

DATA DRIVEN DECISION SUPPORT TO SUPERMARKET LAYOUT

IBRAHIM CIL¹, DERYA AY¹, YUSUF S. TURKAN²

¹Sakarya University, Department of Industrial Engineering, Adapazari, TURKEY

²Beykent University, Department of Industrial Engineering, Istanbul, TURKEY

icil@sakarya.edu.tr, <http://www.icil.sakarya.edu.tr>, deryaayie@gmail.com, ysturkan@gmail.com,

Abstract: - Knowledge is the most valuable asset in today's dynamic business environment. In many organizations, decisions are made based on a combination of judgment and knowledge extracted from databases. Successful business organization to be able to react rapidly to the changing market demands both locally and globally, by utilizing the latest data mining techniques of extracting previously unknown and potentially useful knowledge from vast resources of raw data. We propose a methodological framework for the use of the knowledge discovery process to improve store layout. In this paper, we propose a data driven decision support for store layout and present an empirical study. This paper develops a relational database and uses Apriori algorithm and multidimensional scaling techniques as methodologies for the store layout issue. As the empirical study, a supermarket analysis has done for Migros Turk A.Ş, a leading Turkish retailing company.

Key-Words: - Data Mining, Decision Support Systems, Association Rules, Market Basket Analysis, Apriori Algorithm, Multidimensional Scaling Technique, Store Layout

1 Introduction

Data mining is an exciting and challenging field with the ability to solve many complex scientific and business problems. In recent years, the field of data mining has seen an explosion of interest from both academia and industry [1]. Increasing volume of data, increasing awareness of inadequacy of human brain to process data and increasing affordability of machine learning are reasons of growing popularity of data mining [2]. Data

decision making [3]. Different technologies are invented to meet different decision making goals.

1.1 Data Mining and Decision Support

Data mining is the process of extracting previously unknown, valid and actionable patterns or knowledge from large databases for decision support. In context of this study, relationship between data mining and decision support is shown in Fig.1

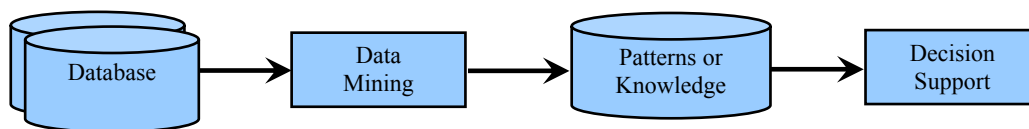


Fig.1 Data Driven Decision Support

mining is a top ten emerging technology.

Data mining is a set of automated techniques used to extract buried or previously unknown pieces of information from large databases, using different criteria, which makes it possible to discover patterns and relationships. This new derived information can be utilized in the areas such as decision support, prediction, forecasting and estimation to make important business decisions, which can help in giving a particular business the competitive edge.

Nowadays, it has been becoming critically important to make a decision that based on evidences than convictions of the authorities. Decision support system is a computer based information system designed to facilitate the decision making process of semi structured tasks. Central issue in DSS support is improvement of

The extraction of hidden predictive information from large databases is a powerful tool with great potential to help organizations to define the information market needs of tomorrow. Data mining tools predict future trends and behaviors, allowing businesses to make knowledge-driven decisions that will affect the company, both short term and long term. The automated prospective analysis offered by data mining tools of today is much more effective than the analysis provided by tools in the past. Data mining answers business questions that traditionally were too time-consuming to resolve. Data mining tools search databases for hidden patterns, finding predictive information that experts may miss because it was outside their expectations.

Data mining is not new. Although data mining is a relatively new term, the technology is not. In the 1960s, Management information Systems and later, in the

1970s, Decision Support Systems (DSS) were praised for their great potential to supply executives with mountains of data needed to carry out their jobs. After 1995s, corporate intranets were developed to support information exchange and knowledge management. The primary decision support tools in use included ad hoc query and reporting tools, optimization and simulation models, online analytical processing and data visualization. On the other hand, a data warehouse is the newest form of decision support system [4]. Data mining was defined as one of the hottest technologies in decision support applications to date.

