DOI: 10.20472/IAC.2015.017.040

OZLEM AKCAY KASAPOGLU
ISTANBUL UNİVERSİTY FACULTY OF BUSINESS ADMINISTRATION , TURKEY

UMMAN TUGBA GURSOY
ISTANBUL UNIVERSITY FACULTY OF BUSINESS ADMINISTRATION, TURKEY

DATA MINING AND ERP: AN APPLICATION IN RETAIL SECTOR

Abstract:

Many medium or large scale organizations with large databases invest on advance data collecting and managing systems. The main point of turning this data into your success is the difficulty of extracting knowledge about the system that you study from the collected data. Enterprise Resource Planning (ERP) software helps companies to put all previously separated data to in single software. ERP has several advantages. Storing whole data in a single place make it possible to analyze data from different business functions. Because a large scale of data are in the same place, new tools are needed to analyze them. In this study customer purchase records from the the biggest computer retailing firm's data in Turkey were analyzed. Association Rules were used to determine the shopping behavior of the customers. According to the results, various rule sets are obtained. This rule sets can be used for purposes such as store layout, shelf arrangement in the store, products can be placed close together to increase sales and other promotional strategies.

Keywords:

ERP, Data Mining, Association Rules, Retailing.

JEL Classification: C88

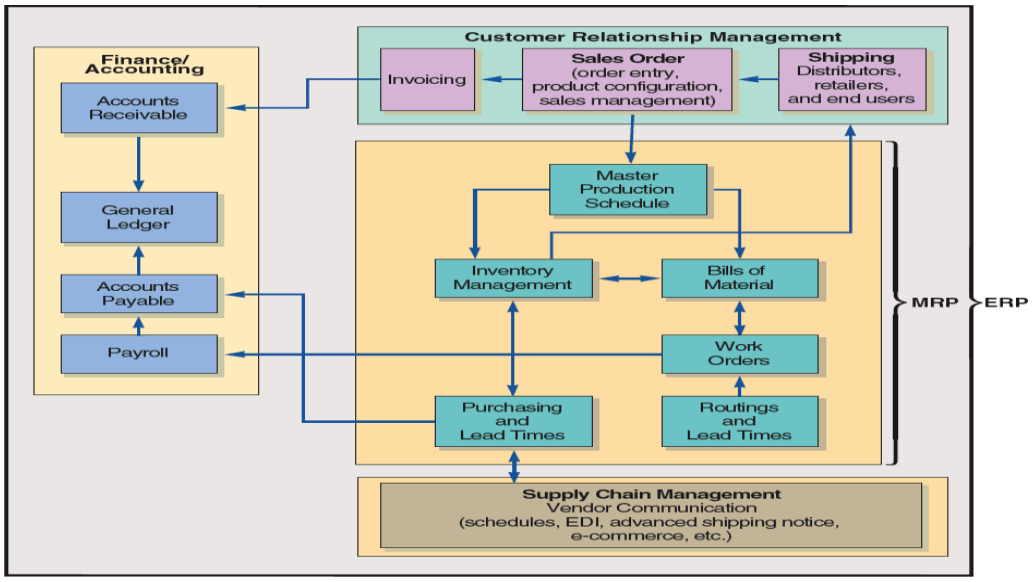
Introduction

The business environment is dramatically changing. Companies today face the challenge of increasing competition, expanding markets, and rising customer expectations. This increases the pressure on companies to lower total costs in the entire supply chain, shorten throughput times, drastically reduce inventories, expand product choice, provide more reliable delivery dates and better customer service, improve quality, and efficiently coordinate global demand, supply, and production. As the business world moves ever closer to a completely collaborative model and competitors upgrade their capabilities, to remain competitive, organizations must improve their own business practices and procedures. Companies must also increasingly share with their suppliers, distributors, and customers the critical in-house information they once aggressively protected. And functions within the company must upgrade their capability to generate and communicate timely and accurate information. To accomplish these objectives, companies are increasingly turning to enterprise resource planning (ERP)systems. ERP provides two major benefits that do not exist in non-integrated departmental systems: (1)a unified enterprise view of the business that encompasses all functions and departments; and (2) an enterprise database where all business transactions are entered, recorded, processed, monitored, and reported. This unified view increases the requirement for, and the extent of, interdepartmental cooperation and coordination. But it enables companies to achieve their objectives of increased communication and responsiveness to all stakeholders. (E.J. Umble et al.,2003, p:241) Enterprise Resource Planning (ERP) software helps companies to put all previously separated data to in single software. ERP has several advantages. Storing whole data in a single place make it possible to analyze data from different business functions. Because a large scale of data are in the same place, new tools are needed to analyze them, data mining is one of the succesfull tools that is used for that purpose. In this paper in the first section enterprise resourse planning and in the second section data mining and association rules will be introduced and in the final section association rules is used to determine the shopping behavior of the customers from the customer purchase records of the biggest computer retailing firm's data in Turkey.

