

MASTER'S THESIS



# Sales Prediction for Pharmaceutical Distribution Companies

*A Data Mining Based Approach*

Neda Khalilzadeh

Master (120 credits)  
Master of Science in Electronic Commerce

Luleå University of Technology  
Department of Business Administration, Technology and Social Sciences



# MASTER'S THESIS

## Sales Prediction for Pharmaceutical Distribution Companies- A Data Mining Based Approach

Supervisors:

Prof. Mohammad Mehdi Sepehri  
Prof. Peter Naude

Referees:

Dr. Amir Albadvi  
Dr. Lars Backstrom

Prepared by:

Neda Khalilzadeh

Tarbiat Modares University Faculty of Engineering  
Department of Industrial Engineering

Lulea University of Technology  
Department of Business Administration and Social Sciences  
Division of Industrial Marketing and E-Commerce

JOINT MSc PROGRAM IN MARKETING AND ELECTRONIC COMMERCE



2008

# ABSTRACT

Due to the tough competitions that exist today, most pharmaceutical distribution companies are in a continuous effort to increase their profits and reduce their costs. Actually, both shortage and surplus of goods can lead to loss of income for these companies. One of the problems in pharmaceutical distribution organizations which deal with public health and pharmaceutical products is how to control inventory levels by means of accurate sales prediction in order to prevent costs of excessive inventory also prevent losing their customers because of drug shortage. Accurate sales prediction is certainly a valuable management tool to meet the mentioned goals since this leads to improved customer service, also, reduced lost sales and costs. However, most pharmaceutical distribution companies in Iran are still using heuristic or traditional statistical techniques to make sales prediction for their products. Thus, the purpose of this research is to apply an innovative and reliable sales prediction method for pharmaceutical distribution companies.

To make sales prediction for a pharmaceutical distribution company, we needed to have past sales records of each drug. Accordingly, we gathered sales data of three years from Pakhsh Hejrat Co. which is one of the leading pharmaceutical distributors in Iran. We chose neural networks as our basic tools for sales prediction since most traditional methods like ARIMA are incapable of modeling nonlinearities that exist in most real data; also, they need forecaster's supervision for the parameter estimation phase. In fact, neural networks are versatile tools for sale prediction since estimation with neural networks can be automatized, and they have proved very effective in order to make prediction by handling non-linear input and output variables. Due to the fact that we did not have enough past sales records of drug items, we came up to a new idea of grouping drugs to find group members and make use of co-members' sales data for each other. Thus, we did a comprehensive network based analysis in order to find clique-sets and group members. Afterwards, we built sales forecasting models with three different approaches: a) ARIMA methodology for time series forecasting, b) Hybrid neural network approach for time series forecasting by means of each drug's past records, and c) Hybrid neural network approach for time series forecasting by means of each drug's past records and its group members' past records. Our evaluations and results indicated that our new methodology (number 3 above) was the best methodology, and the weakest one was ARIMA model.

**Keywords:** *Pharmaceutical Distribution Companies, Data Mining, Time Series Sales Forecasting, Network Based Analysis, Neural Networks*

# **ACKNOWLEDGMENTS**

Actually, this research was carried out from spring of 2007 till spring of 2008 at Tarbiat Modares University (TMU) in fulfillment of the MSc Joint program in marketing and e-commerce with Lulea University of Technology (LTU). While working on this thesis, I have learned many unknown issues, which are really of great value to me.

Here, I would like to express my gratitude to all who helped me with their worthwhile supports during all steps of this thesis.

First of all, I would like to express my appreciation to my supervisors Prof. Mohammad Mehdi Sepehri and Prof. Peter Naude for all their patience, kind supports, encouragements, and helpful and valuable advice while performing this research. Without their encouragements and valuable comments, I was not able to fulfill this project research.

I also would like to thank Prof. Amir Albadvi, the head of Industrial Engineering Department of TMU, and Prof. Esmail Salehi-Sangari, the chairman of Industrial Marketing and E-commerce Division of LTU, for their strong and continuous supports in arranging this joint master program.

Furthermore, I want to express my thanks and regards to Dr. Hamid Farvareh, the PHD candidate in Industrial Engineering Department in TMU, for his valuable and great helps and supports, in all steps of this research.

In addition, I would like to extend my special thanks to the experts and sales managers of pharmaceutical distribution companies especially to Dr. Mostafavi and Prof. Sarkandi who guided me with their valuable comments and ideas. I also want to thank Mr. Haj Sadeghi for providing me with Pakhsh Hejrat's past sales records.

And last but not least, I want to thank my families and friends, specially my parents and my dear brother, for their encouragements, patience, and kind supports while accomplishing this research. I would like to dedicate my thesis to my lovely parents in order to express my deep appreciation towards them for their unending and notable supports in all aspects of my life.

**Neda Khalilzadeh**

**24<sup>th</sup> March 2008**

# TABLE OF CONTENTS

CHAPTER 1 .....	9
INTRODUCTION .....	9
<i>1. Introduction</i> .....	9
1.1. Background of the Study.....	10
1.2. Motivation of the Study.....	15
1.3. Pharmaceutical Distribution Companies in Iran .....	16
1.4. Problem Discussion.....	19
1.4.1. Problem Definition .....	20
1.5. Research Objectives and Question.....	23
1.6. Research Framework.....	24
1.7. Contributions of the Research.....	25
1.8. Outline of the Thesis .....	26
CHAPTER 2 .....	28
DEFINITIONS AND LITERATURE REVIEW .....	28
<i>2. Definitions and Literature Review</i> .....	28
2.1. Data Mining and Knowledge Discovery .....	29
2.1.1. An Overview of Data Mining.....	29
2.1.2. Difference between knowledge Discovery and Data Mining .....	30
2.1.3. Descriptions of the Knowledge Discovery and Data Mining in Literatures .....	31
2.1.4. The Process of Data Mining and Knowledge Discovery .....	32
2.1.5. Data Mining Machine Learning and Statistics .....	34
2.1.6. Data Mining Tasks and Techniques .....	35
2.1.7. Choose Appropriate Data Mining Techniques .....	38
2.1.8. Data Mining Applications .....	38
2.2. Neural Networks .....	41
2.2.1. Neural Networks Architectures .....	43
2.2.2. Learning Process.....	47
2.2.3. Back Propagation Algorithm .....	49

2.2.4. Single-layer Perceptron .....	51
2.2.5. Multilayer Perceptron (MLP) .....	52
2.2.6. Radial basis function network (RBF) .....	54
2.2.7. Time Delay Neural Networks (TDNN) .....	55
2.2.8. Recurrent neural networks.....	56
2.2.9. Benefits and Limitations of Neural Networks .....	56
2.2.10. Applications of Neural Networks .....	58
2.3. Supply Chain .....	63
2.3.1. Inventory and Inventory Management .....	65
2.3.2. Participants in the Supply Chain .....	66
2.3.3. Logistics.....	67
2.3.4. Why Demand Forecasting is Essential in a Supply Chain?.....	68
2.4. Forecasting .....	69
2.4.1. Sales Forecasting .....	71
2.4.2. Time Series Forecasting .....	71
2.4.3. Time Series Forecasting Models .....	79
2.4.4. Comparisons of Different Time Series Forecasting Models .....	81
CHAPTER 3 .....	85
RESEARCH METHODOLOGY .....	85
3. <i>Research Methodology</i> .....	85
3.1. Research Design.....	86
3.1.1. Research Purpose.....	87
3.1.2. Research Approach.....	88
3.1.3. Research Strategy .....	89
3.2. Research Process .....	89
3.2.1. Data Collection and Description.....	92
3.2.2. Data Preparing .....	93
3.3. Exploratory Analysis.....	98
3.4. Network Based Analysis .....	99
3.5. Methods Used for Modeling and Time Series Forecasting .....	99
CHAPTER 4 .....	104

RESULTS AND ANALYSIS.....	104
4. <i>Results and Analysis</i> .....	104
4.1. Data preprocessing .....	105
4.2. Exploratory Analysis.....	107
4.2.1. Data Visualization .....	107
4.2.2. Descriptive Statistics .....	111
4.3. Statistical Analysis of Drugs' Networks .....	113
4.3.1. Network Identification.....	114
4.4. Building Sales Forecasting Model .....	125
4.4.1. ARIMA Methodology for Time Series Forecasting.....	128
4.4.2. A Hybrid Neural Network Approach for Time Series Forecasting.....	138
4.4.3. Model Evaluation .....	144
4.5. Summary .....	147
CHAPTER 5 .....	148
CONCLUSION.....	148
5. <i>Conclusion</i> .....	148
5.1. Conclusion.....	148
5.2. Contributions.....	152
5.2.1. Methodological Contribution .....	152
5.2.2. Empirical Contribution .....	152
5.3. Managerial Implications.....	153
5.4. Research Limitations.....	156
5.5. Further Research .....	157
REFERENCES .....	159
Ho, T.B., (2002), "Introduction to Knowledge Discovery and Data Mining", Available at: ttp://www.netnam.vn/unescocourse/knowlegde/know_frm.htm. ....	164
APPENDIX 1 .....	172
Tables of Prediction Results with Neural Networks.....	172
APPENDIX 2 .....	184
Definitions of Codes in Result Tables of Forecasting .....	184

## **LIST OF TABLES**

Table 1: Applications of data mining in marketing and Business .....	41
Table 2: Applications of Neuronal Networks in Business Forecasting .....	62
Table 3: Related Works on Applications of Neural Networks in Sales and Demand Forecasting	77
Table 4: Sales Data Information Collected from Pakhsh Hejrat Company .....	93
Table 5: Amount of Indices for Different øs .....	97
Table 6: Descriptive Measures of 21 Drugs .....	112
Table 7: Cliques Sets .....	116
Table 8: Descriptive Statistics of Degrees .....	120
Table 9: Frequency and Cumulative Frequency of Degrees.....	121
Table 10: Group Members .....	128
Table 11: Parameter Estimation.....	135
Table 12: Four Months' Forecasting Results for the Original Data .....	135
Table 13: Sensitivity Analysis of Drug 71 for the best 5 models of 20 tried models.....	144
Table 14: Forecasting Results and Accuracy Measures of Test Data for Drug 24 (with Its Own Past Records) .....	146
Table 15: Forecasting Results and Accuracy Measures of Test Data for Drug 24 (with Its Partner's Past Records).....	146
Table 16: Forecasting Results and Accuracy Measures of Test Data for Drug 71 (with Its Own Past Records) .....	146
Table 17: Forecasting Results and Accuracy Measures of Test Data for Drug 71 (with Its Partner's Past Records).....	146

# LIST OF FIGURES

Figure 1: Pharmaceutical Distribution Channel.....	18
Figure 3: A Schematic Representation of the Research Framework .....	25
Figure 4: Outline of the Thesis .....	27
Figure 5: A Typical knowledge Discovery Process.....	30
Figure 6: Combining cognitive psychology with artificial intelligence, databases, and statistics	31
Figure 7: Application of Data Mining for Marketing .....	40
Figure 8: A Simple Neural Network .....	44
Figure 9: A Neural Network with one Hidden Layer .....	44
Figure 10: A Neural Network with a Wide Hidden Layer.....	45
Figure 11: A Neural Network with Multiple Outputs.....	45
Figure 12: The Unit of an Artificial Neural Network .....	46
Figure 13: Three Common Transfer Functions .....	47
Figure 14: A Back-Propagation Network Model.....	51
Figure 15: Structure of the MLP .....	53
Figure 16: The RBF Neural Network Topology .....	55
Figure 17: TDNN Architecture.....	55
Figure 18: Recurrent neural network .....	56
Figure 19: A time-delay neural network.....	60
Figure 20: An Example of a Non-Stationary Time Series.....	73
Figure 21: An Example of a Stationary Time Series .....	74
Figure 22: Research Design of this Study.....	86
Figure 23: Research Process .....	91
Figure 24: Amount of Indices for Different es .....	97
Figure 25: Time Series Plot of Sales Data for Drug 211 .....	108
Figure 26: Time Series Plot of Sales Data for Drug 174 .....	108
Figure 27: Time Series Plot of Sales Data for Drug 24 .....	109
Figure 28: Multiple Time Series Plot of 3 Drugs .....	109
Figure 29: 3D Surface Plot for Drugs 24, 78, 84 .....	110
Figure 30: 3D Surface Plot for Drugs 24, 31, 37 .....	110
Figure 31: 3D Surface Plot for Drugs 24, 174, 86 .....	111
Figure 32: A Part of Output of UCINET for Clique-sets .....	118
Figure 33: Cluster (Tree) Diagram .....	119
Figure 34: A Part of Output of Degree Centrality Measures .....	120
Figure 35: Degree Distribution of the market graph.....	121
Figure 36: Networks of Drugs .....	123
Figure 37: Network of Drugs after Deleting Isolated Nodes and Nodes with Degree 1 .....	123
Figure 38: Network of Clique Indicators .....	125
Figure 39: Time Series Plot of Drug 24 Sales Data.....	130

Figure 40: Autocorrelation Diagram.....	130
Figure 41: Partial Autocorrelation Diagram .....	131
Figure 42: Time series plot of drug 24 sales data after differentiation and log transforming ....	131
Figure 43: Auto-correlation Function of Drug 24 Sales Data.....	133
Figure 44: Partial Auto-correlation Function of Drug 24 Sales Data .....	134
Figure 45: Plot of Forecasting.....	136
Figure 46: Normal Probability Plot of Residuals for ARMA (1, 1, 0) .....	137
Figure 47: Autocorrelation Function of Residuals.....	137
Figure 48: Partial Autocorrelation Function of Residuals .....	138

# **CHAPTER 1**

## **INTRODUCTION**

### ***1. Introduction***

*In this chapter, an introduction of the thesis that offers an overall insight to the research domain is explained. This chapter commences with a background with respect to the research area. Then, it is followed by motivations of the study, a brief presentation of pharmaceutical distribution companies in Iran, and the problem discussion. Then, research objectives, research framework and contributions of the research are presented. At the end of this chapter, the structure of the thesis is demonstrated.*

## **1.1. Background of the Study**

The technologies for generating and gathering data have been progressing promptly. The explosive increase in data and database causes the requirement to offer new technologies and techniques in order to process data into valuable information and knowledge automatically and intelligently (Lee and Siau, 2001). The past decades observed a remarkable growth in information being stored electronically. The quantity of information in the world is anticipated to double every 20 months and the volume and numbers of databases are growing quicker (Frawley et al., 1992). Nowadays large amounts of data are available to companies about their suppliers and customers. As part of their strategy for integrating e-commerce capabilities, numerous organizations are engaged in the development of information systems that will create valuable connections with their suppliers, customers, and other channel partners engaged in distribution, and warehousing activities (Dhond et al., 2000; Shanmugasundaram et al., 2002). In an effort to manage this augment in existing data; also, development of information technologies, the concept of data mining became well known during last few years (Voges and Pope, 2000). Data Mining is the process of exploration and analysis of huge amount of data to make prediction or discover meaningful patterns and rules from large amounts of data (Berry and Linoff, 2004). “Emerging data mining techniques permit the semi-automatic discovery of patterns, models, relationships, associations, changes, anomalies, rules, and statistically significant structures and events in data. Very significant business benefits have been attained through the integration of data mining techniques with current information systems aiding electronic commerce” (Dhond et al., 2000).

Noticeably, the employing of novel technology has resulted in huge customer interaction and transaction databases (Voges and Pope, 2000). For example, McCann and Gallagher, (1990) cited by Voges and Pope, (2000); stated “This data explosion means that the marketing managers have a difficult time capturing the opportunities in the data..... Data must be converted into information by applying marketing and analyzing knowledge.” A main challenge is how to optimally employ the information resources enclosed in these databases to guarantee more effectual marketing decision-making (Voges and Pope, 2000). In fact, the digital revolution has caused digitized information simple to capture, save, process, store, share out, and convey (Fayyad and Uthurusamy, 1996). With noteworthy progress in computing and associated

technologies and their ever-expanding uses in diverse aspects of life, massive amount of data of various characteristics keep being gathered and stored in databases (Mitra and Acharya, 2003). “Data mining is an attempt to make sense of the information explosion embedded in this huge volume of data” (Piatetsky-Shapiro and Frawley, 1991; cited by Mitra and Acharya, 2003). There are various areas where large volumes of data are stored in databases, such as: manufacturing and distribution, business and marketing, finance and investment, telecommunication network, digital library, image archive, scientific domain, health care and pharmaceutics, internet and etc (Mitra and Acharya 2003). Although data mining is the development of a domain with a long history, the term itself was presented recently, in the 1990s. In fact, unprocessed data are hardly ever of direct advantage and its actual value is predicated on the ability to take out information, patterns, and models that are valuable for decision support or discovery and perception of the phenomenon controlling the data source (Mitra and Acharya 2003). In most fields, the analyst processes data manually and, with the help of statistical methods, offer summaries and reports. However, such an approach quickly broke down as the volume of data increased (Mitra and Acharya 2003). When the scale of data processing and exploration surpass human capacities, people necessitate the help of computing technologies for automating the actions. All these have induced the requirement for intelligent data analysis methods, which could find out valuable knowledge from data (Mitra and Acharya 2003). Actually, the term KDD applies to the overall procedure of discovering knowledge from databases (Mitra and Acharya 2003). Although some people deal with data mining as a synonym for KDD, some others treat it as a special phase in Knowledge Discovery (Mitra and Acharya 2003). However, truly, data mining is one phase of the wider process of Knowledge Discovery in Databases (KDD) (Voges and Pope, 2000).

In addition to the successful applications reported above such as such as marketing, finance, banking, manufacturing, distribution, warehousing, security, medicine, multimedia, telecommunications, demand and supply analysis, e-commerce, CRM, investment trend in stocks and real estates; data mining is also available in various forms like text mining, web mining, audio and video data mining, pictorial data mining, relational databases, and social networks data mining (Mitra, Acharya, 2003; Outsourcing Data Entry Services, 2007).. In fact, “data mining is more than just conventional data analysis. It uses traditional analysis tools (like statistics and graphics) plus those associated with artificial intelligence (such as rule induction and neural nets)” (Read, 1999). Actually, “the field of data mining has created an information explosion of

its own, with an ever increasing toolbox of techniques available" (Voges and Pope, 2000). Data mining techniques execute data analysis and may discover important data patterns and relationships, contributing significantly to management and business strategies, and even scientific, pharmaceutical, and medical studies (Hanna, 2004). Many issues of economic, intellectual, scientific, and business interest can be addressed in terms of the following data mining tasks (Berry and Linoff, 2004): Classification, Estimation, Prediction, Associating rules, Clustering, Description and Profiling which will be explained in chapter two. Some of the most significant applications of data mining have been in the field of sales and marketing (Witten and Frank, 2005). These are areas in which companies own immense volumes of recorded data and these data have recently been recognized potentially especially valuable. In these applications, predictions themselves are the key interest (Witten and Frank, 2005). Since sales prediction must be performed with high accuracy and in a short time, it is impossible to do it with manual or traditional techniques. Hence, it is really desired to apply one of the data mining techniques to enhance the accuracy of sales prediction also speed up its process. As stated, the objective of data mining is to generate novel knowledge that the user can act accordingly. "It does this by building a model of the real world based on data collected from a variety of sources, which may include corporate transactions, customer histories and demographic information, process control data, and relevant external databases" (Two Crows Corporation, 1999).

Accordingly, in this research we focus on a special application of data mining techniques in prediction that is use of neural networks in sales prediction. As a matter of fact, prediction of the future's conditions and events is called forecast and the methods and procedures of this act is called forecasting (Bowerman and O'Connell, 1986). Since forecast of future's events acts an important role in decision making process, forecasting is really essential for most organizations and companies (Bowerman and O'Connell, 1986). Each organization has to be able to forecast in order to make intelligent decisions; especially commercial companies in all their procedures need forecasting of future events, sales, and conditions (Bowerman and O'Connell, 1986). "Sales forecasting is a valuable management tool for determining the future magnitudes, timing, and possible effects of uncontrollable influences over a company's future success. It is an estimate of the expected demand for a company's products, and it is an indispensable element of management planning for many major company activities, e.g. marketing, production, finance, research and development, personnel, capital investment determination, purchasing, inventory

and warehousing" (Yip et al., 1997). Instinctively, forecasting models can provide reasonable estimates by using historical data. Consequently, if the marketing department can approximate the sales volume for the following period, the purchasing department can subsequently efficiently manage the inventory to reach just-in-time (JIT) (Kuo, et al., 2002). In fact, precise sales prediction is used in catching the trade-off between customer demand satisfaction and inventory costs (Gupta, et al., 2000). Accurate sales or demand prediction helps to coordinate the replenishment of orders of items while keeping total costs low (transportation and holding costs) and achieving a certain service level at the distribution channel, but breakdown to justify considerable product demand fluctuations may either cause extremely high costs relating to high inventory charges or unfulfilled customer demand and loss of market share (Gupta, et al., 2000). A competent forecasting can lessen inventories, reach greater adaptability to alterations and augment profits. In particular, sales forecasting is very critical as its result is employed by a variety of functions in a distribution company (Mentzer and Bienstock, 1998; cited by Doganis et al., 2006). According to the tough competition that exists nowadays, most companies are in a nonstop try to escalating their profits and decreasing their costs (Doganis et al., 2006). Reliable sales forecasting is certainly an essential and inexpensive way for each organization to meet the abovementioned objectives, since this causes to superior customer service, decreased missing sales and product returns (Doganis et al., 2006). Additionally, Carboneau et al. (2008) proposed that if an increase in forecasting precision can be attained, it would lead to lesser costs due to decreased inventory level as well as higher customer satisfaction that will be according to an enhance in on-time deliveries. Particularly, for the pharmaceutical industry, effective sales forecasting systems can be extremely valuable, due to the short shelf-life of many pharmaceutical products and the importance of the product quality which is closely related to human health (Doganis et al., 2006). In other words, pharmaceutical distribution companies like food companies are further involved in sales forecasting because of their particular attributes, like short shelf-life of their products, the necessity to keep high product quality, and the vacillations and uncertainties in consumer demands (Doganis et al., 2006). Usually, pharmaceutical goods can just be sold for a short period, and both lack and excess of products can cause loss of income for the organization (Doganis et al., 2006). That is why we have chosen pharmaceutical distribution companies to make sales prediction for them. In general, "the variations in consumer demand are caused by factors like price, promotions, changing consumer

preferences or weather changes" (Van der Vorst, et al., 1998; cited by Doganis et al., 2006). However, in pharmaceutical industry (in Iran) distribution companies are not affected by factors such as price, advertisement and changing consumer preferences as price of drugs is determined by government, advertisement is not allowed in this industry, and people cannot have any preference on drugs with different brand names. With the access to past pharmaceutical sales data, distribution centers can make prediction for the future sale to plan for matching with the proper manufacturers and targeting the right regions, retaining right amount of inventories and delivering enough amounts of drugs to different regions according to the people's needs.

Evidently, developing effective techniques for forecasting future sales and customer demands then stock inventories accordingly by using information about past sales would be the finest approach to manage the inventory level and reduced the related costs (Dhond et al., 2000). The methods that have been applied for sales prediction are usually time series forecasting that can be devided to linear and nonlinear, subject to the nature of the model they are derived from (Doganis et al., 2006). Linear models, such as autoregressive moving average (ARMA) (Box and Jenkins, 1976) and autoregressive integrated moving average (ARIMA) (Box et al., 1994) are the most famous methodologies, but their prediction ability is restricted by their supposition of a linear behavior; accordingly, it is not always acceptable (Zhang, 2003; cited by Doganis et al., 2006); in that, if the data are influenced by nonlinear behavior, they are inaccurate and artificial neural networks (ANNs) are better than the conventional statistical methods (Chakraborty et al., 1992; Kumar et al., 1995; Agrawal and Schorling, 1996; Haykin, 1999).

A precise forecasting of future sales by one of the related data mining tools, such as neural networks could lead to considerable savings in inventory expenses (Shanmugasundaram et al., 2002). Neural networks are versatile tools because of their power to generalize trends, their ability to find nonlinear relationships, to store appropriate information about past sales, and to make reliable prediction (Shanmugasundaram et al., 2002). In fact, neural networks are mathematical methods that find out pattern from data. They have confirmed very efficient to make prediction by dealing with non-linear input and output variables, being able to estimate any function in definite circumstances (Hill et al., 1996). Accordingly, in this research we are going to use neural networks to predict future sales for a pharmaceutical distribution company.

## **1.2. Motivation of the Study**

As a matter of fact, choosing this research area has got different reasons:

Firstly, we have chosen pharmaceutical industry as it is a valuable field of study and research. According to Manchanda et al. (2004), and our studies and interviews with experts, we realized that pharmaceutical industry has got many features such as:

- 1) It is one of the most important industries as it directly affiliated to the people's health or death
- 2) It has large inventory costs;
- 3) It receives high quality of data from industry, but still focuses on traditional and heuristic methods;
- 4) It really needs new and improved sales forecasting methods;
- 5) It represents a high percentage of growth every year;
- 6) It is a unique research opportunities for physicians, patients, insurance firms, pharmacists, manufacturers, distributors and wholesalers, marketing and sales managers, and finally for the ministry of health.

Next, pharmaceutical sales forecasting in Iran have not been based on mining the massive amount of data that is available though data mining has considerable value in today's extremely aggressive business environment. As we told in the previous section, the increasing amount of data results to generate and use new tools in order to modify the data to valuable information and knowledge. In Iran, distribution companies started to save their sales data few years ago, so related organizations should take advantage of these growing piles of data by converting them to the worthwhile knowledge. Although mountains of pharmaceutical data have been stored in different companies and organizations (especially distribution companies); still nobody takes advantage of these piles of databases in an effective way. Except for some basic graphs or diagrams and some statistical reports, no one extracts novel and useful knowledge like meaningful rules or patterns from this growing pharmaceutical data in our country. Thus, these companies must improve their ability to collect, extract, analyze and prepare comprehensive reports..

And last but not least, according to the opinion of Dhond et al. (2000), usually, large distribution organizations, particularly geographically separate companies such as

pharmaceutical distributors, are forced to hold large inventories of goods prepared to distribute according to customer demand (Dhond et al., 2000). Therefore, it is crucial for these organizations to have a lot of attention on controlling and managing their inventory levels. If inadequate amount of inventory is kept compared with demand, displeased customers would switch to competitors (Dhond et al., 2000). On the other hand, various types of unnecessary costs are incurred for holding too much inventory, as inventories need a lot of money to keep (Dhond et al., 2000). In addition, most drugs have short expiration periods and shelf lives; therefore, they must be replaced periodically (Dhond et al., 2000). The best way to manage above mentioned problems is to make accurate sales prediction for these companies. Accordingly, in this research we will mine pharmaceutical sales data in order to predict the future sales because these distribution companies are usually forced to carry large inventories of products ready to distribute to hospitals and drug stores. Our research defines the different sales forecasting methods, chooses the most suitable one (which is neural networks) for our goals, describes the process of constructing and choosing an appropriate neural network, and highlights solutions related to mining large quantities of data.

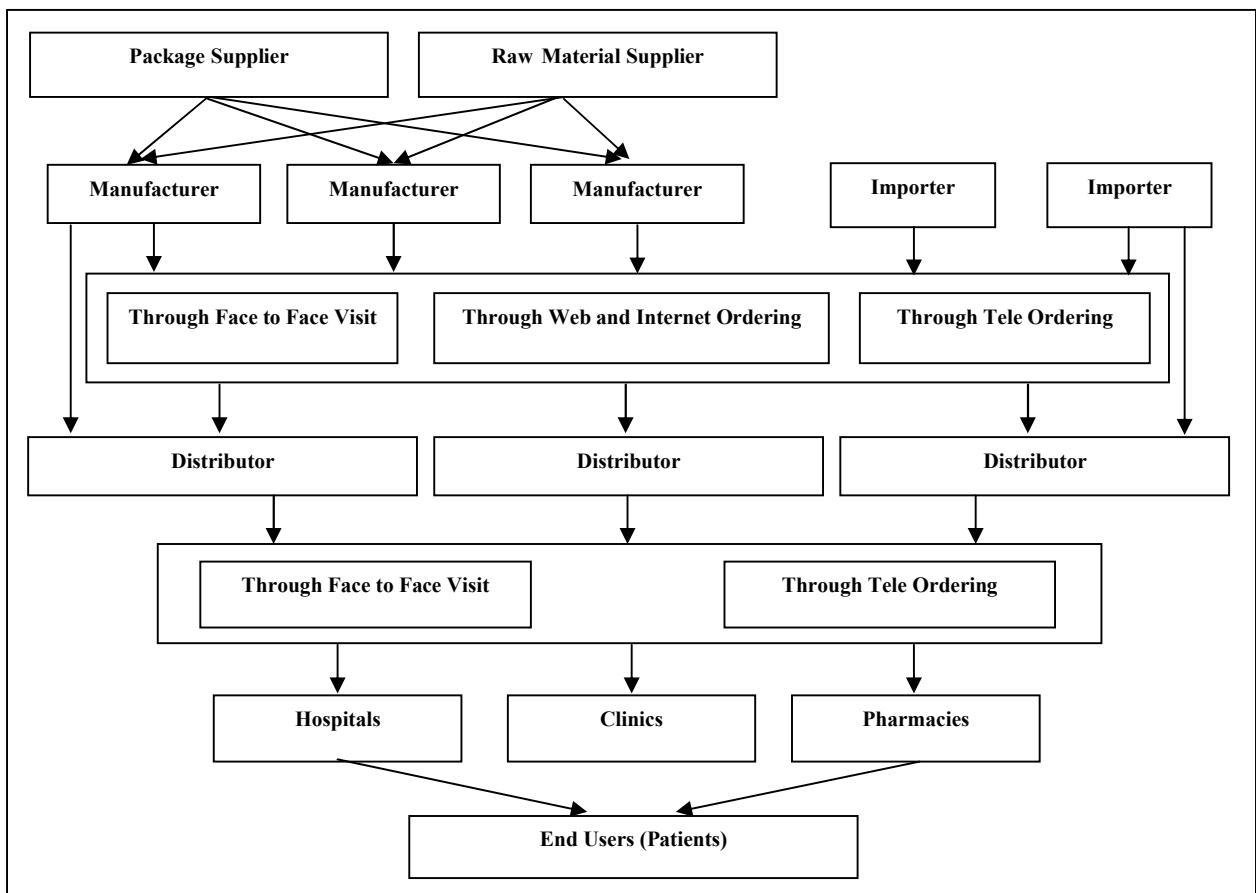
### **1.3. Pharmaceutical Distribution Companies in Iran**

Distribution of pharmaceutical products in Iran is mainly by independent wholesalers, who buy stock for their own account from manufacturers and sell to their customers (hospitals and pharmacies). Distribution and sales of pharmaceuticals is performed either by both full-line and short-line wholesalers. Full-line wholesalers deliver the full collection of available pharmaceuticals and they have some central depots in big provinces, and short-line wholesalers transact just in a selection of pharmaceuticals. In Figure 1, a schematic of pharmaceutical distribution channel is presented. Although many distribution companies and even manufacturers wish to have web and internet ordering, this type of ordering is not current in Iran now. Most orders are made through face to face visits, but in some special cases Tele-ordering may also be carried out. Wholesalers handle their costs and make financial gains based on the difference between the prices at which they buy from the producers and that at which they offer to the retailers. The margin of 10% for distribution companies has been decided by Ministry of Health. Significant changes have taken place in the distribution stage of the pharmaceutical supply chain in Iran during the past decades. In the past, each province has the specific and pre-determined

quota of the total drugs that are produced in the country. For example, Tehran had the 40% percent quota of the total drugs because it had the most population, the most hospitals, and the most physicians among other provinces so it received high percentage of produced drugs. However, nowadays as the production of drugs is unrestrained and there is not any limitation for the volume or extent of producing drugs any more, each province has the will-power of ordering as many as drugs that are needed in that province. In the past, pharmaceutical distributors had to maintain needs of their customers 6 months in advance; it means that their inventory turns lasted 6 months. This fact caused them a lot of undesirable costs and investment in inventory control and warehousing. Today, most pharmaceutical distribution companies can decrease their stock time to 45-60 days, which is still not desirable. In Iran, pharmaceutical distribution companies usually give order to their suppliers (manufactures) and take order from hospitals and drug stores who are their main customers. Each year distributors arrange a meeting with their suppliers to plan and predict their sales amount for next year. Thus, according to their prediction, manufacturers plan for the production of required products, and distributors plan for buying drugs and selling them. Distribution companies also revise their predictions every three months to see whether they have faced excess or shortage of each drug. In other words, they first face the problem and then revise and solve it. However, precise and regular monthly sales prediction would help both distributors and manufacturers in their planning, decision-makings and even cost savings. In addition, accurate and valid sales prediction is truly essential for distributors because of the fact that shortage of drugs may cause them to lose both their suppliers and customers. Furthermore, distributors are obliged to buy all products that they have ordered before, so buying and keeping surplus inventories may cause them many undesired costs, since pharmaceutical drugs have limited shelf lives and expired ones should be thrown away. Consequently, trusty and accurate sales prediction is indispensable for pharmaceutical distributors and it would avoid above-mentioned problems.

Usually there are two kinds of pharmaceutical distributors in Iran which are huge and small distributors, and there is not any medium size distributor. Pakhsh Hejrat Co., from which we collected the sales data, is one of the main and huge size pharmaceutical distribution companies in Iran. After receiving the orders, PakhshHejrat like other pharmaceutical distributors is committed to supply the required drugs to provinces within 24 hours, to cities within 48 hours, and to remote areas within 72 hours. This matter forces them to be always ready

to respond to their customers' needs and always have a huge amount of inventory. Pakhsh Hejrat and most distributors in Iran usually keep inventories for the needs of next 1.5 - 2 months in advance. In comparison with 15 - 20 days of stock time in many developed countries like Japan, 1.5 - 2 months is disappointing! This fact causes many excessive costs and investments for Iranian pharmaceutical distributors. Inventory control, transportation and financial costs; contain a high percentage of total expenses in pharmaceutical distribution companies. For instance, distributors buy products from manufacturers and pay for them, but they sell products and receive the related money gradually. Thus, this gap causes them undesired expenses. Monthly and precise sale prediction would shorten or even eliminate this gap.



