Retailer layout design: a novel hybrid approach with association rules mining and MCRAFT

Ali Osman Kuşakcı*

Department of Management,
School of Business,
Ibn Haldun University,
Istanbul, Turkey
Email: aliosman.kusakci@ihu.edu.tr
*Corresponding author

Elif Karakaya Cesur

Department of Industrial Engineering,
Faculty of Engineering,
Istanbul Medeniyet University,
Istanbul, Turkey
Email: elif.karakaya@medeniyet.edu.tr

Abstract: Spatial layout of a retail store is a crucial decision variable related to both utilisation of store area and purchasing behaviour of the customer. In this respect, the task of optimising the allocation of shelves to specific product segments has become a strategic decision to facilitate a more comfortable shopping environment for customers, which, in turn, increases sales volume. This study proposes a novel hybrid approach to facility layout design problem, which combines association rules mining (ARM) and a facility layout method, MCRAFT. The proposed methodology is composed of two main stages: 1) rule mining; 2) layout design. More specifically, the presented approach exploits the association rules obtained from purchasing records and uses them as a proximity measure input to MCRAFT algorithm to determine the layout of the store. The merit of the proposed methodology is shown with a case study on a prominent Turkish supermarket chain.

Keywords: association rule mining; ARM; facility layout algorithm; supermarket layout; apriori algorithm.

Reference to this paper should be made as follows: Kuşakcı, A.O. and Cesur, E.K. (2020) 'Retailer layout design: a novel hybrid approach with association rules mining and MCRAFT', *Int. J. Industrial and Systems Engineering*, Vol. 35, No. 4, pp.389–409.

Biographical notes: Ali Osman Kuşakcı received his Master's in Industrial Engineering with specification in Financial Engineering from the Vienna University of Technology. Thereafter, he joined into International University of Sarajevo and worked as a Research and Teaching Assistant in Industrial Engineering Department where he completed PhD program in 2013. He joined to the School of Business at the Ibn Haldun University. Among focal points of his research are operations management, decision support systems, heuristics, optimisation with nature inspired algorithms, and design of logistic networks, to where he actively contributes with his works published in selected journals.

Elif Karakaya Cesur is an Assistant Professor in the Department of Industrial Engineering at the Istanbul Medeniyet University, Turkey. She earned her PhD in Industrial Engineering from the Dortmund Technical University, Germany (2015). Her areas of research interests include modelling and simulation, particularly agent-based simulation, inventory management, supply chain management, logistics and neural networks.

1 Introduction

Today, information technology allows to obtain and to store a huge amount of data. However, the bulk data comes along with a difficult task: analysing the tremendously huge data and acquiring the correct information. Obviously, making an analysis by observing the tables with thousands of lines/columns and making useful conclusions are impossible. This necessitates employing computer technologies. Finding patterns, trends, and abnormalities and summarising as simple models in datasets is one of the most important issues in the information age (Gancheva, 2013). Data mining is the discovery of relations and rules which are significant, potentially useful and making predictions about the future through large amounts of available data using computer programs (Bükey, 2014).

One of the data mining application areas that are increasingly widespread in use in many sectors is association rule mining (ARM). In ARM, relationship and the rules are obtained taking advantage of the customer, product and sales information in a retail store. ARM, obtaining of products' sales relationship with another product and finding out the association rules, increases the profits of companies (Sherdiwala and Khanna, 2015).

A case study showed that effective store layout could stimulate demand to the point of doubling the sales rate by making it easier to find items and creating a positive image or feeling. Griffith and Harmgart (2005) examined the influences of store layout in online retailing (Cil, 2012). In contrast, the design and planning of retail facilities have been given little attention despite the importance of the retail sector to the overall economy. Supermarket chains are one of the fastest growing industries in the world. However, nearly all of the published work on facilities layout design has focused on manufacturing facilities (Cil, 2012). The aspect of the spatial layout is especially important (Lu et al., 2011): planning the circulation pattern of the customers and grouping the merchandise. The goal of circulation planning is to provide a path for the customers that expose them to as much of the merchandise as possible while placing any needed services along this path in the sequence they will be needed (Cil, 2012).

Due to aforementioned reasons, supermarket layout design is an increasingly important area in facility layout planning. For instance (Boros et al., 2016) takes supermarket layout issue into consideration in terms of owner aspect and tried to find shortest path using travelling salesman problem. Supermarket layout design is an increasing concern among both retail facilities and practitioners. Dong (2018) carried out multi agent simulation method in attempt to recognise layout pattern on the other hand (Page et al., 2019) configured both design and operational parameters such as aisles, store areas travel duration, etc.

The layout of a retail store plays a key role in its success. An effective layout should consider the following two aspects:

1 efficient use of the available space

2 convincing plan that promotes the sales.

Considering these two points, a good layout design not only reduces the company's storage cost but also boosts its profit since customers are expected to be more easily convinced to make purchasing decisions in a well-conceived store environment. Additionally, a careful arrangement of product groups leads to a change in customers' shopping habit, as they tend to visit more sectors and observe more shelves in the store. One of the very first studies conducted on the effect of store layout on shopping habits reveals that a good layout alone encourages customers to walk through the store and buy more (Farley and Ring, 1966). Another study by Borges (2003) states that finding a good store layout is a complex since the relationship between product groups as well as its impact on consumers' shopping behaviour and traffic in the store is not easy to reveal. Thus, considering its valuable contribution to the success of the firm, the layout problem in a retail store deserves extended attention despite its complicated nature.

In this study, an application ARM is conducted with apriori algorithm (Agrawal and Srikant, 1994) to identify the association rules between items sold in a retail store. Furthermore, the obtained rules are discussed and further implications are highlighted to construct a supermarket layout by taking the associations between specific product groups into consideration.

This work is organised as follows. The current section gives a brief introduction to the problem we focused while Section 2 is devoted to a comprehensive literature review on ARM applications in various fields, retail store layout approaches, and facility layout methods. Section 3 defines the problem and solution approach which is employed. Section 4 presents a case study where the product layout of a supermarket store is optimised. Lastly, Section 5 concludes the work with some remarks on future research directions.

2 Literature review

This section presents a literature review on three topics:

1 applications of ARM with special attention on the apriori algorithm

2 facility layout problem (FLP)

3 supermarket layout design.

ARM algorithms can be classified into three main classes:

1 frequent item set mining

2 sequential pattern mining

3 structured pattern mining.

More details on each class can be found in Yazgan and Kusakci (2016). Among many ARM methods, apriori algorithm, developed by Agrawal and Srikant (1994), has deserved special attention due to the fact it was one of the pioneers of first techniques designed for mining of rules on bulk data. Apriori is designed to work on databases including transactions and can be used to produce all frequent item sets.