Advances in data collection, the widespread use of bar codes for most commercial products, and the computerization of many business transactions have flooded us with information, and generated an urgent need for new techniques and tools that can intelligently and automatically assist us in transforming this data into useful knowledge. Today, there is a huge amount of information locked up in the mountains of data in companies' databases, information that is potentially important but has not yet discovered.

Since almost all mid to large size retailers today possess electronic sales transaction systems, retailers realize that competitive advantage will no longer be achieved by the mere use of these systems for purposes of inventory management or facilitating customer check-out. In contrast, competitive advantage will be gained by those retailers who are able to extract the knowledge hidden in the data, generated by those systems, and use it to optimize their marketing decision making. In this context, knowledge about how customers are using the retail store is of critical importance and distinctive competencies will be built by those retailers who best succeed in extracting actionable knowledge from these data.

Data Mining provides many different techniques to extract knowledge from data. It is an exciting multidisciplinary field of research which has many extremely useful applications. At present the techniques are becoming more commonly used but have not been applied adequately in the store layout. Store layout problem is motivated by applications known as market basket analysis to find relationships between items purchased by customers.

1.2 Store Layout Based on Knowledge

Store layout is an important retailing decision that can help or hurt sales and store profitability. Store layouts are extremely important because they strongly influence in-store traffic patterns, shopping atmosphere, shopping behavior, and operational efficiency. Most retailers know that their retail store layout has great impact on their business. And because of this they contemplate changes

to improve traffic flow, increase their shelf space, and increase their sales. The effects of retail store layouts are too big to overlook.

The store layout has been well studied in both academic and practitioner literature. Merrilees and Miller report that store layout design is one of the more important determinants of store loyalty [5]. Store layout design can play a key role not only in satisfying buyers' requirements but also in influencing their wants and preferences. Store layout affects consumers' price acceptability, which is positively related to purchase intentions. They also report that superstores are currently revolutionizing the nature of retail service, mainly by creating more effective self-service arrangements as a result of improvements in store layout design.

The routine store layout in the supermarkets based on the industrial logic implementation, which means putting products that share some functional characteristics or origins in the same area. So we will find product categories. Despite improvements, the store layout remains organized in product categories as defined by the manufacturers or category buyers. This approach is company oriented and it fails to respond to the needs of the time pressured consumer. Some retailers are trying to move from this organization to consumer oriented [6].

This paper proposes a new store layout approach based on the association rule mining. We assume that attractive store layout, navigational aids, salespeople contact and in-store events induce transitions from recreational shopping to purchase-oriented shopping, whereas retail crowding and time pressure engender shift from purchase-oriented shopping to recreational shopping. In addition, it is predicted that environmental design characteristics have greater impact on shopping path than on purchase decision; marketing interventions exert more influence on purchase decision than on movement; while contextual factors have comparable effect on shopping path and purchase decision. Retailers can utilize the proposed model to dynamically improve their in-store conversion rate.

1.3 Association Rules Mining

The applications of data mining in retail trade enterprises are mainly concentrated in association rules mining. Association rule mining is an initial data exploration approach that is often applied to extremely large data set. An example is grocery store market basket data. Association rules mining provides valuable information in assessing significant correlations. By mining association rules, marketing analysts try to find sets of products that are frequently bought together. They have been applied to a variety of fields.

Market Basket Analysis is used to determine which products sell together, the input data to a Market Basket

Analysis is normally a list of sales transactions, where each has two dimensions, one represents a product and the other represents a customer, depending on whether the goal of the analysis is to find which items sell together to the same person. Apriori algorithm is one of the most widely used and famous techniques for finding association rules [7]. This algorithm was chosen primarily due to the speed of application.

1.4 Multi Dimensional Scaling

Multi Dimensional Scaling (MDS) is a whole class of methods for the visualization of high-dimensional data [8]. Visualization involves the transformation of the high-dimensional geometry into a low-dimensional picture, two or at most three dimensions. The objective of the MDS is to perform this transformation in such a way that distances between objects are preserved as much as possible. The MDS addresses the problem how proximity data can be faithfully visualized as points in a low-dimensional Euclidean space. The quality of a data embedding is measured by a stress function which compares proximity values with Euclidean distances of the respective points. In marketing research practice, meaningful cross correlational structures are merely determined by visual inspection. Thus, the marketing analyst usually aims for a parsimonious representation of the cross-category associations in a compressed and meaningful fashion. The MDS techniques are typically employed to accomplish this task.