**Figure 1: Pharmaceutical Distribution Channel**

## **1.4. Problem Discussion**

Pharmaceutical distribution companies in Iran are facing several challenges, including increased competition, tough regulations that limit advertising, huge amount of inventory as they are forced to be ready to meet their customers. In response, distribution companies are looking for new ways to build market share, decrease their amount of inventories, while retain their valuable consumers, and reducing their costs. In pharmaceutical distribution companies, both lack and surplus of goods can cause loss of income for the organization. Accordingly, one of the problems in pharmaceutical distribution organizations that deal with public health and medical products is that how much quantity of each drug should be kept in the inventory. It is obvious that medical products are sensitive, so should be kept and transported in special conditions. In addition, pharmaceutical drugs have limited shelf lives. Therefore, these companies incur significant financial costs if they carry excessive inventory levels. Consider a large distribution company (Pakhsh Hejrat Co.) that dispenses pharmaceutical drugs to customers in a number of provinces in Iran. Just like any other distribution companies in its situation, it is compelled to hold a huge amount of goods (inventories) ready to send on customer demand (Dhond et al., 2000). The company will have considerable financial costs if it carries surplus quantities of drugs. Furthermore, “pharmaceutical drugs have a short expiration date and must be renewed periodically” (Dhond et al., 2000). On the other hand, the shortage of products may cause unsatisfied customers to shift company loyalties and turn to competing companies and this process gradually makes the company lose potential profits in such cases (Bansal et al., 1998). Therefore, this company has to find a way to keep their inventory levels as small as possible while retain their customers satisfied and be equipped to answer their needs. This can be done only by mining the past sales data in order to make prediction for the future sales. Accurate sales prediction would help distributors to know how much inventory they should keep in order that it is important not to buy too much products from manufacturers and not to buy insufficient amount of drugs since:

- 1) If they buy far more than the needs of customers, they have to invest capital a lot and they will incur excessive costs. They may invest these slept capitals in other parts and increase their profits indirectly.

- 2) If they buy less than the needs of customers, they will lose their customers and potential profits.

Subsequently, in this research we are going to apply one of the data mining techniques that are used for prediction to estimate the future sales of drugs. We will do it by using past sales data in order to help improve sales and inventory management, also decision making at pharmaceutical distribution companies. We will discuss the use of neural network based data mining and knowledge discovery techniques to make sales prediction in a large pharmaceutical distribution company.

#### **1.4.1. Problem Definition**

As we mentioned before, pharmaceutical distribution companies are facing new competition situations and they should survive themselves in these tough conditions. They have to make their customers satisfied and deliver the right amount of drugs to the right place and at the right time. Accordingly, in keeping with its market-leading position, pharmaceutical distribution companies are compelled to keep large standing inventory of medicines ready to distribute on customer demand as shortage of drugs is not acceptable in this industry. However, keeping too much inventory would cost a lot. Consequently, they need to know how much of each drug should be kept as an inventory in a specific period. However, in our country there is not any logical or an effective way of good prediction in order to help distributors manage the drug inventories more properly. Thus, most pharmaceutical distributors have problem in deciding how much quantity of each drug should be kept in the inventory at each warehouse and at a specific time.

After talking with some managers and experts of pharmaceutical distribution companies, we perceived that pharmaceutical distribution companies in Iran are using traditional statistical techniques to make sales prediction and determine inventory levels for each drug item. Usually, they make prediction for next 3 months that is a long period for prediction. Although this system seems simplistic, it neither insures customer happiness and loyalty nor does it optimize inventory stockpiles. The only visible benefit to this system is its ease of management. Nevertheless, in a competitive area such as what the distribution companies are facing to, it is so essential to have a strong model in order to predict the future sales or needs of market in shorter time than 3 months.

In fact, short-interval forecasting can be further practical than long-interval predictions as short-interval predictions let inspector observe sales demand with superior precision than long-interval forecasting (Bansal et al., 1998). However, short-interval forecasting can be extremely complex. The correctness of this prediction is also so important because having the excessive amount of drugs in their inventories or warehouses may cost a lot; also, shortage of drugs is not acceptable in this industry due to the seriousness of their products that directly related to the health or death of the people. Usually, some of these distribution companies in Iran use heuristic methods for inventory management and sales prediction that are not scientific at all, since the managers just use reasoning and their experience. As an illustration, we talked to senior managers of some pharmaceutical distribution companies, we understood that as their next year's sales prediction, some of them just multiply their last years' sales to a percentage (which is usually determined by their suppliers) and add it to last years' sales figures. Some other companies use traditional statistic techniques and they need to make a lot of simplicities for relations such as assuming linear relations and correlations for all situations which are not logical and exact at all.

From pharmaceutical distribution companies, we have selected Pakhsh Hejrat as it is one of the main and leading distributors, its products are mostly produced in Iran, and we had access to its sales data. In Pakhsh Hejrat Co., the common sales prediction process is as follows: First, they inspect the amount of sales in each month of last year. Then, they select the month that had the highest amount of sales; they multiply the obtained figure by 12 and consider it as total sales of next year. They also combine managers' experience with above-mentioned results and consider it as their final sales prediction for next year. After each three months, they revise their prediction and announce it to their suppliers; then order and buy medical products accordingly. It is obvious that this technique is not precise at all and it would result in unacceptable errors. Therefore, they really need a reliable and approved sale prediction method. If they buy surplus products and do not sell them, they would undertake a huge loss. On the other hand, because of increasing financial expenses, Pakhsh Hejrat Co. like other distributors wishes to order optimal quantity of products, keep the inventory in the lowest possible level, and to be just in time. However, manufacturers want to have predefined planning for their production. Thus, providing these co-ordinations is really a challenge for them. We have interviewed experts of six main and pioneer pharmaceutical distribution companies in Tehran. All of them mentioned that these companies extremely need a precise and methodical sale prediction because of the fact that they

really have to keep huge amount of inventory at least 45 days in advance and they incur various undesired financial costs. They believe that inventory control and warehousing include more than 50% of total expenses and with an accurate and reliable sales prediction methods they can reduce their inventory turns from 45-60 days to 15-20 days (like Japan) and accordingly reduce the related costs more than 50%. Hence, by applying a vigorous sales forecasting methodology these companies can reduce their cost at least 50%. Consequently, an exact sales prediction method is the best solution and it would considerably reduce financial expenses. Actually, data mining tools and especially neural networks are intelligent tools that can overcome these problems. Therefore, evolving better methods for predicting future sales then stock inventories appropriately by using data about past sales would be the finest approach to manage the inventory level and reduce related costs (Dhond et al., 2000). An accurate prediction of future sales by one of the vigorous and famous data mining tools which is artificial neural networks could result in considerable savings in inventory costs.

Actually, Pakhsh Hejrat Distribution Co. like other pharmaceutical distributors encounters with different challenges. Firstly, they are committed to meet their customers' needs in 24 to 48 hours and if they could not deliver the right products at the exact time and to the exact place, they would incur a loss. Hence, they have to keep enough amount of inventory not to face shortage of drugs. On the other hand, warehousing and inventory control for these special products are really costly. Thus, they are obliged to carry reasonable and optimized amount of medical products in their warehouses. Accordingly, it is essential for them to predict their future sales in order to keep the right amount of product as their inventories. Secondly, Pakhsh Hejrat Distribution Company like other pharmaceutical distribution companies started to save and computerized their sales records 3-4 years ago. Therefore, still they do not have enough and rich sales databases and we cannot make use of traditional methods like ARIMA that needs at least 50 past records. As a result, we came to the idea of offering a new and effective sales prediction methodology to help them overcome the mentioned problems. In this research, we are going to perform drugs' network based analysis, group drugs to find group members, which have similarity in their sales behavior, and then apply their sales records in our sales prediction with neural networks. In chapter 4, we will see all steps of our network analysis, grouping medical products, and our thorough prediction analysis in detail.

## **1.5. Research Objectives and Question**

The targets of Pakhsh Hejrat Distribution Company like other distributors are to minimize overall cost of keeping inventory and achieve the greatest level of customer satisfaction. To achieve above-mentioned goals, Pakhsh Hejrat company has to control inventory levels by means of precise sales prediction in order to prevent costs of excessive inventory also prevent losing their customers because of drug shortage. Under these particular conditions, it would certainly be valuable for Pakhsh Hejrat Distribution Company to be prepared with an precise sales forecasting model (Doganis et al., 2006). In this way, the volume of the inventory can be optimized regarding changing demands. Accordingly, the objective of this research is concerned with the development of time series sales forecasting models for pharmaceutical products in this company. Thus, in this research we are going to apply data mining techniques that are used for prediction to estimate the future sales of drugs using past sales data to advance sales and inventory management also decision making at pharmaceutical distribution companies. Predictions obtained by applying data mining methods such as neural networks, help the company to set its future policy so that while the company keeps the customers in touch with the company, does not face the financial problem of keeping extra products. Hence, the overall purpose of the study can be summarized in the following question:

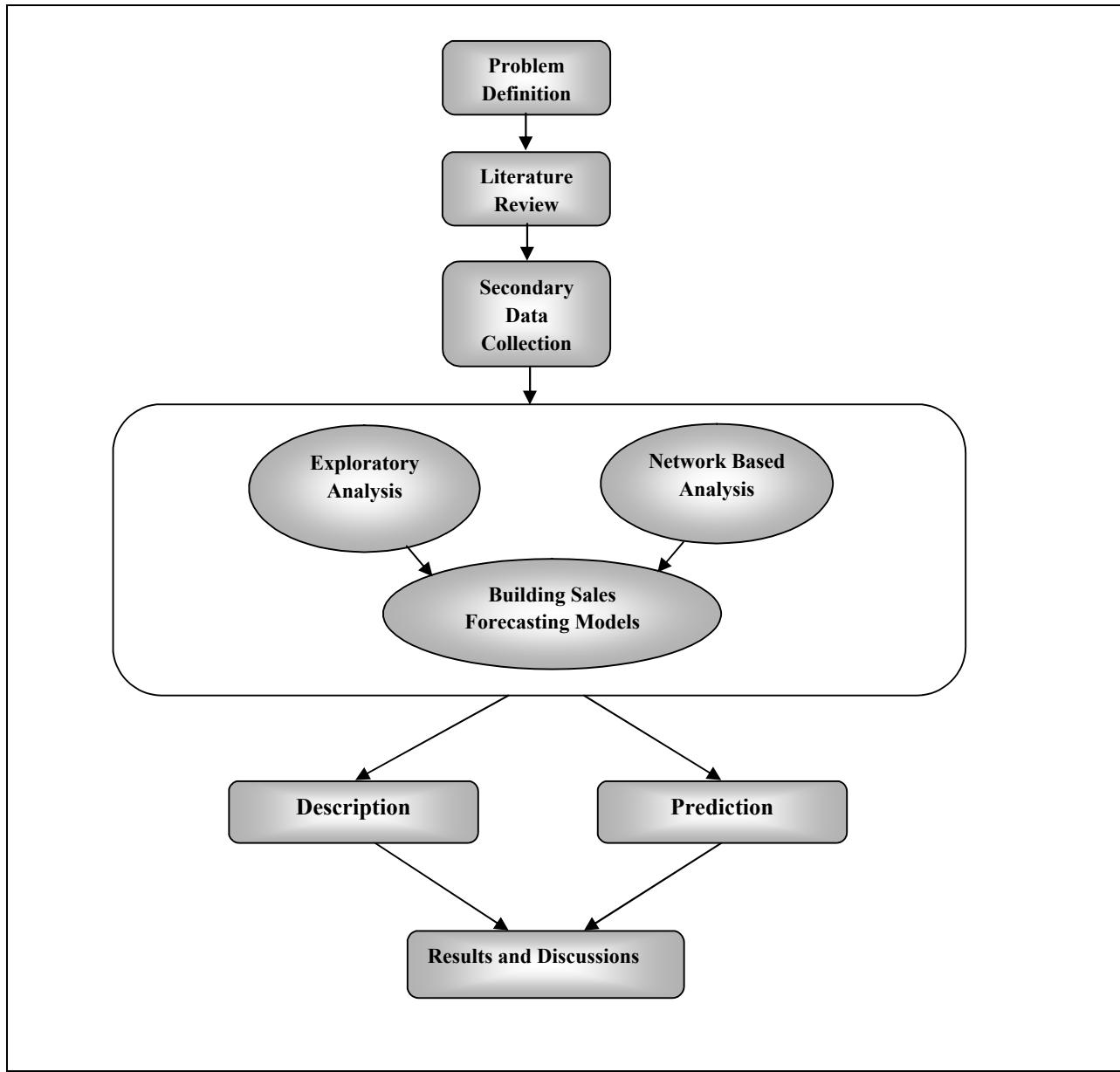
- How sales data and data mining techniques can be used for sales prediction at a pharmaceutical distribution company? In other words, how can we use past sales data of a pharmaceutical distribution company and modern tools, like data mining, in order to predict the future sales of that company?

To find the best solution to the mentioned problems, the sales data are collected from Pakhsh Hejrat pharmaceutical Distribution Company located in Tehran. The data warehouse of Pakhsh Hejrat Co. contains all sales information from 2004 (22th of December) to 2007 (21th of December), which is recorded monthly, so the total number of records are 36 months. These data sets form time series of sales information of the company. However, 36 months sales records are not enough for our goals. Therefore, we have to group drugs to make use of group members' sales information as well. By using these amounts of data, grouping medical products to find group members, and applying neural network based data mining methods for predicting

the future sales, the inventory level problem will be solved and meanwhile customers will not be unsatisfied. We will also examine traditional ARIMA methodology to compare its weak results with our proposed methodology. Consequently, according to instruction and speech of Albadvi, 2007, the goals of our research are: 1) Description, since we attempt to depict the reality in our exploratory analysis and network identification phases in order to understand the nature of our data, also to discover groups of drugs and group members, and 2) Prediction, since we focus on prediction of the future sales for Pakhsh Hejrat Company.

## **1.6. Research Framework**

In this study, after defining the research problem and question, comprehensive literature review with respect to research domain has been carried out. Then, An exploratory analysis, network based analysis, and prediction model building have been conducted. Actually, the goal of both exploratory analysis and network based analysis is description since these phases help us to discover facts and describe realities (we will see the details of these stages in chapter 4). Obviously, the goal of next phase, building sales forecasting model, is prediction. Finally, the results of each step (especially sales prediction phase) will be discussed. Figure 3 is a schematic representation of our research framework.



**Figure 2: A Schematic Representation of the Research Framework**

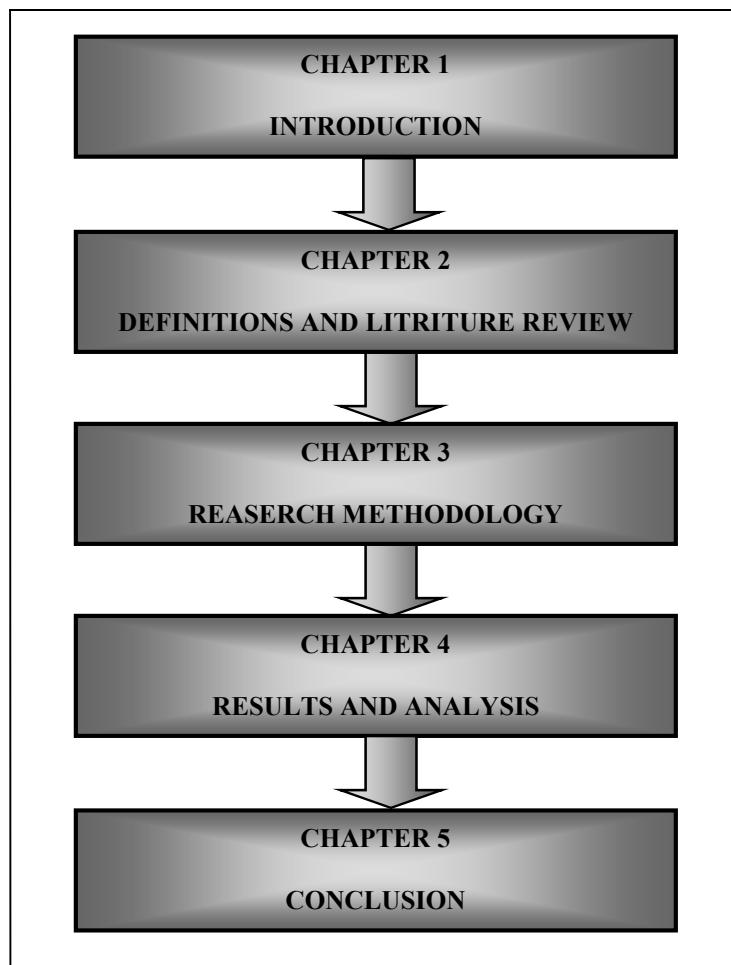
## 1.7. Contributions of the Research

This research has both empirical and methodological contributions. It has empirical contribution since we have define our problem according to the real and existing challenges of pharmaceutical distribution companies in Iran; we have gathered real data of a real company (Pakhsh Hejrat Co.) to achieve our managerial and commercial objectives. This research also has methodological contribution due to the fact that we have introduced a new method of grouping

different drugs in order to make use of co-members' past sales data for each other in our sales prediction (model building). We will see the contributions of this thesis in more detail in chapter 5.

## **1.8. Outline of the Thesis**

This thesis consists of five chapters. In the first chapter, the background of the selected research domain is presented. This chapter is followed by motivation of study, problem discussion, research objectives and question, research framework, and it is ended with the contributions of the research. The literature review in chapter two will give the reader an overall review on definitions, theories, methods relevant to the research area, and surveys related to the topic. In chapter three, research methodology, the design and process of this research will be discussed. Moreover, all the methods and techniques used for each step of network analysis and sales prediction will be mentioned briefly in this chapter. Result obtained by each step of network analysis and sales prediction such as, data preprocessing, exploratory analysis, statistical analysis of drugs' networks, prediction results, and model evaluation will be presented in detail in chapter four which is actually the most important chapter of this thesis. Conclusion, contributions, managerial implications and limitations of the study; also, suggestions for further research will be discussed in chapter five. Figure 4 summarizes the outline of this research.



**Figure 3: Outline of the Thesis**

# **CHAPTER 2**

## **DEFINITIONS AND LITERATURE REVIEW**

### ***2. Definitions and Literature Review***

*The objective of this chapter is to offer explanations and literature review concerning the research domain. It will offer definitions and literature review on data mining and knowledge discovery, neural networks, important issues in a supply chain, definitions and applications of forecasting and more specifically sales forecasting, and finally time series forecasting models and comparisons of them.*

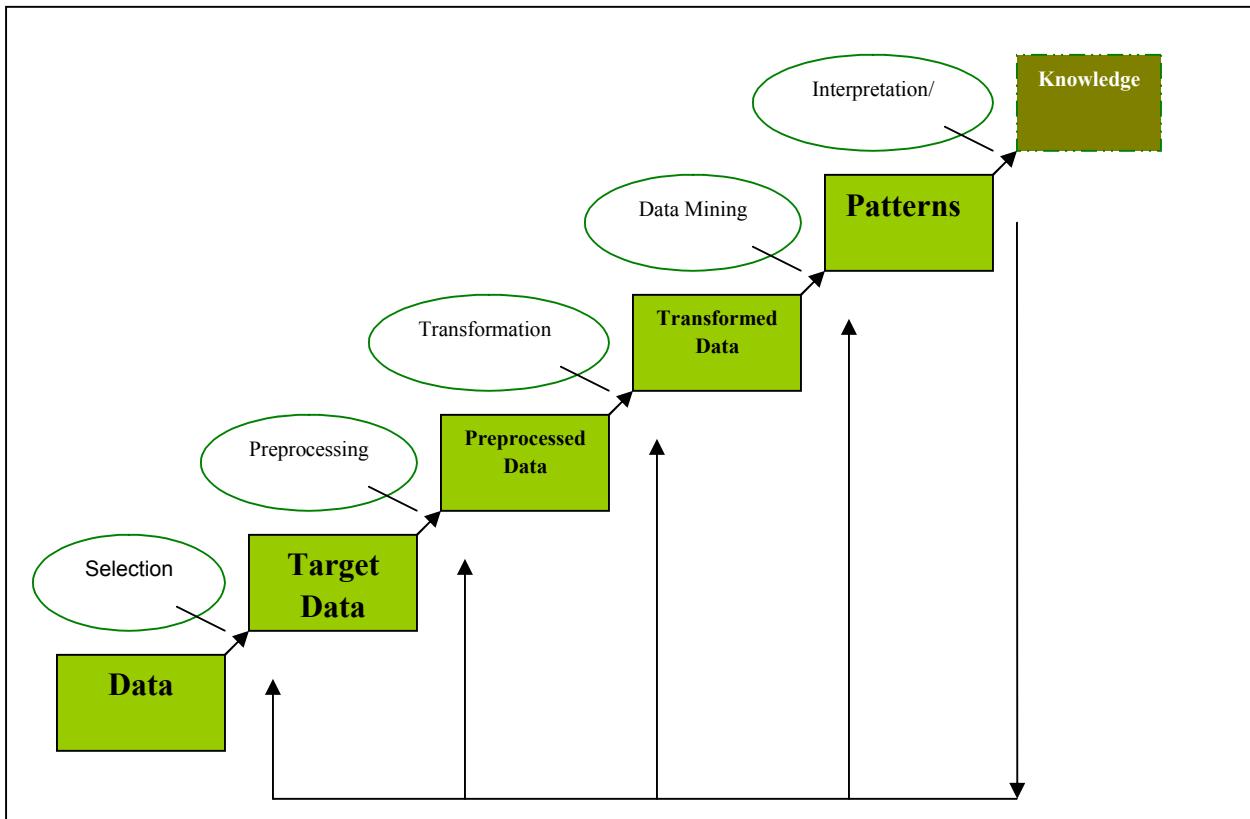
## **2.1. Data Mining and Knowledge Discovery**

### **2.1.1. An Overview of Data Mining**

“Mountainous amounts of data records are now available in science, business, industry and many other areas” (Fu, 1997). To recognize, analyze and ultimately employ this data, a multidisciplinary technique called data mining has been suggested (Fu, 1997). Actually, “the need to understand large, complex, information-rich data sets is common virtually to all fields of business, science, and engineering. In the business world, corporate and customer data are becoming recognized as a strategic asset” (Kantardzic, 2003). The whole procedure of employing a computer-based method, together with novel techniques, for finding out knowledge and information out of data is named data mining (Kantardzic, 2003). “Data mining is most useful in an exploratory analysis scenario in which there are no predetermined notions about what will constitute an "interesting" outcome” (Kantardzic, 2003). In business context, data mining is the process of identifying features, relationships, interesting patterns or models from large databases in order to conduct your business better (Fu, 1997; SPSS White Paper, 1999).

As it is shown in Figure 5, data mining is the core and main part of the Knowledge Discovery in Database (KDD) (Fayyad et al., 1996a). The KDD procedure usually contain the subsequent steps: Data collection, data cleaning, data transformation, pattern exploring (data mining), result interpretation, evaluation and using discovered knowledge (Fayyad et al., 1996; cited by Fu, 1997). “Many vendors, consultants and analysts make data mining appear complex, difficult, vendors and expensive. It may sometimes be complex (involving many parts), but it need not be mysterious or difficult” (SPSS White Paper, 1999). “Data Mining simply means: Finding patterns in your data which you can use to better conduct your business” (SPSS White Paper, 1999).

“Data mining is more than just conventional data analysis” (Read, 1999). It employs traditional analysis means and those connected to artificial intelligence (Read, 1999). Data mining is a unique outlook or approach to data analysis. The goal is more to provide questions rather than responses (Read, 1999). Comings achieved by data mining can subsequently be validated by conventional analysis (Read, 1999).



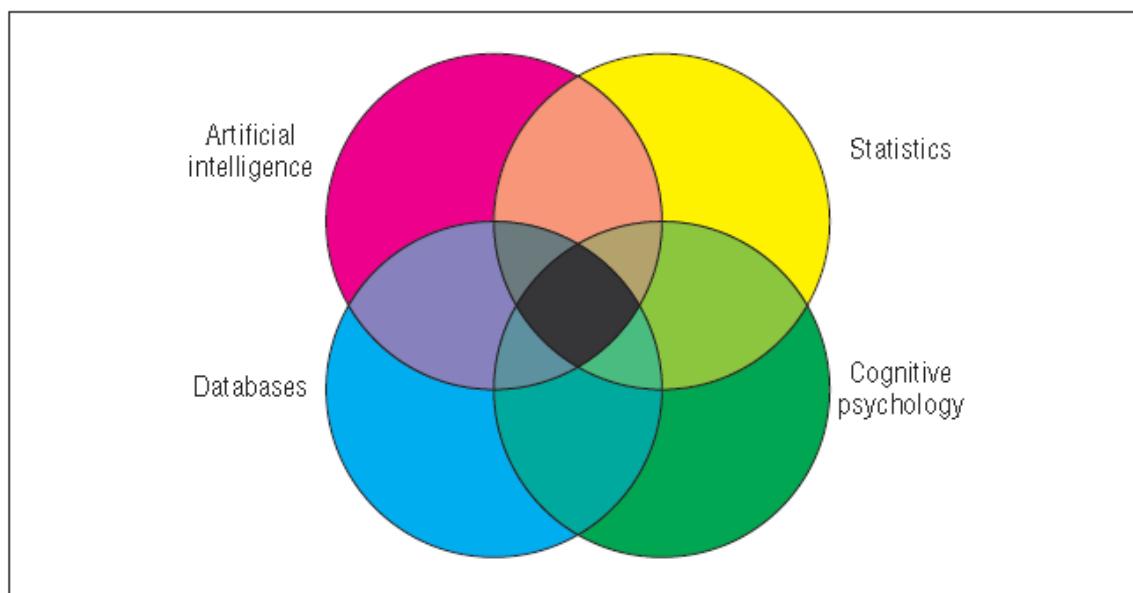
**Figure 4: A Typical knowledge Discovery Process**  
Source: (Fayyad et al., 1996a- 1996b)

### 2.1.2. Difference between knowledge Discovery and Data Mining

Both data mining and knowledge discovery are the techniques to obtain some attributes of an entity from a set of data on an individual component to each of which characteristics/features have previously been acquired by observation (Ohsuga, 2005). The term “Knowledge Discovery in Databases” refers to the whole process of discovering valuable knowledge from data (Fayyad et al., 1996b). Knowledge discovery in databases is the process of recognizing suitable, novel, potentially practical, and eventually comprehensible patterns/models in data (Fayyad et al., 1996a). Data Mining is the fundamental element in the more common procedure of Knowledge Discovery in Databases (Read, 1999).

### **2.1.3. Descriptions of the Knowledge Discovery and Data Mining in Literatures**

- “Data mining is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data” (Fayyad et al., 1996a).
- “Data mining is an interdisciplinary field bringing together techniques from machine learning, pattern recognition, statistics, databases, and visualization to address the issue of information extraction from large databases” (Cabena et al., 1998).
- “Data mining, also known as knowledge discovery in databases, gives organizations the tools to sift through these vast data stores to find the trends, patterns, and correlations that can guide strategic decision making” (Ganti et al., 1999).
- “KDD is the first practical step towards realizing information as a production factor” (Adriaans and Zantinge 1996; cited by Voges and Pope, 2000).
- Pazzani in 2000 stated that “the knowledge discovery and data mining (KDD) field draws on findings from statistics, databases, and artificial intelligence to construct tools that let users gain insight from massive data sets” (Figure 6).



**Figure 5: Combining cognitive psychology with artificial intelligence, databases, and statistics**  
Source: (Pazzani, 2000)

- “Data mining is the exploration and analysis of large quantities of data in order to discover meaningful patterns and rules (Berry and Linoff, 2004).

- “The field of data mining is dedicated to the analysis of data to find underlying connections and the discovery of new patterns” (Vincenti et al., 2005).
- “Data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories (Han and Kamber, 2006).
- “Simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data” (Han and Kamber, 2006).

#### **2.1.4. The Process of Data Mining and Knowledge Discovery**

In fact, KDD is an iterative and interactive procedure involving various phases (Fayyad et al., 1996b) which were shown before in Figure 5 and are explained in detail below:

##### **1- Problem Identification and Understanding the Application Domain:**

The primary step is to comprehend the application area. This step is obviously a precondition for extracting valuable knowledge and for selecting suitable data mining techniques in the third step according to the application objective and the nature of data (Ho, 2002).

##### **2- Selection and Preprocessing Data:**

“The second step is to collect and preprocess the data” (Ho, 2002). Data preprocessing is required to advance the quality of the real data for mining (Mitra and Acharya, 2003; Han and Kamber, 2006). The data we want to analyze by data mining methods are incomplete, noisy, and inconsistent; thus, there is a need for data preprocessing (Han and Kamber, 2006). The basic steps of data preparing are described below (Mitra and Acharya, 2003; Han and Kamber, 2006):

- **Data cleaning:** It consists of some fundamental processes, such as identifying outliers, correct inconsistencies in the data, normalization, noise deletion, handling of missing data, reduction of redundancy, etc (Mitra and Acharya, 2003; Han and Kamber, 2006). Real-world data are regularly incorrect, imperfect, and inconsistent, possibly owing to system

performance flaws, human or operational fault. Low-quality data requires to be cleaned before data mining (Mitra and Acharya, 2003).

- **Data integration and transformation:** This process contains integrating several, heterogeneous data bases produced from diverse sources (Mitra and Acharya, 2003). Moreover, the data may need to be transformed into forms suitable for mining (Han and Kamber, 2006).
- **Data reduction:** Techniques can be applied to acquire a reduced representation of the volume of data set; however, keeps the integrity of the original data. That is, mining on the reduced data set ought to be more efficient yet create the same or roughly the same analytical outcomes (Han and Kamber, 2006).
- **Data discretization:** Techniques can be applied to decrease the number of values for a certain continuous feature by separating the series of the characteristic into intervals. Then, interval labels can be applied to replace real data values (Han and Kamber, 2006).

### **3- Data Mining:**

The third step is data mining that extracts hidden patterns and useful knowledge. This is a crucial procedure where intelligent techniques are used to extract knowledge and patterns from data (Han and Kamber, 2006).

### **4- Result Interpretation and Evaluation:**

This step includes interpreting the achieved knowledge, particularly the interpretation in terms of description or prediction, which are two main objectives of performing discovery systems (Ho, 2002).

### **5- Applying Discovered Knowledge:**

Putting the outcomes into practical use is definitely the final objective of knowledge discovery (Ho, 2002). The information or knowledge obtained by data mining techniques can be applied later to enlighten existing or historical trends or facts, predict the future, and assist decision-makers make strategy from the extracted facts and information.

## **2.1.5. Data Mining Machine Learning and Statistics**

Data mining does not substitute conventional statistical methods. It is an extension of statistical techniques that is partly the outcome of a main alteration in the statistics area (Two Crows Corporation, 1999). The growth of most conventional methods was according to smart theory and analytical techniques that worked relatively well on the diffident amounts of data being examined (Two Crows Corporation, 1999).

According to improved computer power on the massive volumes of accessible data, new methods can estimate approximately any functional outline or interaction unaccompanied (Two Crows Corporation, 1999). Conventional statistical methods depend on the modeler to identify the functional form and interactions (Two Crows Corporation, 1999). The main point is that data mining is the function of AI and statistical methods to general business issues so that make these methods accessible to the experienced knowledge examiner in addition to the skilled statistics expert (Two Crows Corporation, 1999).

- Data Mining and Machine Learning**

Machine learning is applying of computational techniques meant for increasing performance by automating the gaining of knowledge and information from experience (Langley and Simon 1995). Machine learning intends to offer growing levels of mechanization in the knowledge engineering procedure, substituting significant time-consuming human activity with automatic methods that advance precision or efficiency by finding out and applying regularities in training data (Langley and Simon 1995).

### **Data Mining and Statistics**

Statistics and data mining both intend to find out structure in data (Hand, 1999). Because a large amounts of their plans overlap, some people consider data mining as a division of statistics (Hand, 1999). However, that is not a rational evaluation as data mining also utilize thoughts, tools, and techniques from other parts, mainly database technology and machine learning, and is not greatly related to some parts where statisticians are concerned (Hand, 1999). Descriptive and Visualization Techniques, Cluster Analysis, Correlation Analysis, Discriminant Analysis, Factor Analysis, Regression Analysis, Logistic Regression are the basic statistical algorithms.

## **2.1.6. Data Mining Tasks and Techniques**

Actually, the two main objectives of data mining are “prediction” and “description” (Kantardzic, 2003). Prediction engages in applying several fields or variables within the data set in order to predict unidentified or future values of other wanted variables. However, description concentrates on discovering patterns unfolding the interpretable data (Kantardzic, 2003).

The comparative significance of prediction and description for meticulous data-mining applications differ noticeably (Kantardzic, 2003). In general, many problems of logical, academic, economic, and business area can be expressed in terms of the subsequent tasks (Ayad, 2000; Berry and Linoff, 2004):

- Classification
- Estimation
- Prediction
- Associating rules or Affinity Grouping
- Clustering
- Description and profiling
- Mining Sequential Patterns
- Pattern Based Similarity Search
- Change and Deviation Detection

However, fundamental tasks that can be performed with data mining are: Classification, Estimation, Prediction, Affinity grouping or associating rules, Clustering, Description and visualization (Berry and Linoff, 2004). We can see a short explanation of the mentioned functionalities below.

