 relevant scientific journals, over 900 articles on this topic has been published. But these researches in the area of the arrangement of goods, have less focused on customer's buying patterns and more mathematical methods have been used for placement of goods on the shelves. Aloysius and Binu (2013), based on the profits derived from the goods using PrefixSpan algorithm, determine the location of goods on the shelves wich will cause less reliability in applications and don't has emphasis on customer buying patterns.

3. DATA MINING

There are different definitions for data mining in academic texts. A simple definition says data mining is sorting through data to identify patterns and establish relationships. Other says data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. Data mining can predict profitability of potential customers that can be converted into actual customers predict how long the customer will remain loyal and how likely will leave us. Some customers, changing their client to gain competitive advantage (Ozyirmidokuz, 2015). So, we can determine the value of customers through data mining, predict their future behavior and make informed decisions adopted in this regard (Yan et al, 2020).

4. APPLICATION OF DATA MINING IN MARKETING

Today, with the help of data mining capabilities, retailers can make it easier to attract and retain customers, optimize classification and arrangement of goods and gain better understanding of the actual demand. Marketing general process shows in figure 1.

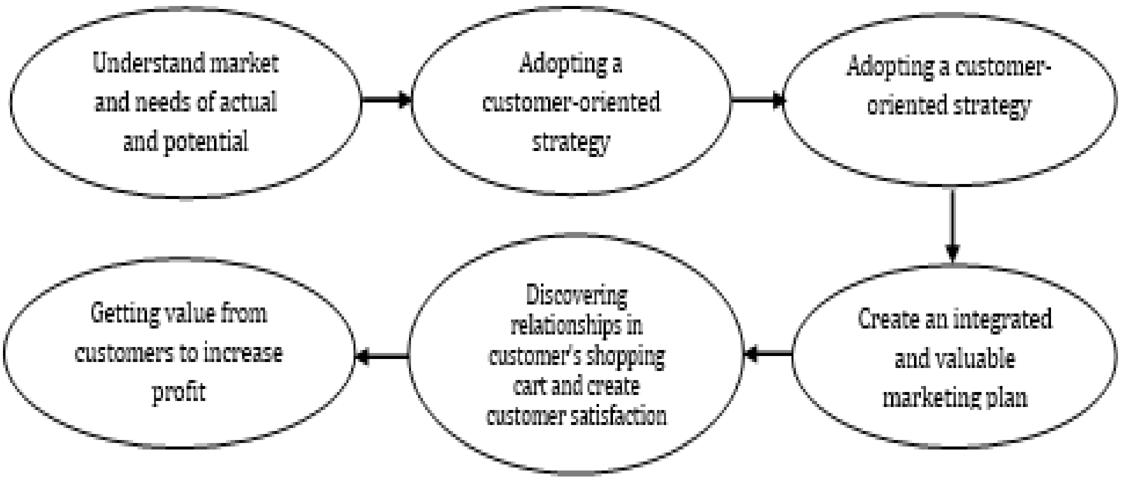

Figure 1 Marketing general process (Edwards, 2020)