2 Methodology

This study initially followed the Cross Industry Standard Practice for Data Mining (CRISP-DM) methodology [9]. CRISP-DM is vendor-independent so it can be used with any DM tool and it can be applied to solve any DM problem. CRISP-DM also defines for each phase the tasks and the deliverables for each task. The CRISP-DM data mining methodology is described by its authors in terms of a hierarchical process model, consisting of sets of tasks described at four levels of abstraction (from general to specific): phase, generic task, specialized task and process instance. At the top level, the data mining process is organized in six phases as follow: 1) context understanding, 2) data understanding, 3) data preparation, 4) modeling, 5) evaluation and 6) deployment.

2.1 System Framework

This paper proposes an integrated framework for the store layout problem. The proposed framework proceeds in a stepwise manner as depicted in Fig.2. The proposed

procedure begins with association rules mining from transaction data.

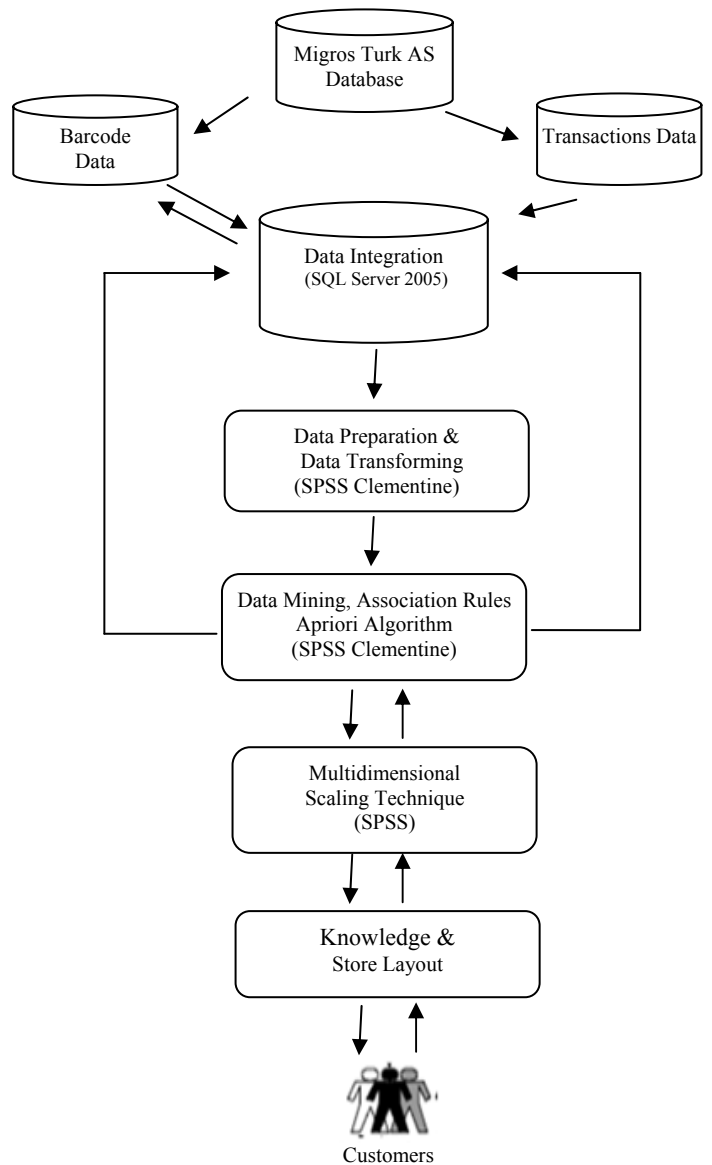


Fig.2 Flowchart of the Proposed Approach

In the first stage, using the CRISP-DM Methodology, market basket analysis has been proceeded to extract knowledge that will be used in store layout design. In the next stage, the MDS is used for visualization, i.e., the procedure execute the MDS to display the set of products in the store space. By assembling categories with strong buying association rules obtained in the first stage, we have tried to propose a new store layout, where consumers find every thing they want in the same store area, maximizing the consumer’s use of the spent in the store. The proposed framework proceeds in the next section with an empirical study.