### **Classification**

Classification encompasses exploration of the characteristics of a recently indicated purpose and conveying it to one of a definite set of categories (Berry and Linoff, 2004). The classification is described by a precise explanation of the classes, and a training set covering pre-classified instances (Berry and Linoff, 2004). Main classification techniques are (Berry and Linoff, 2004):

1. Decision Tree
2. Neural Networks

3. Link Analysis
4. Nearest Neighbor Techniques
5. Support Vector Machines

### **Estimation**

Estimation copes with constantly valued results (Berry and Linoff, 2004). “Estimation Estimation is similar to classification except that the target variable is numerical rather than categorical” (Larose, 2005). Estimation is used to assign a value for several unidentified continuous variable such as income, height, or credit balance (Berry and Linoff, 2004). Well-suited estimation tasks are (Berry and Linoff, 2004):

1. Regression models
2. Neural networks
3. Survival analysis

### **Prediction**

Prediction is identical to classification or estimation, except that for prediction, the results lie in the future. In a forecasting task, the only method to control the precision of the classification is to stay and observe (Berry and Linoff, 2004). The main cause for treating prediction as a different task from classification and estimation is that in predictive modeling there are extra subjects concerning the temporal relationship of the input variables or forecasters to the target variable (Berry and Linoff, 2004). Any of the methods used for classification and estimation can be personalized for applying in prediction by means of training instances where the value of the variable to be forecasted is identified, together with historical data for those instances (Berry and Linoff, 2004). The historical data is employed to make a model that explicates the present observed behavior. While this model is employed to present inputs, the outcome is a forecast of future behavior (Berry and Linoff, 2004). Some examples of prediction tasks in business and research are (Berry and Linoff, 2004):

- Predict the future sales of a company;
- Predicting which customers will disappear in next months;
- Forecasting which cell phone subscribers may want to order a value-added service in future

Some famous prediction techniques are:

1. Linear Regression
2. Nonlinear Regression
3. Decision Tree
4. Traditional Prediction methods like ARMA or ARIMA
5. Neural Networks

Usually, “any of the methods and techniques used for classification and estimation may also be used for prediction, under appropriate circumstances, for prediction” (Larose, 2005). The selection of method relies on the nature of the input data, the kind of value to be forecasted, and the significance related to explicability of the forecasting (Berry and Linoff, 2004). We are going to use neural networks to predict future sales because of their capability to store and use information about past data and make accurate prediction. The descriptions and applications of neural networks in forecasting will be discussed later.

### **Affinity Grouping or Association Rules**

The task of Association rule is to conclude which items go together (Berry and Linoff, 2004). They are usually used in the retail sales society to identify objects that are commonly bought together (Berry and Linoff, 2004). The task of association is to discover rules for quantifying the relationship between two or more features (Larose, 2005). “Association rules take the form If antecedent, then consequent, along with a measure of the support and confidence associated with the rule” (Larose, 2005). Association rules can also be applied to recognize cross-selling prospects and to map out stunning packages or combining products and services (Berry and Linoff, 2004). Some association rule algorithms are:

1. Apriori
2. HPTree
3. AIS
4. SETM
5. RARM

### **Clustering (Unsupervised learning)**

“Clustering is the task of segmenting a heterogeneous population into a number of more homogeneous subgroups or clusters” (Berry and Linoff, 2004). “Clustering refers to the grouping of records, observations, or cases into classes of similar objects” (Larose, 2005). A

cluster is a group of records that are like one another, and different from records in other clusters (Larose, 2005). In fact, what differentiates clustering from classification is that clustering does not depend on predetermined classes (Berry and Linoff, 2004). However, in classification each record is allocated to a predetermined class on the foundation of a model evolved during training on pre-classified instances (Berry and Linoff, 2004). “Clustering is often performed as a preliminary step in a data mining process, with the resulting clusters being used as further inputs into a different technique downstream, such as neural networks” (Larose, 2005).

### **Description and Profiling**

Sometimes the goal of data mining is just to explain what is going on in a complex database through an approach that adds to our understanding of the people, products, or procedures that created the data in the initial place (Berry and Linoff, 2004). Premium description can frequently be achieved by exploratory data analysis, a graphical technique of discovering data in looking for of patterns and trends (Larose, 2005).

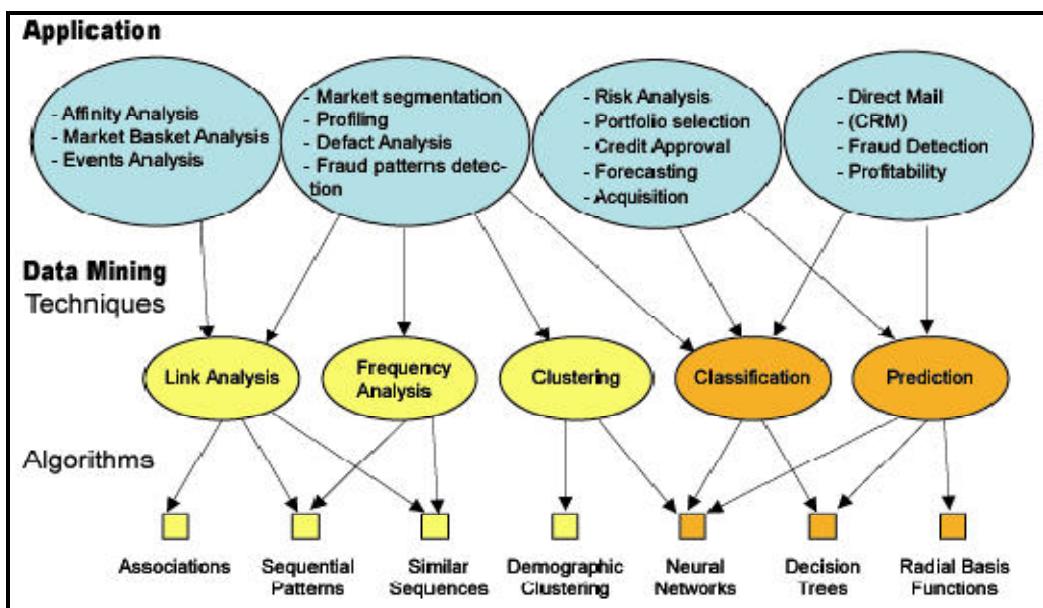
### **2.1.7. Choose Appropriate Data Mining Techniques**

We have seen that there are many data mining techniques that can extract information from data and perform classification, clustering, association, prediction and so on. Before choosing a technology, researcher has to first consider which problem is to be addressed and which results to obtain. Secondly, how much data need to be processed and from which predictions are needed. In this research, we are going to apply neural networks to predict future sales due to their ability to generalize trends and their power to keep applicable information about past sales (Shanmugasundaram et al., 2002) Neural networks are adaptable tools for predictions, since they perform feature selection or sensitive analysis automatically to choose the most important inputs. In chapter 4, we will see how we use this characteristic of neural networks in our sales prediction.

### **2.1.8. Data Mining Applications**

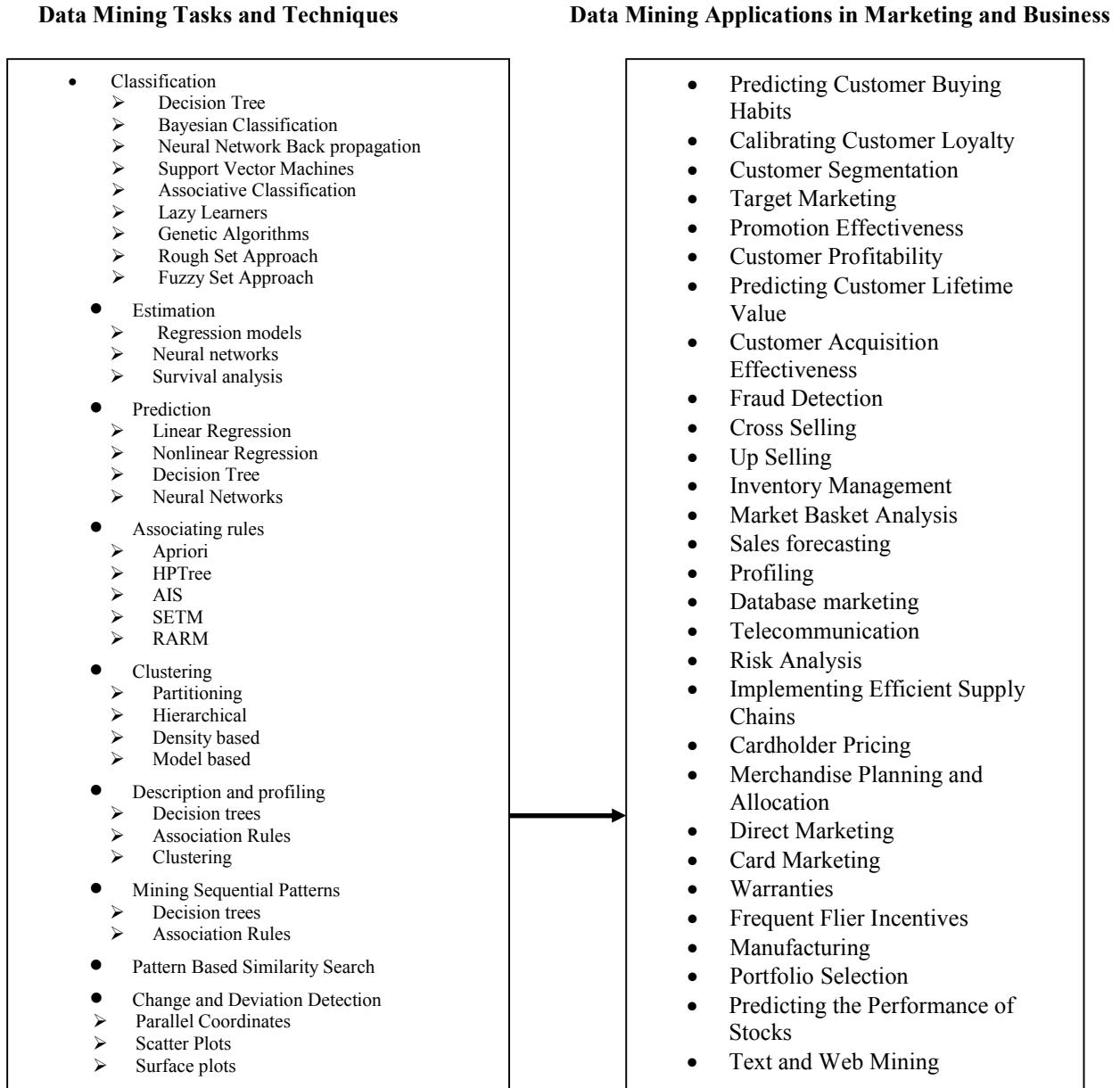
“Data mining is increasingly popular because of the substantial contribution it can make” (Two crows corporation, 1999). Data mining techniques have the capability of supplying companies with competitive advantages in optimizing their employment of information (Dhond

et al., 2000). Artificial neural networks (ANN) has got growing consideration in forecasting area, particularly in time series sales forecasting (Tang and Fishwick, 1991; Thiesing and Vornberber, 1997; cited by Crone, 2003). In addition, “by determining characteristics of good customers (profiling), a company can target prospects with similar characteristics” (Two crows corporation, 1999). “By profiling customers who have left, a company can act to retain customers who are at risk for leaving (reducing churn or attrition), because it is usually far less expensive to retain a customer than acquire a new one” (Two crows corporation, 1999). “Data mining offers value across a broad spectrum of industries” (Two crows corporation, 1999). “Telecommunications and credit card companies are two of the leaders in applying data mining to detect fraudulent use of their services” (Two crows corporation, 1999). “Medical applications are another fruitful area: data mining can be used to predict the effectiveness of surgical procedures, medical tests or medications” (Two crows corporation, 1999). “Companies active in the financial markets use data mining to determine market and industry characteristics as well as to predict individual company and stock performance. Retailers are making more use of data mining to decide which products to stock in particular stores (and even how to place them within a store), as well as to assess the effectiveness of promotions and coupons” (Two crows corporation, 1999). Figure 7 shows applications and data mining techniques and the related algorithms in marketing domain (Dunham, 2002). In addition, Table 1 summarizes data mining tasks and techniques, which contribute to numerous applications in marketing and business domains. In next section, we will go through neural networks in detail since we are going to apply it in our sales prediction.



**Figure 6: Application of Data Mining for Marketing**  
Source: Dunham (2002)

**Table 1: Applications of data mining in marketing and Business**



## 2.2. Neural Networks

“A neural network consists of artificial neurons connected together. Each neuron mimics its biological counterpart, taking various inputs, combining them, and producing an output” (Berry and Linoff, 2004). “Because digital neurons process numbers, the activation function characterizes the neuron. In most cases, this function takes the weighted sum of its inputs and

applies an S-shaped function to it” (Berry and Linoff, 2004). “The result is a node that sometimes behaves in a linear fashion, and sometimes behaves in a nonlinear fashion—an improvement over standard statistical techniques” (Berry and Linoff, 2004).

Artificial neural networks (Hush and Home, 1993; Haykin, 1999; cited by Mitra and Acharya, 2003) are signal processing systems that attempt to imitate the action of biological nervous systems by offering a numerical model of grouping of many neurons linked in a network (Mitra and Acharya, 2003). These are clarified as particularly parallel interconnections of simple processing segments that interact with items of the actual world in a way resembling biological systems (Mitra and Acharya, 2003).

Neural networks are a set of dominant and general-purpose tools that are applied frequently to prediction (especially time-series forecasting), estimation classification, and clustering (Dunham, 2002; Berry and Linoff, 2004). Across a large number of industries and a large number of applications, neural networks have proven themselves repeatedly. However, some general use of neural networks is making a model for classification or prediction (Berry and Linoff, 2004). The phases of this process are (Berry and Linoff, 2004):

- “1. Identify the input and output features.
- 2. Transform the inputs and outputs so they are in a small range, (-1 to 1).
- 3. Set up a network with an appropriate topology.
- 4. Train the network on a representative set of training examples.
- 5. Use the validation set to choose the set of weights that minimizes the error.
- 6. Evaluate the network using the test set to see how well it performs.
- 7. Apply the model generated by the network to predict outcomes for unknown input.”

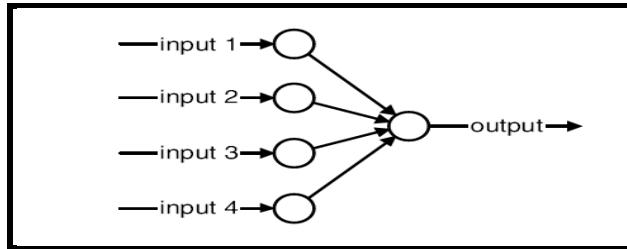
A neural network is similar to a black box that recognizes how to process inputs to generate an output. The computation is rather complicated and hard to realize; however, the outcomes are frequently valuable (Berry and Linoff, 2004). “Neural networks are good for prediction and estimation problems” (Berry and Linoff, 2004). Proper problem has the subsequent three features (Berry and Linoff, 2004):

- ❖ The inputs are properly recognized. We need to have an appropriate idea of which features of the data are noteworthy, but not certainly how to merge them.
- ❖ The output properly recognized. We should realize what we are attempting to model.
- ❖ Experience is presented. We should have many cases where both the inputs and the output are recognized. These known samples are applied to train the network.

### **2.2.1. Neural Networks Architectures**

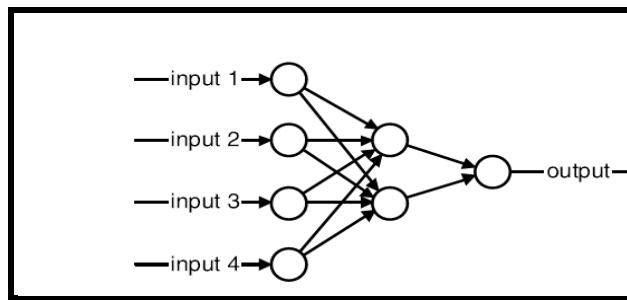
On the basis of the connection architecture (pattern), ANNs can be divided into two general classes (Jain et al., 1996):

- 1) Feed-forward networks: “Feed-forward networks are the simplest and most useful type of network for directed modeling” (Berry and Linoff, 2004). Feed-forward neural networks take inputs on one end and transform them into outputs (Berry and Linoff, 2004). We can divide feed-forward networks to two categories: a) single layer feed-forward network which has only input and output layers, and b) and successfully used for forecasting has been the multilayer feed-forward neural network which differentiates itself by the existence of one or further hidden layers (Haykin, 1999). Some examples of feed-forward networks are single-layer perceptron, multilayer perceptron, and radial basis function networks (Mitra and Acharya, 2003). For predictive modeling, the most regular network is the feed-forward network since the topology, or structure of this network is standard of networks applied in prediction and classification (Berry and Linoff, 2004). During last decades, various kinds of ANN models have been evolved, each intended to solve diverse problems. However, undoubtedly the most broadly and effectively applied for prediction has been the feed-forward kind neural network (Zhang, 2004). That is why we will explain this architecture in more subsequently. Below are different types of feed-forward neural networks (Berry and Linoff, 2004):
- “This simple neural network takes four inputs and produces an output. This result of training this network is equivalent to the statistical technique called logistic regression” (Berry and Linoff, 2004).



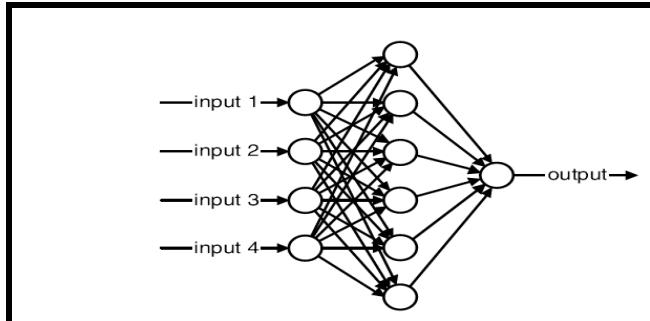
**Figure 7: A Simple Neural Network**  
Source: (Berry and Linoff, 2004)

- “This network has a middle layer called the hidden layer, which makes the network more powerful by enabling it to recognize more patterns” (Berry and Linoff, 2004). Frequently, just one hidden layer is required (Berry and Linoff, 2004).



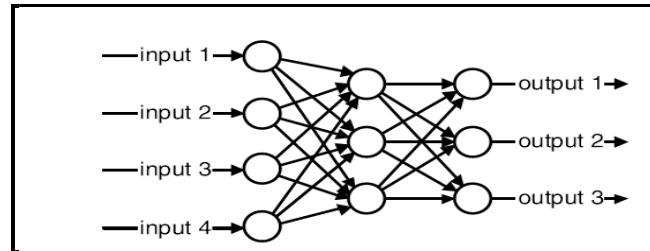
**Figure 8: A Neural Network with one Hidden Layer**  
Source: (Berry and Linoff, 2004)

- “Increasing the size of the hidden layer makes the network more powerful but introduces the risk of over-fitting” (Berry and Linoff, 2004). It is desired that the network generalize on the training set, not remember it. Accordingly, the hidden layer is not supposed to be excessively wide (Berry and Linoff, 2004). Thus, “probably the biggest decision is the number of units in the hidden layer” (Berry and Linoff, 2004). Opportunely, it can be discovered while a network is over-trained. If the network acts superbly on the training set, but performs inferior on the validation set, it is a sign that the network has memorized the training set (Berry and Linoff, 2004). It is recommended that commonly not to employ hidden layers larger than the quantity of inputs (Berry and Linoff, 2004).



**Figure 9: A Neural Network with a Wide Hidden Layer**  
Source: (Berry and Linoff, 2004)

- “A neural network can produce multiple output values” (Berry and Linoff, 2004), but it is not so common to have more than one output.



**Figure 10: A Neural Network with Multiple Outputs**  
Source: (Berry and Linoff, 2004)

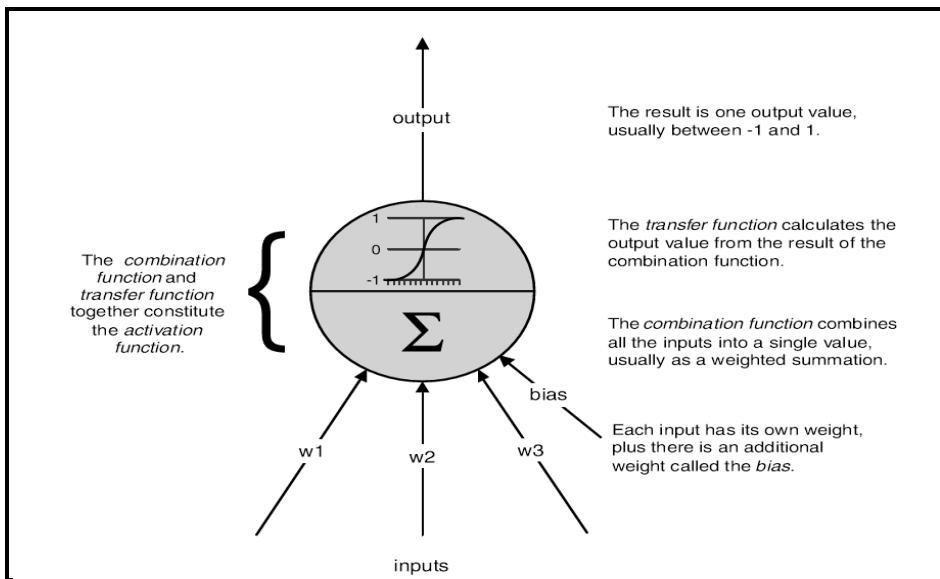
In general, there are three fundamental questions to inquire about this type of neural networks (Berry and Linoff, 2004):

- 1) “What are units and how do they behave? That is, what is the activation function” (Berry and Linoff, 2004)?
- 2) “How are the units connected together? That is, what is the topology of a network” (Berry and Linoff, 2004)?
- 3) “How does the network learn to recognize patterns? That is, what is back propagation and more generally, how is the network trained” (Berry and Linoff, 2004)?

#### ❖ The Unit of a Neural Network

Figure 12 shows the chief attributes of the artificial neuron. The unit merges its inputs into a unique value, which then changes to generate the output; these two activities jointly are named the activation function (Berry and Linoff, 2004). The activation function is comprised of

two elements. The initial part is the combination function that combines all the inputs into a unique value (Berry and Linoff, 2004). According to Figure 12, every input possesses its own weight. Notice that in Figure 12, we have a supplementary input coming, which is constant input and occasionally it is labeled as bias, and is always placed as 1. The bias operates as an overall offset that assists the network to better recognize the patterns (Berry and Linoff, 2004). The training part adjusts the weights on constant inputs exactly the same as it performs on the further weights in the network (Berry and Linoff, 2004). “The most widespread combination function is the weighted sum, where each input is multiplied by its weight and these products are added together. Other combination functions are sometimes useful and include the maximum of the weighted inputs, the minimum, and the logical (AND or OR) of the values” (Berry and Linoff, 2004).

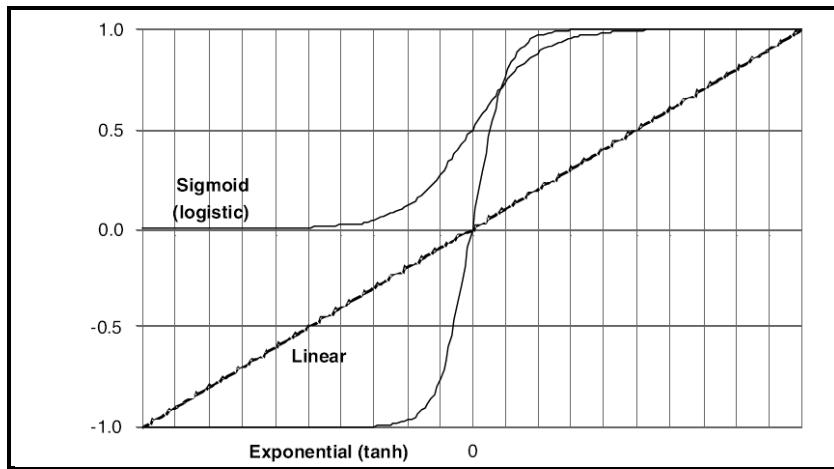


**Figure 11: The Unit of an Artificial Neural Network**

Source: (Berry and Linoff, 2004)

The subsequent element of the activation function is the transfer function (Berry and Linoff, 2004), which “transfers the value of the combination function to the output of the unit” (Berry and Linoff, 2004). Three common transfer functions are the sigmoid logistic, linear, and hyperbolic tangent functions that are depicted below (Berry and Linoff, 2004). According to Figure 13, “The sigmoid activation functions are S-shaped curves that fall within bounds” (Berry and Linoff, 2004). For instance, the sigmoid logistic function [ $\text{logistic}(x) = f(x) = 1 / (1 + e^{-x})$ ] generates values between 0 and 1, and the hyperbolic tangent [ $\text{tang}(x) = f(x) = (e^x - e^{-x}) / (e^x + e^{-x})$ ]

$x - e^{-x}) / (e^x + e^{-x})$ ] generates values between -1 and 1 for all potential outputs of the summation function (Berry and Linoff, 2004).



**Figure 12: Three Common Transfer Functions**

Source: (Berry and Linoff, 2004)

- 2) Feedback (interactive or recurrent) networks: in which loops take place due to feedback connections (Mitra and Acharya, 2003). A recurrent neural network “distinguishes itself from a feed-forward neural network in that it has at least one feedback loop” (Haykin, 1999). Hopfield network and adaptive resonance theory (ART) models are some examples of feed-back networks (Mitra and Acharya, 2003).

### 2.2.2. Learning Process

The assets that is of most important worth for a neural network is the capability of the network to learn from its environment, and to advance its performance via learning (Haykin, 1999). A neural network finds out about its environment during an interactive procedure of adjustments employed to its synaptic weights and bias levels (Haykin, 1999). Perfectly, the network comes to be more well-informed regarding its environment following each iteration of the learning process (Haykin, 1999). The three major learning paradigms are supervised, unsupervised, and reinforcement (Bigus, 1996).

- **Training of Artificial Neural Networks**

A very significant issue of using neural networks intended for data mining is how to cope with our raw material, the past data (Bigus, 1996). The most widespread approach is to randomly split the source data into two or further data sets. One subset of the data is employed to train the neural network, and another subset is applied in testing the precision of the neural network (Bigus, 1996). It is essential to understand that the neural network by no means learns or adjusts its weights via the test data (Bigus, 1996). A number of people imply that a third subset is needed, so the developer employs a train-and-test data set to generate the neural network model and a third party separately tests the network by means of the validation data (Bigus, 1996). However, there are some cases when this usual method is not appropriate such as when the data has temporal or time-series characteristics (Bigus, 1996). Random selection from this data set would be disastrous. In this situation, it is common to employ data from a definite time period for training and the most current data for the validation and testing steps (Bigus, 1996), which is exactly like our case that we have time series data (sales data).

- **Input and Output Encoding (Normalizing)**

One shortcoming of neural networks is that all feature values ought to be encoded in a standardized approach (Larose, 2005). However, some softwares, like SATISTICA, do it systematically. The motivation of this encoding has to do with how these functions work near 0. In this range, they act in a nearly linear manner Small alters in x ends in minute changes in the output (Berry and Linoff, 2004). In addition, needing that the entire inputs to be in the identical range, stops one set of inputs controlling other inputs as well (Berry and Linoff, 2004). Even though it is suggested that inputs be in the range from -1 to 1, this ought to be taken as an instruction, not a firm regulation (Berry and Linoff, 2004). For example, normalizing variables, subtracting the mean and dividing by the standard deviation, is a regular transformation of variables. This ends in minute enough values to be practical for neural networks (Berry and Linoff, 2004). In addition, we may simply apply the min–max normalization (Larose, 2005):

$$X^* = \frac{X - \min(X)}{\text{range}(X)} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

### **2.2.3. Back Propagation Algorithm**

The most widespread kind of neural network model is identified as the back propagation network (Moshiri and Cameron, 2000). That is why we explain this model in more details below. Back propagation algorithm was the initial practical method for training networks (Berry and Linoff, 2004). At the heart of back propagation are the following three steps (Berry and Linoff, 2004):

1. “The network gets a training example and, using the existing weights in the network, it calculates the output or outputs” (Berry and Linoff, 2004).
2. “Back propagation then calculates the error by taking the difference between the calculated result and the expected (actual result)” (Berry and Linoff, 2004).
3. “The error is fed back through the network and the weights are adjusted to minimize the error. Hence, the name back propagation is due to the fact that errors are sent back through the network” (Berry and Linoff, 2004).

Actually, “back propagation algorithm measures the overall error of the network by comparing the values produced on each training example to the actual value. It then adjusts the weights of the output layer to reduce, but not eliminate, the error” (Berry and Linoff, 2004). We should remember that “the goal is to generalize and identify patterns in the input, not to memorize the training set” (Berry and Linoff, 2004).

- **How Does a Unit Adjust Its Weights?**

Through the training stage, the general objective is to find out the most precise weights to be assigned to the connector lines. In addition, through training, the output is calculated frequently and the outcome is compared to actual value (Sivanandam and Deepa, 2006). Any variance is assumed as a training error and it is significant for this training error to be as small as feasible in order that the predicted output is trustworthy (Sivanandam and Deepa, 2006). Through adjusting the weights, the error is reduced repeatedly until a point is achieved that signifies the least amount of reasonable error. At this position, a precise prediction can be generated (Sivanandam and Deepa, 2006). In other words, it is approximated that whether altering the

weight on each input would augment or reduce the error. The unit subsequently adjusts each weight to decrease the error as much as possible (Berry and Linoff, 2004). Subsequent to being revealed enough training instances during sufficient generations, the weights on the network no longer alter considerably; also, the error does not reduce any longer (Berry and Linoff, 2004). This is the position where training discontinues; the network has found out to identify patterns in the input data (Berry and Linoff, 2004). “This technique for adjusting the weights is called the generalized delta rule” (Berry and Linoff, 2004).

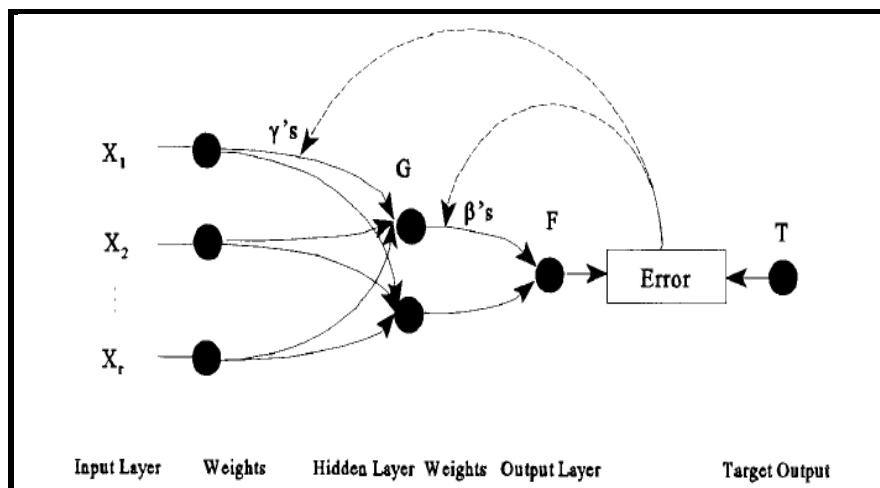
- **Back-Propagation Network (BPN)**

“Back-Propagation Network (BPN) models are one of the most popular types of ANN models. They are static or feed-forward-only (input vectors are fed through to output vectors, with no feedback to input vectors again); they are hetero-associative (the output vector may contain variables different from the input vector); and their learning is supervised” (Moshiri and Cameron, 2000). A classic BPN model employs three vectors: input vector, one or more concealed vectors, and an output vector (Moshiri and Cameron, 2000). Later than the input and output vectors are read into the BPN model, the network initially chooses parameters arbitrarily and processes the inputs to produce a forecasted output vector (Moshiri and Cameron, 2000). Subsequent to computing the error between its forecasted outputs and the monitored outcomes, the network alters the parameters in directions that will decrease the error, produces a new output vector, computes the error, and adjusts its parameters another time, and so on (Moshiri and Cameron, 2000). The iteration or learning process persists until the network achieves a firm particular error (Moshiri and Cameron, 2000). Figure 14 is a sketch of the stages of a classic BPN model with  $r$  input units ( $X$ ), two hidden or middle vectors ( $G$ ), and one output vector ( $F$ ) ( $\gamma$ 's and  $\beta$ 's are the middle vector and output vector weights, respectively) (Moshiri and Cameron, 2000). In a common structure, the ANN output vector generated by a model or network including  $r$  input units,  $q$  hidden units, and one output unit is presented as (Moshiri and Cameron, 2000):

$$F(x, w) = F \left[ \beta_0 + \sum_{j=1}^q G(x\gamma_j)\beta_j \right]$$

“Where  $F(x,w)$  is the network's final output,  $F$  is the activation function for the final step,  $G$  is the activation function for a hidden or intermediate unit,  $X = [1, X_1, \dots, X_r]$  is the input vect or(including the intercept constant), and  $W = (y_1, y_2, \dots, y_q, \beta_j)$  is the parameters or weights matrix” (Moshiri and Cameron, 2000). “Each term  $y_i$  stands for a  $r*1$  vector of weights relating to the  $r$  input variables to one of the  $q$  intermediate units.  $\beta_j$  refers to a  $q*1$  vector of weights relating each intermediate output vector to the final output vector.  $F$  and  $G$  can take any functional form, but the non-linear sigmoid function is a popular one, particularly for  $G$ ” (Moshiri and Cameron, 2000).