Studies on dataset originating from various areas employed the apriori algorithm to extract association rules. Since its first introduction, it has been successfully applied in medicine (Abdullah et al., 2008; Chen et al., 2005; Zhang et al., 2014), education (Ahmed et al., 2009; Angeline, 2013), banking (Aggelis and Christodoulakis, 2003; Yang, 2013), e-commerce (Sharif et al., 2005), telecommunication (Jiang, 2011), finance (Xu and Zhang, 2009), marketing (Cil, 2012; Gancheva, 2013; Gürgen, 2008), tourism (Liu and Fan, 2013; Mu et al., 2009).

A considerable amount of literature has been published on FLPs that are referred to locating of departments in the facilities (Sharma and Singhal, 2016). Tompkins et al. (2010) state that it could be possible to reduce approximately 50% of total operation cost by supporting well-positioned departments. Preliminary work on layout problems was undertaken by Koopmans and Beckmann (1957). They considered the FLPs as a typical industrial problem that aims to create a facility in an attempt to minimise the transportation cost of materials, documents between workstations or departments. In another major study, Meller et al. (1998) investigated that FLP dealing with determining a non-overlapping arrangement in order to minimise the distance-based measure. Lee and Lee (2002) published a paper in which they described that FLPs include an arrangement of different size of locations within the total area. It could be restricted by the length and width of area to optimise total cost notably by taking transportation and slack area cost into consideration. A remarkable work by Altuntas et al. (2013) employed fuzzy association rules to determine facility layout in a cellular manufacturing system.

Up to date, many mathematical methods are exploited for solving layout design problems including metaheuristics (Balakrishnan et al., 2013; Moslemipour et al., 2018), integer programming (Abbasi et al., 2017), as well as fuzzy AHP-TOPSIS (Agarwal and Singholi, 2018). Some complex layout design problems could not be solved with analytical methods because of high number of variables and nonlinearity that the problems possess. In order to handle such NP-hard problems, a number of metaheuristics methods such as simulated annealing, ant colony and genetic algorithm are suggested (Kundu and Dan, 2012; See and Wong, 2008).

Another creative topic apart from mentioned studies is dynamic layout calculations. In practice, some facilities do not need absolutely separated zones by walls or barriers such as airport cargo warehouses. Hence, specific areas should be rearranged based on change (Krishnan et al., 2008; Lim and Noble, 2006; McKendall and Shang, 2006). Lastly, a relatively new concept on flexible and cellular manufacturing requires a novel layout strategy which can be solved by heuristics (Ariafar et al., 2011; Kumar et al., 2015). Essentially, while proposing a solution approach to the layout design problem, the functional units are placed such that a cost function measuring a certain overall proximity measure is minimised.

At this point, we note that the essential challenge is defining a proper proximity measure while dealing with a FLP. Specifically, for the retail industry, we propose that the degree of association between product groups can be an appropriate indicator of proximity measure, which, in turn, can be exploited by facility layout algorithms. In this work, a simple but effective facility layout algorithm, MCRAFT, is employed to find an improved layout for a retail store by using confidence values of association rules. A similar study was conducted by Cil (2012) where association rules were used in combination with multidimensional scaling technique (Cil, 2012). Differently from this work, our approach integrates association rules directly with a layout design tool, MCRAFT, and considers two alternative views, in which the calculated confidence values between each product groups are interpreted as a measure of proximity from customer's perspective and a measure of remoteness from retailer's perspective.

3 Problem definition and background

3.1 Methodology

Although, to date, some methods have been developed and introduced to arrange supermarket layout, in this study a novel hybrid method is proposed which is a combination of association rules mining and facility layout algorithms because of the non-integrated and unproductive layout solutions of existing methods. This study is composed of three main stages. First, pre-processing and transformation of data level which deals with obtained receipt data preparation before ARM application. Second, finding association rules phase utilises data mining principles in order to find out significant relations between product groups, and lastly, layout development phase tries to arrange an appropriate supermarket order by using MCRAFT one of computer-aided facility layout tool. The compact form of this aimed study is represented visually in Figure 1.

3.2 Association rules mining and apriori algorithm

Apriori algorithm has been one of the most famous of all ARM algorithms (Agrawal and Srikant, 1994). Apriori is designed to operate on databases containing transactions and can be used to generate all frequent item set.

Each transaction consists of a set of items. Considering a threshold C, the apriori algorithm defines the item sets that are subsets of at least C transactions in the database. Apriori makes multiple passes over the database. The first pass of the algorithm simply counts item occurrences to identify the large 1-itemsets, L_1. Then, it produces the candidate item sets in C_1 and saves the frequent item sets in L_1.

A subsequent pass, say pass k, has two phases. In the first phase, the algorithm generates the candidate item sets in C_k from large item sets L_{k-1} using the apriori-gen function. This function takes as an argument L_{k-1}, the set of all large $(k-1)$-itemsets. In the joining step, this function first joins L_{k-1}.

Next, it generates all $(k-1)$-subsets from the candidate item sets in C_k and deletes all candidate item sets from C_k where some $(k-1)$-subset of the candidate item set is not in the frequent item set L_{k-1}. Then, it scans the transactional database and examines C_k for determining which candidates are frequent and the support for each candidate item set in C_k. Saves the frequent item sets in L_k. The algorithm terminates when L_k. becomes empty. Figure 2 shows the steps of the apriori algorithm (Strickland, 2015).

The FLP focuses generally on manufacturing systems and deals with material flows among various departments inside the facility. Therefore, the basic aim of FLP is to find an optimal arrangement of manufacturing or service facilities, which ensures to minimise moves of workers and to complete material handling operations with minimum cost. Thus, the end-product of a facility layout study is an optimised physical organisation of a production environment.