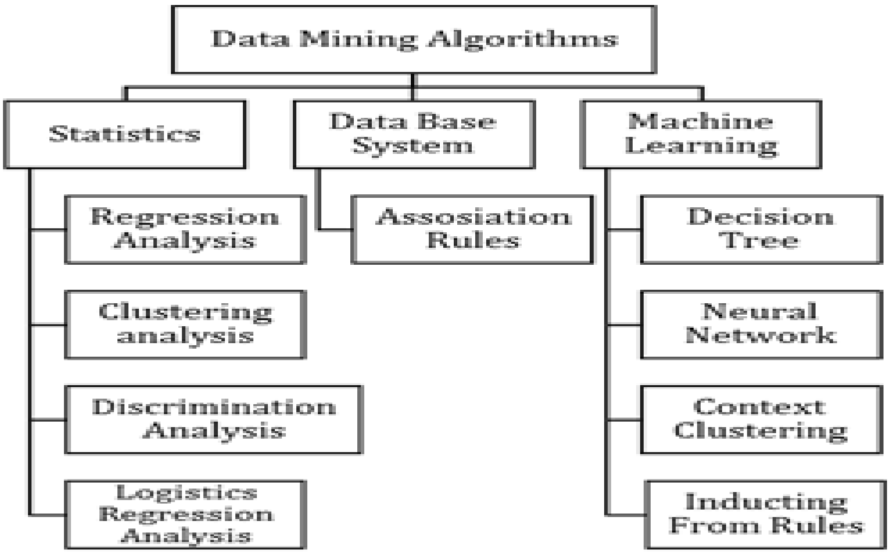

Figure 2 data mining algorithms

As shown in the figure, in cases where the use of database technology is required, the discovery of association rules will be the best and most applicable data mining algorithm. Recently most data mining algorithms used in the field of marketing, includes classification, clustering, regression and association rules (Femina Bahari, 2015). Table 2 shows the number of researches in this area by various data mining algorithms in the period 2010 to 2019.

Table 1 using of data mining algorithms in CRM domain

CRM approaches	Data mining algorithm	Number of researches
Identifying customers	Classification	5
	Clustering	6
	Regression	1
	Assosiation rules	0
Attracting customers	Classification	5
	Clustering	1

	Regression	1
	Assosiation rules	0
Customer maintenance	Classification	14
	Clustering	7
	Regression	2
	Assosiation rules	13
Customer development	Classification	1
	Clustering	2
	Regression	1
	Assosiation rules	5

As mentioned above, since this research related to the two pillars of maintaining and developing cuustomers, discovery of association rules would be appropriate algorithm to achieving its objectives.

5. DISCOVERING ASSOCIATION RULES

This method of data mining is to find a set of dependency rules according to that can be said wich set of data affecting other set of data. In other words, these rules predict the occurrence of an event based on the occurrence of other events (Lee, 2013). This research seeks to discover the rules by which can say what each person buys goods with what goods. The aim of discovering association rules is to find the order of the goods in carts. Here, for each cart, one law will be found and explored that whether this law applies in the other carts or not. Finally, a set of rules that apply in the greatest number of carts are provided as a set of association rules.

6. RESEARCH METHOD AND CASE STUDY

The data used in this study, has been collected from “Hamyaran Sepehr Rastin” company’s database, which is exploiting General Provisions Hypermarkets of Mashhad Municipality. Data is related to the customer’s purchases of the largest hypermarket of the company named Piroozi store. The reason for choosing this store as a case study, is the large number of customers and extent diversity of them and much higher sales volume than other stores. In the process of research, several tools used at different stages such as: Oracle Database, Excel for preparing data and Rapidminer for making and evaluating models. Two methods Crisp-DM and SEMMA that are most known data mining process, was reviewed. Analysis showed that Crisp-DM method has higher flexibility than the SEMMA method. According to the data available for this research and imminent need to implement pre-processing on raw data sets, the crisp-DM method was more effective.

7. RESEARCH STEPS

A. Understand the business problem: This step includes business goals and plan appropriate correspondence between operational objectives and functions of data mining.

B. Understanding data: This step involves the collection, characterization and quality control of data that have been obtained from transactions of customers with the organization. This data gathered from sales of goods to different customers. Total factors (number of customers) in over the course of a week to the number 9477, for each factor, 15 fields identified and extracted as define in Table 3.

Table 2 Data fields of Hamyaran Sepehr transactions

Name in database	Variable	Description
Action_ID	Factor Number	Is a unique identification code that is registered in system for each payment. It is a number start from 20000000000 and the first number shown store code.
SAL	Year	Hijri date from 1300 to 1400
MAH	Month	From 1 to 12
Rooz	Day	From 1 to 31
SITE_NAME	Store name	This name derived from location of store
STUFF_CODE	Barcode of goods	It is 13-digit code for Iranian goods and beginning with 626, 5 numbers then is the manufacturer code in the standard organization
STUFF_NAME	Product name	Nature of product + packing type + size + brand
SHOP_NAME	Store stand	
LOAN_SUP_ NAME	Bonded goods supplier name	These suppliers bring their goods to store and they're responsible for their sales.
SUPPLIER_ NAME	Definite supplier name	These suppliers sell their goods to the store.
Sum(SALE_ QUANTITY)	The number of customer purchase	
Sum(STUFF_ PURE_PRICE2)	Commodity prices after reducting discounts	
STUFF_ LOCATION	Goods location	
NAME_LVL2	Product group level 2	
NAME_LVL3	Product group level 3	