3 Empirical Study and Results

This paper develops a relational database and uses Apriori algorithm and Multi Dimensional Scaling techniques as methodologies for the store layout. Knowledge extraction by association rule mining results is illustrated as knowledge patterns/rules and clusters in order to propose suggestions and solutions to the case firm for store layout. This paper investigates the store layout issue based on market basket analysis in Migros Turk AS.

Migros Turk AS, one of Turkey's largest supermarket chains. Founded in 1954, Migros Turk, the leader of the Turkish retail sector, has 7,000 employees and serves 160 million customers a year. Listed in the Istanbul Stock Exchange under the symbol MGRS, Migros Turk serves Turkish consumers with seven different retailing operations: Migros Stores, Sok Stores, Shopping Centers, Ramstores, Online Shopping, Bakkalim and Wholesale Stores. Migros Turk operates 150 Migros stores, 292 Sok stores, 3 Ramstores in Baku, 3 Ramstore Shopping Centers in Moscow, and Migros Shopping Centers in Beylikduzu, Ankara and Antalya. For more information on Migros Turk, visit (<http://www.migros.com.tr>).

First of all, we need to measure the relationship among products. To do so, we have got six month database from a Turkish supermarket. The database has 1 million transactions during the period (from 1.1.2005 to 31.1. 2005.). they gave as two text file. The first file contains transactions during the period. Each transaction has the date, cache no, receipt no, and barcode. Fig.3 shows the raw data.

Dosya	Düzen	Bicim	Görünüm	Yardım
20030101	001	7004	7290000193717	
20030101	001	7059	8690605230103	
20030101	001	6937	8690528101000	
20030101	001	7022	8690690252517	
20030101	001	7112	8691365000013	
20030101	001	7078	8690536800087	
20030101	001	7018	8690642000203	
20030101	001	7034	8690188010278	
20030101	001	7079	8690530552845	
20030101	001	7074	8691365000013	
20030101	001	7082	8690101112270	
20030101	001	7078	8690526096124	
20030101	001	7076	8690521025174	
20030101	001	6999	1120	
20030101	001	7019	8690101123603	

Fig.3 A sale transaction table

The second database file contains product data. The first step toward a data mining is to transform text files into SQL Server 2005. This study established relational data base tables and transferred them on MS SQL Server within ODBC environment in order to implement the data table on SPSS Clementine. The text formatted data of transaction records was loaded into a relational database for querying. For the repeat of cache no reasons, we created a new filed, customer no (Mid), which contains date, cache no and receipt no by using SQL query and in total, 186886 customer have been identified.

The store carries 12077 different products grouped in 558 product categories. In the product data file there are barcode, product name, and group item code (Fig.4). Migros Turk has used grid layout. According to current layout, Migros Turk has hierarchy categories like item category and sub item category. In this context we have determined 35 main categories.

Barcode	Product Name	Price
8690521017575	CLEAR SAMPUAN 750 GR YAGLI SAÇLAR	3426
8690521017650	OMO TOZ 1500 GR EKO	3022
8690521018350	ELIDOR SAÇ KREMI 300 GR SERAMIDLİ	3434
8690521018442	CIF LUX 435 GR.	3061
8690521018466	CIF LUX 900 GR.	3061
8690521018480	CIF LUX SIVI 1900 GR.	3061
8690521018527	CIF LUX 900 GR FERAH	3061
8690521018541	CIF LUX SIVI ECO 450 GR	3061
8690521018596	DOMESTOS TOZ 950 GR.	3050
8690521018718	ELIDOR SAMPUAN 400 GR INCE TELLİ SAÇLAR	3425
8690521018732	ELIDOR SAÇ KREMI 300 GR INCE TELLİ SAÇLAR	3434
8690521018916	CIF OKSI-JEL 1000 GR. FLORAL	3071
8690521019036	CIF LUX 5 KG. BULASIK SIVISI	3061
8690521019593	DOMESTOS 810 GR*2 % 20 IND.	3056
8690521019883	RINSOMATİK 15 KG GELİN ÇIÇEĞİ	3012

Fig.4 A sale item description relation

At the end of this stage, we have constituted a pivot table, columns are considered item categories, and rows are customer numbers (Fig.5). We have chosen all the categories, which are 35 different categories, to construct a correlation matrix.