Figure 1 Flow diagram of the proposed methodology (see online version for colours)

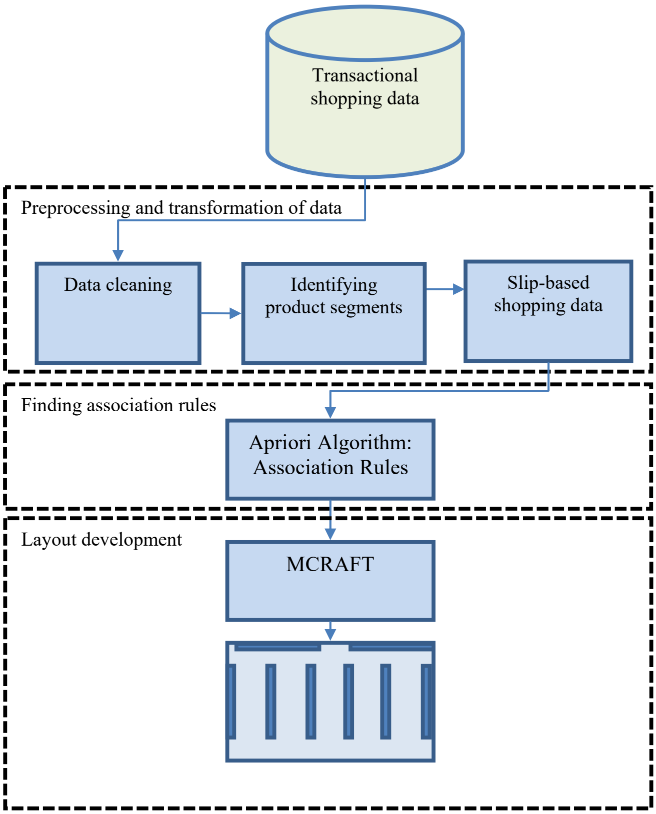
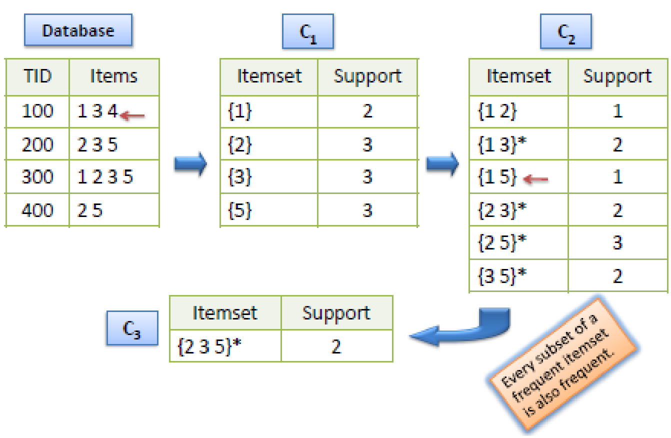


Figure 2 Main steps of the apriori algorithm (see online version for colours)



Source: Strickland (2015)

3.3 MCRAFT

Within the concept of this paper, MCRAFT is used as a computer-aided layout technique, which produces a variety of placement by using flows between departments/workstations. MCRAFT was firstly presented by Hosni et al. (1980). MCRAFT is an enhanced version of CRAFT. Thus, it shows approximately the same characteristics with CRAFT. In order to clarify the difference between them, some basic information about CRAFT is given in the following.

In the first place, the initial layout is entered into the system and the algorithm evaluates the distance between the centres of each department to find out the cost of the initial layout. The material/document/human of the flow and the distance between departments are used to calculate the cost of the initial layout.

Secondly, CRAFT starts with an initial layout and executes two-way and three-way interchanges. Each interchange iteration causes cost reduction. The next iteration begins with the new layout and it is proceeding until no further cost reduction can be encountered by the pairwise exchange. The department pairs for the pairwise interchange are accepted only if they are located in the same area or they are contiguous. This is the most challenging restriction of CRAFT algorithm.

However, in the MCRAFT, the constraints about the pairwise interchange are loosened. CRAFT is able to only make changes between the departments, which are a neighbour or available in the same area. MCRAFT loosens this limitation and makes all changes possible. In other words, the MCRAFT removes the pairwise change restriction that is occurred in the CRAFT method. This improvement embedded in MCRAFT leads very significant contribution while figuring out for the best layout structure that the CRAFT cannot operate. MCRAFT separates the facility area into bands and allocates the bands to one or more departments. The number of bands in the facility layout is determined by the researchers or practitioners.

4 Application on a retailer chain

In this work, products associated with each other have been determined by ARM to the dataset consist of shopping records in a supermarket chain in the Turkish retail industry. Considering apriori algorithm's good record of successfully extracting association rules, the same is employed on the dataset we mine for reasonable consequences.

4.1 Data collection and pre-processing

During a six-month period, 170,810 receipts were collected. Considering these 170,810 receipts, it was observed that the supermarket's hierarchical classification of the product is highly controversial. In other words, some entries in the dataset were grouped into some subclasses where, for instance, kasseri cheese and white cheese were classified into different classes. Thus, a pre-processing of the dataset was necessary where some subclasses were combined together to build new higher classes. After the pre-processing stage, 36 main product groups have been determined.

Before we introduce the methods used and the results we obtain, we should note that the results of this study must be considered with some cautiousness:

- Data are collected during a specific time interval will give different results in different periods.
- Campaigns and promotions of products will affect purchases in a period of collection of receipts.
- Shopping habits will vary due to target audiences and region of the markets where the study was performed.

4.2 Apriori algorithm's solution steps

In this part, the solution procedure of apriori algorithm is shown and association rules are found on an illustrative example to demonstrate how the algorithm works. In the illustrative dataset above, I_k = {bakery products, milk products, meat products, cheese, hot drinks} represents product set where sample size is D = 5 (see Table 1).

Table 1 Sample dataset

ID	*Bakery products*	*Milk products*	*Meat products*	*Cheese*	*Hot drinks*
1	0	0	1	1	0
2	1	1	0	0	0
3	0	1	1	0	0
4	1	0	1	1	1
5	1	0	0	0	0
Sum	3	2	3	2	1

Table 2 C1 candidate set

Item set	*Support*
Bakery products	0.60
Milk products	0.40
Meat products	0.60
Cheese	0.40
Hot drinks	0.20

Next, the supports of products are found. support of bakery products: 3/5 = 0.6. The same process is repeated for other products and all support values are found. Thus, the C1 candidate set is obtained as in Table 2.

Assuming that the minimum support threshold is set as 20%, we observe that all instances in Table 2 are above that level. Accordingly, L1 is generated as shown in Table 3.

Supports of dual combination of products in L1 are calculated and C2 is generated. Support value for {bakery products, milk products} = 1/5= 0.2 and confidence value for {bakery products, milk products} = 1/3 = 0.33. After finding support and confidence values of other combinations C2 table is formed in Table 4.