In the management information system, products group level one, are all of hypermarket goods (Fentoni et.al, 2017). In other words, any goods that are essential in a hypermarket. Due to the high volume of product group level one, product group level two defined as a subset of it. To put it simply, if the level one is considered as all products available in store, product leve two, examining the goods to find the adjacent relationship with each other and we shelve.

C. Data preparation (Selecting, Cleaning & converting data): Data cleaning is an operation where problems with poor data are resolved. Data quality problems include noise, outliers, missing values, and duplicate data (Hasani et al, 2016). There were 15 noise data in this dataset due to a cash register sales error, all of which were detected & deleted. There were also three factors in the dataset exceeding three million tomans related to the organizational purchase that was outliers and removed from the dataset. In this study, the data size for the software is very large, therefore there is a need for vertical deleting of some data. Also, due to the diverse range of data in the data tables, at this stage, many of the sales table fields that were not relevant to the target, were omitted from the modeling phase. The number of fields of level 3 commodity group was 368 types in the time period. By refining and pruning this commodity classification according to the shelves available in the store, this number was reduced to 125 commodity groups. There were 32 types of level-2 commodity group fields, which were reduced to eight commodities with the amendments. This was due to the mistakes in the grouping of goods by users. Table 4 shows Level 2 commodity groups and the number of its subgroups.

Table 3 Frequency of commodity group level 2

Level 2 Commodity groups	Level 2 Commodity subgroups
Foodstuffs	57
Detergent and sanitation	47
Refrigerated products	12
Stationery	3
Others	3
Plastics	1
protein products	1
Fruitage	1
Total	125

D. Modeling and evaluation: To fulfill this step we need to determine suitable data mining algorithm and model evaluation functions and methods according to research problem. Given the purpose raised in the first phase of the CRISP-DM process as a business issue, in this step we use Rapidminer software for Association Rules Discovery and main process of this model shows in figure 3.