The BPN model can learn all kinds of constant functions if sufficient intermediate steps are permitted (Rumelhar et al., 1986; White, 1990; cited by Moshiri and Cameron, 2000). “However, the learning in BPN is slow, particularly if the size of the network and the training data set are large” (Moshiri and Cameron, 2000).



**Figure 13: A Back-Propagation Network Model**

Source: (Moshiri and Cameron, 2000).

## 2.2.4. Single-layer Perceptron

“The concept of perceptron (Rosenblatt, 1958; Rosenblatt, 1962) was one of the most exciting developments during the early days of pattern recognition” (Mitra and Acharya, 2003). The typical (single-layer) perceptron, with two classes of patterns, tries to discover a linear decision boundary unraveling the two classes. A perceptron comprise a single neuron with

adaptable weights,  $w_j$ ,  $j = 1, 2, \dots, n$ , and threshold  $\theta$ . Given an input vector  $x = [x_1, x_2, \dots, x_n]^T$ , the net input to the neuron is (Mitra and Acharya, 2003):

$$v = \sum_{j=1}^n w_j x_j - \theta.$$

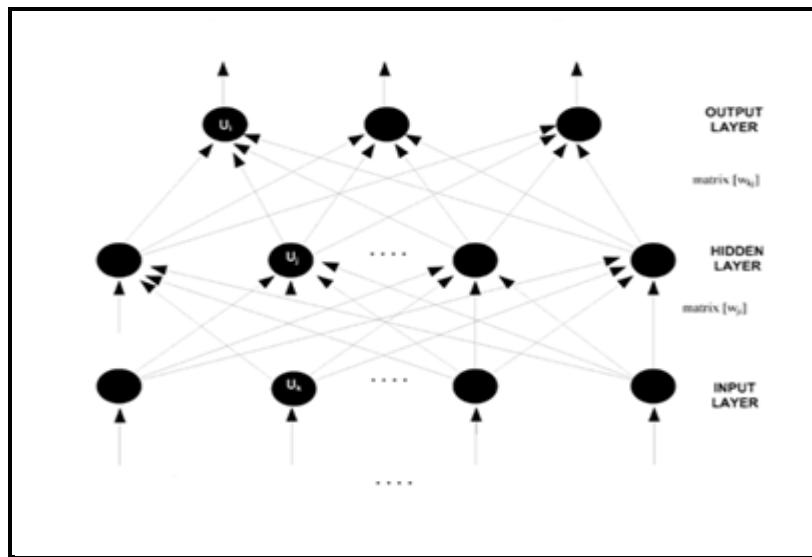
“The output  $y$  of the perceptron is +1 if  $v > 0$  and is 0 otherwise. In a two-class classification problem, the perceptron assigns an input pattern to one class if  $y = 1$  and to the other class if  $y = 0$ . The linear equation  $\sum_{j=1}^n w_j x_j - \theta = 0$  defines the decision boundary that halves the space” (Mitra and Acharya, 2003). Rosenblatt, (1962) built up a learning process to find out the weights and threshold in a perceptron, given a group of training models. This algorithm is outlined as follows (Mitra and Acharya, 2003):

1. “Initialize the weights and threshold to small random numbers” (Mitra and Acharya, 2003).
2. “Present a pattern vector  $x = [x_1, x_2, \dots, x_n]^T$  and evaluate the output of the neuron” (Mitra and Acharya, 2003).
3. “Update the weights according to  $w_j(t+1) = w_j(t) + \varepsilon(d - y)x_j$ , where  $d$  is the desired output,  $t$  is the iteration number, and  $\varepsilon$  ( $0.0 < \varepsilon < 1.0$ ) is the learning rate (step size)” (Mitra and Acharya, 2003). In fact, learning takes place just while the perceptron produces an error (Mitra and Acharya, 2003). A single-layer perceptron is not appropriate for conditions that present nonlinearity and multiple classes (Mitra and Acharya, 2003). Therefore, the multilayer perceptron network has been introduced (Mitra and Acharya, 2003).

### **2.2.5. Multilayer Perceptron (MLP)**

One of the most widely implemented neural network architecture is the multilayer perceptrons (MLP) model (Haykin, 1999). “This network has a three-layer, feed forward, hierarchical structure” (Partovi and Anandarajan, 2002). “The total number of neurons, number of neurons on each layer, as well as the number of layers determine the accuracy of the network model” (Partovi and Anandarajan, 2002). A classic MLP is shown in Figure 15. The neurons

in the input layer signify the features or stimuli in a data set. These inputs ( $x_1, x_2, \dots, x_n$ ) begin the activations into the network (Partovi and Anandarajan, 2002). The upper part of the neuron obtains this sum and computes the degree to which the sum is significant by means of a transfer function ( $J$ ), generating an distinct output, where,  $w$  is weight vector  $w [w_1, w_2, \dots, w_n]$ ; and  $x$  is the input vector  $x [x_1, x_2, \dots, x_n]$ ; for a definite neuron (Partovi and Anandarajan, 2002). The preference of the transfer function normally relies on the nature of the output of the network (Fausett, 1994). Incidentally, there are a number of options, comprising the step function, sigmoid function, hyperbolic tangent function, and linear function among others (Partovi and Anandarajan, 2002). To sum up, each bond in the ANN has a weight that is produced from the input values and subsequently altered to an output value through a transfer function (Partovi and Anandarajan, 2002). The output value of a neuron is a function of the weighted sum of its inputs (Berry and Linoff, 2004). “The weights represent both the strength and nature of the connection between neurons” (Partovi and Anandarajan, 2002). The diagnosis of these weights is a significant part of the learning procedure, and is produced by a repetitive training procedure that case instances with identified decision outputs are constantly demonstrated to the network (Partovi and Anandarajan, 2002). A normally applied learning process in ANN is the BP algorithm (Partovi and Anandarajan, 2002).



**Figure 14: Structure of the MLP**  
Source: (Partovi and Anandarajan, 2002)

- **Multilayer Perceptron (MLP) Using Back-propagation of Error**

The real meaning of the BP learning algorithm is to fill the input- output relations inside the MLP topologies in order that it learns sufficiently regarding the earlier period to generalize the future (Partovi and Anandarajan, 2002).

- **Sensitive Analysis**

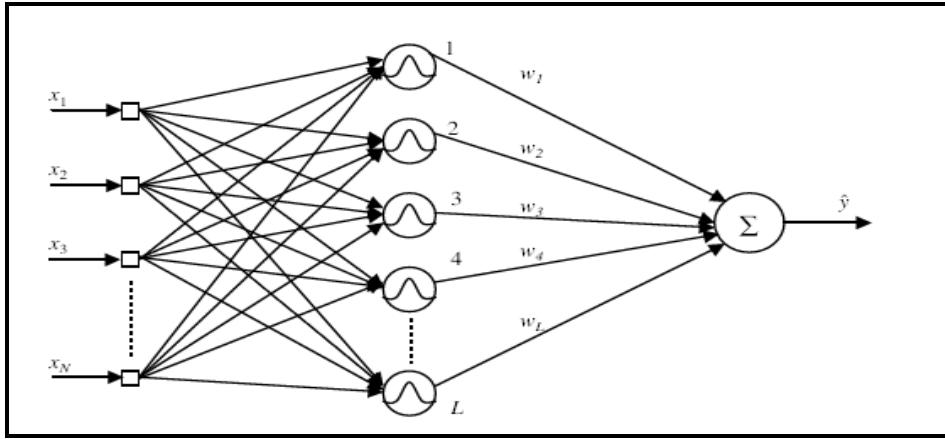
“A technique called sensitivity analysis can be used to get an idea of how opaque models work” (Berry and Linoff, 2004). “Sensitivity analysis uses the test set to determine how sensitive the output of the network is to each input” (Berry and Linoff, 2004). The following are the fundamental phases (Berry and Linoff, 2004):

1. “Find the average value for each input. We can think of this average value as the center of the test set” (Berry and Linoff, 2004).
2. “Measure the output of the network when all inputs are at their average value” (Berry and Linoff, 2004).
3. “Measure the output of the network when each input is modified, one at a time, to be at its minimum and maximum values (usually –1 and 1, respectively)” (Berry and Linoff, 2004).

For a number of inputs, the output of the network alters slightly for the three values (minimum, average, and maximum). The network is not responsive to these inputs (at least when all other inputs are at their average value). Other inputs lead to a huge consequence on the output of the network. The network is perceptive to these inputs. The extent of alteration in the output calculates the sensitivity of the network for each input. (Berry and Linoff, 2004).

## **2.2.6. Radial basis function network (RBF)**

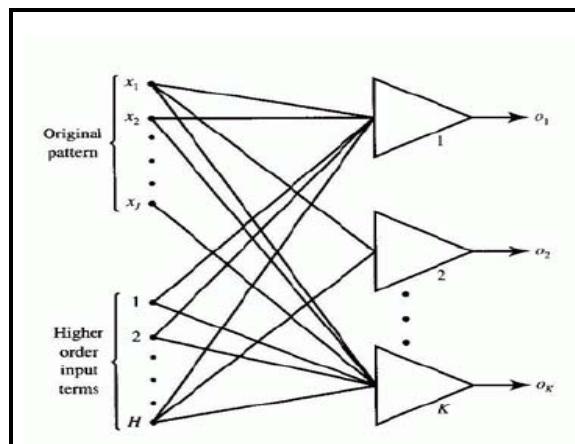
“RBF networks form a special neural network architecture that consists of three layers. The input layer is only used to connect the network to its environment” (Doganis et al., 2006). “The hidden layer contains a number of nodes, which apply a nonlinear transformation to the input variables, using a radial basis function” (Doganis et al., 2006). “The output layer is linear and serves as a summation unit” (Doganis et al., 2006). The classic structure of an RBF neural network with only one output node is illustrated in Figure 16 (Doganis et al., 2006).



**Figure 15: The RBF Neural Network Topology**  
Source: (Doganis et al., 2006)

### 2.2.7. Time Delay Neural Networks (TDNN)

TDNN is an active neural network that is created by inserting local memory into input and hidden layers of a multi-layer feed-forward neural network (Zhang et al., 2003; cited by Ture and Kurt, 2006). A Time Delay Neural Network (TDNN) model is presented in the Figure 17. This kind of neural networks is particularly proficient at managing temporal patterns and it is mostly used for forecasting. In this situation, the input data are registered into the network individually Bansal et al, 1998). The network post pones passing the data to the subsequent node earlier than the new data value is presented (Bansal et al, 1998). Thus, the TDNN keeps a record of the time sequence inside the network. This history increases its sales forecasting abilities (Bansal et al, 1998).



**Figure 16: TDNN Architecture**  
Source: (Bansal et al, 1998)

## 2.2.8. Recurrent neural networks

Other than the most accepted feed-forward ANNs, numerous other kinds of neural networks can be intended for forecasting objectives as well (Zhang, 2004). Specifically, recurrent neural networks (Husken and Stagge, 2003; cited by Zhang, 2004) that clearly explain the dynamic nonlinear pattern can be a proper substitute to feed-forward kind ANNs for definite time series forecasting issues (Zhang, 2004). Recurrent neural networks let output signals of several of their neurons to move back and act as inputs for the neurons of the same layer or those of the prior layers (Carboneau et al., 2008). The training technique named “back-propagation through time” could be employed in train a RNN on a certain training set (Werbos, 1990; cited by Carboneau et al., 2008). Figure 18 indicates the formation of RNN for the supply chain demand forecasting issue. As it is indicated in Figure 18, just the output of the hidden neurons is applied to act as the input to the neurons of the same layer (Carboneau et al., 2008).

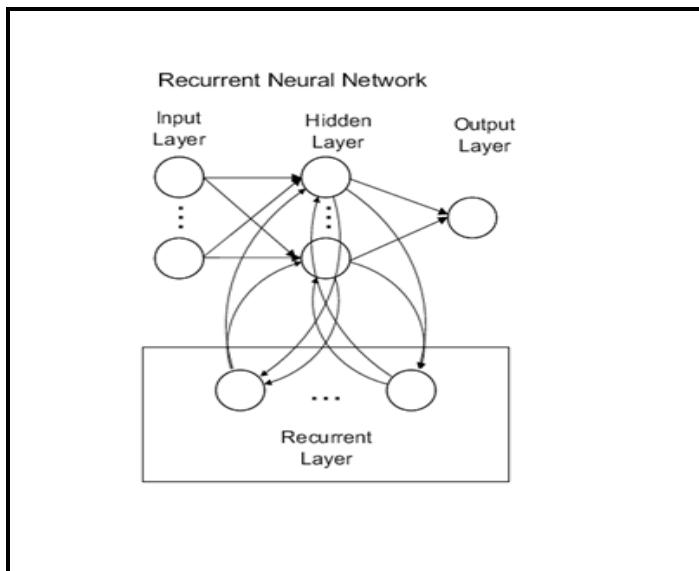


Figure 17: Recurrent neural network  
Source: (MathWorks, 2000; cited by Carboneau et al., 2008)

## 2.2.9. Benefits and Limitations of Neural Networks

- **Advantages of Neural Networks**

It is obvious that a neural network gets its analyzing power by, primarily, its extensively parallel distributed formation and, next, its capability to study and consequently generalize. These

two information-processing abilities cause it probable for neural networks to answer difficult (large-scale) issues that are presently uncontrollable (Haykin, 1999). Actually, the most important reason of applying neural networks is to be able to forecast data patterns that are considerably complicated for the conventional statistical models (Sivanandam and Deepa, 2006). Furthermore, the learning capability of neural networks lets them to adapt to dynamic and shifting market environments and is a far more adaptable forecasting system than conventional statistical models (Sivanandam and Deepa, 2006). Moreover, neural network tools are capable of finding out patterns and trends in every set of data they are provided, inclusive of significantly unorganized and uneven data (Sivanandam and Deepa, 2006). Numerous businesses are persisting in extending the applications of neural networks in diverse fields of their procedures to amplify and advance their competitiveness (Sivanandam and Deepa, 2006). One of the significant benefits of NNs is that individuals with modest information of either forecasting or ANNs can set up logical forecasts in a short space of time (Yip et al., 1997). These networks have at least two possible potencies over the more conventional model-fitting methods such as regression (Bishop, 1995). Primarily, ANNs are able to perceive and take out nonlinear connections and relationships among forecaster variables. Subsequently, the inferred patterns and related approximations of the accuracy of the ANN are not based on the diverse hypothesis regarding the distribution of variables (Bishop, 1995). A significant advantage of ANNs is that researchers with slight information of forecasting or ANNs can set up logical forecasts in a satisfactory amount of time. The application of neural networks presents the following practical skills and abilities (Haykin, 1999):

1. Nonlinearity
2. Input–Output Mapping
3. Adaptive Capacity
4. Evidential Response
5. Contextual Information
6. Fault Tolerance
7. Uniformity of Analysis and Design
8. Neurobiological Analogy

- **Disadvantages of Neural Network Models**

There are also few detractors who indicate the drawbacks of applying neural networks as prediction tools. First, the design of the neural network is an extremely complicated process that still counts generally on trial and error (Sivanandam and Deepa, 2006). The training procedure is time-consuming, and ought to be constantly recurred to justify alters in variables' values (Sivanandam and Deepa, 2006). "One of the dangers with any model used for prediction or classification is that the model becomes stale as it gets older, and neural network models are no exception to this rule" (Berry and Linoff, 2004). The problem of maintaining a neural network model state-of-the-art is made further complexity by two issues (Berry and Linoff, 2004). "First, the model does not readily express itself in the form of rules, so it may not be obvious when it has grown stale. Second, when neural networks degrade, they tend to degrade gracefully making the reduction in performance less obvious. In short, the model gradually expires and it is not always clear exactly when to update it" (Berry and Linoff, 2004). Moreover, "neural networks, although able to generate a solution to many problems, are unable to explain how they arrive at their results. As such, neural networks are considered as black boxes whose rules of operation are completely unknown" (Sivanandam and Deepa, 2006). "It is also important for the network designer to recognize the possibility of over-training (over-fitting) when designing the neural network" (Sivanandam and Deepa, 2006). The answer is to integrate more current data into the neural network. One method is to retract the same neural network to training style and begin providing it with novel values (Berry and Linoff, 2004). A different method is to get going again by inserting novel cases into the training set and training a completely new network, conceivably even by a diverse topology (Berry and Linoff, 2004). In summary, a neural network is just as appropriate as the training set applied to generate it. The model is steady and must be clearly updated by inserting further brand new examples into the training set and retraining the network (or training a fresh network) in order to maintaining it novel and helpful (Berry and Linoff, 2004).

## **2.2.10. Applications of Neural Networks**

For instance, "in the past decade, ANNs have emerged as a technology with a great promise for identifying and modeling data patterns that are not easily discernible by traditional statistical methods in fields as diverse as cognitive science, computer science, electrical engineering and finance" (Alon et al., 2001). Qi, (1996) did an inclusive study of neural

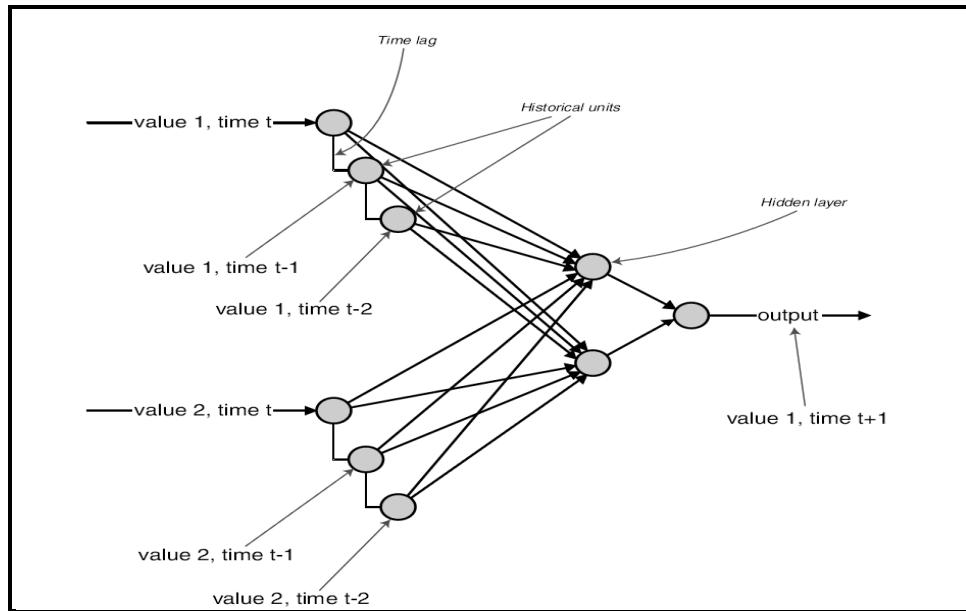
networks' applications in finance. In addition, Qi and Maddala (1999) and Qi (1999) demonstrated that numerous researches in the finance are proofing that forecasting of stock returns using linear regression could be enhanced by neural networks. "ANNs have also been increasingly used in management, marketing and retailing" (Alon et al., 2001). The kinds of applications consist of (Alon et al., 2001):

Market response forecasting (Curry and Moutinho, 1993; Dasgupta et al., 1994; Van Wezel and Baets, 1995; Agrawal and Schorling, 1996; Ainscough and Aronson, 1999; cited by Alon et al., 2001), consumer choice forecasting (West et al., 1997; Davies et al., 1999; cited by Alon et al., 2001), tourism marketing (Mazanec, 1992, 1994, and 1999; Davies et al., 1999; cited by Alon et al., 2001), buyer and seller relationship analysis (Wray et al., 1994), and market segmentation analysis (Fish et al., 1995; Mazanec, 1999; Natter, 1999; cited by Alon et al., 2001). Nowadays, neural networking has begun making significant progress in various areas within the social sciences, most notably, the field of business. Neural networks studies have led to advances in the precision of predictions and forecasts in business (Adya and Collopy, 1998). Krycha and Wagner (1999) did a comprehensive survey of ANN applications in management and marketing. According to these capabilities, neural networks have been applied to solve concerns in various fields, such as time series forecasting (Aburto and Weber, 2007). "It has been shown that such networks provide competitive results in forecasting stock exchange indexes" (McCluskey, 1993; cited by Aburto and Weber, 2007), "corporative bonds" (Moody, 1994; cited by Aburto and Weber, 2007), and sales forecasting (Thiesing and Vornberber, 1997). "A combination of neural networks with fuzzy logic has been proposed for modeling electricity demand" (Abraham, and Nath, 2001; cited by Aburto and Weber, 2007). "This hybrid intelligent system outperformed pure neural networks and also ARIMA models" (Aburto and Weber, 2007). Other successful applications of neural networks have been evolved in relation to inventory control problems (Bansal et al., 1998; He et al., 2002). He et al., in 2002, merged data mining and knowledge discovery and neural networks to figure out inventory issues. "Accurate prediction of demand is the key to reduce the cost of inventory for an enterprise in Supply Chain" (Dong and Wen, 2006). In 2006, Dong and Wen made an initial effort to apply recurrent neural networks as forecasting tool to decrease the ambiguity of inventory management. Founded on recurrent neural networks, a novel forecasting model of demand in supply chain is proposed by them. They presented a new forecasting model for inventory management. In addition, Cheng et all.,

(2006) proposed BPN-DFS (Back-propagation Neural networks-Demand Forecasting System) as an intelligent method of demand modeling. “It is capable of mapping nonlinear relationships between the marketing demand and demand affecting factors, learning those relationships from incomplete and uncertain data, and enabling easy inclusion any relevant factors into it” (Cheng et all., 2006). Finally, for supply chain demand forecasting, Carboneau et al., (2008) presented that the novel methods, like neural networks, usually outperform traditional time series methods, since neural networks are able to mode nonlinear relations.

- **Neural Networks for Time Series**

“Increasing research efforts have been directed at applying NNs to business situations” (Adya and Collopy, 1998). “In many business problems, the data naturally falls into a time series” (Berry and Linoff, 2004). “Although predominant in the financial industry, time series appear in other areas, such as forecasting and process control” (Berry and Linoff, 2004). Neural networks are simply adjusted for time-series studies (Berry and Linoff, 2004). For example, Figure 19 shows a time delay neural network that is trained on the time-series data, beginning at the oldest position in the data (Berry and Linoff, 2004). We will see the application of neural networks in time series prediction in detail later.



**Figure 18: A time-delay neural network**  
Source: (Berry and Linoff, 2004)

- **Neural Networks in Business Forecasting**

“The recent upsurge in research activities into artificial neural networks (ANNs) has proven that neural networks have powerful pattern classification and prediction capabilities” (Zhang, 2004). ANNs have been effectively applied for various tasks in numerous areas of business, industry, and science (Widrow et al., 1994). One of the main application fields of ANNs is forecasting since artificial neural networks have appeared as a major quantitative modeling technique for business forecasting (Zhang, 2004). Recently, there has been a growing attention to forecasting applying ANNs (Zhang, 2004). “The ability to accurately predict the future is fundamental to many decision processes in planning, scheduling, purchasing, strategy formulation, policy making, and supply chain operations” (Zhang, 2004). “As such, forecasting is an area where a lot of efforts have been invested in the past. Yet, it is still an important and active field of human activity at the present time and will continue to be in the future” (Zhang, 2004). Forecasting sets up the basis for strategic, tactical, plus operational decisions in numerous business firms. Its position in winning planning in marketing, finance, manufacturing, workforce, and other practical fields is fixed (Zhang, 2004). ANNs are a fairly new and capable tool for business forecasting (Zhang, 2004). “Unlike most traditional model-based forecasting techniques, ANNs are data-driven, self-adaptive, and nonlinear methods that do not require specific assumptions on the underlying data generating process” (Zhang, 2004). These attributes are especially attractive for functional forecasting conditions where data are huge or simply accessible, but the theoretical model or the hidden relationship is not recognized (Zhang, 2004). Furthermore, ANNs have worldwide practical estimation capability and are able to find out any kind of complicated relationship. In view of the fact that the quantity of probable nonlinear relationships in business data is usually vast, ANNs have the superiority in estimating them properly (Zhang, 2004).

Table 2 provides a sample of recent business forecasting applications with ANNs reported in the literature from 1991 to 2008.

**Table 2: Applications of Neuronal Networks in Business Forecasting**

<b>Problems</b>	<b>Studies</b>
<b>Accounting earnings, earnings surprises</b>	Callen et al., (1996), Dhar & Chou, (2001).
<b>Business failure, bankruptcy,</b>	Yang, (1999), McKee & Greenstein, (2000), Atiya, (2001), Zhang, (2004).
<b>Consumer choice, market segments, market share, marketing trends, stock returns</b>	Agrawal & Schorling, (1996), Wang & Leu, (1996), West et al., (1997), Aiken & Bsat, (1999), Vellido et al., (1999), Zhang, (2004).
<b>Electricity demand</b>	Hippert et al., (2001).
<b>GDP growth, inflation, industrial production</b>	Tkacz, G. (2001), Chen et al., (2001), Tseng et al., (2001).
<b>International airline passenger volume, tourist demand, travel demand</b>	Nam & Schaefer, (1995), De Carvalho et al., (1998), Law, (2000), Zhang, (2004).
<b>Business cycles and recessions</b>	Qi, (2001).
<b>Inventory control, forecasting demand, product sales, and retail sales</b>	Tang et al., (1991), Kong and Martin, (1995), Thiesing et al., (1995), Ansuj et al. (1996), Luxhoj et al., (1996), Yip et al., (1997), Thiesing and Vornberber, (1997), Bansal et al., (1998), Shanmugasundaram et al., (2002), Charytoniuk et al., (2000), Alon et al., (2001), Kuo, (2001), Snyder, (2002), Zhang and Qi, (2002), Zhang, (2004), Chang and Wang, (2006), Cheng et al., (2006), Dong and Wen, (2006), Doganis et al., (2006), Carbonneau et al., (2008).
<b>Exchange rate</b>	Nag and Mitra, (2002).
<b>Forecasting service request</b>	Balaguer et al., (2008).

From all above-mentioned applications of neural networks, we have chosen its application in sales forecasting and we will talk over it in next sections.

## 2.3. Supply Chain

Due to the fact that in this research a) our target is sales forecasting for pharmaceutical distribution companies, and b) we referred to a pharmaceutical distribution company to gather sales data, we can say that we should focus on logistics aspect of the supply chain. However, because supply chain management (SCM) contains all logistic activities we are going to talk about different aspects of it in this section.

“The term "supply chain management" arose in the late 1980s and came into widespread use in the 1990s. Prior to that time, terms such as "logistics" and "operations management" were used in businesses instead” (Hugos, 2006). Here are some explanations of a supply chain:

- “The supply chain is the network of organizations that is involved, through upstream and downstream linkages, in the different processes and activities that produce value in the form of products and services in the hands of the ultimate consumer” (Christopher, 1992).
- “The supply chain is the network of entities through which material flows. Those entities may include suppliers, carriers, manufacturing sites, distribution centers, retailers, and customers” (Lummus and Alber, 1997).
- “The supply chain is the functions within and outside a company that enable the value chain to make products and provide services to the customer” (Cox, et al., 1998).
- “A supply chain consists of all stages involved, directly or indirectly, in fulfilling a customer request. The supply chain not only includes the manufacturer and suppliers, but also includes transporters, warehouses, retailers, and customers themselves” (Chopra and Meind, 2003; cited by Hugos 2006).

Finally, a summary definition of the supply chain (Quinn, 1997; cited by Lummus and Vokurka, 1999) can be stated as “all the activities involved in delivering a product from raw material through to the customer, including sourcing raw materials and parts, manufacturing and assembly, warehousing and inventory tracking, order entry and order management, distribution

across all channels, delivery to the customer, and the information systems necessary to monitor all of these activities”.

Actually, “there are three kinds of flow in any supply chain: material flow, information flow, and cash flow. The material flows in the downward direction and cash flows in upward direction of the supply chain, whereas the information flows in both the directions and inventory exists at all stages of the supply chain” (Routroy, and Kodali, 2005). The main objective of these inventories is to protect from ambiguity emerging from demand, process (Strader et al., 1998; cited by Routroy, and Kodali, 2005) and supply (Routroy, and Kodali, 2005). Provided that this is what a supply chain is, subsequently we can describe supply chain management as the things we do to affect the performance of the supply chain and acquire the outcomes we desire (Hugos, 2006). Supervising the flow of material from supply resources to the final client includes design, planning and the managing of supply chains. During previous decade, supply chain management has evolved into an extensive concept (Appelqvist et al., 2004). Houlihan, (1987) mentioned that SCM tries to poise contradictory actions like production, sales, distribution, and promotion. The Supply Chain Council, (1997) defined SCM as “every effort involved in producing and delivering a final product or service, from the supplier’s supplier to the customer’s customer”. Effective supervision of these actions presents chances in terms of cost and lead-time lessening and superior quality (Appelqvist et al., 2004). However, it comes into views that on explanation of “supply chain management” there is slight agreement (New, 1997; Lummus et al., 2001; Mentzer et al., 2001; Kauffman, 2002; cited by Burgess et al., 2006). Kathawala and Abdou, (2003) presented that SCM “has been poorly defined and there is a high degree of variability in people’s minds about what is meant”. Some definitions of supply chain management are:

- “An integrating philosophy to manage the total flow of a distribution channel from supplier to ultimate customer” (Ellram and Cooper, 1993).
- “A set of approaches to integrate suppliers, manufacturers, warehouses, and stores efficiently, so that merchandise is produced and distributed at right quantities, to the right locations, and at the right time, in order to minimize system-wide costs while satisfying service level requirements” (Simchi-Lev et al., 2000; cited by Routroy, and Kodali, 2005).

- “The systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole” (Mentzer et al, 2001). Obviously, this definition is general and it is not limited to any specific scope or field.
- “The coordination of production, inventory, location, and transportation among the participants in a supply chain to achieve the best mix of responsiveness and efficiency for the market being served” (Hugos, 2006).

Essentially, “the objective of SCM is to reduce or minimize total cost, improve total quality, maximize customer service and increase profits” (Boubekri, 2001; cited by Wagner and Alderdice, 2006). Efficient supply chain management needs simultaneous developments in both client service levels and the inside operating effectiveness of the corporations in the supply chain (Hugos, 2006). In fact, organizations in any supply chain ought to decide both independently and cooperatively concerning their activities in five fields (Hugos, 2006):

- Production
- Inventory
- Location
- Transportation
- Information

### **2.3.1. Inventory and Inventory Management**

As our basic purpose in this research refers to sales forecasting in order to help inventory control, we are going to talk about inventory further than the other areas (production, location, transportation, and information).

Inventory is extended all over the supply chain and comprises everything as of raw material to finished products that are kept by the producers, distributors, and retail dealers inside a supply chain (Hugos, 2006). “Holding large amounts of inventory allows a company or an entire supply chain to be very responsive to fluctuations in customer demand” (Hugos, 2006). “However, the creation and storage of inventory is a cost and to achieve high levels of efficiency, the cost of inventory should be kept as low as possible” (Hugos, 2006). “Managers must decide where they want to position themselves in the trade-off between responsiveness and efficiency”

(Hugos, 2006). Accordingly, the two key goals of supply chain inventory planning are (Routroy, and Kodali, 2005):

- (1) Keeping the particular service level to the client, that is accessibility of supplies at the proper time and at the appropriate position (Routroy, and Kodali 2005).
- (2) Decreasing the stock levels in the supply chain to reduce the capital invested in inventory (Routroy, and Kodali, 2005).

In fact, “Inventory management is a set of techniques that are used to manage the inventory levels within different companies in a supply chain” (Hugos, 2006). “The aim is to reduce the cost of inventory as much as possible while still maintaining the service levels that customers require” (Hugos, 2006). “The inventory management content refers to the ongoing provision of standard items with independent demand, where some speculative quantity should always be on hand” (Buxey, 2006). Businesses keep these supplies for diverse causes, inclusive of defense against broad lack of products or probable difficulties with providers; otherwise, since unit price increases may be coming up (Buxey, 2006).

### **2.3.2. Participants in the Supply Chain**

In its easiest outline, a supply chain includes a corporation, the providers and clients of that company (Hugos, 2006). Comprehensive supply chains include three further kinds of contributors (Hugos, 2006). Primarily, there is the supplier's provider or the final provider at the commencement of an extended supply chain. Then there is the customer's customer or ultimate customer at the end of an expanded supply chain. Ultimately, there are firms who provide services in marketing and advertising, IT, logistics, and finance (Hugos, 2006). As our main goal is sales forecasting for a pharmaceutical distribution company, we have to define distributors more specifically below:

- **Distributors**

Distributors (wholesalers) are corporations that catch inventory in large quantities from manufacturers and distribute a mass of connected product lines to clients (Hugos, 2006). “A distributor is typically an organization that takes ownership of significant inventories of products

that they buy from producers and sell to consumers" (Hugos, 2006). A wholesaler can also be a company that only acts as broker between the manufacturer and the client and on no account obtains possession of those goods. This type of wholesaler carries out mostly the actions of marketing and sales (Hugos, 2006). In both these instances, as the demands of clients develop and the variety of accessible goods modifies, the distributor is the representative that constantly follows customer demands and accords them with goods that are reachable (Hugos, 2006). However, in Iran, pharmaceutical distribution companies usually buy products from manufacturers and they are not brokers.

### **2.3.3. Logistics**

Early resources of logistics were come up mainly for military uses (Lummus et al., 2001). In an article in 1898, logistics was defined as: "the art of handling troops in the theatre of war; tactics that of handling them on the field of battle" (Simpson and Weiner, 1898; cited by Lummus et al., 2001). Ultimately, the applications of logistics have shifted into the business field (Lummus et al., 2001). Even though numerous business arenas have individually described logistics, one organization, APICS, described logistics in the industrial, and business outlooks (Cox et al., 1998):

In an industrial context, logistics refers to "the art and science of obtaining, producing, and distributing material and product in the proper place and in proper quantities. In a military sense (where it has greater usage), its meaning can also include the movement of personnel" (Cox et al., 1998).