Table 3 L1 large item set

Item set	*Support*
Bakery products	0.60
Milk products	0.40
Meat products	0.60
Cheese	0.40
Hot drinks	0.20

Table 4 Dual combinations for C2 candidates, support and confidence values

Item set	*Support*	*Confidence*
Bakery products, milk products	0.20	0.33
Bakery products, meat products	0.20	0.33
Bakery products, cheese	0.20	0.33
Bakery products, hot drinks	0.20	0.33
Milk products, meat products	0.20	0.50
Milk products, cheese	0	0
Milk products, hot drinks	0	0
Meat products, cheese	0.40	0.66
Meat products, hot drinks	0.20	0.33
Cheese, hot drinks	0.20	0.50

Table 5 L2 large item sets

Item set	*Support*	*Confidence*
Bakery products, milk products	0.20	0.33
Bakery products, meat products	0.20	0.33
Bakery products, cheese	0.20	0.33
Bakery products, hot drinks	0.20	0.33
Milk products, meat products	0.20	0.50
Meat products, cheese	0.40	0.66
Meat products, hot drinks	0.20	0.33
Cheese, hot drinks	0.20	0.50

Table 6 C3 candidate set

Item set	*Support*	*Confidence*
Bakery products, meat products, cheese	0.20	0.33
Bakery products, meat products, hot drinks	0.20	0.33
Bakery products, cheese, hot drinks	0.20	0.33
Meat products, cheese, hot drinks	0.20	0.33

Duals below the support threshold are removed from C2 and L2 large item set is created. L2 item sets are presented in Table 5. Next, the new support values are determined to get the triple combinations from L2 and C3 is generated. L3 large items sets consist of the C3 candidate set as in Table 6. All item sets are above the support threshold value in candidate set C3. Thus, the L3 large item set is created as in Table 7. Similarly, C4 candidates and L4 are obtained as in Table 8 and Table 9, respectively.

Table 7 L3 large item set

Item set	*Support*	*Confidence*
Bakery products, meat products, cheese	0.20	0.33
Bakery products, meat products, hot drinks	0.20	0.33
Bakery products, cheese, hot drinks	0.20	0.33
Meat products, cheese, hot drinks	0.20	0.33

Table 8 C4 candidate set

Item set	*Support*	*Confidence*
Bakery products, meat products, cheese, hot drinks	0.20	0.33

Applying the same procedure as before, L4 can be extracted from C4 (see Table 9).

Table 9 L4 large item set

Item set	*Support*	*Confidence*
Bakery products, meat products, cheese, hot drinks	0.20	0.33

Finally, association rules were found. Thus, for the given dataset, the following rules can be generated, Table 10.

Table 10 Rules obtained with apriori algorithm for the given threshold values

Antecedent	*Consequent*	*Support*	*Confidence*
Bakery products	Milk products	0.20	0.33
Bakery products	Meat products	0.20	0.33
Bakery products	Cheese	0.20	0.33
Bakery products	Hot drinks	0.20	0.33
Milk products	Meat products	0.20	0.50
Meat products	Cheese	0.40	0.66
Meat products	Hot drinks	0.20	0.33
Cheese	Hot drinks	0.20	0.50
Bakery products	Meat products	0.20	0.33
Bakery products	Meat products	0.20	0.33
Bakery products	Cheese, hot drinks	0.20	0.33
Meat products	Cheese, hot drinks	0.20	0.33
Bakery products	Meat products, cheese, hot drinks	0.20	0.33

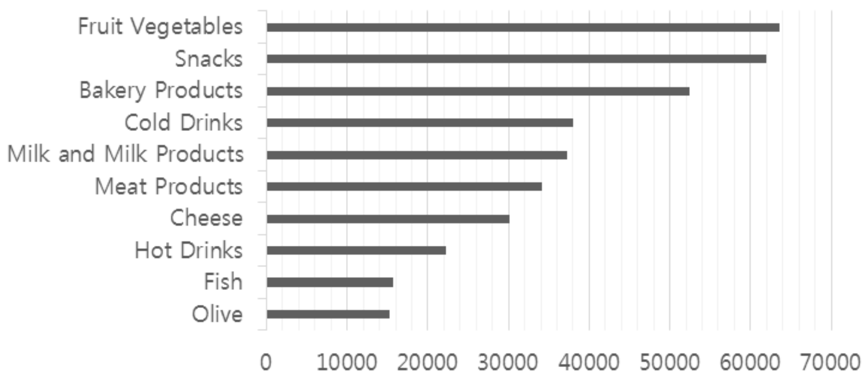
For instance, the obtained rule for bakery products => milk products indicates that; the probability of observing the concurrent sales of bakery products and milk products in total receipt transactions is 20% and customers purchasing bakery products will also buy milk products with a probability of 33%.

4.3 ARM with KNIME

To find associations in the dataset, KNIME software is used. The software has hundreds of modules for data integration, data transformation, methods for data analysis and visualisation.

In the database subject to this study, we classified the products into 36 product groups. Figure 3 shows the most frequently sold ten product groups. Main product groups dominating the dataset are fruits vegetables (63,531), snacks (61,920), and bakery products (52,488).

Figure 3 The most frequently sold product groups



Applying KNIME's Association Rule Lerner which is implemented based on the work by Borgelt (2012) association rules for 36 product groups are obtained. To restrict ourselves to the most valuable association rules, the minimum support value is assigned as 10% and the minimum confidence value is assigned as 20%. As a result of the first stage of the analysis, 12 association rules are extracted from the dataset. These rules are shown in Table 11.

Table 11 Association rules for product groups

	Antecedent	*Consequent*	*Support*	*Confidence*
1	Snacks	Milk and milk products	0.102	0.28
2	Milk and milk products	Snacks	0.102	0.47
3	Fruits and vegetables	Milk and milk products	0.103	0.28
4	Milk and milk products	Fruits and vegetables	0.103	0.47
5	Snacks	Bakery products	0.107	0.29
6	Bakery products	Snacks	0.107	0.35
7	Snacks	Cold drinks	0.108	0.30
8	Cold drinks	Snacks	0.108	0.49
9	Snacks	Fruits and vegetables	0.124	0.34
10	Fruits and vegetables	Snacks	0.124	0.33
11	Bakery products	Fruits and vegetables	0.128	0.42
12	Fruits and vegetables	Bakery products	0.128	0.42

Some association rules created for the product groups can be listed as:

- The rule for snacks => milk and milk products: the possibility of being purchased together of snacks and milk and milk products is 10.2% and customers of snacks get milk and milk products with 28.3% probability.
- The rule for milk and milk products => snacks: the possibility of being purchased together of milk and milk products and snacks is 10.2% and customers of milk and milk products get snacks with 47% probability.
- The rule for fruits – vegetables=> milk and milk products: the possibility of being purchased together of fruits – vegetables and milk and milk products is 10.3% and customers of fruits – vegetables get milk and milk products with 27.8% probability.
- Rules for cold drinks=> snacks: the possibility of being purchased together of cold drinks and snacks is 10.8% and customers of cold drinks get snacks with 48.7% probability.