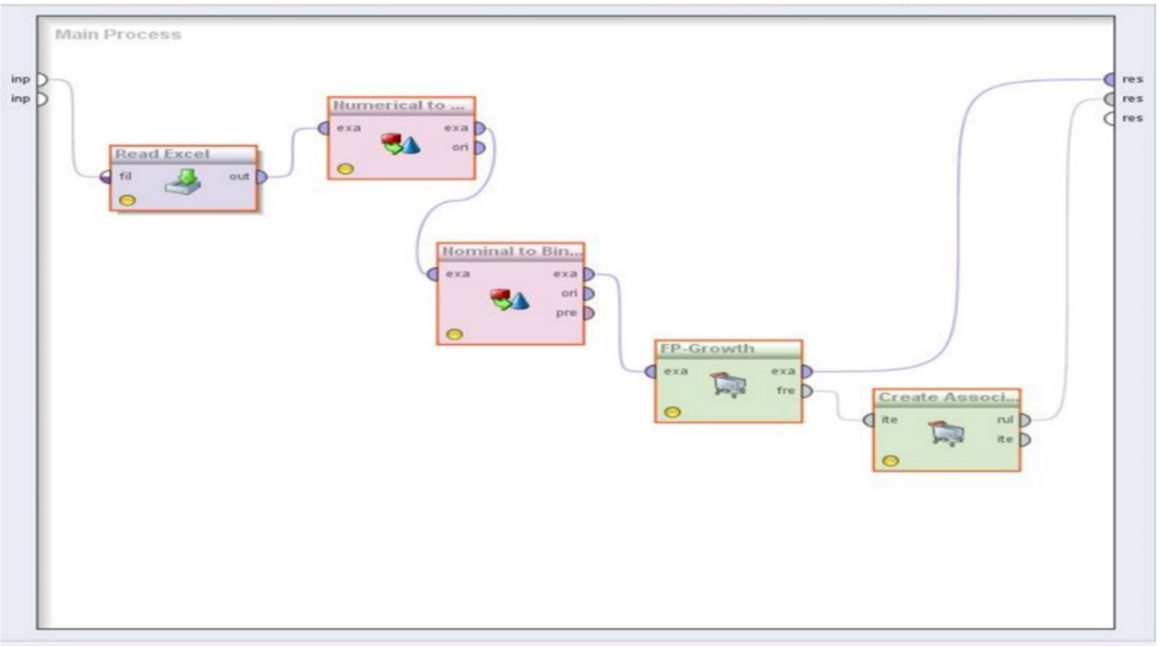

Figure 1 Association Rules model of this research

Given that a data mining process produces many patterns, there is a need for criteria to measure the suitability of constructed patterns. Each criterion is assigned an acceptance threshold that is set by the user. Patterns that do not meet this threshold are considered inappropriate (Moshkani & Nazemi, 2014). Different criteria have been proposed for this purpose including support factor, confidence factor, accuracy and time spent. But in many studies support, confidence & lift factor have been cited as the most appropriate criteria for evaluating patterns produced (Ahn, 2012). These two criteria are obtained through equation 1, 2 and 3.

$$(1)\ Support = \frac{frq\ (X,Y)}{N}\quad (2)\ Confidence = \frac{frq\ (X,Y)}{frq\,(X)}\quad (3)\ Lift = \frac{support}{supp(x).supp(y)}$$

In this study, due to the high volume and variety of commodity groups, these three criteria were used for evaluation. We insert the data obtained in the time period into the model in the rapidminer software to extract the association rules. For level 2 commodity group, 17 rules are extracted and with the criteria of support and confidence factor evaluation are presented in table 5. From study of Gupta et. Al (2006) A minimum support criterion 10% and a minimum confidence criterion 80% and lift criterion bigger than 1, are considered.

Table 4 Association rules for commodity level 2

Row	Rule condition	Rule result	Support	Confidence	Lift
1	Foodstuff, plastics	Detergent and sanitation	8.17%	82.12%	1.74
2	Refrigerated products, plastics	Detergent and sanitation, foodstuff	6.26%	87.16%	2.23
3	Refrigerated products, foodstuff	Detergent and sanitation	6.30%	87.61%	1.98
4	Detergent and sanitation, proteins	Foodstuff	4.48%	87.74%	2.66
5	Detergent and sanitation	Foodstuff	37.74%	87.94%	3.12
6	Refrigerated products, foodstuff, Detergent and sanitation	Detergent and sanitation	6.26%	88.37%	1.89
7	Refrigerated products	Foodstuff	42.65%	90.05%	4.31
8	Refrigerated products, fruitage	Foodstuff	6.3%	90.09%	1.57
9	Refrigerated products, proteins	Foodstuff	4.71%	91.08%	2.88
10	Detergent and sanitation, fruitage	Foodstuff	5.14%	93.41%	3.28
11	Plastics	Foodstuff	9.95%	94.08%	1.73
12	Refrigerated products, proteins, Detergent and sanitation	Foodstuff	3.13%	95.00%	3.01
13	Refrigerated products, Detergent and sanitation	Foodstuff	25.02%	96.56%	2.43

14	Detergent and sanitation, plastics	Foodstuff	8.17%	97.25%	2.69
15	Refrigerated products, Detergent and sanitation, fruitage	Foodstuff	3.59%	97.32%	2.04
16	Refrigerated products, plastics	Foodstuff	7.09%	98.62%	2.45
17	Refrigerated products, Detergent and sanitation, plastics	Foodstuff	6.26%	99.48%	1.39

According to the definitions of lift, confidence and support criteria, the most important rules in this set are the seventh-row law and then the fifth-row. Different rules were analyzed using RapidMiner software to discover the adjacency of commodity groups to each other regarding their combined sales and the isometric graph for each commodity groups in the rule's result were plotted. As the two subgroups of foodstuffs and detergent and sanitation commodities are in the field of the results, two diagrams can be plotted. Figure 4 shows the graph for the foodstuff subgroup in which all the rules for this subgroup are specified. This figure also shows the adjacency of the commodity groups that are imortant to being close to each other.