1.Enterprise Resource Planning (ERP)

Advances in MRPII systems that tie customers and suppliers to MRP II have led to development of ERP systems. Erp is a software that allows companies to automate and integrate many of their business prosesses, share a common database and business practices Throughout the enterprise and produce information real time. The objective of an Erp system to coordinate a firm's whole business is seldom achieved. This is accomplished by using centralized database to asist he flow of information among business functions. (*Heizer, Render, 2009, p488-489)* A schematic showing some of the relationships for a manufacturing firm appears in Figure 1.

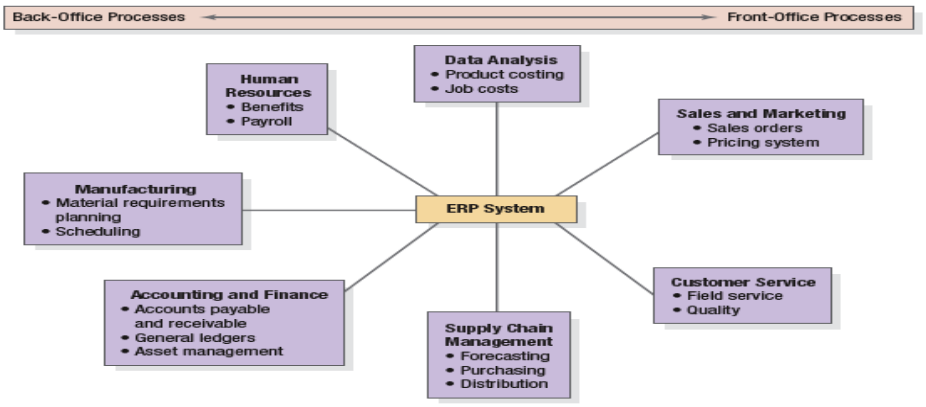
Figure 1: Erp İnformation flows



Source: Heizer J., Render B., Operations Management, Pearson International Edition, New Jersey 2009, p:488

ERP revolves around a single comprehensive database that can be made available across the entire organization (or enterprise).The database collects data and feeds them into the various modular applications (or suites). As new information is entered as a transaction in one application, related information is automatically updated in the other applications.The ERP system streamlines the data flows throughout the organization and provides employees with direct access to a wealth of real-time operating information. ERP eliminates many of the cross-functional coordination problems older nonintegrated systems suffered from. (Krajewski et al.,2010, p.562)

Figure 2: Erp application modules



Source: Krajewski L.J.,Ritzman L.P., Malhotra M. K., Operations Management, Pearson Global Edition, 2010, p:563

ERP systems provide an increased level of integration to support core business processes and is an amalgamation of three most important components - Business Management Practices, Information Technology and Specific Business Objectives. Till recent past operational and transactional needs and not information were the focus of most ERP implementations. At the core of ERP is a well managed centralized data repository which acquires information from and supply information into the fragmented applications operating on a universal computing platform. Information in large business organizations is accumulated on various servers across many functional units and sometimes separated by geographical boundaries . Such information islands can possibly service individual organizational units but fail to enhance enterprise wide performance, speed and competence. In these cases, it is sometimes necessary to gather the scattered data into a single database called a Data Warehouse (DW), before submitting it to data mining activity. The key objective of an ERP system is to integrate information and processes from all functional divisions of an organization and merge it for effortless access and structured workflow. The integration is typically accomplished by constructing a single databaserepository that communicates with multiple software applications providing different divisions of anorganization with various business statistics and information.