We conduct another trial where we look for subtler relations between product groups. In the second trial, the minimum support value is decreased to 5% while the minimum confidence value is kept the same as 20%. As a result of the second analysis, 53 association rules are obtained. These rules are shown in Table 12.

Table 12 Association rules for product groups extracted with support = 5% and confidence = 20%

Antecedent		*Consequent*	*Sup.*	*Conf*
1	Olive	Cheese	0.05	0.665
2	Cheese	Olive	0.05	0.284
3	Bakery products, fruits – vegetables	Milk and milk product	0.051	0.493
4	Bakery products, milk, and milk products	Fruits – vegetables	0.051	0.567
5	Milk and milk products, fruits – vegetables	Bakery products	0.051	0.396
6	Bakery products, snacks	Fruits – vegetables	0.053	0.429
7	Snacks, fruits – vegetables	Bakery products	0.053	0.497
8	Bakery products, fruits – vegetables	Snacks	0.053	0.413
9	Snacks, fruits –vegetables	Milk and milk product	0.054	0.523
10	Snacks, milk and milk products	Fruits – vegetables	0.054	0.434
11	Milk and milk products, fruits – vegetables	Snacks	0.054	0.519
12	Hot drinks	Fruits – vegetables	0.054	0.414
13	Cold drinks	Meat products	0.057	0.284
14	Meat products	Cold drinks	0.057	0.256
15	Cold drinks	Milk and milk product	0.065	0.298
16	Milk and milk products	Cold drinks	0.065	0.292
17	Hot drinks	Snacks	0.066	0.506
18	Cold drinks	Bakery products	0.072	0.323
19	Bakery products	Cold drinks	0.072	0.233
20	Meat products	Milk and milk product	0.073	0.337

Table 12 Association rules for product groups extracted with support = 5% and confidence = 20% (continued)

	Antecedent	*Consequent*	*Sup.*	*Conf*
21	Milk and milk products	Meat products	0.073	0.368
22	Bakery products	Cheese	0.074	0.42
23	Cheese	Bakery products	0.074	0.241
24	Milk and milk products	Cheese	0.077	0.355
25	Cheese	Milk and milk product	0.077	0.439
26	Meat products	Cheese	0.078	0.39
27	Cheese	Meat products	0.078	0.442
28	Snacks	Cheese	0.078	0.216
29	Cheese	Snacks	0.078	0.444
30	Cold drinks	Fruits – vegetables	0.08	0.215
31	Fruits – vegetables	Cold drinks	0.08	0.361
32	Snacks	Meat products	0.081	0.224
33	Meat products	Snacks	0.081	0.406
34	Meat products	Bakery products	0.082	0.409
35	Bakery products	Meat products	0.082	0.266
36	Bakery products	Milk and milk product	0.09	0.412
37	Milk and milk products	Bakery products	0.09	0.292
38	Fruits – vegetables	Cheese	0.09	0.243
39	Cheese	Fruits – vegetables	0.09	0.511
40	Meat products	Fruits – vegetables	0.098	0.489
41	Fruits – vegetables	Meat products	0.098	0.263
42	Snacks	Milk and milk product	0.102	0.283
43	Milk and milk products	Snacks	0.102	0.47
44	Fruits – vegetables	Milk and milk product	0.103	0.474
45	Milk and milk products	Fruits – vegetables	0.103	0.278
46	Snacks	Bakery products	0.107	0.294
47	Bakery products	Snacks	0.107	0.347
48	Snacks	Cold drinks	0.108	0.298
49	Cold drinks	Snacks	0.108	0.487
50	Snacks	Fruits – vegetables	0.124	0.341
51	Fruits – vegetables	Snacks	0.124	0.332
52	Bakery products	Fruits – vegetables	0.128	0.345
53	Fruits – vegetables	Bakery products	0.128	0.418

4.4 Layout design with proposed approach

Results obtained from the above data mining analysis can be used to make more effective supermarket layout design and organise different product campaigns. In previous studies, supermarket organisation is arranged by looking some product features such as fresh fruit and vegetables should be located to entering gates or to provide convenience replenishment process, for instance, meat products must be placed the right or left corner of the market to service meat and fish products from freezer storage to marketplace. However, this study utilises MCRAFT, one of facility layout technique, to rearrange product departments.

In order to carry out the layout study with MCRAFT, some initial assumption was made:

- A number of departments (number of product groups): even though 36 product groups are selected to find significant relationships between them, but for supermarket layout study only nine essential groups are determined.
- The total area of the supermarket: in the real case study, the measurement of supermarket area is 300 square metres and the dimension of rectangular in other words length and width of the area is 20 metres' × 15 metres, respectively. In addition to this, the area of each cell is accepted as five square metres while department width is composed of only two cells.
- Bands width: as mentioned before that MCRAFT separates the whole area into bands and allocated the band to one or more departments. For this case, the bands with are specified as two cells.
- Department dimensions: to arrange the specific height and width of each zone of product groups, distributions of each group are used as a base. Then, the real distribution values are normalised according to total area and lastly, departments' dimensions are rounded up or down based on the multiples of five. All mentioned calculation steps and results are given in Table 13.

Table 13 Main product departments and their dimensions

Dep.	*Product groups*	*Product groups distribution*	*The rate for 300 m^2*	*The multiples of five round up/round down*
1	Cheese	30,139	25,47	25
2	Olive	15,294	12,93	15
3	Milk products	37,222	31,46	30
4	Fruit vegetables	63,531	53,70	55
5	Bakery products	52,488	44,36	45
6	Snacks	61,92	52,34	50
7	Meat product	34,133	28,85	30
8	Cold drinks	37,94	32,07	30
9	Hot drinks	22,275	18,83	20
Sum.		354,942	300	300

In order to place product groups, two different strategies are constructed from the point of both customer and retailer in terms of total length of customer stay inside the supermarket. First one is a customer-oriented strategy, which aims to position products groups with a strong relationship with them. The principle behind this thinking is that customers are inclined to buy the second product that reminds the first product if these two products are close to each other on the shelf. On the other hand, retailer-oriented strategy wishes to increase the average amount of time spent on each customer inside the market. Thus, the owner of the facility intends to locate product groups that have high support rate far away from each other.

4.4.1 Consumer-oriented strategy

In this strategy, the products with higher confidence value than the other products must be placed on neighbouring shelves. The following flow matrix, given in Table 14, is established by taking these confidence values into accounts. In other words, the relation values between products are considered as a flow between departments which are required by MCRAFT software.