Mid	14	13	32	31	30	6	15	29	3	28	7	18	3
200301010016931	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016932	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016939	\$	\$	\$	\$	1	\$	\$	\$	\$	\$	\$	\$	\$
200301010016940	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016944	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016947	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016948	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016949	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016950	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016952	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016953	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016954	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016955	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016958	\$	\$	\$	\$	\$	\$	\$	4	\$	\$	1	1	\$
200301010016959	\$	\$	\$	\$	1	\$	\$	\$	\$	\$	\$	\$	\$
200301010016960	1	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016963	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$
200301010016964	\$	\$	\$	\$	\$	1	\$	\$	\$	\$	\$	\$	\$
200301010016969	\$	\$	\$	\$	\$	1	\$	\$	\$	\$	\$	\$	\$
200301010016972	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$	\$

Fig.5 Pivot Table

Data modeling is where the data mining software is used to generate results for various situations. In the study, SPSS Clementine is employed as a data mining tool for analysis. The data processing in Clementine is done through the use of nodes, which are then connected together to form a stream frame. In addition, data visualization can be presented to users after the mining process has been done. Fig.6 shows our final model.

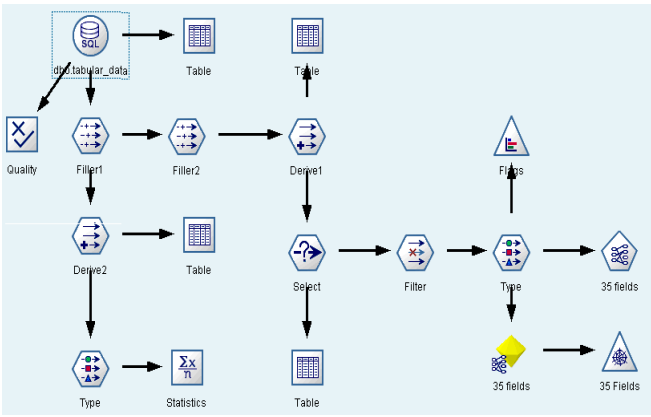


Fig.6 Created Model in SPSS Clementine

Association rules are discovered by the Apriori algorithm, in which minimum support and minimum confidence are set to 0.05% and 20.0%, respectively. 518 rules then are identified and presented in Fig.7.

Consequent	Antecedent	Confidence %	Rule Support %
5	25	76.703	5.578
5	8	70.612	9.211
5	35	70.577	6.552
5	24	70.566	1.942
19	9	68.918	4.997
5	15	68.417	0.787
5	9	67.986	4.930
5	18	67.386	9.683
5	4	67.064	5.455
21	7	65.985	4.104
5	16	65.188	6.016
5	3	63.820	4.877
5	7	63.799	3.968
19	20	63.517	11.179
5	14	63.283	6.791
19	24	63.229	1.740
19	4	62.627	5.094
21	22	62.529	8.905
5	22	61.596	8.773
19	15	61.507	0.708
19	35	61.346	5.695
19	8	61.330	8.000
19	11	61.199	25.610

Fig.7 Association rules of basic product (min sup=0.05%, min conf=20%)

A diagram produced in determining product-purchasing relationships. In the process of discovering affinities in the basket contents, Fig.8 illustrates associations between the products, with the darker lines indicating the strongest association between products. The Web diagram suggests a strong association between 15, 110 and 121 purchases. Further analysis on this group using an association rule detection technique was performed to find out more information on this group.

We use the buying association measure to create a category correlation matrix and we apply the multi dimensional scale technique to display the set of products in the store space.