Erp systems undergo significant changes over the past several years. One direction points to their interoperability- the ability of one piece of software to interact with others. Electronic data interchange, a system that allows data to be transfered between companies on a batch basis and XML (Extensive Markup Language) let companies structure and Exchange information without rewriting their existing software or having to purchase new software or hardware. The goal of these and other such methods is to automate, in almost real time, the sharing of information across enterprise boundaries.(Krajewski et al.,2010, p.563)

There are advantages of the Erp Systems. These are**:** Erp systems provides integration of the supply chain, production, and administration,creates commonality of databases,can incorporate improved best processes,it increase communication and collaboration between business units and sites, it has an off-the-shelf software database and may provide a strategic advantage and the d sadvantages of the Erp Systems are**:** they are very expensive to purchase and even more so to customize, implementation may require major changes in the company and its processes,it is so complex that many companies cannot adjust to it, involves an ongoing, possibly never completed, process for implementation and expertise is limited with ongoing staffing problems.(*Heizer, Render, 2009, p491)*

ERP vendors such as SAP,Oracle,BaaN,PeopleSoft,JD Edwards, etc. have produced ERP applications, such as Business Intelligence, Customer Relationship Management, Supply Chain Management, etc. They are good for the business but it is a big challenge for IT professionals to drive information out of the complex, large data volumes, report and analyze them. To able to understand current business activities, most large companies use ERP applications to run their critical applications. Building a separate data and store them by extracting data from ERP applications is not an easy task, the data captured from ERP applications are analyzed detailly to make decisions for business critical applications with the requirement of the detection of the hidden anomalies and patterns.

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the need for turning such data into useful information and knowledge. The information and knowledge gained can be used for applications ranging from market analysis, fraud detection and customer retention, to production control and science exploration. (Han, Kamber, 2006,p.1)
Data mining can be viewed as a result of the natural evolution of information technology. The database system industry has witnessed an evolutinoary path in the development of the following functionalities: data collection and database creation, data management and advanced data analysis.
Data mining is an integral part of knowledge discovery in databases, which is the overall process of converting raw data into useful information. This process consists of a series of transformation steps, from data preprocessing to postprocessing of data mining results. (Tan, Steinbach, Kumar, 2006, p.3)

DATA MINING AND ERP

During the past decade, the two major factors that have revolutionized the Information Age are the advent of the Internet and Integrated Enterprise Package Applications, commonly known as ERP (Enterprise Resource Planning) systems. Both factors require the decision-makers to sift through tons of complex data to analyze hidden anomalies and detect patterns to understand current business activities.
The data mining process is very time consuming, and the entire process is confined to the back-office. Traditional data mining products are not intuitive, nor integrated with On-Line Analytical Processing (OLAP) activities. Today, most data mining tools require a separate database designed specifically for data mining processes, which takes too much time and resources to construct. It is a costly proposition.
Also the traditional data mining tools are not developed to work against multi-dimensional data structures, which are widely used to deploy data marts in organizations. Building a separate data store by extracting data from ERP applications is a difficult and a very important task. Most large companies use ERP applications to run their business critical applications. (www.decisionsystems.com)
Enterprise resource planning (ERP) software allows a business to consolidate previously separate data into a single application. This means that finance, sales, production, and other business functions use the same application, and the data for that application are in a single database. This has several implications. Because all of the data are stored in the same place, it is much easier to analyze data from multiple business functions. Because so much more data are in the same place, new tools are needed to analyze the data. (www.ittoolbox.com)

ASSOCIATION RULES ANALYSIS

Many business enterprises accumulate large quantities of data from their day-to-day operations. For example, huge amounts of customer purchase data are collected daily at the checkout counters of grocery stores. Such data, commonly known as market basket transactions. Retailers are interested in analyzing the data to learn about the purchasing behavior of their customers. Such valuable information can be used to support a variety of business-related applications such as marketing promotions, inventory management, and customer relationship management. (Tan, Steinbach, Kumar, 2006, p.327)

Each association rule is usually evaluated by a support and a confidence measure.

The Support of an association rule is the ratio of the number of transactions having the value “yes-true” for all items in both the set X and Y divided by the total number of transactions.