Table 14 Relationship matrix chart for customer oriented structure

	Cheese	*Olive*	*Milk and milk products*	*Fruit vege.*	*Bakery*	*Snack*	*Meat product*	*Cold drink*	*Hot drink*
Cheese		0.284	0.439	0.511		0.444	0.442		
Olive	0.665								
Milk and milk products	0.355			0.278	0.292	0.47	0.368	0.298	
Fruit vegetables	0.243		0.474		0.418	0.332	0.263	0.361	
Bakery	0.42		0.412	0.345		0.347	0.266		
Snack	0.216		0.289	0.341	0.294		0.224	0.298	
Meat product	0.39		0.337	0.489	0.409	0.406		0.256	
Cold drink				0.215	0.323	0.487	0.284		
Hot drink				0.414		0.506			

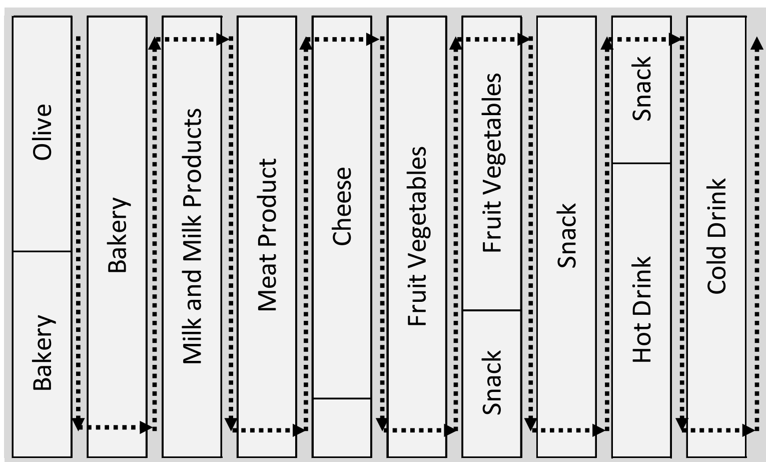
As stated in Section 3.3, the MCRAFT loosens the pairwise change restriction between the departments that are neighbours or located in the same area and it renders all kind of changes possible. Firstly, the initial cost is calculated by MCRAFT based on the initial layout. Then, it obtains any new layout by changing departments' layouts, calculates the new cost, takes the optimum cost options and continues to switch departments positions iteratively until finding the minimum cost value. For customer-oriented strategy, software completed this process after seven iterations and reached the minimum cost as 95.29 € (see Table 15).

MCRAFT uses the sweeping pattern as can be realised from the layout design depicted in Figure 4. According to the confidence value obtained from customer receipts, the most relevant products are aligned beside each other.

Table 15 Steps of improving the initial layout for consumer-oriented strategy

Iteration no.	*Type*	*Action*	*Cost (€)*
0	Initial cost		118.53
1	Switch:	7 and 5	112.31
2	Switch:	5 and 4	109.30
3	Switch:	5 and 1	100.01
4	Switch:	7 and 1	97.57
5	Switch:	4 and 6	96.14
6	Switch:	2 and 5	95.64
7	Switch:	9 and 8	95.29

Figure 4 Supermarket layout in terms of consumer-oriented strategy



4.4.2 *Retailer-oriented strategy*

In this strategy, the main effort is to place two product groups which have a close relationship with remote regions in an attempt to go up the time that customers spend in the market. In order to generate inverse proportion between products, the existing relations are subtracted from 1 and Table 16 is obtained.

MCRAFT facility layout software is run once more again for the retailer-oriented strategy and the all calculations of each iteration and changes between departments are presented in Table 17.

Based on the altered flow matrix, the positions of product groups are relocated as can be seen in Figure 5. To attain a better understanding the difference between two strategies, providing an example could be very beneficial at this point. In customer-oriented strategy Olive is put besides bakery because of the strong relationship, whereas in the following Figure 5, four various product groups which are milk and milk products, meat products, cheese, and snacks are placed between olive and bakery to have customers walk around the market to ensure that customers see more products during walking.

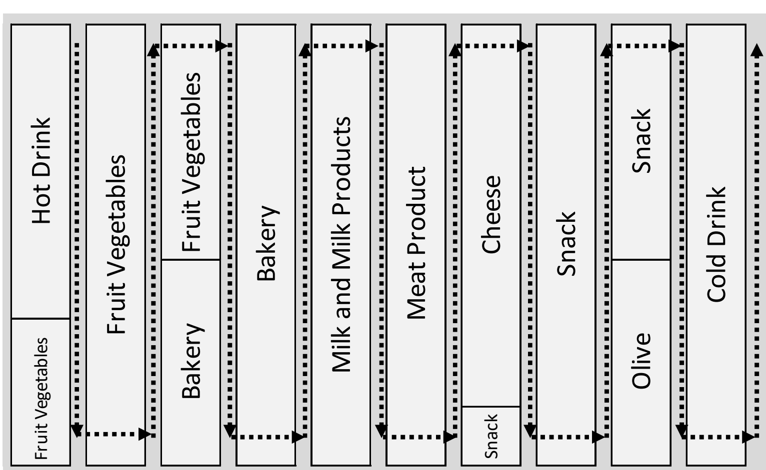
Table 16 Relationship matrix chart for retailer-oriented strategy

	Cheese	*Olive*	*Milk and milk products*	*Fruit veget.*	*Bakery*	*Snack*	*Meat product*	*Cold drink*	*Hot drink*
Cheese		0.716	0.561	0.489		0.556	0.558		
Olive	0.335								
Milk and milk products	0.645			0.722	0.708	0.53	0.632	0.702	
Fruit vegetables	0.757		0.526		0.582	0.668	0.737	0.639	
Bakery	0.58		0.588	0.655		0.653	0.734		
Snack	0.784		0.711	0.659	0.706		0.776	0.702	
Meat product	0.61		0.663	0.511	0.591	0.594		0.744	
Cold drink				0.785	0.677	0.513	0.716		
Hot drink				0.586		0.494			

Table 17 Steps of improving the initial layout for retailer-oriented strategy

Iteration no	*Type*	*Action*	*Cost*
0	Initial cost	-	208.72
1	Switch:	9 and 3	193.32
2	Switch:	9 and 1	184.21
3	Switch:	4 and 1	178.11
4	Switch:	3 and 8	175.11
5	Switch:	6 and 5	172.79
6	Switch:	3 and 7	171.51

Figure 5 Supermarket layout in terms of retailer-oriented strategy



5 Conclusions

The primary purpose of this work is to find meaningful relationships between products in a major retailer chain in Turkey. Taking advantage of these relationships, not only identification of effective advertising campaigns and promotions as well as the allocation of the shelf is possible but also the arrangement of supermarket layout could be possible. Thereby, an increase in the sales volume and sales revenue of the retail store is possible.