However, in a business context, logistics was defined as: "the management of all inbound and outbound materials, parts, supplies, and finished goods" (Cavinato, 1982; cited by Lummus et al., 2001). Logistics has been considered as one of the key issues in the field of classical marketing theory (McCarthy, 1964; Borden, 1964; cited by Svensson, 2002). Council of Logistics Management (CLM) in 1998 defined logistics as:

"The process of planning, implementing, and controlling the efficient, effective flow and storage of goods, services and related information from the point of origin to the point of consumption for the purpose of conforming to customer requirements".

- **The relationship of logistic and supply chain management**

In fact, there is a distinction between the theory of supply chain management and the customary theory of logistics (Hugos, 2006). Actually, Logistics usually refers to “activities that occur within the boundaries of a single organization and supply chains refer to networks of companies that work together and coordinate their actions to deliver a product to market” (Hugos, 2006). In other words, supply chain management includes the managing of supplies, information and finance in network containing providers, producers, wholesalers and customers (Stanfield, 2002). Basically, “all these activities are intended to deliver the optimal result to the end-user via procurement of raw materials, manufacturing, distribution, and customer services” (Kim et al., 2004; cited by Symeonidis et al., 2006). Johnson and Wood, (1996) stated that “supply chain management is somewhat larger than logistics.” Cooper et al., (1997) mentioned that a “contemporary understanding of SCM is not appreciably different from the understanding of integrated logistics management”.

#### **2.3.4. Why Demand Forecasting is Essential in a Supply Chain?**

Demand forecasts have a fundamental position for supply chain management. The future demand of a specific product is the basis for the particular replenishment structures (Aburto and Weber, 2007). While the entire area of SCM has many different aspects, we want to focus in this research on topics related to sales forecasts for distribution companies. Precise sales prediction is utilized for capturing the trade-off between customer demand satisfaction and inventory costs (Gupta, et al., 2000). Failure to be responsible for considerable product demand variations may cause excessively high manufacturing charges relating to high inventory costs or displeased clients’ demand and dropping market share (Gupta, et al., 2000). Actually, accurate forecasting is significantly important in a supply chain performance since growing forecast mistakes will add to the cycle time and reduce perceived service status in a supply chain (Ganeshan et al., 2001). Due to tough existing competition, companies are working very hard to decrease their costs in order to boost their competitive advantages. Actually, “one of the major purposes of supply chain collaboration is to improve the accuracy of forecasts” (Raghunathan, 1999) as superior quality of forecasts would finally cause general cost savings (Carboneau et al., 2008). Thus, analysis of forecasting techniques is of considerable value for firms. In our research, we aim to make precise

sales prediction for pharmaceutical distribution companies in order that they can make correct decision on how much drugs they should buy from manufacturers in a specific period and this issue would also help manufacturers to plan for their production and decide how much drugs they should produce in a given period. This matter would increase the level of forecasts in pharmaceutical supply chain and eventually cause general cost reductions (Carboneau et al., 2008). Carboneau et al. in 2008 suggested that if an increase in forecasting accuracy can be achieved, it will result in lower costs because of reduced inventory as well as increased customer satisfaction that will result from an increase in prompt deliveries.

## 2.4. Forecasting

In this section, we will review forecasting and especially sales forecasting methods and application of neural networks in this area. We also make comparison between neural networks and other time series models.

Prediction of the future's conditions and events is called forecast and the methods and procedures of this act is called forecasting (Bowerman and O'Connell, 1986). As Thomassey et al., 2005 stated, forecasting is widespread in numerous arenas including energy production, economics, meteorology, business, finance, sociology, etc. Since forecast of future's events acts an important role in decision making process, forecasting is really essential for most organizations and companies. Each organization has to be able to forecast in order to make intelligent decisions; especially commercial companies in all their procedures need forecasting of future events and conditions. People, especially managers, need to make critical decisions based on what happened before (Bowerman and O'Connell, 1986). "To enhance the commercial competitive advantage in a constantly fluctuating environment; an organization's management must make the right decision in time depending on the information at hand" (Kuo, 2001). "The decision lead-time ranges from several years to several hours based on the types of business. Thus, making an accurate decision plays a prominent role" (Kuo, 2001). We bring some examples of conditions that need predictions below (Bowerman and O'Connell, 1986):

- **In marketing management:** For planning the methods of selling, a reliable demand forecasting (sales forecasting) should be done. For degree of needed efforts and activities, overall amount of demand should be predicted. Finally, for running effective advertising

campaign, demand of different markets and different consumer groups should be predicted (Bowerman and O'Connell, 1986).

- **In strategy management:** For long term decision making, a company should predict general conditions of economics, fluctuations of prices and expenses, advances in technology, market grow, etc. Prediction of these cases plays an important role in determining whether new investment and equipment will be needed or not (Bowerman and O'Connell, 1986).
- **In production planning:** Product demand forecasting in special periods is essential in each production line as it helps production planning and inventory control. For estimating amount of raw material that should be bought, a manufacturer can extend product demand foresting to forecast of required raw material (Bowerman and O'Connell, 1986).

Forecasting is also essential in financial management, personnel management, process control, and so on. If we want to predict future's events, we should rely on previous and past events. It means that we should analyze past data and information in order to make prediction for the future; also, we should assume that the prediction model would not be changed in future and if any change happens, we should make required changes in the model (Bowerman and O'Connell, 1986). Usually, there are two possible roads to victory (Herrera, 1999):

1. “One can predict the future by extrapolating directly from past observations” (Herrera, 1999).
2. “One can try to make a model that explains relationships between observations made in the past, and then use that model in a simulation to make predictions of the future” (Herrera, 1999).

In fact, “modeling is the better approach, since it allows to correlate different observations with each other, thereby improving the chances of making correct predictions (Herrera, 1999)”. Also, “the modeling approach enables the user to work with input variables, thereby allowing him or her to formulate different scenarios and observe the consequences that might result when implementing any one of these scenarios” (Herrera, 1999).

As mentioned before, this thesis is concerned with the development of time series sales forecasting models for short shelf-life medical products in distribution companies. Accordingly, we are going to talk about sales forecasting and time series forecasting in next sections.

### **2.4.1. Sales Forecasting**

“Sales forecasting is a valuable management tool for determining the future magnitudes, timing, and possible effects of uncontrollable influences over a company's future success” (Yip et al., 1997). “It is an estimate of the expected demand for a company's products, and it is an indispensable element of management planning for many major company activities, e.g. marketing, production, finance, research and development, personnel, capital investment determination, purchasing, inventory and warehousing” (Yip et al., 1997). Sales forecasting is extremely complicated due to the effects of inside and outside environments. Nevertheless, trustworthy forecasting of sales can advance the level of business strategy (Kuo, 2001). “In order to enhance the competitive advantage of a commercial enterprise in a constantly fluctuating environment, the manager must make the right decisions on time depending on the information at hand” (Kuo, 2001). “Theoretically, if a marketing department can estimate the sales quantity for the next period, the materials department can then effectively control the inventory to achieve just-in time (JIT) delivery” (Kuo, 2001). Sales forecasting always plays a prominent role in a decision support system. Efficient sales forecasting in advance can facilitate the decision maker to estimate amount and cost of production and inventory, even find out the sale price (LeVee, 1992; cited by Kuo, 2001) which will result in a lower inventory level and achieve the objective of just-in-time (Kuo, 2001).

### **2.4.2. Time Series Forecasting**

A time series is a chain of measurements, usually got at consecutive positions in time (STATISTICA 7, Electronic Manual). Time series analysis has two key objectives: (a) “identifying the nature of the phenomenon represented by the sequence of observations, and (b) forecasting (predicting future values of the time series variable)” (STATISTICA 7, Electronic Manual). Time series analysis and prediction have been successfully used in a wide range of knowledge fields (Box and Jenkins, 1976; Makridakis et al., 1998; Brockwell and Davis, 2002). There is abundant literature on techniques of examining time series (Herrera, 1999). There even were organized competitions on a worldwide level in order to advance the state of the art of methodologies for time series analysis and prediction (Makridakis and Hibon 1997; Makridakis et al. 1984; Weigend and Gershenfeld 1994; cited by Herrera, 1999). Among many existing

methods, two famous and efficient models that are often used for time series prediction are the following ones:

1. **Linear models:** Auto-regressive integrated moving average (ARIMA) models (Box and Jenkins, 1976; Makridakis et al., 1998). ARIMA processing has been presented to be the most successful approach to model a broad variety of time series (Makridakis et al., 1998). Despite the fact that the Box-Jenkins is an old approach, it is still a dominant player in the forecasting arena (Alon et al., 2001). However, these models do not execute properly while there are elements of the time series that demonstrate a nonlinear behavior. In this case, other models, such as neural networks, must be applied.

- **ARIMA Modeling**

ARIMA models (p, d, q) “can be viewed as linear filters from the point of view of digital signal processing” (Balaguer et al., 2008). Specifically, the three types of parameters in the model are: the autoregressive parameters (p), the number of differencing passes (d), and moving average parameters (q). The time structure of these filters is the following models (Box and Jenkins, 1976; Makridakis et al., 1998; cited by Balaguer et al., 2008):

$$y(k) = a_1 \cdot y(k-1) + \dots + a_n \cdot y(k-n) + e(k) + b_1 \cdot e(k-1) + \dots + b_{m-1} \cdot e(k-m+1) + C$$

“where  $y(k)$  is the variable to be predicted using previous samples of the time series,  $e(i)$  is a sequence of i.i.d. (Independent and identically distributed) terms which have zero mean, and  $C$  is a constant” (Balaguer et al., 2008). ARIMA modeling is made up of two processes (Box and Jenkins, 1976; Balaguer et al., 2008):

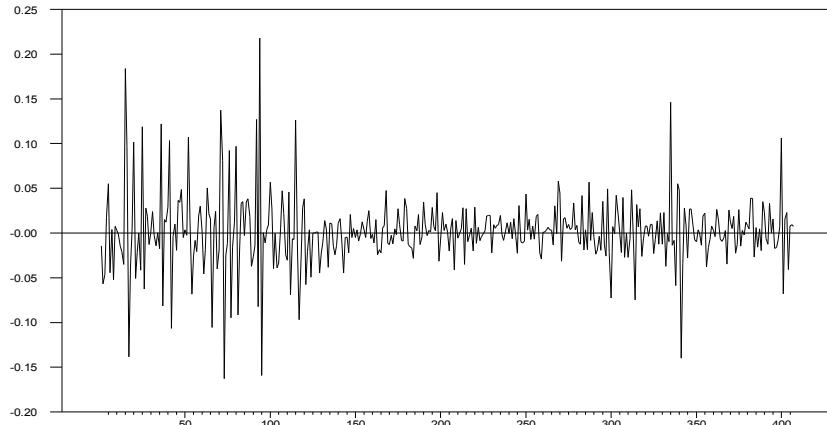
- **Autoregressive process:** “Those terms involving coefficients  $a_i$  ( $i = 1, \dots, n$ ) carry out a linear relationship between the value predicted by the model at time  $k$  and the past values of the time series; this part is known as auto-regressive (AR)” (Balaguer et al., 2008).
- **Moving average process:** “Those terms involving coefficients  $b_i$  ( $i = 1, \dots, m$ ) carry out a linear relationship between the value predicted by the model at time  $k$  and a Gaussian distribution of i.i.d. samples. This part is known as moving average (MA) and can be considered as the autoregressive structure of the residuals” (Balaguer et al., 2008).

Accordingly, “each observation is made up of a random error component (random shock) and a linear combination of prior observations” (STATISTICA 7, Electronic Manual). Various phases must be followed to achieve the coefficients (Makridakis et al., 1998; cited by Balaguer et al., 2008):

**(a) Stationary Series:** A time series is supposed to be ‘strictly stationary’ provided that the joint distribution of  $(y(1), \dots, y(k))$  is the same as that of  $(y(1+t), \dots, y(k+t))$  for all  $t$  (Tsay, 2005). In other words, ‘strictly stationary’ needs that the joint distribution of  $(y(1), \dots, y(k))$  is invariant under time shift. A time series  $\{y(t)\}$  is ‘weakly stationary’ provided that both the mean of  $y(t)$  and the covariance between  $y(t)$  and  $y(t-l)$  are time-invariant, where  $l$  is an arbitrary integer (Tsay, 2005). More exclusively,  $\{y(t)\}$  is weakly stationary provided that (a)  $E(y(t)) = \mu$ , which is a constant, and (b)  $Cov(y(t), y(t-l)) = \gamma_l$ , which only depends on  $l$  (Tsay, 2005). For example, Figure 20 shows a nonstationary time series and Figure 21 belongs to a stationary time series.



**Figure 19: An Example of a Non-Stationary Time Series**



**Figure 20: An Example of a Stationary Time Series**

“Therefore, the first step is based on transformations of the time series in order to achieve the desired effect in the autocorrelation” (Balaguer, 2008). Usually, a differentiation is enough to make the series stationary (Balaguer, 2008). If we perform the differentiation between consecutive values, it is named simple differentiation. If not, we perform the differentiation between periods of the time series, it is identified as seasonal differentiation (Balaguer, 2008). Generally, considerable alters in level (strong upward or downward changes) require first order non seasonal (lag=1) differencing; tough alters of slope typically need second order non seasonal differencing. Seasonal patterns need respective seasonal differencing (STATISTICA 7, Electronic Manual).

**(b) Identification:** In this phase, “the order of both the AR and the MA part must be estimated. Autocorrelation and partial autocorrelation computation is used to obtain these orders” (Balaguer, 2008). In other word, at this stage, we should make a decision about numbers autoregressive (p) and moving average (q) parameters that are essential to give in an effectual but parsimonious model of the process (parsimonious means that it has the fewest parameters and greatest number of degrees of freedom among all models that fit the data) (STATISTICA 7, Electronic Manual). In practice, the numbers of the p or q parameters rarely need to be greater than 2.

- **Autocorrelation Function (ACF)**

The correlation coefficient between  $y(t)$  and  $y(t-l)$  is named the lag-l autocorrelation of  $y(t)$  and is generally indicated by  $\rho_l$ , that in the process of the weak stationary assumption is just a function of  $l$  (Tsay, 2005). ACF is defined as follows (Tsay, 2005):

$$\rho_l = \frac{Cov(y(t), y(t-l))}{\sqrt{Var(y(t))Var(y(t-l))}} = \frac{\gamma_l}{\gamma_0}$$

This function is used to determine the order of the MA part of the model (Tsay, 2005). However, the main tool applied in the identification step to determine the order of MA is the correlograms of autocorrelation (ACF) (STATISTICA 7, Electronic Manual).

- **Partial Autocorrelation Function (PACF)**

“The PACF of a time series is a function of its ACF and is a useful tool for determining the order of the AR model. A simple yet effective way to introduce PACF is to consider the following AR models in consecutive orders” (Tsay, 2005):

$$\begin{aligned} y(t) &= \phi_{0,1} + \phi_{1,1}y(t-1) + e_{1t}, \\ y(t) &= \phi_{0,2} + \phi_{1,2}y(t-1) + \phi_{2,2}y(t-2) + e_{2t}, \\ &\vdots \quad \vdots \end{aligned}$$

Where  $\phi_{0,j}$ ,  $\phi_{i,j}$  and  $\{e_{jt}\}$  are, in that order, the constant term, the coefficient of  $y(t-i)$ , and the error term of an AR ( $j$ ) model. The approximation  $\hat{\phi}_{1,1}$  of the first equation is named the lag-1 sample PACF of  $y(t)$ . The estimation  $\hat{\phi}_{2,2}$  of the second equation is described as the lag-2 sample PACF of  $y(t)$ , and so on (Tsay, 2005). The main tool applied to recognize the order of the AR model is the correlograms of partial autocorrelation (PACF) (STATISTICA 7, Electronic Manual).

**(c) Estimation:** In this stage, the parameters are calculated applying function minimization processes, in order that the sum of squared residuals is kept to a minimum (STATISTICA 7, Electronic Manual). The estimates of the parameters are applied in the forecasting phase to estimate novel values of the series (STATISTICA 7, Electronic Manual). “In fact, the estimation

process is performed on transformed (differenced) data; before the forecasts are generated, the series needs to be integrated (integration is the inverse of differencing) so that the forecasts are expressed in values compatible with the input data" (STATISTICA 7, Electronic Manual). "In addition to the standard autoregressive and moving average parameters, ARIMA models may also include a constant" (STATISTICA 7, Electronic Manual). "The interpretation of a (statistically significant) constant depends on the model that is fit" (STATISTICA 7, Electronic Manual).

**(d) Selection:** The most appropriate models are selected. These models ought to present sufficient predictions (Balaguer, 2008). With the purpose of assessing models, data are divided into two groups: training and validation. The training set is applied to create the model, and the validation set to assess it (Balaguer, 2008). Furthermore, the model ought to have a small number of numbers of parameters (as few as probable). Usually, "Akaike information criterion (AIC) based on information theory is used to achieve a trade-off between both requirements (an adequate prediction and a few numbers of parameters)" (Weigend and Gershenfeld, 1996; cited by Balaguer, 2008).

**(e) Diagnosis:** A proper model provides statistically autonomous residuals that include just noise and no systematic components i.e. the auto-correlogram of residuals is not supposed to display any sequential dependencies (STATISTICA 7, Electronic Manual). "A good test of the model is (a) to plot the residuals and inspect them for any systematic trends, (b) to examine the auto-correlogram of residuals (there should be no serial dependency between residuals)" (STATISTICA 7, Electronic Manual), and (c), "Autocorrelation and partial autocorrelation of the residues are calculated to test whether they are statistically relevant. Ideally, residues should be i.i.d" (Balaguer, 2008).

"ARIMA processing has shown to be the most effective tool to model a wide range of time series" (Makridakis et al., 1998; cited by Balaguer et al., 2008). However, it should be noticed that in time series forecasting issues, it is recommended that as a minimum 50 or even 100 observations are required to provide linear ARIMA models (Box and Jenkins, 1976). In addition, "these models do not work properly when there are elements of the time series that

show a nonlinear behavior. In this case, other models, such as time processing neural networks, must be applied” (Balaguer, 2008).

**2. Nonlinear models:** Neural networks models and in particular feed forward neural networks that have been explained in detail previously. One of the main application fields of ANNs is forecasting (Zhang et al., 1998; Hippert, et al., 2001). In Table 3, we can see a comprehensive list of neural networks’ applications in sales or demand forecasting. Actually, ANNs have obtained growing notice in forecasting theory, resulting in victorious applications in time series sales forecasting (Thiesing and Vornberber, 1997; Makridakis et al., 1998; Reed and Marks, 1999; cited by Crone, 2003). In fact, artificial neural network has recently captivated a great deal of notice as a novel method for prediction in finance, business and economics (Moshiri and Cameron, 2000). “The chief advantages of this new approach are that such models can usually find a solution for very complex problems, and that they are free from the assumption of linearity that is often adopted to make the traditional methods tractable” (Moshiri and Cameron, 2000). Abundant researches have certificated the victory of ANNs in prediction of financial data (De Gooijer and Hyndman, 2006).

**Table 3: Related Works on Applications of Neural Networks in Sales and Demand Forecasting**

Author	Year	Article
Tang et al.	1991	Times series forecasting using neural networks vs. Box–Jenkins methodology
Kong and Martin	1995	A back-propagation neural network for sales forecasting
Thiesing et al.	1995	Short term prediction of sales in supermarkets
Ansuj et al.	1996	Sales forecasting using time series and neural networks
Luxhoj et al.	1996	A hybrid econometric-neural network modeling approach for sales forecasting
Yip et al.	1997	Application of Artificial Neural Networks in Sales Forecasting
Thiesing and Vornberber	1997	Sales forecasting using neural networks

<b>Bansal et al.</b>	1998	Neural Networks Based Data Mining Applications for Medical Inventory Problems
<b>Bansal et al.</b>	1998	Neural Networks Based Forecasting Techniques for Inventory Control Applications
<b>Charytoniuk et al.</b>	2000	Neural-network-based demand forecasting in a deregulated environment
<b>Alon et al.</b>	2001	Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods
<b>Kuo</b>	2001	A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm
<b>Shanmugasundaram et al.</b>	2002	Use of Recurrent Neural Networks for Strategic Data Mining of Sales Information
<b>Snyder</b>	2002	Forecasting sales of slow and fast moving inventories
<b>Zhang and Qi</b>	2002	Predicting consumer retail sales using neural networks
<b>Zhang</b>	2004	Tourism Demand Forecasting for the Tourism Industry: A Neural Network Approach. A Chapter In: Neural Networks in Business Forecasting
<b>Chang and Wang</b>	2006	Fuzzy Delphi and back-propagation model for sales forecasting in PCB industry
<b>Cheng et al.</b>	2006	Implementation of a Back-Propagation Neural Network for Demand Forecasting in a Supply Chain-A Practical Case Study
<b>Dong and Wen</b>	2006	An Improved Neural Networks Prediction Model and Its Application in Supply Chain
<b>Doganis et al.</b>	2006	Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing
<b>Carboneau et al.</b>	2008	Application of machine learning techniques for supply chain demand forecasting

- **Differences between ARIMA and NNs:**

“There are many contrasting features in ARIMA and ANNs. First, ARIMA is linear and ANNs are inherently nonlinear” (Zhang, 2004). “ARIMA is a very comprehensive linear model and can represent many types of linear relationships, such as autoregressive (AR), moving average (MA), and mixed AR and MA time series structures” (Zhang, 2004). Furthermore, several exponential smoothing models can be signified by ARIMA models (Mckenzie, 1984). In contrast, ANNs are widespread practical approximators and are able to model any kind of nonlinear relationship (Zhang, 2004). Next, ARIMA is fundamentally parametric though ANN

is naturally non-parametric. Consequently, in applying ARIMA, the common model outline is known though in applying ANNs, we do not require to identify a special model outline (Zhang, 2004). Certainly, linear ARIMA is reasonably easy to perceive and apply in comparison with ANNs. Lastly, ARIMA models are evolved according to approved statistical theory, and statistical methods are frequently applied in model evolution and sufficiency examination (Zhang, 2004). However, there is not any methodical technique in neural network model building and ANNs are frequently behaved as black boxes (Zhang, 2004).

### **2.4.3. Time Series Forecasting Models**

Prior to early 1920s, predictions were conducted by extrapolating time series (Weigend, and Gershenfeld, 1993). What might be named as “modern forecasting” started in 1927, once Yule demonstrated auto-regressive techniques to predict the yearly quantity of sunspots (Yule, 1927). His model calculated predictions as a weighted sum of earlier data (Diez and Fernández, 2005). Provided that acceptable performance was to be reached via this linear method, an external element named noise had to be provided for, since this noise influences the linear method (Diez and Fernández, 2005). This linear method with noise was broadly applied for the next 50 years, when study concluded in the ARIMA method suggested by Box and Jenkins (Box and Jenkins, 1976; cited by Diez and Fernández, 2005). From that time, the methods that have been applied in sales forecasting have normally been time series techniques that can be divided to linear and nonlinear, influenced by the nature of the model they are dependent on (Doganis et al., 2006). Linear models, such as autoregressive moving average (ARMA) (Box and Jenkins, 1976) and autoregressive integrated moving average (ARIMA) (Box et al., 1994) are the most famous methodologies, but their prediction ability is restricted by their supposition of a linear behavior; consequently, it is not always acceptable (Zhang, 2003; cited by Doganis et al., 2006).

During the 1980s, two critical progresses occurred that influenced the evolution of time series study two crucial developments took place that affected the evolution of time series research. On the one hand, the advancements of computer science resulted in using far more complex algorithms. On the other hand, the evolution of machine learning techniques, such as artificial neural networks leaded to significant progress in time series domain (Diez and Fernández, 2005). Actually, researchers introduced a number of nonlinear methodologies such

as artificial neural networks (which are the most famous ones in time series forecasting domain) in order to address possible nonlinearities in time series modeling (Darbellay and Slama, 2000).

De Gooijer and Hyndman, (2006) offered a literature concerning time series forecasting models, covering the period 1982–2005. First, they divided time series models to linear models and nonlinear ones. Then, they became narrow on different kinds of each model and discussed about strength and weakness of each method. Among many introduced linear models in their paper, we can name Holt Winters' Exponential Smoothing methods, State Space models, ARIMA models, ARARMA models, Vector ARIMA (VARIMA) models, Vector Autoregression (VAR) models, BVAR models, and ARMAX models. They also presented different nonlinear models such as, Nonlinear Exponential Smoothing methods, SETAR models, CTAR models, STAR models, FCAR models, VFCAR models, Artificial Neural Networks (ANNs), Fractionally Integrated ARMA (ARFIMA) models, ARCH models, and Generalized ARCH (GARCH) models.

Among many kinds of existing methods for forecasting, we should choose the best one for our approach. Accordingly, in the following part, we bring a review of the evaluation of different time series forecasting methods. In fact, most of them compare performance of neural networks with linear traditional methods. Since both ARIMA and exponential smoothing models are vigorous time series forecasting methods, they are usually employed more than other statistical methods as benchmarks of comparison to neural networks.

Noticeably, the main question concerns the precision of each modeling technique. Accordingly, a number of studies have been performed to evaluate different techniques and the outcomes are not obviously supporting one specific method (Doganis et al., 2006). Although various comparative studies between traditional models and neural networks have been carried out, outcomes are varied with regard to whether ANNs are better than the linear methods in forecasting or not (Adya and Collopy, 1998; Zhang et al., 1998). Several studies have offered practical confirmation on the comparative superiority of one model to the other in diverse forecasting cases (Hill et al., 1996; Prybutok et al., 2007). Several researches have concluded that ANNs are superior to conventional techniques (Weigen et al., 1991; cited by Kuo et al., 2002), whereas others have achieved a contradictory result (Tang et al., 1991; cited by Kuo et al., 2002). Following survey displays comparisons of various time series forecasting methods that developed.

#### **2.4.4. Comparisons of Different Time Series Forecasting Models**

Following survey displays comparisons of diverse time series forecasting methods that developed during 1991-2008.

Since generation of neural networks, only few studies have concluded that traditional methods have superiore to or at least the same performance as neural networks. For instance, Tang et al. (1991) reported the results of a study that compared the performance of neural networks and traditional methods in time series forecasting. Their research demonstrated that for time series with long memory both ANNs and Box-Jenkins models had the same performance. Furthermore, Foster et al. (1992) noticed that exponential smoothing is better than neural networks in forecasting yearly data, and comparable in forecasting quarterly data. Although it appears that neural networks may outperform conventional statistical methods in forecasting time series with trend and seasonal patterns, Nelson, et al. (1994) found that neural networks cannot properly model the seasonal patterns in their data. Also, Callen et al. (1996) showed that forecasting performance of linear time series models was better than that of neural networks even if the data were nonlinear. Moreover, Church and Curram (1996) and Ntungo and Boyd (1998) demonstrated that neural networks performed nearly the same as econometric and ARIMA approaches. Afterwrads, Heravi et al. (2004) showed that linear models generate more reliable prediction results than neural networks for European industrial production series.

In contrast, many reserches concluded that neural networks had better forecasting performance than linear methods. As an illustration, in a comparative study of the performance of neural networks and conventional methods in forecasting time series, Tang et al. (1991) found that ANNs surpassed the Box–Jenkins in short term forecasting. Then, Chakraborty et al. (1992) applied neural networks approach to multivariate time-series analysis. They precisely forecasted the flour prices in three cities in USA. According to their outcomes, neural networks approach was better than classic methods.

In a succeeding study, Ansuj et al. (1996) compared the ARIMA model with interventions and ANN model in examining the behavior of sales in a medium size corporation. The outcomes proved that ANN model is more precise. Hill et al. (1996) also demonstrated that neural networks were considerably superior to traditional techniques when predicting quarterly

and monthly data. Agrawal and Schorling (1996) compared the test results of neural networks with multinomial logit model of forecasting and they proved that neural networks could predict brand shares more accurately. Furthermore, Yip et al. (1997) examined application of neural networks in sales forecasting. Applying several measures of accuracy, the result of the evaluation confirmed that the neural networks predict better than the time series smoothing methods of forecasting.

Afterwards, Zhang et al. (1998) did an inclusive review of the literature regarding the employment of ANNs in various forecasting domains. In most cases, ANNs were better than linear methods. Kuo and Xue (1998), proved that their developed artificial intelligent could find non-linear relationships better than conventional time series methods. Elkateb et al. (1998) also compared neural networks with ARIMA models in peak load forecasting. The results of their study demonstrated that neural networks had better forecasting performance than ARIMA models. In fact, neural networks have surpassed conventional forecasting methods in definite conditions (Hornik et al., 1989; cited by Gruca et al. 1999) for example, when there are nonlinearities (White and Stinchcombe, 1992; cited by Gruca et al. 1999), also when there are considerable interactions among inputs (Rumelhart and McClelland, 1986; cited by Gruca et al. 1999).

Subsequently, Ainscough and Aronson (1999) compared neural network and regression analysis, as a linear model, in modeling and predicting the results of retailer activity on the sales of definite products applying scanner data. According to the results of their study, neural networks had better performance than regression model. Then, Qi (2001) reported that ANNs are very likely to do better than other methods when the required data are saved as new as probable. Although statistical techniques have been proven effective for a long time, they still have definite drawbacks (Kuo and Xue 1998; Kuo and Cohen 1998; Kuo et al. 2002; Chen and Ou 2008). For example, when the data are influenced by particular conditions, like promotion, the prediction results of conventional methods are undesirable (Kuo and Cohen 1998; Kuo et al. 2002). In addition, it is always logical to anticipate to have a noteworthy level of non-linearity in sales behavior, (Carboneau et al. 2008). However, traditional methods are not capable of modeling nonlinear relationships, so most researchers apply neural network models to cope with forecasting problems. Although the contribution of neural networks compared to the traditional

methods seems degraded in some cases (Hill et al. 1994), their vigorous capability to model nonlinear relations and their adaptation are highly attractive for most forecasting subjects (Zhang, 2004). Therefore, advanced methods like neural networks could serve as more appropriate approximator (Zhang, 2004) and would be more suitable for the time series sales forecasting than linear models like ARIMA.

However, Adya and Collopy (1998), in their review paper, presented that regardless of growing applications of neural networks in prediction and more specifically in business prediction, outcomes and opinions concerning their contribution are varied. Adya and Collopy (1998) concluded that assessing studies in this area is complicated, owing to lack of obvious criteria. It means that, although neural networks have acquired growing attention in forecasting domain, resulting in competent applications in time series sales forecasting (Crone, 2003), some studies point out that no single approach works best in every condition, and combining diverse methods is an efficient and effectual way to advance forecasting accuracy (Zhang, 2003). Consequently, some recent researches offered good explanations of the hybrid ARIMA–ANN models or combination of other conventional and ANN techniques (Zhang, 2003; Aburto and Weber 2007; Aslanargun et al. 2007; Jain and Kumar, 2007; Diaz-Robles et al. 2008). As an illustration, Zhang (2003) recommended a hybrid ARIMA- ANN approach in which ARIMA was applied to model the linear part and ANNs was used to model the prediction errors. He showed that the hybrid method outperformed both separate methodologies. Furthermore, in a research by Kuo, et al. (2009), a hybrid algorithm founded on radial basis function neural network for sales prediction. They presented “hybrid of particle swarm and genetic algorithm based optimization (HPSGO) algorithm” congregated advantages of “particle swarm optimization (PSO) and genetic algorithm” to advance the learning performance of RBF neural network (Kuo, et al., 2009). Outcomes of the research proved that the presented HPSGO algorithm had better performance than PSO, genetic algorithm, and Box-Jenkins model (Kuo, et al., 2009). Also, Khashei and Bijari (2010) proposed a new hybrid neural networks model. They applied an artificial neural network (p,d,q) model for time series forecasting to have a more precise forecasting model than neural networks. The results of three real data sets pointed out that the proposed model was an appropriate way to advance forecasting accuracy accomplished by neural networks (Khashei and Bijari 2010). Both theoretical and empirical results have confirmed that combination of various models is a successful way of enhancing the performance

of forecasting models (Khashei and Bijari 2010). Accordingly, it appears to be logical to apply hybrid models in most sales forecasting domains.

According to the presented survey, numerous prediction methods have been offered and each method has its specific advantages and disadvantages in comparison with other techniques. However, none of the accomplished studies described the applications of hybrid linear and nonlinear neural networks in forecasting. They also did not offer a novel technique for handling the problem of not having enough past records for forecasting. However, owing to the specific constraints of the pharmaceutical products, numerous and new items with few historical data, existing forecasting methods are generally inappropriate. This motivates the evolution of a novel hybrid approach, which combines both linear and nonlinear methods and their relevant strengths. Accordingly, in this research, the use of hybrid neural network by using each medicine's past records and its group members' past records to make precise sales prediction for pharmaceutical distribution companies is examined.

After getting the result in chapter 4, the accuracy of our proposed methodology will be proved.

# **CHAPTER 3**

## **RESEARCH METHODOLOGY**

### ***3. Research Methodology***

*In this chapter, the research methodology that is used to achieve the objectives and respond the research question is explained. It begins with research design; it demonstrates research purpose, research approach, and research strategy. Then, it is followed by research process that was used in this thesis. All the methods applied in each phase of modeling: data collection, data preprocessing, ARIMA modeling, hybrid ANNs approach, gaining prediction results of the models used in this research and model evaluation are stated in this chapter.*

### 3.1. Research Design

The research design is a framework for accomplishing marketing research and it is the foundation for performing the project (Malhotra, 2006). Accordingly, it is a fundamental plan that directs the data gathering and analysis parts of the research. It identifies data by the kind of information to be gathered, the sources of the data, and the data collection process (Javaheri, 2007). “A good research design will ensure that the information collected will be consistent with the objectives of the study and that the procedures regarding data collection is accurate and efficient” (Javaheri, 2007). The research design of this study is demonstrated in Figure 22 and it is described in more detail below.