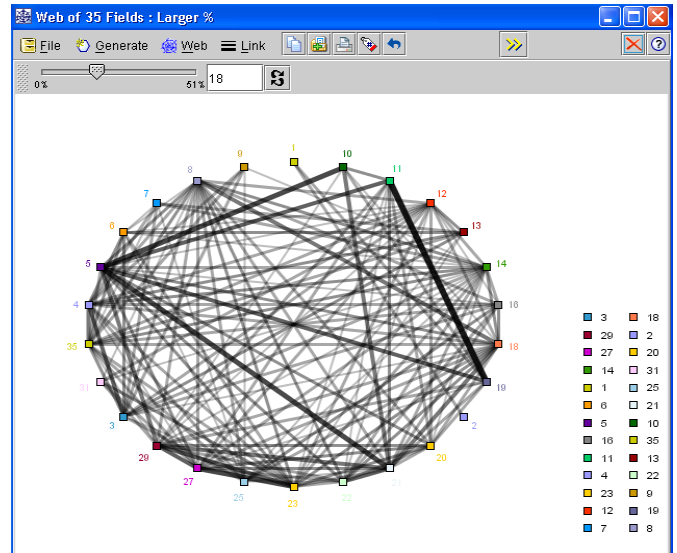


Fig.8 Association diagram (Line Val. >18%)

Here, the MDS technique is used to provide us with a tool for the visualization of the database at different levels of details. We have chosen 35 different categories to construct a correlation matrix. Once we have established the correlation matrix, we are able to calculate the spatial representation of these relationships through the MDS. We have used data as distance and an asymmetric matrix to produce the results. In order to use the correlation matrix as distances among categories, we have inversed the values by subtracting 1 from all values. So, if two products have a strong correlation the proximities will be small, this means that those categories are similar and should be represented in a nearby space on the map.

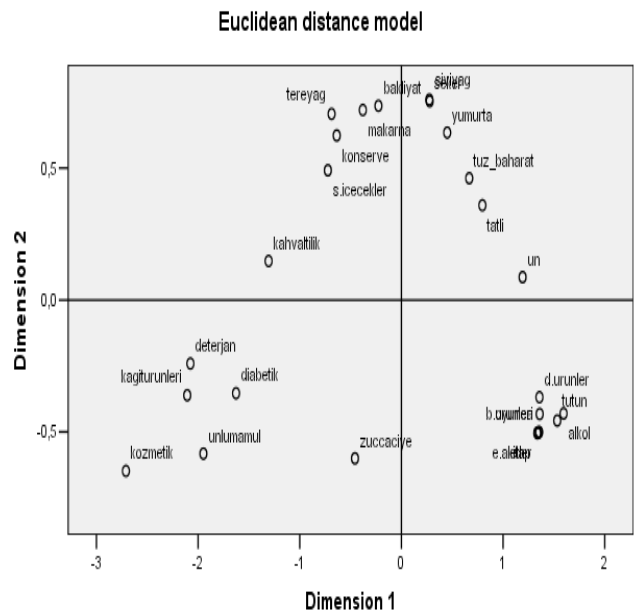


Fig.9 MDS on Association Matrix

However, the model stress is 0,093 and the square correlation is 0,96, which means that these results are acceptable and significant about accepting this model. We represent all categories in the multidimensional space as showed in the Fig.9.

As shown from Fig.9, Four Clusters can be formed. The first cluster in the Figure 9 is comprises dried beans, spaghetti, canned food, hot drinks, and breakfast products. Cluster 2 shows oil, sugar, egg, and salt, spices, flour, and dessert products being bought together in many shopping. Cluster 3 comprises paper products, detergent, cleanness products, and the personnel care products floury and diabetic products. Cluster 4 represents tobacco, alcoholic drinks, books, magazine, writing material toy, baby products, and textile. These clusters also mean four consumption universes for the new store layout.

Migros Turk' current store layout is shown in figure 10. According to obtained results from this study, the products having strong association have been located in the same area and generally customer visit this area for their shopping. The visited area surround in red rectangular in Fig.10. For space reasons, we will not show the other layout maps.

layout, where consumers find everything they want in the same store area, maximizing the consumer's use of time spent in the store.