To this end, we used ARM to gain valuable insights on customer receipts data collected from the retailer. After the dataset was pre-processed, it was analysed by means of a data mining software tool, KNIME where association rules were found with scanning the dataset with apriori algorithm. Ultimately, dominating rules were presented. Apriori algorithm, as the most frequently applied tool for ARM, was employed to extract association rules from a dataset of one of the supermarket chains operating in the retail sector.

The second aim of this work was to construct a supermarket layout on the basis of the results obtained from the rules mining stage. Besides that, two fundamental approaches were created by taking customer's and retailer's point of views into account and all calculations were rearranged based on these principles. The originality of this study can be specified that firstly, the link between association rules and a facility layout technique, MCRAFT was rarely exploited. Secondly, facility layout methods have been utilised for rearranging the departments of the company

However, within the context of this study, it has used for supermarket retailer layout design by using an MCRAFT.

The paper offers some theoretical and practical benefits which can be used both by researchers and practitioners of the retail sector. Firstly, the methodology successfully integrates association rules with a simple layout approach providing a practical ground for layout design. Secondly, this study relies on the real market basket data which increases the reliability of the obtained results related to the suggested layout design. Thirdly, the suggested layout can be easily updated based on the feedbacks on sales volume for a given layout arrangement with up-to-date transaction data gathered from the store. Lastly, this study experiences a hurdle related to practical results of the proposed layout. Obviously, its implementation requires the approval of the top management which may be the subject of further research.

References

Abbasi, E., Ahmadi, S.H., Naderi, S. and Vahdani, F.A. (2017) 'Modelling of layout design and selection of appropriate design with a case study', *International Journal of Industrial and Systems Engineering*, Vol. 25, No. 2, pp.251–264 [online] https://doi.org/10.1504/IJISE.2017.081520.

Abdullah, U., Ahmad, J. and Ahmed, A. (2008) 'Analysis of effectiveness of apriori algorithm in medical billing data mining', in *2008 4th International Conference on Emerging Technologies*, IEEE, pp.327–331 [online] https://doi.org/10.1109/ICET.2008.4777523.

Agarwal, D. and Singholi, A.K.S. (2018) 'Performance analysis of a FLP problem using AHP-TOPSIS and FAHP-FTOPSIS', *International Journal of Industrial and Systems Engineering*, Vol. 30, pp.401–424 [online] https://doi.org/10.1504/IJISE.2018.096159.

Aggelis, V. and Christodoulakis, D. (2003) 'Association rules and predictive models for e-banking services', in *1st Balkan Conference on Informatics*.

Agrawal, R. and Srikant, R. (1994) 'Fast algorithms for mining association rules', in *Proc. 20th Int. Conf. Very large data bases*, VLDB, Vol. 1215, pp.487–499.

Ahmed, A.M., Norwawi, N.M. and Ishak, W.H.W. (2009) 'Identifying student and organization matching pattern using apriori algorithm for practicum placement', in *2009 International Conference on Electrical Engineering and Informatics*, IEEE, Vol. 1, pp.28–31 [online] https://doi.org/10.1109/ICEEI.2009.5254821.

Altuntas, S., Dereli, T. and Selim, H. (2013) 'Fuzzy weighted association rule based solution approaches to facility layout problem in cellular manufacturing system', *International Journal of Industrial and Systems Engineering*, Vol. 15, No. 3, pp.253–271 [online] https://doi.org/10.1504/IJISE.2013.056676.

Angeline, D.M.D. (2013) 'Association rule generation for student performance analysis using apriori algorithm', *The SIJ Transactions on Computer Science Engineering and Its Applications* (*CSEA*), Vol. 1, No. 1, pp.12–16.

Ariafar, S., Ismail, N., Tang, S.H., Ariffin, M.K.A.M. and Firoozi, Z. (2011) 'Design of a facility layout model in hybrid cellular manufacturing systems under variable demand', *International Journal of Industrial and Systems Engineering*, Vol. 9, No. 4, pp.373–387 [online] https://doi.org/10.1504/IJISE.2011.043482.

Balakrishnan, J., Cheng, C.H. and Lau, C.M. (2013) 'Modelling rearrangement constraints in dynamic layout', *International Journal of Industrial and Systems Engineering*, Vol. 15, No. 4, pp.373–393 [online] https://doi.org/10.1504/IJISE.2013.057475.

Borgelt, C. (2012) 'Frequent item set mining', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 2, No. 6, pp.437–456 [online] https://doi.org/10.1002/widm.1074.

Borges, A. (2003) 'Toward a new supermarket layout: from industrial categories to one stop shopping organization through a data mining approach', in Babin, B.J. (Ed.): *Symposium on Retail Patronage and Strategy*, 4–5 November, p.28, USM Jean-Charles Chebat, HEC, Montreal.

Boros, P., Fehér, O., Lakner, Z., Niroomand, S. and Vizvári, B. (2016) 'Modeling supermarket re-layout from the owner's perspective', *Annals of Operations Research*, Vol. 238, Nos. 1–2, pp.27–40 [online] https://doi.org/10.1007/s10479-015-1986-2.

Bükey, F.Ö. (2014) *Data Mining Applications in Customer Relationship Management and a Comparative Study in the Banking Sector*, Marmara University.

Chen, A-P., Chen, Y-C. and Tseng, W-C. (2005) 'Applying extending classifier system to develop an option-operation suggestion model of intraday trading – an example of Taiwan index option', in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, Springer, pp.27–33 [online] https://doi.org/10.1007/11552413_5.

Cil, I. (2012) 'Consumption universes based supermarket layout through association rule mining and multidimensional scaling', *Expert Systems with Applications*, Vol. 39, No. 10, pp.8611–8625 [online] https://doi.org/10.1016/j.eswa.2012.01.192.

Dong, J. (2018) 'Isovist based analysis of supermarket layout', *Verification of Visibility Graph Analysis and Multi-Agent Simulation*, July 2017, pp.1–12.

Farley, J.U. and Ring, L.W. (1966) 'A stochastic model of supermarket traffic flow', *Operations Research*, Vol. 14, No. 4, pp.555–567.

Gancheva, V. (2013) *Market Basket Analysis of Beauty Products*, Erasmus University Rotterdam.

Griffith, R. and Harmgart, H. (2005) 'Retail productivity', *The International Review of Retail, Distribution and Consumer Research*, Vol. 15, No. 3, pp.281–290 [online] https://doi.org/10.1080/09593960500119481.

Gürgen, G. (2008) 'Birliktelik Kuralları ile Sepet Analizi ve Uygulaması', Marmara University.