Support (X implies Y) = P(X, Y)/ N (1)

The Confidence of an association rule is the ratio of the number of transactions having the value “yes-true” for all items in both the set X and the set Y divided by the number of transactions having the value “yes-true” for all items in the set Y.

Confidence (X implies Y) = P(X, Y) / P(Y) (2)

The association rule discovery task consists of extracting from the data being mined all rules with support and confidence greater than or equal to user-specified tresholds called minimum support and minimum confidence. (Freitas, p. 23)

APPLICATION

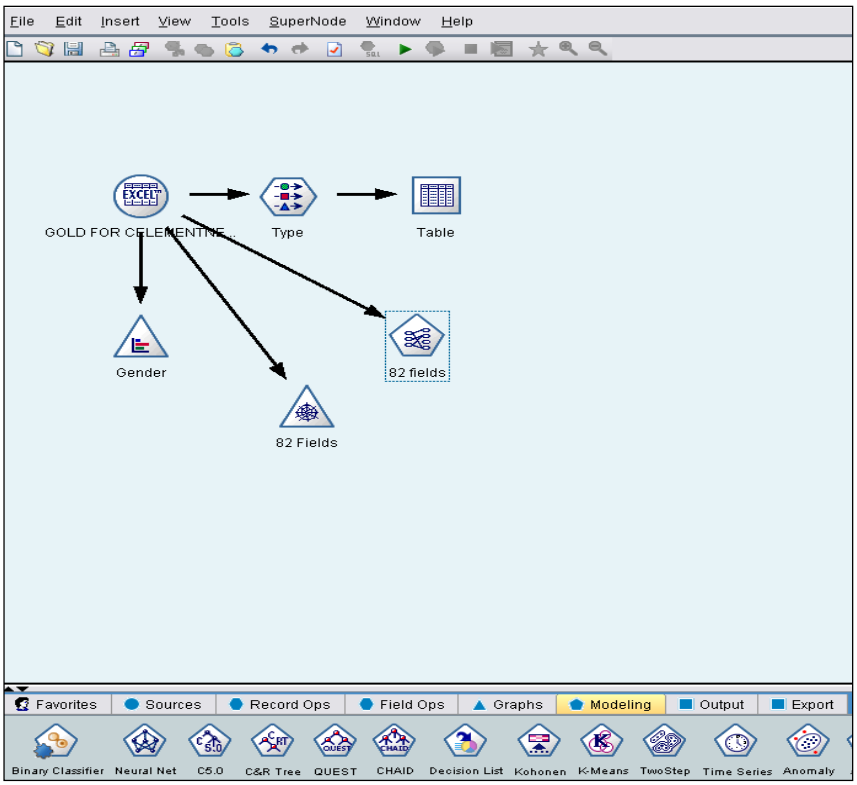
In this analysis, we applied the data of the biggest computer-electronic retailer in Turkey. Data consists of 85 fields and 4875 records.(Table 1) Those are from May 2014. Variables are products like adaptor, main board etc. Data is coded as 0 and 1

Table 1: Data set

	A	C	D	E	F	G	H	I	J	K
1	Number	Gender	Quantity	Adaptor	Main board	Light	Connection Device	Barkode Reader	Hard Disc	Computer Parts
2	4.251	3	1,00	0	0	0	0	0	0	0
3	782	3	1,00	0	0	0	0	0	0	0
4	4.600	3	1,00	0	0	0	0	0	o	0
5	9.529	3	2,00	0	0	0	0	0	0	0
6	3.840	3	1,00	0	0	0	0	0	0	0
7	6.326	3	1,00	0	0	0	0	0	0	0
8	3.024	3	1,00	0	0	0	0	0	0	0
9	2.058	3	1,00	0	0	0	0	0	0	0
10	9.493	3	1,00	0	0	0	0	0	0	0
11	6.677	3	1,00	0	0	0	0	0	0	0
12	6.392	3	1,00	0	0	0	0	0	0	0
13	6.403	3	1,00	0	0	0	0	0	0	0
14	2.572	3	1,00	1	0	0	0	0	0	0
15	4.761	3	1,00	0	0	0	0	0	0	0

Figure 3 shows the SPSS Clementine model for the analysis.