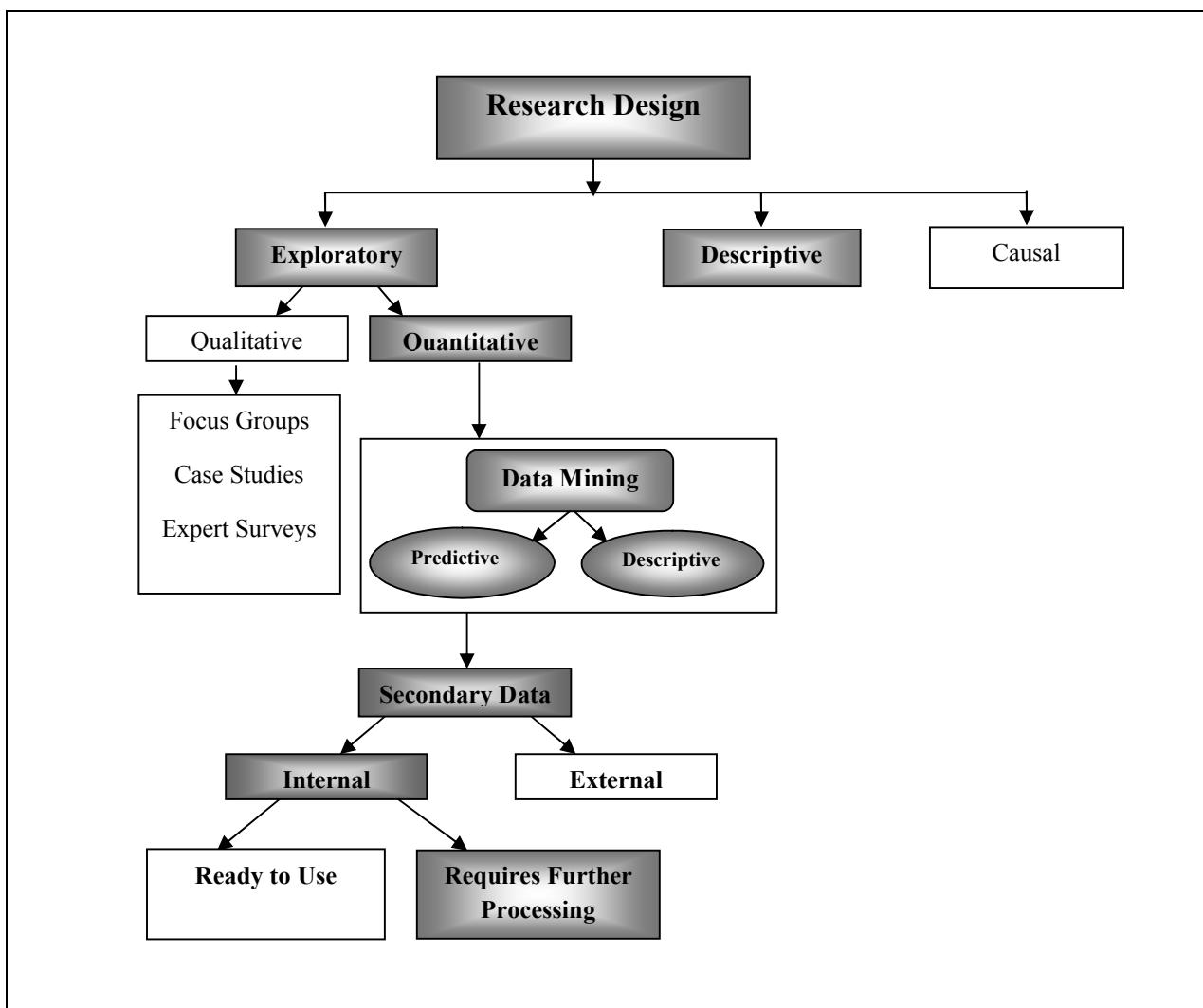


Figure 21: Research Design of this Study

### **3.1.1. Research Purpose**

Researches can be categorized according to their purpose. For instance, researches are most often classified as Exploratory and Conclusive. Conclusive research designs can be categorized to Causal (Explanatory) and Descriptive (Saunders et al., 2000; Malhotra, 2006). The different types are explained below:

As its name indicates, exploratory research aims to explore or search through a problem or situation to provide insight and understanding (Malhotra and Peterson, 2006). In other words, “primary purpose of exploratory research is to shed light on the nature of a situation and identify any specific objectives or data needs to be addressed through additional research” (Malhotra, 2006; cited by Javaheri, 2007). “Exploratory research is most useful when a decision maker wishes to better understand a situation and/or identify decision alternatives. Exploration is particularly useful when researchers lack a clear idea of the problems they will meet during the study” (Malhotra, 2006; cited by Javaheri, 2007). However, the purpose of descriptive research is to explain market characteristics or functions (Malhotra, 2006; cited by Javaheri, 2007). Descriptive research can be conducted to make prediction or to find out the degree to which marketing variables are correlated (Malhotra, 2006). “Description is to make complicated things understandable by reducing them to their component parts. Descriptive research could be in direct connection to exploratory research, since researchers might have started off by wanting to gain insight to a problem and after having stated it; their research becomes descriptive” (Saunders et al., 2000; cited by Javaheri, 2007). Finally, causal studies set up causal relationship among variables. In these studies, the emphasis is on analyzing a situation or a dilemma to clarify the relationships among variables (Saunders et al., 2000).

Actually, the expression “data mining” explains the procedure of discovering knowledge or information from databases stocked in data warehouses (Cooper and Schindler, 2003). The purpose of data mining is to discover and extract valid, novel, and meaningful patterns or models in data (Agrawal and Srikant, 1995; cited by Cooper and Schindler, 2003). Data mining is a valuable tool, an approach that merges exploration and discovery with confirmatory statistical study to find out and certify relationships (Cooper and Schindler, 2003). However, we can set data-mining tasks into one of two classes (Kantardzic, 2003). Firstly, prediction data mining that

“involves using some variables or fields in the database to predict unknown or future values of other variables of interest” (Fayyad et al. 1996b). Secondly, descriptive data mining that concentrates on discovering human-interpretable patterns and characterizing the general properties of the data in the database (Fayyad et al. 1996b). Descriptive tasks may also help researchers in their prediction (like our research). We will see in next chapter that we conduct both descriptive and prediction tasks. Since our tool for this study is data mining, and we will apply both descriptive and prediction tasks of data mining, the purpose of this research is mainly exploratory but with some descriptive aspects that help us in our prediction phase.

### **3.1.2. Research Approach**

The purposes of a research may be achieved by means of diverse methods: qualitative and quantitative approaches (Cooper and Schindler, 2003). Qualitative research offers insights and perceptive of the research problem (Malhotra, 2006). Characteristics of qualitative study are that they are based largely on the researcher's own description, emotions and reaction (Holme and Slovang, 1991). On the other hand, quantitative research seeks to quantify the data. It seeks conclusive evidence by applying some form of statistical analysis (Holme and Solvang, 1991; (Malhotra, 2006). It means that, in the quantitative approach, results are based on numbers and statistics that are presented in figures, whereas in qualitative approaches outcomes are based on explaining an event with the applying of words (Holme and Solvang, 1991). In addition, qualitative research lies at the unstructured end of continuum, whereas quantitative research is highly structured (Holme and Solvang, 1991; Malhotra, 2006; Malhotra and Peterson, 2006).

Despite the fact that exploration depends more greatly on qualitative techniques, in several cases there exists enough data to permit data mining or exploration of relationships among individual measurements to occur. The concept of data mining lets decision-makers to be supported by means of exploratory quantitative research (Malhotra and Birks, 2003; cited by Javaheri, 2007). Since the approach of this study is data mining, this research can be considered as a quantitative research.

### **3.1.3. Research Strategy**

“Research strategy will be a general plan of how you will go about answering the research question you have set” (Sounders et al., 2000; cited by Javaheri, 2007). A research strategy is a definite technique that the researcher desires to gather the required data. Based on the character of the research question, the researcher can decide among an experiment, a survey, history, secondary data analysis and case study (Yin, 1994; cited by Javaheri, 2007).

Actually, there are two classes of data applied in researches: primary and secondary data (Zikmund, 2003). In fact, a researcher generates primary data for the particular purpose of dealing with the problem at hand. However, secondary data have already been gathered for objectives other than the problem at hand. Secondary data contain data produced within an organization, information made accessible by business and government sources, commercial marketing research companies and computerized databases (Malhotra and Birks, 2003; Malhotra, 2006). Secondary data can be categorized as internal and external data. The internal data are generated inside the company for which the research is being performed and the external data are produced by sources outside the company (Malhotra and Birks, 2003; Malhotra, 2006). For this research, internal secondary data were gathered from a pharmaceutical distribution company's database (Pakhshe Hejrat Company's database); more details are mentioned in the data collection section. Since the approach of this thesis is data mining and data were collected from a distribution company's databases, the strategy of this study is a secondary data analysis. As a matter of fact, for secondary data exploration, we should begin initially with the company's own data records (Cooper and Schindler, 2003).

Accordingly, the major purpose of this research is exploratory with some descriptive aspects and the approach of this research is quantitative (data mining). The research strategy is the analysis of secondary data.

## **3.2. Research Process**

In this research, we aim to find the best neural network model that fits the sales series of each drug and compare its efficiency with ARIMA models. The goal is to predict the future amount of sales for a specific drug, e.g. Acetaminophen, using past sales data that were recorded

monthly for three years. To reach the goal, three different methods such as ARIMA models and two kinds of neural networks (building ANNs by using each drug's own past records, building ANNs by using each drug's own past records and its group members' past records as well) were performed and compared. Our results will show that our new methodology (the last one above) outperformed the other ones.

After defining the problem and our goals, to reach our objective (sales prediction for a pharmaceutical distribution company) different studies related to forecasting and especially sales forecasting were thoroughly reviewed in chapter 2. Based on literature, which was discussed in chapter 2, we found that there are different methods for sales forecasting. Among many existing methods, two different models that are frequently used for time series prediction are: 1) Linear models: ARIMA modeling, 2) Nonlinear models: neural networks. As can be seen in related works on time series forecasting, to make a prediction for a special product most researches used only past records of that product. In this research, some parts of the process were constructed based on previous methodologies (works) on ARIMA modeling and neural networks. However, due to the lack of data in this research, we also introduced a new approach which is to do a network analysis in order to find the groups of drugs and group members, then make a prediction by using past records of each drug and its group members as input variables. The overall procedure of our thesis consists of multiple steps such as problem definition, literature review, data collection and description, data pre-processing, exploratory analysis, network based analysis, model building or sales prediction and model evaluation. Each Step of above procedure and especially our sales prediction procedure consists of many different courses of actions that are shown in Figure 23 in more detail. In order to implement the overall research process, working out with different softwares was required. Our network analysis was done be means of UCINET and NetDraw and our model building phase was performed by using STATISTICA. In the following sections, each step (except for problem definition and literature review that were covered in previous chapters) and the methods associated with each step are explained briefly.

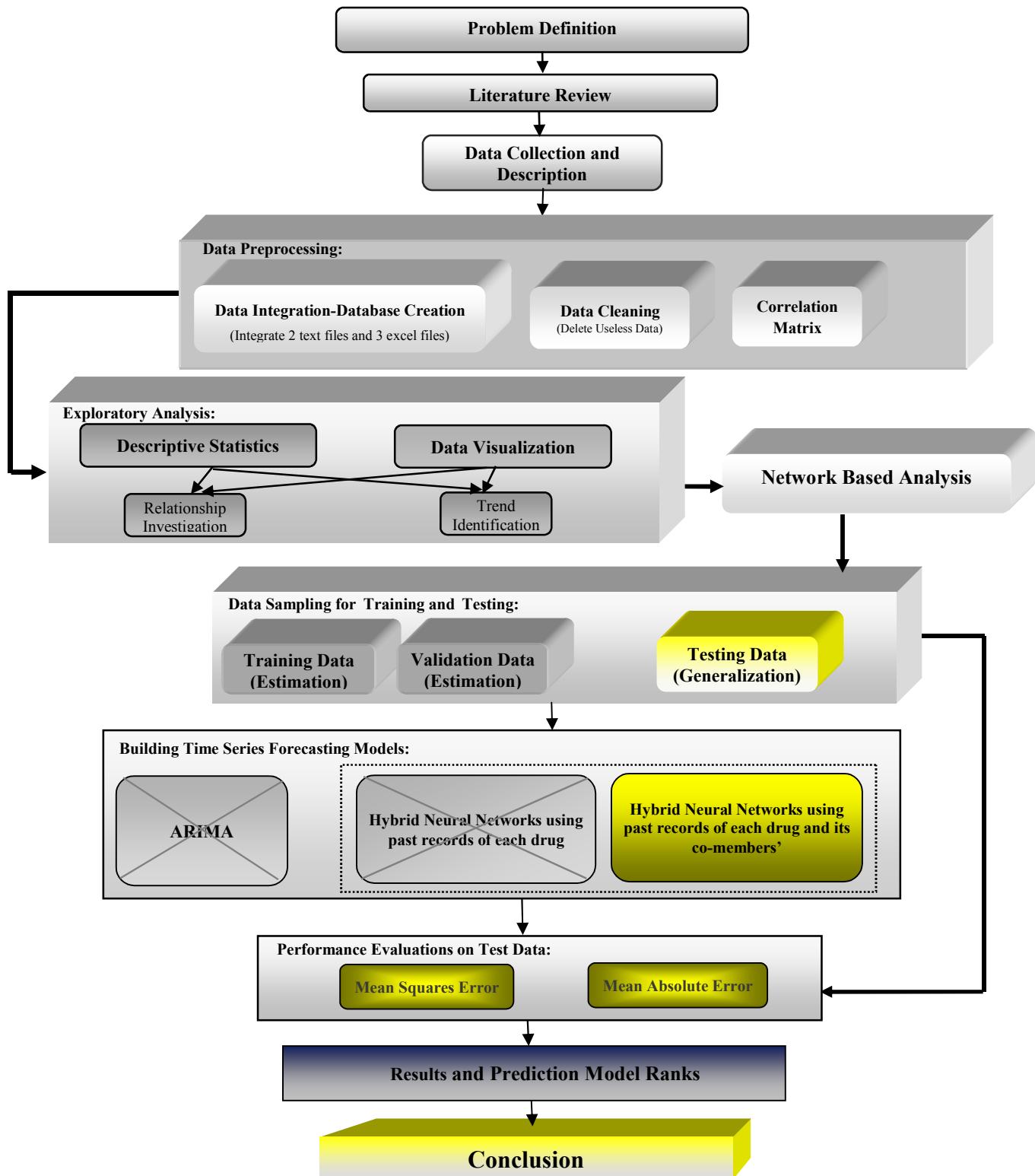


Figure 22: Research Process

### **3.2.1. Data Collection and Description**

As mentioned before, there are two types of data usually applied in researches: primary and secondary data (Zikmund, 2003). Primary data are the data that someone collects on his own with a definite purpose in mind, whereas secondary data have already been collected by other researchers with diverse purposes in mind (Javaheri, 2007). Secondary data are categorized as internal and external data. The internal data are produced inside the organization for which the research is being performed and the external data are generated by sources outside the company (Malhotra and Birks, 2003). As stated, for data mining approaches secondary data are usually used.

As we stated before, pharmaceutical distribution companies are facing new competition situations and they should survive themselves in these tough conditions. They have to make their customers satisfied and deliver the right amount of drugs to the right place and at the right time. Thus, they have to optimize their inventory levels and improve their inventory management. This problem can be solved by mining past sales data to make better decision for the future. Making good decisions needs a proper model to make appropriate predictions. Consequently, these companies are forced to have a reliable sales forecasting that leads to inferior costs owing to decreased inventory as well as growing customer satisfaction that will be due to an boost in on-time deliveries (Carboneau et al., 2008). On this account, we needed to have their past sales records. Hence, sales data information of a selected pharmaceutical distribution company is collected from that organization's database (internal databases). Thus, our work was mainly based on internal secondary data. Actually, we gathered internal secondary data from Pakhsh Hejrat Co. which is one of the main drug distribution companies in our country.

As a matter of fact, Pakhsh Hejrat has got near 1200 kinds of drugs which are bought from different manufacturers and sold to different regions (18 main cities) of our country. In fact, Pakhsh Hejrat Co. commenced recording sales information from December 2004 to December 2007 every month, so the total number of records is 36 months for each medical product. Table 3 displays the sales information consisted in the company database. As can be

seen, database of Pakhsh Hejrat Co. includes name and code of medicines, sales number, name and code of centers, name of manufacturers, price and monthly date of sales. To approach the objective of the thesis, not all these information are required. We just selected three items: Goods' Codes, Date and Number of Sold Products. Therefore, we had to do data cleaning and reduction that are going to be explained in next section.

**Table 4: Sales Data Information Collected from Pakhsh Hejrat Company**

Name of Drug	Code of Drug	Name of Center	Code of Center	Producer Name	Producer Code	Year/Month	Sales Number	Price
Naproxin	065171004	Zahedan	15	Modava	669	83/11	320	101817

### 3.2.2. Data Preparing

After collecting the required data for sales prediction process, data must be converted to appropriate format. Therefore, gathered data have to be integrated, cleaned, reduced and transformed to meet the requirements of sales prediction (Han and Kamber, 2006). Some book says that if data mining is considered as a process then data preparation is at the heart of this process. Actually, this step is inevitable in data mining researches since databases are highly sensitive to noisy, missing, and inconsistent data; databases are typically in huge sizes and they are likely originated from multiple heterogeneous sources. In addition, some database may include fields that are outmoded or redundant, missing values, outliers, data in a shape that is not appropriate for data mining models and values that are not consistent with strategy or common sense (Larose, 2005).. There are a number of data preprocessing techniques: data cleaning, data integration, data transformation and data reduction. Data cleaning can be applied to remove noise, supply missing values, and correct inconsistencies in the data (Han and Kamber, 2006). Data integration merges data from multiple sources into a coherent data store, such as a data warehouse. Data transformations involve data normalization/scaling, and feature construction. In data transformation, the data are transformed or consolidated into structures suitable for mining (Han and Kamber, 2006). Data reduction can decrease the data volume by combining, eliminating redundant features (feature selection), or clustering (Han and Kamber, 2006). These

methods are not mutually exclusive; they could work mutually. Data preprocessing methods, when used prior to mining, can considerably improve the general quality of the mined patterns and/or the time necessary for the real mining (Han and Kamber, 2006). Data preprocessing step usually takes the most time needed for the whole KDD process; Pyle (1999) approximated that data preparation be responsible for 60% of all the time and attempt developed in the whole data mining procedure. In following sections data integration, data cleaning and reduction, data transformation, also data normalization/scaling are explained in more detail.

- **Data Integration**

As stated above, data integration merges data from multiple sources into a coherent data store. In fact, we gathered secondary data from Pakhsh Hejrat Co. which is one of the main drug distribution companies in our country. As mentioned before, Pakhsh Hejrat has got more 1200 kinds of drugs which are bought from different manufacturer and sold to different regions (18 main cities) of our country. Furthermore, database of Pakhsh Hejrat pharmaceutical distribution company includes drug's name, drug's code, number of sold drugs, center's name, center's code, name of manufacturers, prices and monthly date of sales. Pakhsh Hejrat pharmaceutical distribution company provided us with their sales data for 36 months which are stored in two text files for years of 84 and 85, also three excel files for last 3 months of 83, first three months of 86 and next 6 months of 86 in Farsi calendar. It means that overall we had 5 separate files (with different formats) related to different years and different months from December 2004 to December 2007. From a time series forecasting point of view, we needed to have all these information in a single file such that each row of the file displays the name and amount of the specific product sold during 36 months. Moreover, the codes of some drugs were different with each other in different files and we had to correct them. It means that in two text files the codes contained three excessive zeros at the beginning. We had to delete these zeros in order to have a uniform database.

- **Data Cleaning and Reduction**

After data was integrated into one Excel file, data has to be cleaned to fulfill the requirements of the data mining process which is sales prediction in our research. Data cleaning can be applied to remove noise or unnecessary data. Actually, the main problem of our data was not

selling some kind of products in some months. This problem made many zeroes in data matrix. Many records had these zeros and they should be deleted. In addition, we had to select some drugs among these 1200, as it was impossible to work with all of them. Instead of choosing some kinds of drugs randomly, we extracted 217 kinds of drugs, which were sold in all 36 months; in other words which were sold permanently. By this job, we also made the same sales condition for all 217 drugs. And we changed the drugs' code to 1-217 for simplicity. The logic of this selection is that we could have an accurate sales forecasting only for the products that we have enough past sales data, so for the drug which fore example were sold only 10 months we cannot make good forecasting as we do not have enough past sales information about it. Thus, we use these 217 kinds of drugs for our grouping and model building in next phases since the methodology is the same for any number of drugs. In fact, just the time that is needed for sales prediction would increase for more kinds of drugs. In addition, many data (information) were only overheads and were not helpful to us in any way, such as names and manufacturer's code, name of drugs, name of provinces. To approach the objective of the thesis not all these information are required; hence, they were deleted from table. Therefore, we built a database of three features: Goods' Codes, Date and Number of Sold Items. Briefly, data cleaning was done in Excel 2007 as follow:

- Extracting 217 kinds of drugs that were sold in all 36 months.
- Deleting all over heads which are not helpful to us in any way.

In this way, after data cleaning, there is a great data reduction in dataset from nearly 14381 rows to nearly 220 rows. In addition, kinds of drug reduced from more than 1200 to 217.

- **Data Transformation**

After, we did data integration, cleaning and reduction by means of Microsoft Office Excel 2007; we selected the total amount (sums) of sales of different drugs to all regions and copied them in a new sheet. After that, we got correlation from sums of sales in all months for all drugs. Since we had 217 numbers of different drugs, a matrix with 217 rows and 217 columns (217\*217) that showed the correlation among all kinds of monthly sold drugs was built. Then, we examined all correlations which are higher than  $\theta = 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85$  and 0.9 separately to see the behavior and position of these drugs according to

each other. In other words, we aimed to extract different groups of drugs that had the highest similarity or correlation then drew the related graphs. Thus, we built incident matrixes which contains of 0, 1 values. We gave value 0 if the correlation between two drugs was less than selected  $\theta$  and gave 1 if the cross-correlation was greater than  $\theta$ . We selected  $\theta = 0.65$  as the basis of our work because if we consider  $\theta$  very small such as 0.3 all drugs would be correlated and if we consider it too high such as  $\theta = 0.9$  drugs would not be correlated at all and we would not have any network. Hence, we wish to choose the best threshold that gives us appropriate number of cliques and the best combination of cliques (drugs' groups). Generally, it is desirable to have all the following criteria simultaneously:

1. There are large numbers of cliques.
2. The variance or standard deviation of the cliques' size is not too much, that is cliques have nearly the same sizes.
3. Mean of degree for the nodes is high.
4. Distribution of Nodes' degrees is nearly identical. It means that the variance or standard deviation of degrees is small.

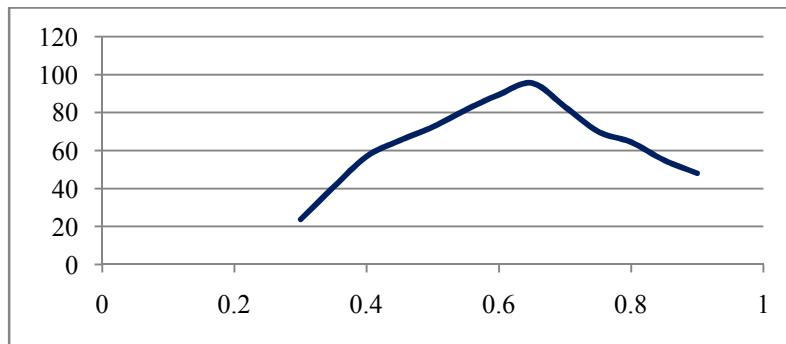
Accordingly, we want to have the max of the following index:

$$\text{Index} = [(\text{Number of Cliques}) \cdot (\text{Mean of Degree})] / [(\text{Standard Deviation of clique size}) \cdot (\text{Standard Deviation of Degree})]$$

We calculated above index for different threshold s from  $\theta = 0.3, \theta = 0.35, \theta = 0.4 \dots$  to  $\theta = 0.9$  and it is presented in Table 4. We also bring Figure 24 that demonstrated different amount of indices for different thresholds. As it is obvious, horizontal axis indicates  $\theta$ s and vertical axis shows indices.

**Table 5: Amount of Indices for Different  $\Theta$ s**

Threshold ( $\Theta$ )	Index
<b>0.3</b>	23.8896
<b>0.35</b>	40.8742
<b>0.4</b>	57.0096
<b>0.45</b>	65.2974
<b>0.5</b>	72.4539
<b>0.55</b>	81.4536
<b>0.6</b>	89.3765
<b>0.65</b>	<b>95.4661</b>
<b>0.7</b>	83.1195
<b>0.75</b>	70.0228
<b>0.8</b>	64.4491
<b>0.85</b>	55.0083
<b>0.9</b>	48.0996



**Figure 23: Amount of Indices for Different  $\Theta$ s**

Actually, the cross-correlation matrix for  $\Theta = 0.65$  is the basis of our network analysis for sales data and now we are able to group drugs according to their correlation or their mutual relationship. This grouping would help us to run more accurate sales forecasting with better performance.

- **Normalize/Scale Data**

“Data normalization involves scaling the attribute values to make them lie numerically in the same small interval/scale, and thus have the same importance” (Javaheri, 2007). Since neural network models produce better models when the data is normalized and normalizing the input data will help speed up the learning phase, all data should be normalized or standardized before modeling. In general, there are three normalization techniques and there is no significant difference among these three techniques. These three methods are (Han and Kamber, 2006):

- **Min-max normalization:** min-max normalization performs a linear transformation on the original data. The advantage of this method is that it preserves all relationships of the data values exactly. It does not introduce any potential bias into the data.
- **Z-score normalization** (or zero-mean normalization): in z-score normalization the values for an attribute, A, are normalized based on the mean and standard deviation of A.
- **Normalization by decimal scaling:** normalizes by moving the decimal point of values of attribute A. The number of decimal points moved depends on the maximum absolute value of A.

However, in this research, there was no need to normalize data since STATISTICA software does it automatically.

### 3.3. Exploratory Analysis

After data preprocessing, in order to better understanding the nature of our data, we did a thorough exploratory analysis. This phase was beneficial to know that our series have sales trends or not, have seasonal patterns or not, and to know whether our products have linear sales behavior or non linear behavior. Accordingly, we drew the sales plots and surface plots of our time series data for different drugs; in this part, we used STATISTICA software. Most of our drugs had considerable various sales trends. Some of them showed seasonality, the variance was also altering through time for most drugs. Thus, most series were not stationary and they needed differencing or even de-trending to become stationary. Subsequently, we computed some descriptive measures and we understood that most medicines really have various natures and characteristics. Therefore, from our exploratory analysis we concluded that most drugs have different characteristics and most of them have specific behavior of sales. Furthermore,

relationships among drugs are mostly nonlinear. Accordingly, we have to put linear models aside and make use of nonlinear or hybrid models for our sales forecasting. In general, from this section we understood important points that guided us in our modeling phase. For instance, we perceived that most drugs show strong nonlinear relationships, so techniques that model nonlinear relationships would be more successful than linear methods in this research. However, since we can see both linear and nonlinear relationships for some drugs, we had better use hybrid models.

### **3.4. Network Based Analysis**

Pakhsh Hejrat Co. has got only 3 years of their sales records. This matter was really an obstacle for our correct prediction of their sales as for time series prediction we should have at least 50 past records. Accordingly, we came to the idea, which is really a new idea, to find groups of drugs, which have similar sales behavior and make prediction by means of past records of group members. Hence, we found clique sets, degree of each drug, group members, and we visualized the networks of drugs. In this section, we carried out all our analysis by means of UCINET and NetDraw softwares, which are useful for network analysis. The output of this section helped us greatly in our prediction phase as we noticed that which drugs could help each other in sales forecasting. This phase is explained in more detail in chapter 4.

### **3.5. Methods Used for Modeling and Time Series Forecasting**

- Data Sampling for Training and Testing**

In this phase, data should be divided to training and test sets. Actually, training data set is applied to build the model, and test data set is applied to test the model. Researchers have built up training methods that decrease the possibility of over-fitting (fitting the training data extremely exactly that leads to unsatisfactory results on new data) of training data (Ye, 2003; Javaheri, 2007). An effective method to evaluate outcomes is to put aside some data and compare training results with unseen test results. Certainly, one might stay until novel data appears through application of the answer, but it is logical to check performance before actual application. It avoids unforeseen poor outcomes and gives the researchers time to mine the greatest performance from the application system (Ye, 2003; Javaheri, 2007). “Training is

typically done on a large proportion of the total data available, whereas testing is done on small percentage of the data that has been held out exclusively for this purpose" (Javaheri, 2007).

As told before, we were able to gather sales data of 36 months and we had to divide our dataset to train and test sets. In this thesis to avoid over-fitting and to do model evaluation, the dataset was divided into two subsets for ARIMA model building: one for training (32 records) and one for testing and performance evaluation (4 records), and dataset were divided into three subsets for ANN: 4 months for testing the results, 4 months for cross-validation and the rest for training the network.

- **ARIMA Methodology**

ARIMA (p,d,q) models (explained in chapter 2) are auto-regressive integrated moving average models, also called Box–Jenkins models (Ture and Kurt, 2006). ARIMA model predicts a variable's future values from its past values and it needs at least 50 past records. Although ARIMA models are quite flexible in modeling a wide range of time series patterns, their major limitation is the presumed linear form of the model. Thus, ARIMA methodology was not useful for our case as we had only 36 past sales record for each drug; also, in our exploratory analysis we found that the relationships among drugs were both linear and nonlinear, but mostly nonlinear. Accordingly, we had to leave out this methodology. Although we had to ignore ARIMA methodology because of its linear property and because of not having enough past sales records, we brought it to see its performance on our time series in comparison with our new methodology.

As was mentioned in chapter 2, ARIMA modeling involves four stages (Box and Jenkins, 1976; Berthouex and Brown, 2002; STATISTICA 7, Electronic Manual):

- (1) Transforming the series to make it stationary ( finding d parameter)
- (2) Identification of the initial p and q parameters, using auto-correlation and partial auto-correlation methods;
- (3) Estimation of the p (autoregressive) and q (moving average) components to see if they contribute significantly to the model or if one or the other should be dropped;

- (4) Diagnosis of the residuals to see if they are random and normally distributed, indicating a good model (Box and Jenkins, 1976; Berthouex and Brown, 2002; STATISTICA 7, Electronic Manual).

To fit ARIMA to the available time series for drug 24 (as an example), four-stage procedure of transforming, model identification, estimation of model parameters and diagnostic checking of the estimated parameters were adopted in STATISTICA 7 software and presented in chapter 4. To determine the possible persistence structure in our sales data, the auto-correlation function (ACF) and the partial auto-correlation function (PACF) were used. Based on the visual examination of these plots, all possible ARIMA models were identified for the data series. We used the exact maximum likelihood (Melard) method according to Melard et al., (2006) to compute the SS (sum of squares) for the residuals as typically just a few iterations with the exact maximum likelihood method are required to complete the parameter estimates (STATISTICA 7, Electronic Manual). According to this method, we found parameters in STATISTICA. Then, we examined our model on the test data and we calculated the values of residuals, Mean Squared Error (MSE) and Mean Absolute Error (MAE) to evaluate its performance. Although we chose the best ARIMA model, the result was not satisfactory enough. In ARIMA methodology we may not use past records of other drugs, and we could just use the past records of each drug for itself as inputs. Thus, in this section we did not use group members' past records for our forecasting. All steps of our ARIMA modeling for drug 24 (as an example) are described in detail in chapter 4.

- **A Hybrid Neural Network Approach**

Neural networks (discussed thoroughly in chapter 2), are a class of dominant, general-objective tools used in prediction (Berry and Linoff, 2004). Actually, in this research, we carried out hybrid type of neural networks. The origin of this idea was related to Zhang, (2004) that combined ARIMA methodology with ANNs in order to model linear parts with ARIMA and nonlinear parts with ANNs. However, the basic idea of our hybrid approach is to let linear ANN model the linear parts and let nonlinear ANN model the nonlinear parts and then combine the results from both linear and nonlinear models. Our model building process carried out with STATISTICA and involved the following three steps. First, a linear ANN was built based on the

sales data to estimate the linear components. The residuals from the linear model contained some nonlinearities (since we had done a comprehensive exploratory analysis and saw that in our sales data the relationships were mostly nonlinear; besides, it is widely accepted by researchers that a real data set always contains more or less nonlinear relationships). Then, we applied ANNs to the residuals to estimate the nonlinear components. Finally, we combined these two parts together to make a forecast. In this section, we introduced a new and effective methodology to improve the performance of our forecasting. Due to the fact that we had access to only 36 records (which is not enough for time series prediction), we came to the new idea of grouping drugs according to their sales cross-correlation to increase our records and use group members' sales records for each other. By this work, we solved the problem of not having enough past records for each drug. We examined our methodology for 21 drugs with different degree and for making a comparison, we built hybrid neural network model with two different approaches: 1) Just using each drug's own past records, 2) Using each drug's own past records and its group members' past records as well. In chapter 4, we will observe that how we have implemented the results of our network analysis in our sales forecasting with hybrid ANNs. We also will see the results and all steps of our methodology in more detail in next chapter.