Conclusions

This paper discusses association rules for data mining extract knowledge from a database and a new supermarket store layout based on the association among categories. This approach allows supermarkets to cluster products around meaningful purchase opportunities related to use association.

Acknowledgments

We would like to thank the Commission for the Scientific Research Projects of Sakarya University. We also would like to thank Mrs. Tiryal Demirkılınc, Manager of Data Warehouse in Migros Turk T.A.Ş.

References

- [1] Olafson, S.X. Li and S. Wu, Operations research and data mining, *European Journal of Operational Research* 187 (2008), pp. 1429–1448.
- [2] Marakas, G.M. *Decision support systems in the 21st Century*, Prentice-Hall of India, Second Edition, 2004.
- [3] Cil, I., Alptürk, O., Yazgan, H.R., A New Collaborative System Framework Based on Multiple Perspectives Approach: *InteliTeam, Decision Support Systems* 39, 4, 2005, 545-685.
- [4] Brohman, M. K, The Business Intelligence Value Chain: Data-Driven Decision Support in a Data Warehouse Environment: An Exploratory Study, *Proceedings of the 33rd Hawaii International Conference on System Sciences – 2000*,
- [5] Merrilees, B. Miller, D. Superstore interactivity: a new self-service paradigm of retail service? *International Journal of Retail & Distribution Management*, Volume 29, Number 8, 2001.
- [6] Borges, A., Toward a new supermarket layout: from industrial categories to one stop shopping organization through a data mining approach,. *Proceedings of the 2003 Society for Marketing Advances Annual Symposium on Retail Patronage and Strategy*, Montreal, November 4-5, 2003
- [7] Agrawal Rakesh, R.S.: Fast algorithms for mining association rules. In: *Proceedings of the 20th VLDB Conference*, Santiago, Chile, 1994.
- [8] Borg, I., P. Groenen, Modern Multidimensional Scaling, *Springer Series in Statistics*, Springer, Berlin, 1997.
- [9] Shearer, C. The crisp-dm model: The new blueprint for data mining. *Journal of Data Warehousing*, 5, 2000.

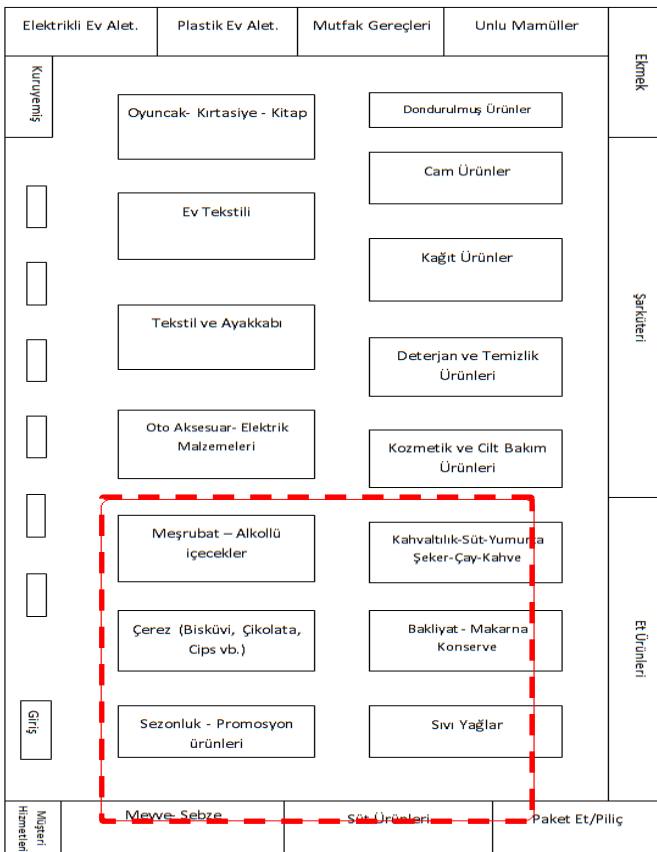


Fig.10 Current store layout of the branch of the Migros

By assembling categories with strong buying associations, we have tried to propose a new store