Figure 3: Model



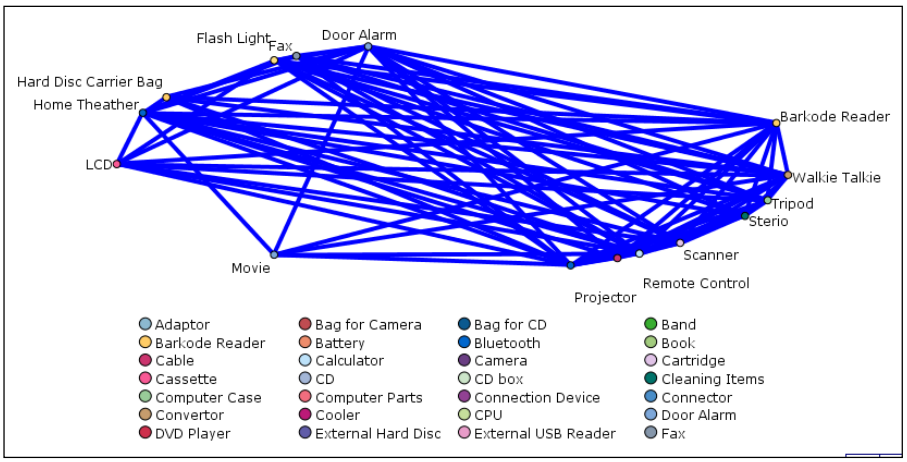
65,23% of the customers are man, 23,86% of are firms and the rest is woman.

Figure 4: Distribution of gender

Value	Proportion	%	Count
2.000		65,23	3180
3.000		23,86	1163
1.000		10,91	532

Web graph shows which items are bought together. (Figure 5)

Figure 5: Web graph



In data preparation phase variables are organized for Association Rules. Apriori which is the most commonly known algorithm is applied to the data set.

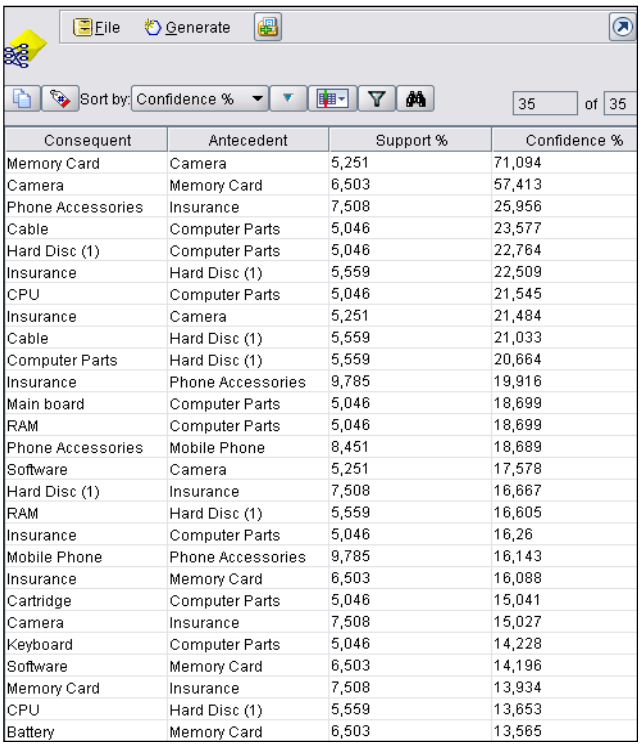
Table 2: Apriori algorithm

Field	Type	Values	Missing	Check	Direction
Number	Typeless			None	None
Gender	Set	1.0,2.0,3.0		None	In
Adaptor	Flag	1.0/0.0		None	Both
Main board	Flag	1.0/0.0		None	Both
Light	Flag	1.0/0.0		None	Both
Connection D...	Flag	1.0/0.0		None	Both
Barkode Rea...	Flag	1.0/0.0		None	Both
Hard Disc	Flag	o/"0.00000...		None	Both
Computer Pa...	Flag	1.0/0.0		None	Both
Bluetooth	Flag	1.0/0.0		None	Both
CD	Flag	1.0/0.0		None	Both
Bag for CD	Flag	1.0/0.0		None	Both

Rulesets can be seen in Table 3. Rulesets can be interpreted as below:

- 71,094% of the customers who bought camera, also bought memory card. The support rate is 5,251%.
- 23,577% of the customers who bought computer parts, also bought cable. The support rate is 5,046%.
- 21,545% of the customers who bought computer parts, also bought CPU. The support rate is 5,046%.
- 18,689% of the customers who bought mobile phone, also bought phone accessories. The support rate is 8,451%.