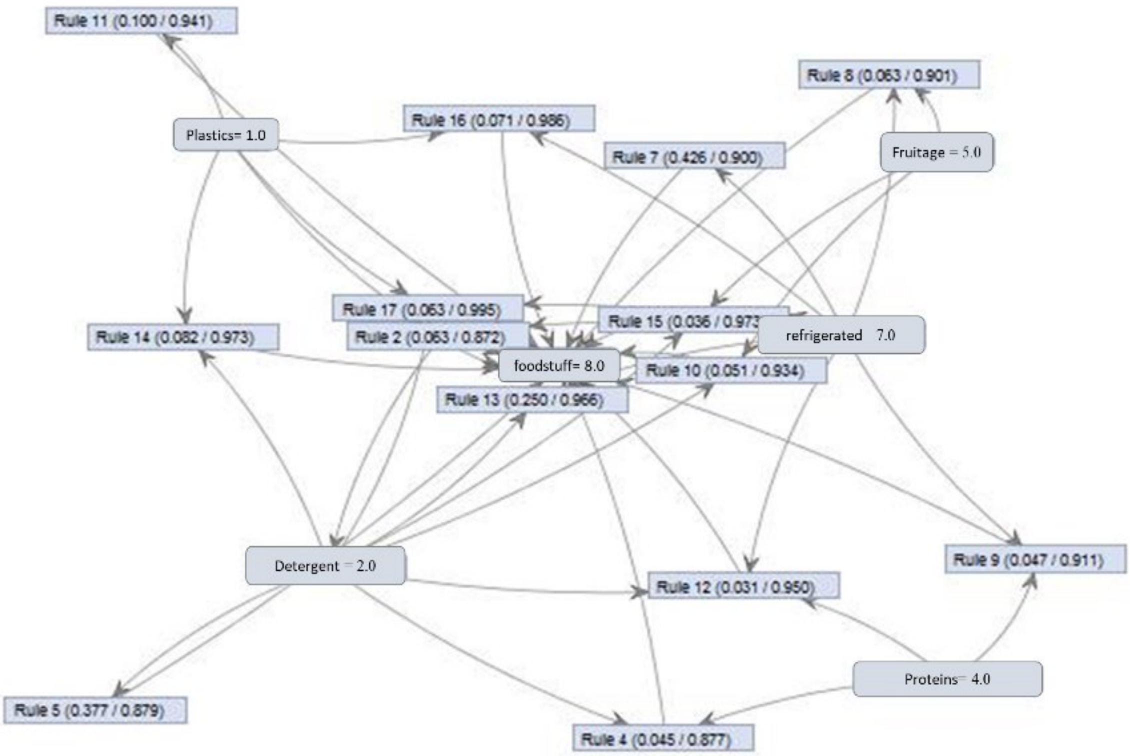

Figure 2 isometric graph of foodstuff subgroup

A careful analysis of Figure 3 concludes that the high coles proximity of refrigerated products and foodstuffs is important for the sale of them together, followed by the detergent, hygiene and food groups. This is due to the 42.6% support for Rule 7 and 37.7% for Rule 5. The graph also shows that meat and protein and fruitage stands are equally important for the food group. Similarly, the relationships of the detergent and sanitations subgroups are determined.

Figure 5 also shows the relationships between all Level 2 commodities in which the relationships among all subgroups are specified.