- **Model Evaluation**

Once a model has been generated and tested, its performance should be evaluated. In this study two forecast error measures, Mean Squared Error (MSE), and Mean Absolute Error (MAE) were employed for model evaluation and model comparison. We compared accuracy measures (MAE and MSE) of both approaches for 19 drugs (we dropped drug 31 and 86 since they do not have any group member), we saw that in 16 cases the prediction performance was improved greatly when we used partners' past records as well. We also compared these result with results of ARIMA methodology and we saw the difference between our new methodology and ARIMA. Accordingly, we proved that our methodology outperformed noticeably both ARIMA methodology and ANN sales forecasting without use of partner's records.

In summary, in this research we performed three different techniques for sales forecasting. 1) ARIMA modeling, 2) Hybrid ANNs by means of each drug's own past records, and 3) Hybrid ANNs by means of each drug's own past records and its group members' past records. As the

methodology is the same for all drugs, it was not necessary to examine these techniques on all drugs. Therefore, we selected 21 drugs and predicted their future sales with three above techniques. We brought just one example for the first approach as it was rejected even theoretically at first, and we brought 21 examples for the second and third approaches. By comparing their results and performances, we evidenced that our proposed methodology (the third one above) was superior to both first and second ones.

# **CHAPTER 4**

## **RESULTS AND ANALYSIS**

### ***4. Results and Analysis***

*This chapter commences with data preprocessing phase and then goes through an exploratory analysis in order to better understanding the nature of our data. Next, the networks of drugs are shown and the network based analysis is explained by using NetDraw and UCINET softwares. Afterwards, ARIMA modeling is performed using STATISTICA software. Then, we built hybrid neural network model with two different approaches by using STATISTICA software and the prediction results are presented. Eventually, the model evaluation is explained.*

## **4.1. Data preprocessing**

As explained in data collection phase in chapter 3, we gathered secondary data from Pakhsh Hejrat Co. which is one of the main drug distribution companies in our country. In fact, Pakhsh Hejrat has got near 1200 kinds of drugs that are bought from different manufacturers and sold to different regions (provinces) of our country. Thus, database of Pakhsh Hejrat pharmaceutical distribution company includes drug's name (code), number of sold drugs, center's name (code), name of manufacturers, price and date (month) of sales. Pakhsh Hejrat pharmaceutical distribution company provided us with their sales data for 36 months which are stored (in different formats) in two text files for years of 84 and 85, also three excel files for last 3 months of 83, first three months of 86 and next 6 months of 86 in Farsi calendar. It means that overall we had five separate files related to different years and different months from December 2004 to December 2007. For sales prediction, we needed to have all these information in a single file such that each row of the file presents the name and amount of the specific product sold during 36 months. After data was integrated in to one Excel file, data had to be cleaned and to meet the requirements of our modeling process. Actually, the main problem of this data was not selling some kind of products to some provinces from a month onward. This problem made many zeroes in data matrix. Many records had these zeros and they were to be deleted. Then, we had to select some drugs among these 1200 kinds to make sales prediction for them. Instead of choosing randomly, we extracted 217 kinds of drugs, which were sold in all 36 months in other words which were sold permanently. By this job, we also made the same condition for all 217 drugs. And we changed the drugs' code to 1-217 for our easiness. The logic behind this selection is that we could have an accurate sales forecasting if we have enough past sales data, so for the drugs which for example were sold only 10 months we cannot make good forecasting as we do not have enough past information about it. Thus, we used these 217 kinds of drugs for our grouping, and model building as the methodology does not differ for any numbers of drugs and it is the same for all of them. Just the needed time increase for more kinds of drugs.

In addition, many data were only over-heads and was not helpful to us in any way, such as names and manufacturer's code, name of drugs, name of provinces. To approach the objective of the thesis not all these information are required; hence, they were deleted from table. Thus, we built a database of three items in Excel: Products' Codes, Date and Number of

Sold Items. After data cleaning, there is a great data reduction in dataset from nearly 14381 rows to nearly 220 rows. In addition, kinds of drug reduced from more than 1200 to 217.

Subsequently, we built a correlation matrix (217\*217) in an Excel file for sales of drugs in 36 months, which would be very useful in network analysis, feature selection and finally sales prediction. Accordingly, we first built a cross-correlation matrix (adjacent matrix) by use of their sales amount. Then, we examined different thresholds from  $\theta = 0.3$ ,  $\theta = 0.35$ ,  $\theta = 0.4\dots$  to  $\theta = 0.9$ . Afterwards, we built incident matrixes which contains of 0, 1 values. We give value 0 if the correlation between two drugs was less than selected  $\theta$  and give 1 if the cross-correlation was greater than  $\theta$ . We select  $\theta = 0.65$  as the basis of our work because if we consider  $\theta$  very small such as 0.3 all drugs would be correlated and if we consider it too high such as  $\theta = 0.9$  drugs would not be correlated at all and we would not have any network. Hence, we wished to choose the best threshold. Generally, it is desirable to have all the following criteria simultaneously. It means we will have the best combination of the cliques (drugs' groups) when:

- There are large numbers of cliques.
- The variance or standard deviation of the cliques' size is not too much, that is cliques have nearly the same sizes.
- Mean of degree for the nodes is high.
- Distribution of Nodes' degrees is identical. It means that the variance or standard deviation of degrees is small.

Accordingly, we want to have the max of the following index:

$$\text{Index} = [(\text{Number of Cliques}) \cdot (\text{Mean of Degree})] / [(\text{Standard Deviation of clique size}) \cdot (\text{Standard Deviation of Degree})]$$

As we presented in previous chapter (in data preparing phase), we calculated above index for different threshold s from  $\theta = 0.3$ ,  $\theta = 0.35$ ,  $\theta = 0.4\dots$  to  $\theta = 0.9$  and we found that the above index is max (95.4661) for  $\theta = 0.65$ .

Therefore, the cross-correlation matrix for  $\theta = 0.65$  is the basis of our network representation of the sales data and now we are able to group drugs according to their cross-correlation or their mutual relationships. This grouping would help us in our prediction phase to have better and more reliable forecasting.

Although it is not needed for building ARIMA models to make data normal, before applying neural networks, the data should be normalized. However, in this research, there was no need to normalize data since STATISTICA software does it systematically.

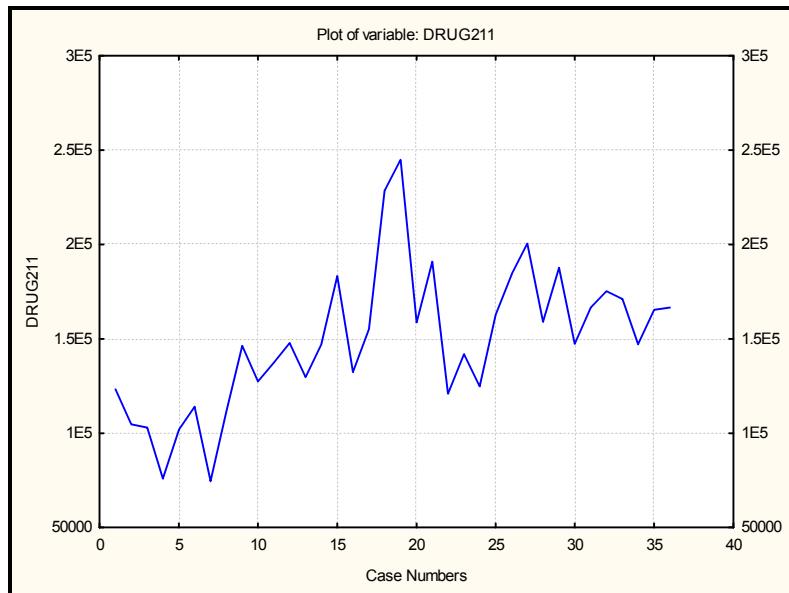
## **4.2. Exploratory Analysis**

After data preprocessing, in order to better understanding the nature of our data, we did an exploratory analysis. In this phase, we searched through data to see what is happening through these data. Is there any trend? Is there any cycle? Do they have linear behavior or non linear behavior? Variance for number of sales is high or not? And so on. Our exploratory analysis consists of two steps: data visualization and descriptive statistics.

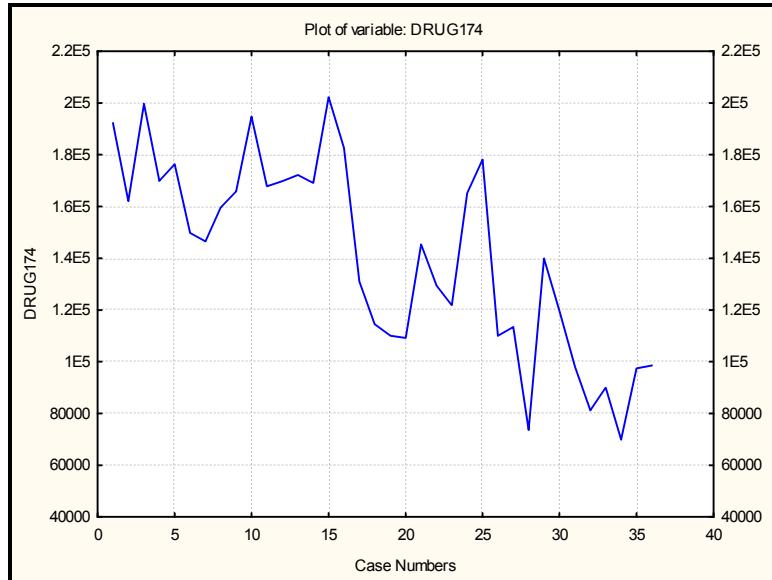
### **4.2.1. Data Visualization**

In this part, we use STATISTICA software to show the time series plots of some drugs for more comprehension. For example, Figures 24, 25, and 26 show the time series plot of drugs 211, 174 and 24. The horizontal axis indicates months and the vertical one presents the amount of sales. As the figure shows, there are considerable trend in all of them. For instance, in first plot for drug 211, we can see strong nonlinearity and upward trend, it also shows a weak seasonality, and its mean increases as the number of months increases. It means that sales figures are not fluctuated over fixed mean. The variance is also altering. Thus, this series is not stationary and it would need differencing or even de-trending to become stationary. We can explain time series plot of drugs 24 and 174 like previous one, except for the point that the trends are downward for both of them. Therefore, based on the visual examination of the time series plot of sales data in Figure 25, 26, and 27, there is sign of non-stationary in all of them and if we want to make ARIMA model we have to first make them stationary. In other words, the nature of each data is different from other; that is, one may have upward trend, one may have downward trend, one may not have trend at all. One may be stationary but one may not. In Figure 28, we show multiple plot of drugs 54, 56, and 61 to see that how the sales behavior of these drugs differ with each other. For instance, drug 56 has not follow special trend and its mean are nearly fixed, number of sales fluctuates approximately around 20000 and it has little deviation from mean. Drug 61 has a gradual downward trend, but with almost slight deviation and finally drug 54 presents a slight upward trend, but dramatic changes through time and it fluctuates

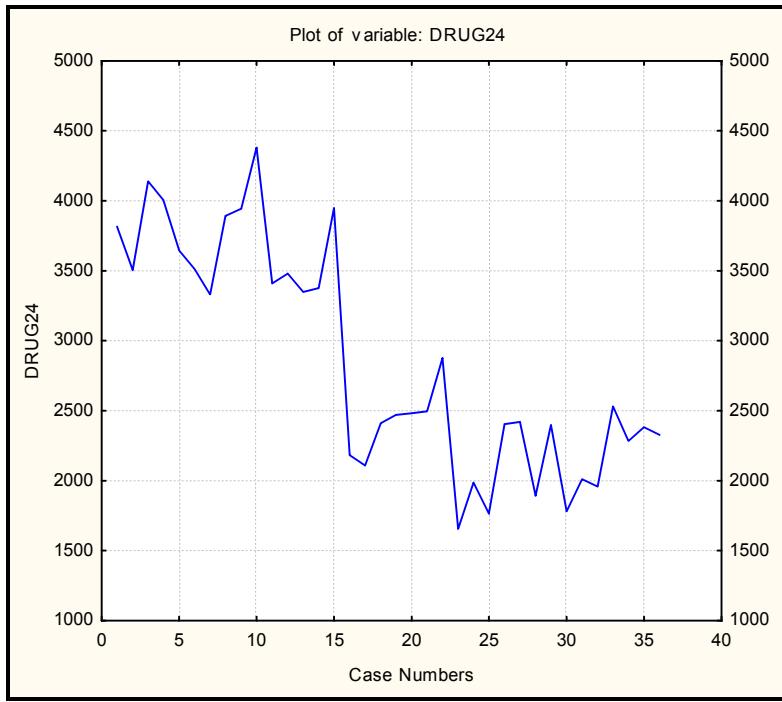
substantially. As a result, we cannot assume a unique model for all drugs and we have to build a model for each of them separately. It means that these drugs are dissimilar products and have different characteristics; only the name of them is the same. However, in next chapters we would see that how we make use of similar sales behavior of some of them in our sales prediction.



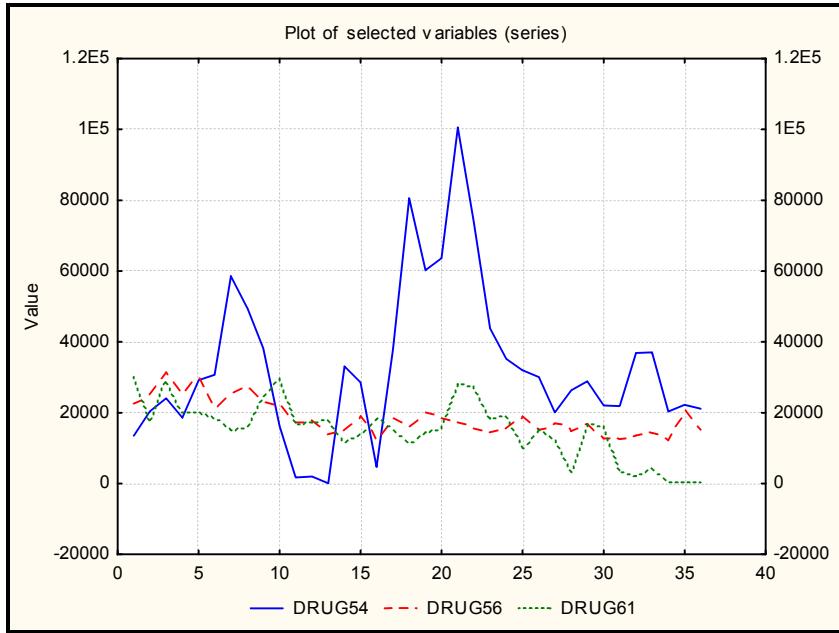
**Figure 24: Time Series Plot of Sales Data for Drug 211**



**Figure 25: Time Series Plot of Sales Data for Drug 174**



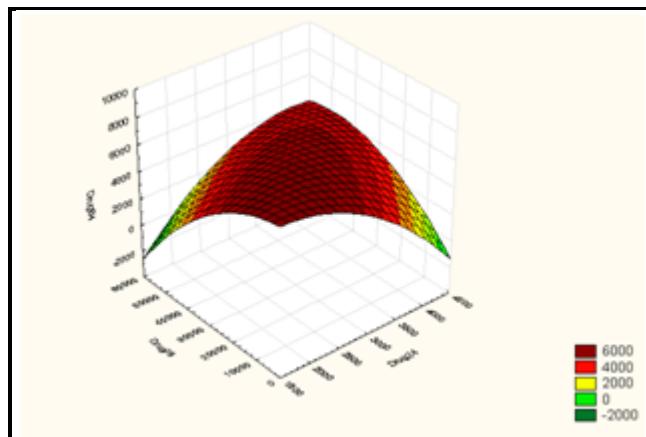
**Figure 26: Time Series Plot of Sales Data for Drug 24**



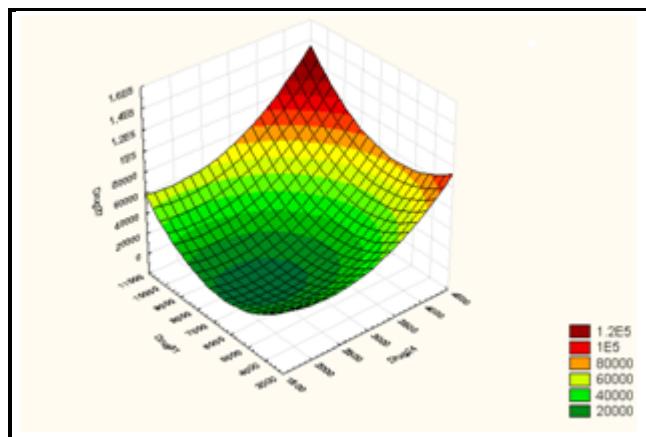
**Figure 27: Multiple Time Series Plot of 3 Drugs**

We could also depict the surface plot for each three drug in order to see that whether they have linear or nonlinear relationships. As an illustration, Figure 29 shows strong non-linearity

among these three drugs. It is the same for plot 30. However, as we can see from plot 31, these three drugs, that is 24, 174 and 86 displays approximately linear relationships. Since drug 24 is repeated in all plots, we can conclude that this drug has liner relationships with some drugs and nonlinear relationships with other ones, so we cannot build a single model for this drug and we have to make use of both linear and nonlinear models. Consequently, after examining the relationships of more than 20 drugs we can say that although in some cases there are some linear relationships, in most cases the relationships among drugs are mainly nonlinear. We would make use of this characteristic of most drugs in following sections to explain why only linear models like ARIMA are not enough for our prediction.



**Figure 28: 3D Surface Plot for Drugs 24, 78, 84**



**Figure 29: 3D Surface Plot for Drugs 24, 31, 37**

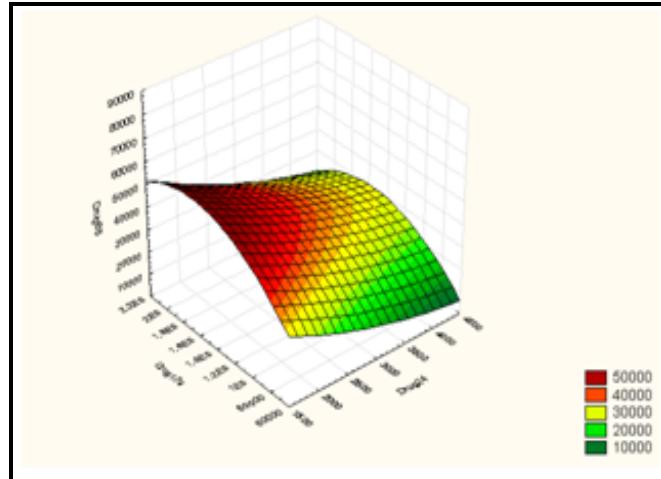


Figure 30: 3D Surface Plot for Drugs 24, 174, 86

#### 4.2.2. Descriptive Statistics

In this section, we are going to find some measures of tendency and dispersion like mean, median, variance, and standard deviation to better perceive the characteristics of drugs. Since we have many kinds of drugs, here we choose some of them to present the descriptive measures. Table 5 shows these descriptive measures for 21 selective drugs.

**Table 6: Descriptive Measures of 21 Drugs**

Descriptive Statistics of 21 Drugs									
Drugs Code	Mean	Median	Sum in 36 Months	Minimum	Maximum	Variance	Std. Dev.	Standard Error	
<b>Drug12</b>	10819.0	10819.0	400302	5226.00	17354.0	1.019239E+07	3192.55	524.852	
<b>Drug22</b>	4808.1	4663.0	177899	2198.00	7541.0	1.730166E+06	1315.36	216.243	
<b>Drug24</b>	2849.6	2496.0	105434	1657.00	4381.0	6.311993E+05	794.48	130.612	
<b>Drug30</b>	20596.4	18785.0	762066	6358.00	39117.0	7.160466E+07	8461.95	1391.137	
<b>Drug31</b>	5837.1	5762.0	215972	3561.00	9824.0	2.228769E+06	1492.91	245.432	
<b>Drug37</b>	24506.4	17364.0	906737	-3246.00	93946.0	7.011937E+08	26480.06	4353.295	
<b>Drug51</b>	6009.0	5481.0	222332	35.00	17590.0	1.313668E+07	3624.46	595.857	
<b>Drug61</b>	15235.7	15992.0	563722	277.00	30105.0	6.781183E+07	8234.79	1353.792	
<b>Drug66</b>	4411.3	1598.0	163217	-114.00	27018.0	3.680639E+07	6066.83	997.380	
<b>Drug71</b>	860.9	1074.0	31854	-44.00	1721.0	2.409805E+05	490.90	80.703	
<b>Drug78</b>	23982.8	20671.0	887365	8535.00	55849.0	8.372386E+07	9150.07	1504.263	
<b>Drug84</b>	5704.1	5735.0	211053	3661.00	8261.0	1.563402E+06	1250.36	205.558	
<b>Drug86</b>	38346.0	38852.0	1418802	15585.00	78052.0	1.744982E+08	13209.78	2171.674	
<b>Drug95</b>	15355.9	15788.0	568168	1061.00	45495.0	9.442650E+07	9717.33	1597.519	
<b>Drug103</b>	3744.0	3666.0	138528	5.00	9318.0	8.602782E+06	2933.05	482.190	
<b>Drug124</b>	411.6	417.0	15231	42.00	622.0	1.206934E+04	109.86	18.061	
<b>Drug127</b>	1441.4	1161.0	53330	86.00	3286.0	8.751081E+05	935.47	153.791	
<b>Drug 132</b>	9421.6	9421.6	348601	1550.00	17055.0	1.893737E+07	4351.71	715.416	
<b>Drug167</b>	21081.3	19912.0	780009	8118.00	54196.0	1.166155E+08	10798.87	1775.323	
<b>Drug168</b>	24008.6	21077.0	888320	8398.00	61742.0	1.878208E+08	13704.77	2253.051	
<b>Drug174</b>	141043.3	145323.0	5218603	69885.00	202207.0	1.391763E+09	37306.34	6133.125	

From above table, it is clear that these medicines really have various natures and we can extend this matter to all drugs. Their mean and sum of sales are differed greatly. Some of them were sold highly like drug 174; some of them have not sold a lot like drug 71 and 124. Some drugs have high variance such as drugs 174, 37 and 95 and also they have a high range; however, some of them have not. Negative figures indicate the shortage of drugs, so some drug presents shortage and Pakhsh Hejrat Company have to try to avoid it because it would lead to customer loss.

Therefore, from our exploratory analysis we can conclude that most drugs have different characteristics and most of them have specific behavior of sales. Furthermore, relationships

among drugs are mostly nonlinear. Accordingly, we have to put linear models aside and make use of nonlinear or hybrid models for our sales forecasting. In summary, from this section we understand important points that would guide us in our modeling phase. As most of drugs show strong nonlinear relationships, techniques that model nonlinear relationships would be more successful in this research. However, since we can see both linear and nonlinear relationships for some drugs, we had better use hybrid models.

### **4.3. Statistical Analysis of Drugs' Networks**

Most pharmaceutical distribution companies in Iran have not got more than 3-4 years of their sales records. For instance, Pakhsh Hejrat Company that is among main distribution companies in pharmaceutical industry has got only 3 years of their sales records, which we have access to them. This matter is really an obstacle for correct prediction of their sales. On the other hand, as we mentioned in our literature review it is advised many times that for time series prediction we should have at least 50 recodes. However, we have got only 36 records. Thus, we cannot make use of traditional methods for our sales forecasting. Accordingly, we came to the idea, which is really a new idea, to find groups of drugs that have similar sales behavior and make prediction by means of past records of group members. That is why in this section we are going to find clique sets, group members, and networks of drugs.

As mentioned before, we had near 1200 kinds of drugs. However, we selected 217 kinds of them that were sold permanently. We claim that we can cluster them in order to have some groups of drugs which have similar sales or demand behavior. In other words, each cluster or groups contains of some drugs, which have similar sales behavior. We would make use of this point in our model building in next section. For example, if we wish to predict the sales of drug 103 which is located in a group of 15 drugs, we will also make use of other drugs' (the drugs that are in the same group with drug 103) past sales information since we have only 36 records of its past sales. As a result, we would increase our past sales records with this vital grouping. We will illustrate in prediction section that this matter would increase the performance or accuracy of our sales prediction. Therefore, in this part we are going to demonstrate our network analysis.

#### 4.3.1. Network Identification

Actually, it is beneficial to find proficient methods of summarizing and visualizing the drugs' sales data that would permit one to achieve valuable information about the behavior of the market (Boginski et al., 2006). The analysis of time series plots becomes more and more complex as the number of drugs grows (Boginski et al., 2006). An alternative way of summarizing the sales data is based on representing the drugs as a graph (network). It should be noticed that the method of signifying a dataset as a network becomes more and more broadly applied in diverse practical applications (Boginski et al., 2006). "This methodology allows one to visualize a dataset by representing its elements as vertices and observe certain relationships between them" (Boginski et al., 2006). Studying the structure of a graph signifying a dataset is significant for perceiving the internal properties of the application it demonstrates, as well as for advancing storage and information recovery. One can simply visualize a graph as a set of dots (vertices) and links (edges) connecting them, which in numerous cases makes this demonstration suitable and easily comprehensible (Boginski et al., 2006). A natural graph representation of our drugs is based on the cross-correlations of their sales. The graph is made as below: a vertex stands for each drug, and two vertices are associated by an edge if the correlation coefficient of the corresponding couple of drugs (estimated over a certain period of time) surpasses a particular threshold (Boginski et al., 2006).

Let  $G = (V, E)$  be an undirected graph with the set of  $n$  vertices  $V$ , and the set of edges  $E = \{(i, j) : i, j \in V\}$ . We say that the graph  $G = (V, E)$  is connected if there is a path from any vertex to any vertex in the set  $V$  (Boginski et al., 2006). If the graph is separated, it can be decomposed into numerous connected sub-graphs, which are referred to as the linked components of  $G$ . The degree of a vertex is the number of edges emerging from it (Boginski et al., 2006). "A clique in a graph is a set of completely interconnected vertices" (Bomze et al., 1999; cited by Boginski, et al, 2005). "An independent set is a set of vertices without connections" (Boginski, et al, 2005).

In this section, we carry out all our analysis by means of UCINET and NetDraw softwares, which are useful for network analysis. Firstly, we need to have a correlation matrix for sales of drugs that we talked about them in data preprocessing phase. We built different incident matrixes

for various values of correlation threshold ( $\theta$ s). In fact, low figure for  $\theta$  is not accepted as most drugs would be connected to each other and we would have almost a full network. In addition, with high figures for  $\theta$  we would not have any logical network at all. Because of the fact that we wish to have a few groups with high similarity inside them, both of high and low figures are not acceptable. Thus, as mentioned before, we chose  $\theta = 0.65$  as our threshold because it gives us appropriate number and combination of cliques. Actually, this cross- correlation matrix acts as our input to build the sale's network.

- **Clique Sets**

First, we detected cliques (groups) in our network by means of UCINET software and threshold of 0.65. In our analysis of the drugs' graph, we chose a relatively high correlation that would ensure that we consider only the edges corresponding to the pairs of drugs, which are significantly correlated with each other. The results are demonstrated below. As it is obvious, 63 cliques or groups were found which contain the drugs that have cross-correlations more than 0.65 with each other. We set minimum clique size to 3 which is logical. From the result, it is found that each drug belongs to which clique or cliques. The cliques may have overlaps, i.e. some drugs may belong to more than one cliques. For instance, we can observe the fact that some drugs like 24, 174, and 103 exists in most cliques. It is not a draw back at all; otherwise, it indicates that these drugs can be very useful for prediction of many other drugs, also their future sales can be predicted by use of many other drugs. That is why these kinds of drugs are important to us. Clique-sets that we found are shown in Table 6 and it stated that which members (drugs) are in each clique. Figure 32 shows a part of UNICET software's results for finding clique-sets.

**Table 7: Cliques Sets**

Cliques	Members in each Clique					
1	24 30 56 95 102 103					
2	19 24 30 56 102 103					
3	22 24 30 32 56 102					
4	24 30 32 38 102					
5	24 30 38 102 103					
6	24 30 95 102 103 174					
7	17 20 23 24					
8	17 22 23 24 30					
9	22 23 24 30 32					
10	23 24 30 32 38					
11	17 23 24 30 38					
12	17 22 24 30 56					
13	22 24 37 39					
14	20 24 37 39					
15	22 24 39 127					
16	24 41 103 132					
17	20 24 41					
18	24 95 102 103 127 174					
19	22 24 102 127					
20	24 132 199					
21	2 56 95 102					
22	2 27 56 102					
23	2 66 95					
24	12 14 15 106					
25	12 14 15 204					
26	12 13 14 106					
27	12 14 169					
28	14 15 16					
29	22 26 27					

<b>30</b>	26	27	79			
<b>31</b>	22	27	32	56	102	
<b>32</b>	40	71	96	216		
<b>33</b>	40	71	122			
<b>34</b>	42	43	132			
<b>35</b>	41	43	132			
<b>36</b>	45	49	50	145	167	168
<b>37</b>	45	48	49	50		
<b>38</b>	45	50	145	167	168	214
<b>39</b>	49	167	168	209		
<b>40</b>	41	57	138			
<b>41</b>	61	71	122			
<b>42</b>	61	71	174			
<b>43</b>	71	95	96	174	216	
<b>44</b>	71	95	103	174		
<b>45</b>	71	95	114	216		
<b>46</b>	71	121	216			
<b>47</b>	71	114	122			
<b>48</b>	71	130	174			
<b>49</b>	56	78	103			
<b>50</b>	78	103	127			
<b>51</b>	39	78	127			
<b>52</b>	22	32	56	134		
<b>53</b>	145	167	214	215		
<b>54</b>	145	167	201	215		
<b>55</b>	150	151	156	157	174	
<b>56</b>	151	156	157	158	159	
<b>57</b>	154	155	161	162		
<b>58</b>	156	157	161	162		
<b>59</b>	156	157	159	161		
<b>60</b>	127	157	174			

<b>61</b>	167	168	209	214
<b>62</b>	95	102	175	
<b>63</b>	127	174	201	

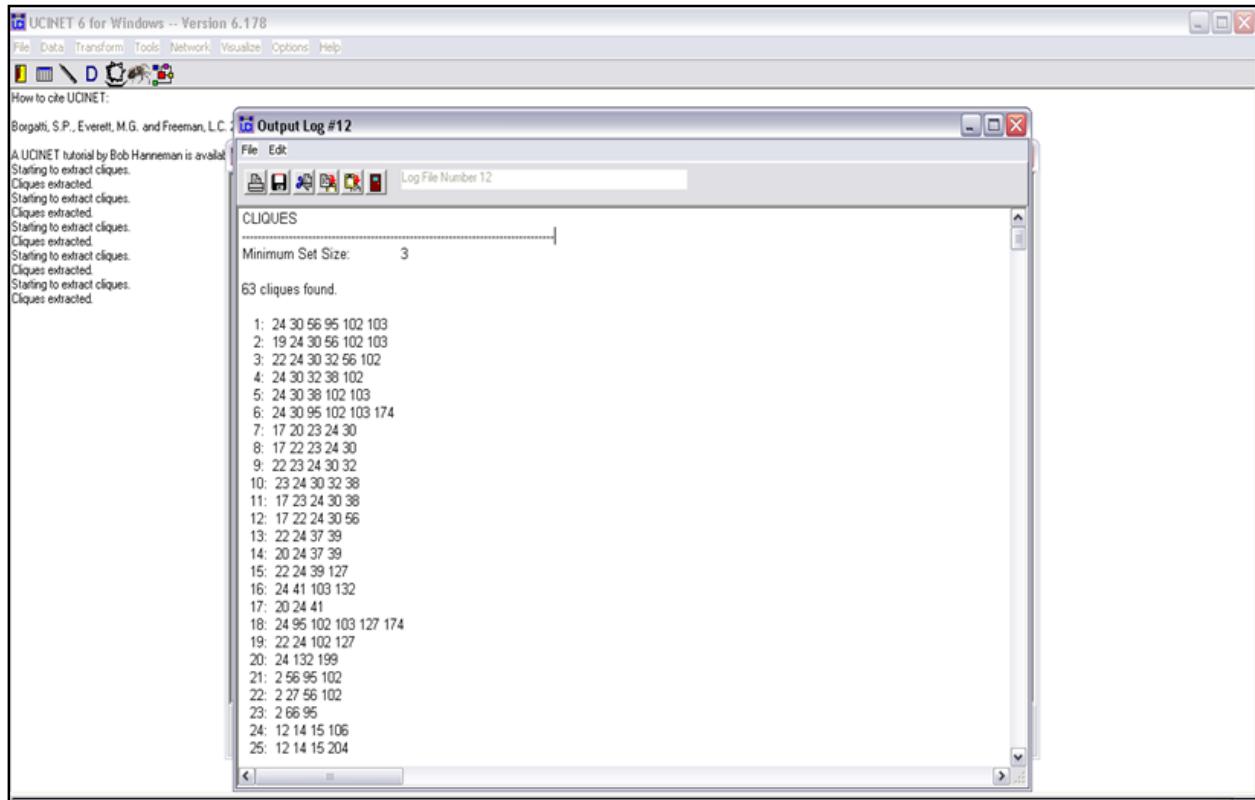
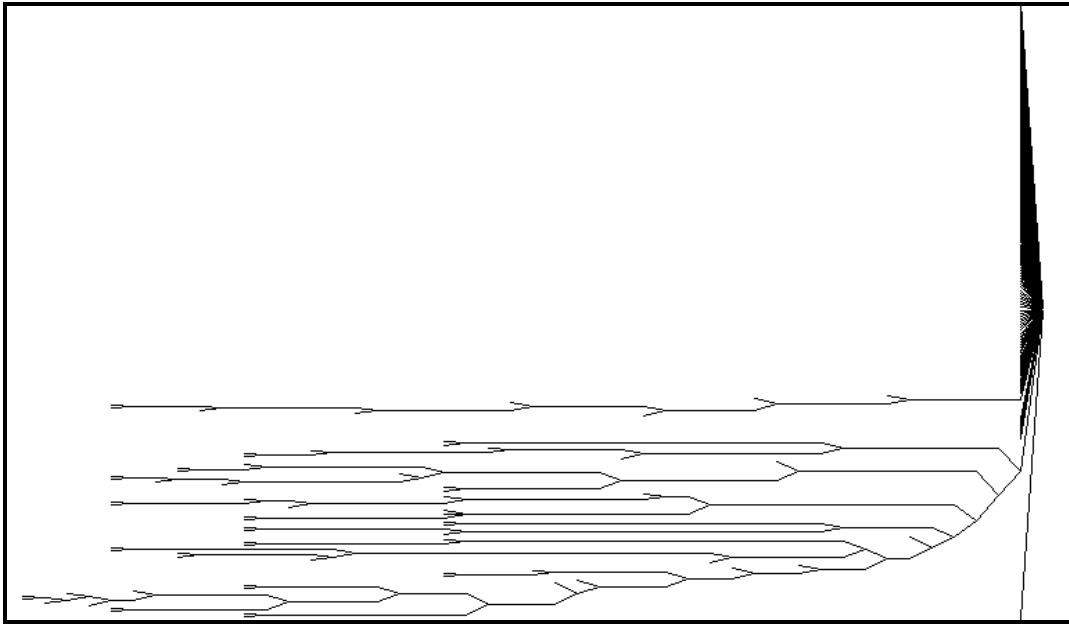


Figure 31: A Part of Output of UCINET for Clique-sets

- **Tree Diagram**

With UCINET, we draw the Tree or Cluster Diagram of our cliques, which is shown in Figure 33. This diagram is like a tree with many branches and leaves. Actually, the branches are signs of each clique and the leaves in each branch illustrate the drugs in the related clique (branch). The black and dense part shows that there are many cliques with small number of members such as 3, 2, 1 or even 0.