Hosni, Y. A. Whitehouse, G. E. and Atkins, T. S. (1980) 'Micro-CRAFT Program Documentation', Institute of Industrial Engineers, Norcross, GA.

Jiang, M. H. (2011) 'Mining Mobile Telecommunication Services Using Apriori Algorithm', Advanced Materials Research, Vol. 291–294, pp.2369–2373. https://doi.org/10.4028/www.scientific.net/AMR.291-294.2369

Koopmans, T. C. and Beckmann, M. (1957) 'Assignment Problems and the Location of Economic Activities', Econometrica, Vol. 25, No.1, pp.53. https://doi.org/10.2307/1907742

Krishnan, K. K. Cheraghi, S. H. and Nayak, C. N. (2008) 'Facility layout design for multiple production scenarios in a dynamic environment', International Journal of Industrial and Systems Engineering, Vol. 3, No. 2, pp.105-133. https://doi.org/10.1504/IJISE.2008.016740

Kumar, R., Haleem, A., Garg, S.K. and Singh, R.K. (2015) 'Automated guided vehicle configurations in flexible manufacturing systems: a comparative study', *International Journal of Industrial and Systems Engineering*, Vol. 21, No. 2, pp.207–226 [online] https://doi.org/10.1504/IJISE.2015.071510.

Kundu, A. and Dan, P.K. (2012) 'Metaheuristic in facility layout problems: current trend and future direction', *International Journal of Industrial and Systems Engineering*, Vol. 10. No. 2, pp.238–253 [online] https://doi.org/10.1504/IJISE.2012.045182.

Lee, Y.H. and Lee, M.H. (2002) 'A shape-based block layout approach to facility layout problems using hybrid genetic algorithm', *Computers and Industrial Engineering*, Vol. 42, Nos. 2–4, pp.237–248 [online] https://doi.org/10.1016/S0360-8352(02)00018-9.

Lim, H.H. and Noble, J.S. (2006) 'The impact of facility layout on overall remanufacturing system performance', *International Journal of Industrial and Systems Engineering*, Vol. 1, No. 3, pp.357–371 [online] https://doi.org/10.1504/IJISE.2006.009793.

Liu, D. and Fan, S. (2013) 'Tourist behavior pattern mining model based on context', *Discrete Dynamics in Nature and Society*, Vol. 2013, pp.1–12 [online] https://doi.org/10.1155/2013/108062.

Lu, X., Jia, Z., Yang, J. and Liu, H. (2011) 'Design and implementation of lean facility layout system of a production line', *International Journal of Industrial Engineering: Theory, Applications and Practice*, Vol. 18, No. 5, pp.1260–1269.

McKendall, A.R. and Shang, J. (2006) 'Hybrid ant systems for the dynamic facility layout problem', *Computers and Operations Research*, Vol. 33, No. 3, pp.790–803 [online] https://doi.org/10.1016/j.cor.2004.08.008.

Meller, R.D., Narayanan, V. and Vance, P.H. (1998) 'Optimal facility layout design', *Operations Research Letters*, Vol. 23, Nos. 3–5, pp.117–127 [online] https://doi.org/10.1016/S0167-6377(98)00024-8.

Moslemipour, G., Lee, T.S. and Loong, Y.T. (2018) 'Solving stochastic dynamic facility layout problems using proposed hybrid AC-CS-SA meta-heuristic algorithm', *International Journal of Industrial and Systems Engineering*, Vol. 28, No. 1, pp.1–31 [online] https://doi.org/10.1504/IJISE.2018.088561.

Mu, Z., Yi, C., Xiaohong, Z. and Junyong, L. (2009) 'Study on the recommendation technology for tourism information service', in *2009 Second International Symposium on Computational Intelligence and Design*, IEEE, pp.410–415 [online] https://doi.org/10.1109/ISCID.2009.111.

Page, B., Trinh, G. and Bogomolova, S. (2019) 'Comparing two supermarket layouts: the effect of a middle aisle on basket size, spend, trip duration and endcap use', *Journal of Retailing and Consumer Services*, Vol. 47, pp.49–56 [online] https://doi.org/10.1016/J.JRETCONSER.2018.11.001.

See, P.C. and Wong, K.Y. (2008) 'Application of ant colony optimisation algorithms in solving facility layout problems formulated as quadratic assignment problems: a review', *International Journal of Industrial and Systems Engineering*, Vol. 3, No. 6, pp.644–672 [online] https://doi.org/10.1504/IJISE.2008.020679.

Sharif, M.N.A., Ching, N.M., Bakri, A. and Zakaria, N.H. (2005) 'Using a priori algorithm for supporting an e-commerce system', *Journal of Information Technology Impact*, Vol. 5, No. 3, pp.129–138.

Sharma, P. and Singhal, S. (2016) 'A review of objectives and solution approaches for facility layout problems', *International Journal of Industrial and Systems Engineering*, Vol. 24, No. 4, pp.469–489 [online] https://doi.org/10.1504/IJISE.2016.080289.

Sherdiwala, K.B. and Khanna, S.O. (2015) 'Association rule mining: an overview', *International Multidisciplinary Research Journal*, Vol. 2, No. 4, pp.1–3.

Strickland, J. (2015) *What the Heck are Association Rules in Analytics?* [online] https://bicorner.com/2015/07/22/what-the-heck-are-association-rules-in-analytics/ (accessed 9 March 2017).

Tompkins, J.A., White, J.A., Bozer, Y.A. and Tanchoco, J.M.A. (2010) *Facilities Planning*, John Wiley and Sons, Hoboken.

Xu, Z-M. and Zhang, R. (2009) 'Financial revenue analysis based on association rules mining', in *2009 Asia-Pacific Conference on Computational Intelligence and Industrial Applications* (*PACIIA*), IEEE, Vol. 1, pp.220–223 [online] https://doi.org/10.1109/PACIIA.2009.5406454.

Yang, G.X. (2013) 'The research of improved apriori mining algorithm in bank customer segmentation', in *Advanced Materials Research*, Vol. 760, No. 1, pp.2244–2249, Trans. Tech. Publ.

Yazgan, P. and Kusakci, A.O. (2016) 'A literature survey on association rule mining algorithms', *Southeast Europe Journal of Soft Computing*, Vol. 5, No. 1, pp.5–14 [online] https://doi.org/dx.doi.org/10.21533/sejsc.v5i1.102.

Zhang, W., Ma, D. and Yao, W. (2014) 'Medical Diagnosis data mining based on improved apriori algorithm', *Journal of Networks*, Vol. 9, No. 5, pp.1339–1345 [online] https://doi.org/10.4304/jnw.9.5.1339-1345.