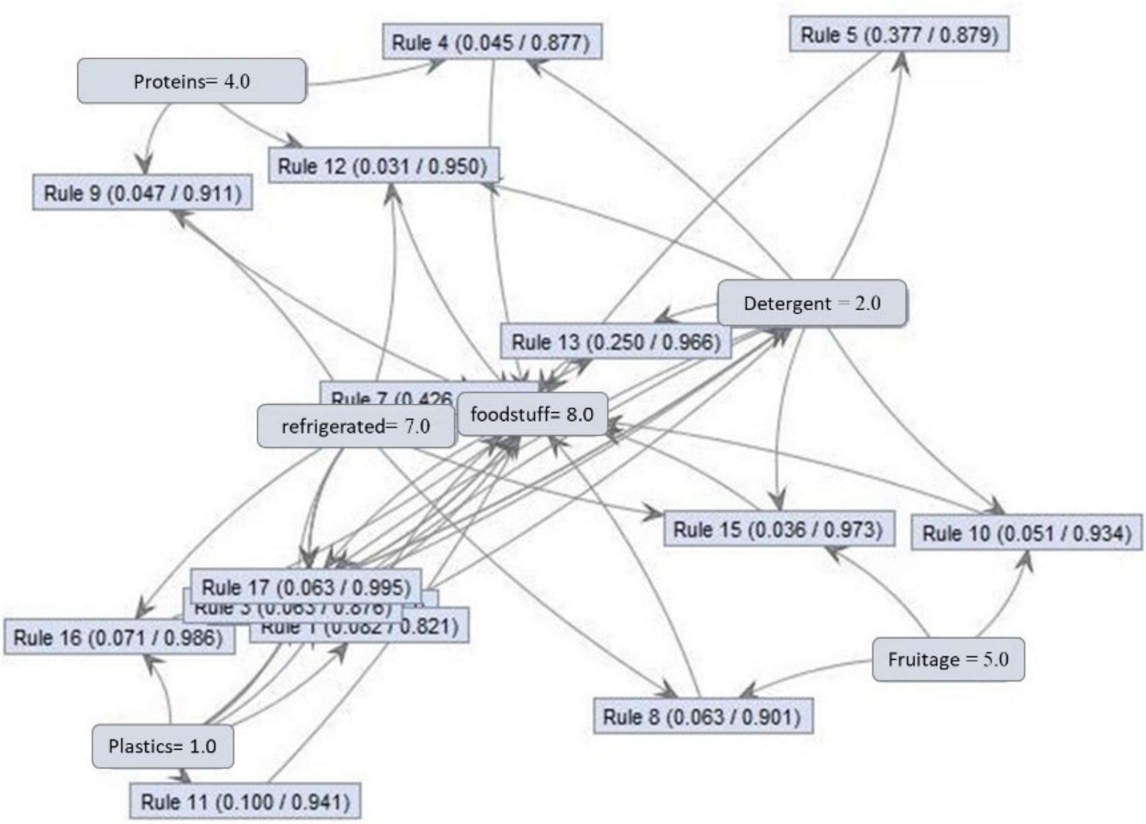

Figure 3 isometric graph for all of level 2 commodity group

Likewise, for the level 3 commodity group and its subgroups the associative rules were discovered and the model output containing 165 rules, including biscuits 65 rules, cheese 3 rules, macaroni 51 rules, tissue 6 rules, coreal 28 rules, tomato paste 11 rules & Spices 1 rule. For example, the rules extracted for the cereal subgroup as a food subgroup and a member of the Level Three commodity group included 28 rules which resulted imortant in the adjacency of the commodity groups of spices, tomato paste, pasta and cakes and biscuits. For all the subgroups of commodities, associative laws and their relationships were extracted, and their results analyzed, including:

- The macaroni commodity group and then the spice and liquid oil are important in the vicinity of the tomato paste group.
- Toothpaste, shampoo, soap and toilet paper are important in the vicinity of the paper towel group.
- The cream, jam, yogurt and milk commodities in the fridge are important in the vicinity of the cheese group.

8. DISCUSSION

Unfortunately, today, attention to blind and costly advertising has taken the place of intra-organizational issues which not only imposing heavy costs on the organization but also do not produce the desired results. While, increasing productivity within the organization at a lower cost and greater efficiency would be a better alternative to costly advertising. The core of this research is to investigate and identify sales patterns of commodity groups based on their proximity.

Given that shelf space in stores is limited, and smart and optimal layout of goods has a significant impact on sales. The results show that customers usually follow a certain shopping pattern. Understanding these patterns can significantly improve the store layout and responds to customer needs and finally maximize sales. Based on this goal, the discovery of any knowledge that indicates a relationship between the features extracted from the database is important to the researcher and can be effective in organization planning to enhance customer relationships and increase product sales.