Table 3: Rulesets



Consequent	Antecedent	Support %	Confidence %
Memory Card	Camera	5,251	71,094
Camera	Memory Card	6,503	57,413
Phone Accessories	Insurance	7,508	25,956
Cable	Computer Parts	5,046	23,577
Hard Disc (1)	Computer Parts	5,046	22,764
Insurance	Hard Disc (1)	5,559	22,509
CPU	Computer Parts	5,046	21,545
Insurance	Camera	5,251	21,484
Cable	Hard Disc (1)	5,559	21,033
Computer Parts	Hard Disc (1)	5,559	20,664
Insurance	Phone Accessories	9,785	19,916
Main board	Computer Parts	5,046	18,699
RAM	Computer Parts	5,046	18,699
Phone Accessories	Mobile Phone	8,451	18,689
Software	Camera	5,251	17,578
Hard Disc (1)	Insurance	7,508	16,667
RAM	Hard Disc (1)	5,559	16,605
Insurance	Computer Parts	5,046	16,26
Mobile Phone	Phone Accessories	9,785	16,143
Insurance	Memory Card	6,503	16,088
Cartridge	Computer Parts	5,046	15,041
Camera	Insurance	7,508	15,027
Keyboard	Computer Parts	5,046	14,228
Software	Memory Card	6,503	14,196
Memory Card	Insurance	7,508	13,934
CPU	Hard Disc (1)	5,559	13,653
Battery	Memory Card	6,503	13,565
Cable	Battery	9,128	13,034
Battery	Camera	5,251	12,891
Main board	Hard Disc (1)	5,559	12,546
Cartridge	Hard Disc (1)	5,559	11,808
Computer Case	Computer Parts	5,046	10,976
Cable	Head Set	5,046	10,976
Computer Parts	Insurance	7,508	10,929
Cable	CD	5,169	10,317

CONCLUSION

In this study the biggest computer retailing firm's data in Turkey is analyzed to identify shopping behavior of the customers like the items bought together by the customers, etc..Enterprise Resource Planning (ERP) software is one of the biggest information source for a company. Various functions of business, from production planning, to purchasing, inventory management and supply chain management are data sources for ERP system. Data mining tools are applied to the scattered data across multiple databases.of ERP. In this analysis, the data are extracted from the biggest computer-electronic retailer's ERP System. Association rules under Data Mining is used to analyze data. Various rule sets are gathered from the data. These rule sets can be used for promotional strategies, store layout.

Acknowladgement:
This paper is funded by Istanbul University Scientific Research Projects Foundation UDP-54699

Reference

FREİTAS A.A. (2002). *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer, p: 23.

HAN J.; KAMBER M. (2006) *Data Mining Concepts and Techniques*. Morgan Kauffman, p:1.

HEİZER J., RENDER B. (2009) *Operations Management*. Pearson International Edition, New Jersey, p:488

KRAJEWSKİ L.J.,RİTZMAN L.P., MALHOTRA M. K. (2010) *Operations Management*. Pearson Global Edition, p:563.

SASTRY S.H., BABU M.S.P. (2013) Implementation of CRISP Methodology for ERP Systems. *International Journal of Computer Science Engineering* (IJCSE) ISSN : 2319-7323 Vol. 2 No.05, p:203.

TAN P.N.; STEİNBACH M.; KUMAR V. (2006). *Introduction To Data Mining*. Pearson Education. p: 3 and 327.

UMBLE E. J.; HAFT R. R., UMBLE M. M. (2003) Enterprise Resource Planning: Implementation Procedures and Critical Success Factors*, European Journal of Operational Research*. Vol. 146 p. 241–257.

Discovering Knowledge Hidden in ERP Data, http://www.decisionsystems.com/, 28.07.2015.

What Big Data Means for ERP? , http://it.toolbox.com/, 28.07.2015.