All charts and tables related to the distribution of data can reveal unknown points about customer behavior and even the supplier behaviour. These tables and charts were provided to store managers, and they were welcomed to understand the hidden relationships between product groups and supplier customer sales patterns. For example, out of 190 suppliers, 36 of them have a major role in the store's sales. The discovery of associative rules in the supplier domain can also minimize the lost sales of each supplier.

In addition, the discovery of associative rules at level 2 (the main commodity group) can be used to design an activity relationship diagram, which is one of the inputs to systemic layout planning algorithms such as CRAFT and CORELAP.

Given the knowledge gained from the discovery of associative rules, it is important that orders should be based on the interrelationship of goods. Also, if associative rules are considered in hypermarket layout, lost sales will be reduced, and customer purchases are targeted. On the other hand, with knowing the relationship between commodities using associative rules, it is possible to sell low-sales goods in the store with a top-selling goods as a package with some offers. Also, using the association rules can find a suitable location for promotion sales goods that are in separate boxes from shelves, and by knowing the adjacency relationships with other goods, increase their sales.

REFERENCES

[1]. Aloysius, G., Binu, D. (2013)"An approach to products placement in hypermarkets using PrefixSpan algorithm", Journal of King Saud University - Computer and Information Sciences Volume 25, Issue 1, Pages 1-116

[2]. Ahn. K.Ill, (2012) “Effective product assignment based on association rule mining in retail”, Expert Systems with Applications, Volume 39, Issue 16, 15 November 2015, Pages 12551-12556, ISSN 0957-4174.

[3]. Berry, M.J., Linoff (2016). Data Mining Techniques: For Marketing, Sales, and Customer Support. Book. Second edition, Wiley Publishing Inc.

[4]. Castelli. M, Vanneschi. L, (2018) “Genetic algorithm with variable neighborhood search for the optimal allocation of goods in shop shelves”, Operations Research Letters, Volume 42, Issue 5, July 2018, Pages 355-360, ISSN 0167-6377.

[5]. Chen. MC, Chia-Ping Lin, (2007) “A data mining approach to product assortment and shelf space allocation”, Expert Systems with Applications, Volume 32, Issue 4, May 2012, Pages 976-986, ISSN 0957-4174.

[6]. Cil. I, (2012) Consumption universes based hypermarket layout through association rule mining and multidimensional scaling, Expert Systems with Applications, Volume 39, Issue 10, August 2015, Pages 8611-8625, ISSN 0957-4174.

[7]. Ciprian, t.c., Alexandra (2013). The effect of shelf display on purchasing decision, Romanian journal of experimental applied psychology, vol.4, issue 4.

[8]. Edwards. Chase J, Joshua S. Bendickson, Brent L. Baker, Shelby J. Solomon, Entrepreneurship within the history of marketing, Journal of Business Research, Volume 108, 2020, Pages 259-267, ISSN 0148-2963.

[9]. Femina Bahari. T, Sudheep Elayidom. M, An Efficient CRM-Data Mining Framework for the Prediction of Customer Behaviour, (2015) Procedia Computer Science, Volume 46, 2015, Pages 725-731, ISSN 1877-0509.

[10]. Frontoni. E, Marinelli. F, Rosetti. R, Zingaretti. P, Shelf space re-allocation for out of stock reduction, (2017) Computers & Industrial Engineering, Volume 106, 2017, Pages 32-40, ISSN 0360-8352,

[11]. Gerrikagoitia. J.K, Castander. I, Rebón. F, Alzua-Sorzabal. A, (2019) New Trends of Intelligent E-marketing Based on Web Mining for E-shops, Procedia - Social and Behavioral Sciences, Volume 175, 12 February (2019), Pages 75-83, ISSN 1877-0428.

[12]. Goswami, S., Gupta, N. (2013). A look at implications of changing store layout and design in India. Journal of Management Research, Volume-5, Issue-1, April 2013

[13]. Gupta, N., Mangal, N., Tiwari, K., Mitra, P. (2006) “Mining Quantitative Association Rules in Protein Sequences”, In Proceedings of Australasian Conference on Knowledge Discovery and Data Mining – AUSDM.

[14]. Hadi. H, El-Khalili. N, May AlNashashibi, Ghassan Issa, Abed Alkarim AlBanna, Application of data mining algorithms or improving stress prediction of automobile drivers: A case study in Jordan, Computers in Biology and Medicine, Volume 114, (2019), 103474, ISSN 0010-4825,

[15]. Hasani. M, Jam poor. M, Rahimi zadeh. K, (2015), Using data mining to investigate customer behavior and semantic relevance of shopping cart products, 2nd International Conference on Soft Computing and Information Technology, Mahshahr

[16]. Kinsey, Jean D. & Wolfson, Paul J. & Katsaras, Nikolaos & Senauer, Benjamin, 2001. "Data Mining: A Segmentation Analysis of U.S. Grocery Shoppers," Working Papers 14335, University of Minnesota, the Food Industry Center.

[17]. Kotu. V, Bala Deshpande, (2014) Chapter 2 - Data Mining Process, In Predictive Analytics and Data Mining, edited by Vijay KotuBala Deshpande, Morgan Kaufmann, Boston, 2019, Pages 17-36, ISBN 9780128014608.

[18]. Lee. D, Park. S.H, Songchun Moon, Utility-based association rule mining: A marketing solution for cross-selling, Expert Systems with Applications, Volume 40, Issue 7, (2013), Pages 2715-2725, ISSN 0957-4174,

[19]. Liang Chen, Y., Ming Chen J., Wen Tung C. (2006). A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales, Decision Support Systems.

[20]. Massara. F, Porcheddu. D, Melara. R, (2014) “Asymmetric Perception of Sparse Shelves in Retail Displays, Journal of Retailing”, Volume 90, Issue 3, September 2018, Pages 321-331, ISSN 0022-4359.

[21]. Moncer A. Hariga, Abdulrahman Al-Ahmari, Abdel-Rahman A. Mohamed, (2007) “A joint optimisation model for inventory replenishment, product assortment, shelf space and

display area allocation decisions", European Journal of Operational Research, Volume 181, Issue 1, 16 August 2012, Pages 239-251, ISSN 0377-2217.

[22]. Mu-Chen. C, Ai-Lun Chiu, Hsu-Hwa, (2010), mining changes in customer behavior in retail marketing, Expert Systems with Applications, Volume 28, Issue 4, May 2010, Pages 773-781, ISSN 0957-4174.

[23]. Nafari. M, Jamal Shahrabi, (2010) "A temporal data mining approach for shelf-space allocation with consideration of product price", Expert Systems with Applications, Volume 37, Issue 6, June 2014, Pages 4066-4072, ISSN 0957-4174.

[24]. Ngai, E.W.T, Xiu, Li, Chau, D.C.K, (2019), "Application of data mining techniques in customer relationship management: A literature review and classification", Expert systems with applications, Vol36, P2592-2602.

[25]. Ozyirmidokuz. E, Uyar. K, Ozyirmidokuz. M.H, A Data Mining Based Approach to a Firm's Marketing Channel, Procedia Economics and Finance, Volume 27, 2015, Pages 77-84, ISSN 2212-5671.

[26]. Peppers D., Roger M. (2016). "Managing Customer Experience and Relationships: A Strategic Framework, Edition 3", October 25, New York john Wiley, P. 5.

[27]. Suryana. A, Yulianto. E, (2019) "Application of Data Mining with Association Rules to Review Relationship between Insured, Products Selection and Customer Behavior" Universal Journal of Electrical and Electronic Engineering 6(3A): 45-61, 2019.

[28]. Valenzuela. A, Raghubir. P, Mitakakis. C, "Shelf space schemas: Myth or reality?", Journal of Business Research, Volume 66, Issue 7, July 2013, Pages 881-888, ISSN 0148-2963.

[29]. Van Nguyen. T, Zhou. L, Chong A.Y.L, Boying Li, Pu. X, Predicting customer demand for remanufactured products: A data-mining approach, (2020) European Journal of Operational Research, Volume 281, Issue 3, 2020, Pages 543-558, ISSN 0377-2217.