**Figure 32: Cluster (Tree) Diagram**

- **Degree Centrality Measures**

We also found the Table of Freeman's Degrees Centrality Measures that presents the degree of each drug (node), which means that each drug is connected to how many drugs. The results of the degree measures were demonstrated in a table that contains code of 217 drugs and degrees of each of them are shown. Unfortunately, we could not bring the table of degrees here as it would be a table with 217 columns. However, we bring just a part of this table in UCINET environment in Figure 34. From this table, we saw that there are some drugs that are not located in any cliques such as drugs that have only one or two co-members; also, there are also some drugs that are located alone and we cannot make use of other drugs' sales records for them. We also obtained some descriptive statistics for degrees that are presented in Table 7.

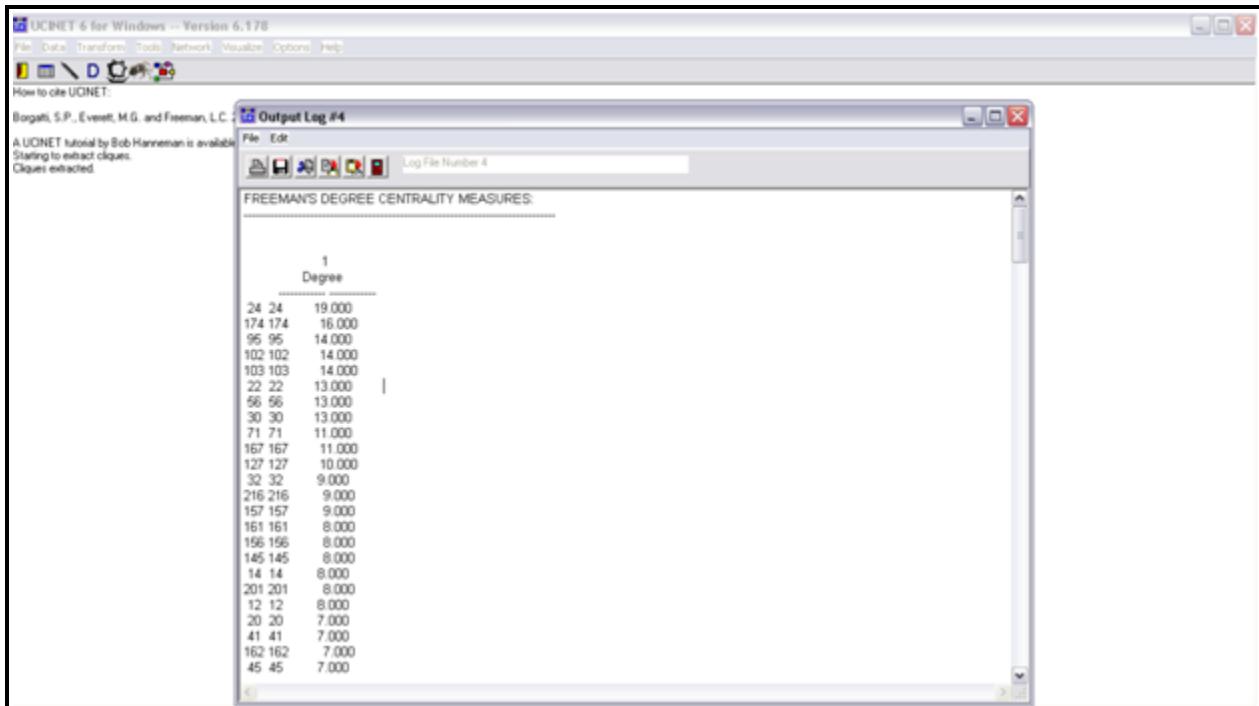


Figure 33: A Part of Output of Degree Centrality Measures

Table 8: Descriptive Statistics of Degrees

Mean of Degree	2.516
Std Dev of Degree	3.561
Sum of Degree	546.000
Variance Degree	12.683
Minimum of Degree	0
Maximum of Degree	19.000

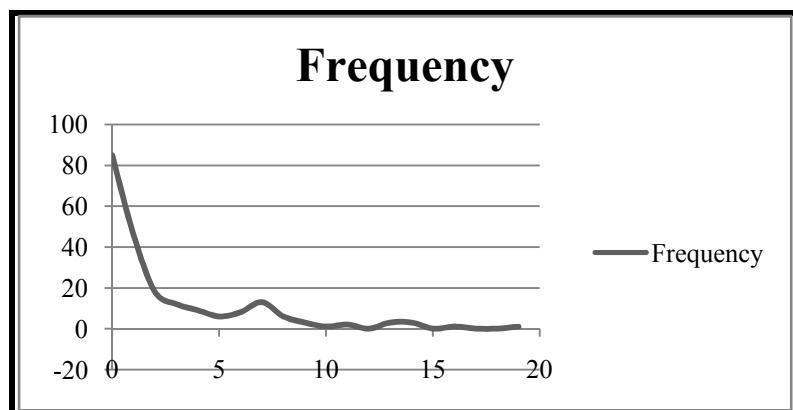
- Frequency and Cumulative Frequency of Degrees

We also found degree frequency and cumulative frequency in Table 8. Then, we draw Figure 34 that displays degree distribution  $P(k)$  which is a function describing total number of vertices in a graph with given degree (number of connection to other vertices). In figure 35, we observe that frequency falls sharply till degree 5 and then it grows slightly till degree 7, and then it drops

again with slight fluctuations. Hence, the degree distribution is not normal at all. Although it falls and rises (fluctuates) in some point, we can say that it is approximately skewed to the right.

**Table 9: Frequency and Cumulative Frequency of Degrees**

Degree	Frequency	Cum-Frq
0	85	85
1	46	131
2	18	149
3	12	161
4	9	170
5	6	176
6	8	184
7	13	197
8	6	203
9	3	206
10	1	207
11	2	209
12	0	209
13	3	212
14	3	215
15	0	215
16	1	216
17	0	216
18	0	216
19	1	217
$\Sigma=217$		

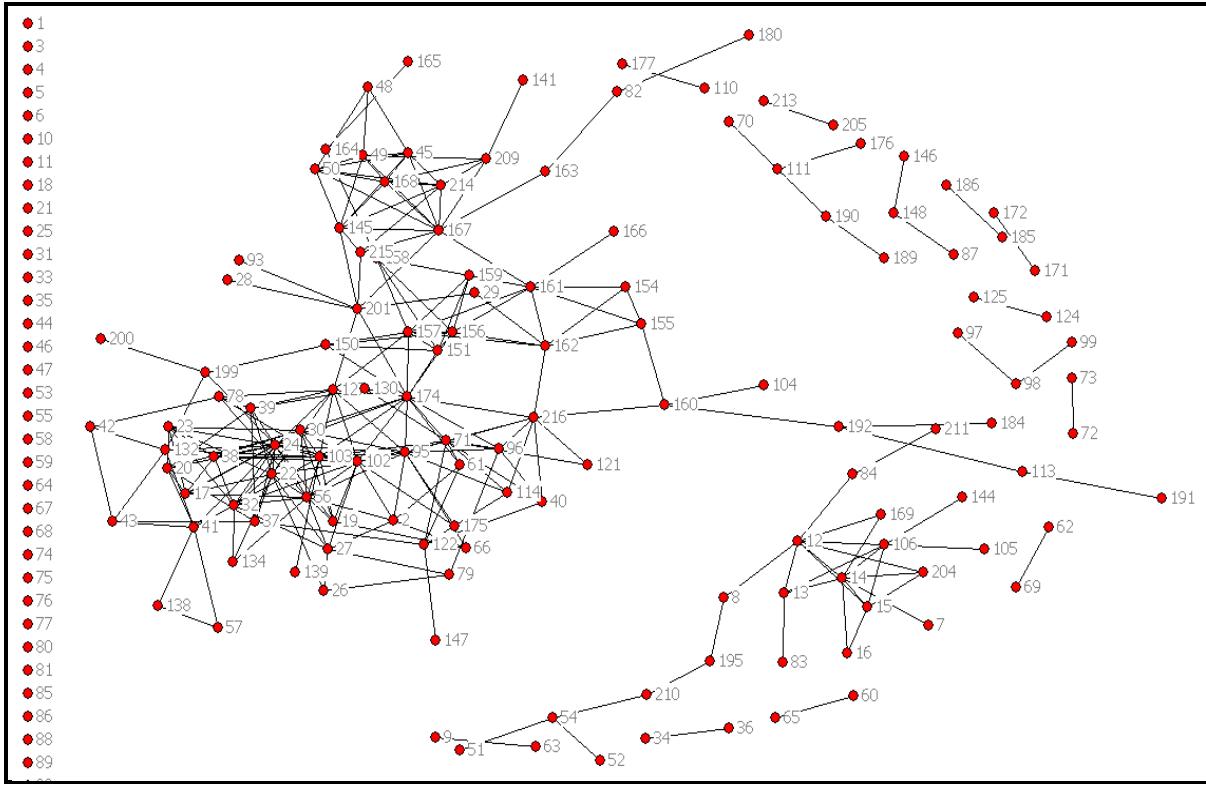


**Figure 34: Degree Distribution of the market graph**

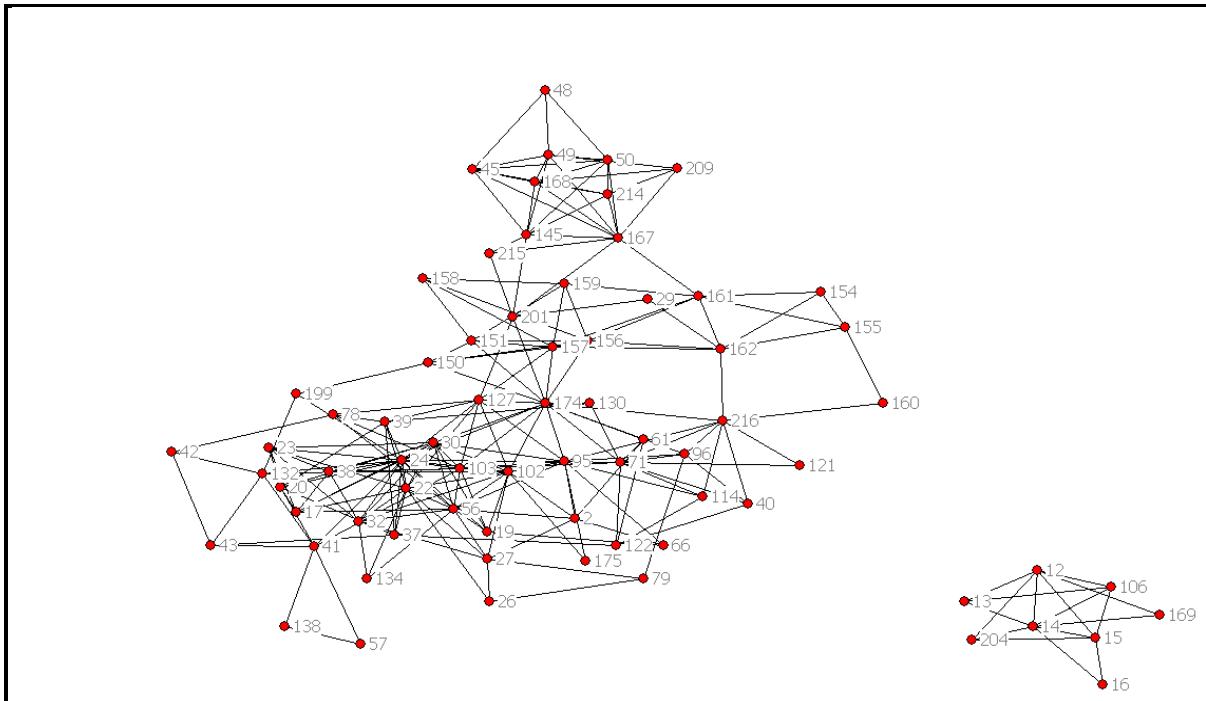
- **Visualizing the Networks of Drugs**

After detecting clique-sets and finding degree centrality measures, we visualized the network of our drugs with NetDraw software in order to summarize our findings in previous section. Figure 36 shows us an overall network for drugs. In this figure, we can see that which drugs are connected to each other. For instance, some drugs like drug 1 are alone and are not connected to any other drugs, so we cannot predict their sales by using sales records of other drugs. However, some drugs like 24, 174 and 103 are pivotal drugs as they are connected to many other drugs, so for their sales prediction we can make use of many other drugs. In addition, we can use of their records for prediction of many drugs which are connected to these drugs. Accordingly, these kinds of drugs are very beneficial for us. After deleting isolated nodes and nodes with degree 1, Figure 37 is obtained. As it is obvious, in this network there are only two main components.

Obviously, by studying the characteristics of the drugs that are in the same cliques, pharmacists and managers may extract many useful hints that would help them in their marketing planning.



**Figure 35: Networks of Drugs**



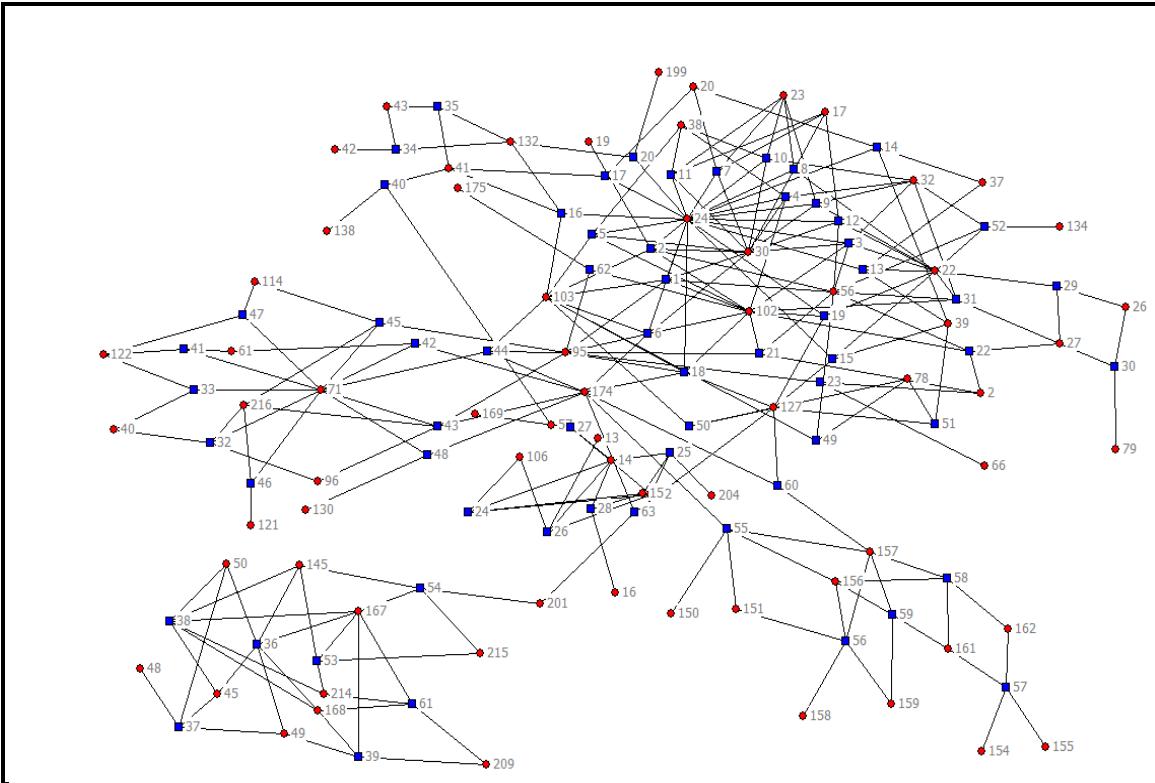
**Figure 36: Network of Drugs after Deleting Isolated Nodes and Nodes with Degree 1**

- **Visualizing the Network of Clique Indicators**

Subsequent to drawing drugs' network, we draw networks of clique set with NetDraw software and it is presented below in Figure 37. As it is clear, red circles (nodes) display drugs and blue squares demons tare cliques. Thus, we have 217 circles and 63 squares. Solid lines connect drugs to their cliques. For example, clique 30 is connected to drug 26, drug 79, and drug 27. It means that clique 30 contains these three drugs. It is also possible that one drug belongs to some cliques like drug 27 that belongs to clique 30, clique 22, clique 29, and even clique 31. Maximum number of drugs in each clique set is 6, and minimum of it is 3. In previous section, we found the degree of each drug or each node; we can also see the same matter in the following figure. For instance, output of degree centrality measures indicated that the degree of node 24 was 19 and if we observe Figure 38 as well, we can see that node (drug) 24 is connected to 19 cliques (squares).

Actually, Managers or pharmacists can benefit a lot if they study the structure of these graphs carefully since they would give them many useful massages. For instance, they would have clusters or groups of drugs that are not based on healing characteristic of medicines, instead, are based on their sales behavior. Thus, they could package these drugs and allocate a special visitor and marketing planning to each package.

In summary, our outputs in this section would be of many benefits to us because the value of a graph is much more than many sentences. After visualization and network analysis, we found the importance of each drug. If we have not draw these graph, we could not understand which drugs (like 24 and 174) are so important to us, and what are dense groups or cliques. The output of this section would help us in prediction phase as we know that we should select the past records of which drugs to forecast the future sales of each drug. In the succeeding part, we would better perceive the significance of our network analysis.



**Figure 37: Network of Clique Indicators**

#### 4.4. Building Sales Forecasting Model

Forecasting has a significant position in business planning (Zhang, 2004). The capability to precisely forecast the future is essential to numerous decision actions in retail, marketing, manufacturing, inventory control, personnel, and many other business functional fields (Zhang, 2004). Increasing forecasting precision can ease the savings of a lot of money to a company and is a main incentive for applying formal systematic forecasting techniques and for investigating novel and superior forecasting techniques. Time series modeling method is one of the main approaches broadly applied in practice (Zhang, 2004). Many efforts have been dedicated to expanding and advancing time series forecasting techniques (Zhang, 2004). As stated before, there are two approaches to time series modeling and forecasting linear approach and nonlinear approach. There are a large number of linear forecasting models such as moving average, exponential smoothing, and time series regression and so on. One of the most important and popular linear models is the autoregressive integrated moving average (ARIMA) that was popularized by Box and Jenkins in the 1976. “Although ARIMA models are quite flexible in

modeling a wide range of time series patterns, their major limitation is the presumed linear form of the model" (Zhang, 2004). To be precise, a linear autocorrelation structure is supposed sooner than the model is fitted to the data (Zhang, 2004). "Therefore, an ARIMA model is not able to capture nonlinear patterns that are commonly seen in many business and economic time series" (Zhang, 2004). "Linear models have been used for a long time and are still very useful, but the linear assumption underlying these models may be too restrictive. The nonlinear approach to time series modeling is perhaps more appropriate for most real world problems" (Zhang, 2004).

. "Nevertheless, the nonlinear world is much more complex than the linear world since there are so many possible nonlinear relationships or structures. Most nonlinear models developed during the last two decades are parametric in nature" (Zhang, 2004). "To use these models, the model forms have to be specified first. Therefore, these models may not be useful if the data characteristics do not match model assumptions" (Zhang, 2004). A more adaptable nonlinear model is the artificial neural networks (ANNs) which have obtained more considerations last decade (Zhang et al., 1998; cited by Zhang, 2004). The main advantage of neural networks is that they are data driven and do not need limiting suppositions about the structure of the basic model. This adaptable nonparametric method is proper for numerous problems with complicated nonlinear structures (Zhang, 2004). In this research, we used both linear models (ARIMA) and nonlinear models (hybrid neural networks) and compared their results.

As mentioned before, pharmaceutical distribution companies are facing many difficulties in Iran. Firstly, they have to be always ready to meet their customers' needs. Thus, they are usually forced to carry a large standing inventory of products ready to deliver to their customers (drugstores or hospitals). These companies like Pakhsh Hejrat, incur significant financial costs if they carry excess quantities of drugs relative to the customer demand and they meet with financial costs if they carry excessive inventory levels, especially because pharmaceutical drugs have limited shelf lives. On the other hand, due to the importance of their products that are directly related to health or death of people, shortage is not accepted in this industry at all. Because of negative experiences, unsatisfied customers (drugstores or hospitals) may switch company loyalties and turn to competing distributors to serve them, so they would loss potential profits in such cases. In addition, this company, like many other distribution companies, suffers

from shortage of past records (which is only 36 months' sales data) so they cannot make prediction by means of traditional time series forecasting methods, like ARIMA that needs at least 50 past records. Furthermore, in our exploratory analysis we found that the relationships among drugs were both linear and nonlinear, but mostly nonlinear. Moreover, it is widely accepted by a large number of researchers in empirical tests that Box-Jenkins is not a precise technique for post-sample time-series forecasting, at least in the areas of business and economic applications where the level of randomness is high and where uniformity of pattern, or relationships, cannot be certain (Makridakis and Hibon 1997). Hence, we could not rely on only linear model such as ARIMA and even on a single nonlinear model. As a result, we had to solve both problems (shortage of past records and nonlinearity). These problems caused us to come to the idea of an accurate and innovative sales prediction for Pakhsh Hejrat Company. Although we had to ignore ARIMA methodology because of its linear property and because of not having enough past sales records, we are bringing an example of it to see its performance on our time series in comparison with our newly offered methodology.

As stated before, we had sales records of near 1200 drugs, but we could not make a single model for all these drugs since they are not the same products and many of them have different characteristics. We selected 217 of these drugs that were sold in all 36 months to be able to consider equal condition for them; also make use of more records. Then, we decided to make use of similarity in sales behavior of some drugs by means of their cross-correlation matrix to both increase our past sales records and improve the performance of our model. Consequently, we grouped them in 63 cliques as it was explained in detail in previous section. To build our forecasting model, we had to examine our methodology on many different drugs in order to be able to prove the validity of it. However, there is no need to try all 217 kinds of drugs, as the methodology is the same for all of them. Accordingly, by use of Freeman's degree centrality table that presents degree of each drug and clique-set matrix; we extracted 21 of drugs from various degrees and build Table 9 that explicitly presents each drug with its group members. This table would be one of the main tools for our prediction in this section. For instance, to predict the future sales of drug 12 in addition to its own past sales records, we would use the sales records of drugs 8, 13, 14, 15, 84, 106, 169, and 204. We are explaining the steps and results of our prediction with ARIMA and hybrid neural networks models in detail in next sections. All steps of our modeling phase were done with STATISTICA (version 7) software.

**Table 10: Group Members**

Drugs	Group Members																			
<b>24</b>	17	19	20	22	23	24	30	32	37	38	39	41	56	95	102	103	127	132	174	199
<b>174</b>	24	30	61	71	95	96	102	103	127	130	150	151	156	157	174	201	216			
<b>95</b>	2	24	30	56	66	71	95	96	102	103	114	127	174	175	216					
<b>103</b>	19	24	30	38	41	56	71	78	95	102	103	127	132	139	174					
<b>22</b>	17	22	23	24	26	27	30	32	37	39	56	102	127	134						
<b>30</b>	17	19	20	22	23	24	30	32	38	56	95	102	103	174						
<b>71</b>	40	61	71	95	96	103	114	121	122	130	174	216								
<b>167</b>	45	49	50	145	161	163	167	168	201	209	214	215								
<b>127</b>	22	24	39	78	95	102	103	127	157	174	201									
<b>132</b>	22	23	24	27	30	32	38	56	102	132	134									
<b>12</b>	8	12	13	14	15	84	106	169	204											
<b>168</b>	45	49	50	145	167	168	209	214												
<b>37</b>	20	22	24	37	39	43	122													
<b>78</b>	39	42	56	78	103	127														
<b>61</b>	2	61	71	122	174															
<b>66</b>	2	19	66	95																
<b>84</b>	12	84	211																	
<b>124</b>	124	125																		
<b>51</b>	51	54																		
<b>31</b>	Empty																			
<b>86</b>	Empty																			

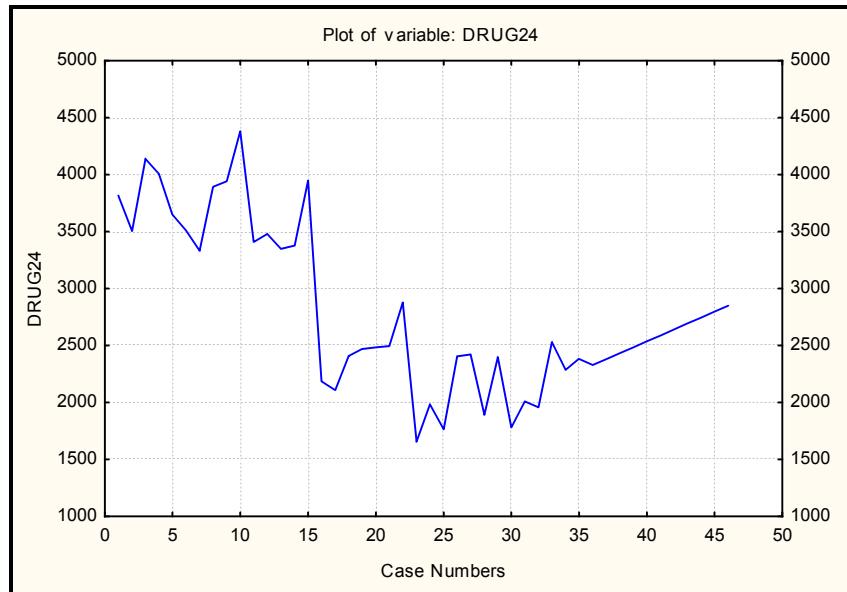
#### 4.4.1. ARIMA Methodology for Time Series Forecasting

In this section, we carried out all steps with STATISTICA software and we used only the past sales data of each drug, as it is not acceptable to use past records of each drug's group members in ARIMA methodology. As stated in chapter 2, general model of autoregressive moving average model introduced by Box and Jenkins, (1976) includes autoregressive as well as moving average parameters, and explicitly includes differencing in the formulation of the model. Specifically, the three types of parameters in the model are: the autoregressive parameters (p), the number of differencing passes (d), and moving average parameters (q). In the notation introduced by Box and Jenkins, models are summarized as ARIMA (p, d, q); so, for example, a

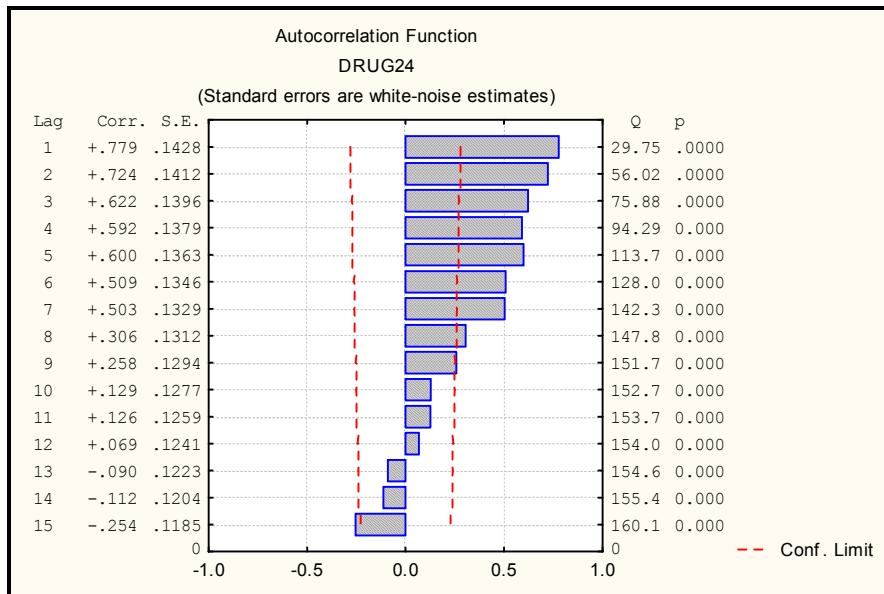
model described as (0, 1, 2) means that it contains 0 (zero) autoregressive (p) parameters and 2 moving average (q) parameters which were computed for the series after it was differenced once. We bring results of all steps of sales prediction with ARIMA methodology and with STATISTICA software for drug 24 as the methodology is the same for all drugs. However, this methodology is not our main method in this research. All steps of ARIMA methodology are described below:

- **Step 1: Transforming the Data**

As stated before, the input series for ARIMA requires to be stationary, i.e., it ought to have a constant mean, variance, and autocorrelation through time (STATISTICA 7, Electronic Manual). Consequently, typically the series initially requires to be differenced until it is stationary (it also frequently needs log transforming the data to make the variance steady) (STATISTICA 7, Electronic Manual). The number of times the series requires to be differenced to accomplish stationarity is demonstrated in the d parameter. With the aim of finding out the essential level of differencing, we should examine the plot of the data and autocorrelogram (STATISTICA 7, Electronic Manual). “Significant changes in level (strong upward or downward changes) usually require first order non seasonal (lag=1) differencing; strong changes of slope usually require second order non seasonal differencing. Seasonal patterns require respective seasonal differencing. If the estimated autocorrelation coefficients decline slowly at longer lags, first order differencing is usually needed” (STATISTICA 7, Electronic Manual). However, several time series may need little or no differencing, and that over differenced series generate less constant coefficient approximations (STATISTICA 7, Electronic Manual). If we consider the time series plot of this drug which is shown in Figure 39, we see that it is not stationary at all as it has not constant mean and variance through time. Thus, it needs both log transforming and differencing. We also should consider its autocorrelogram (Figures 40 and 41) to see that whether it has seasonal pattern or not. Since it has not sinuous fluctuations, it does not need seasonal differencing. After two times differencing and log transforming the series became approximately stationary (Figure 42).



**Figure 38: Time Series Plot of Drug 24 Sales Data**



**Figure 39: Autocorrelation Diagram**

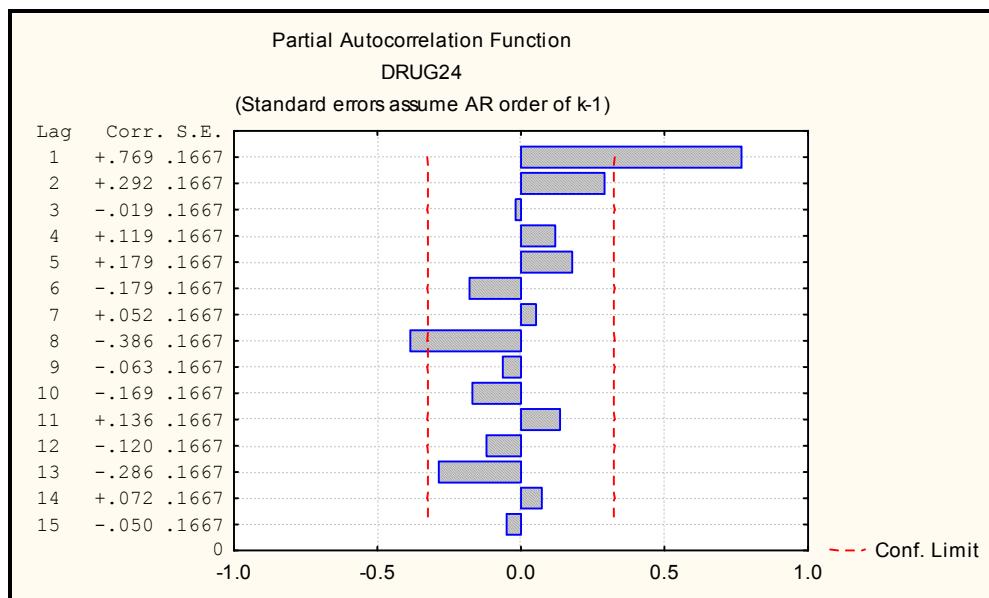


Figure 40: Partial Autocorrelation Diagram

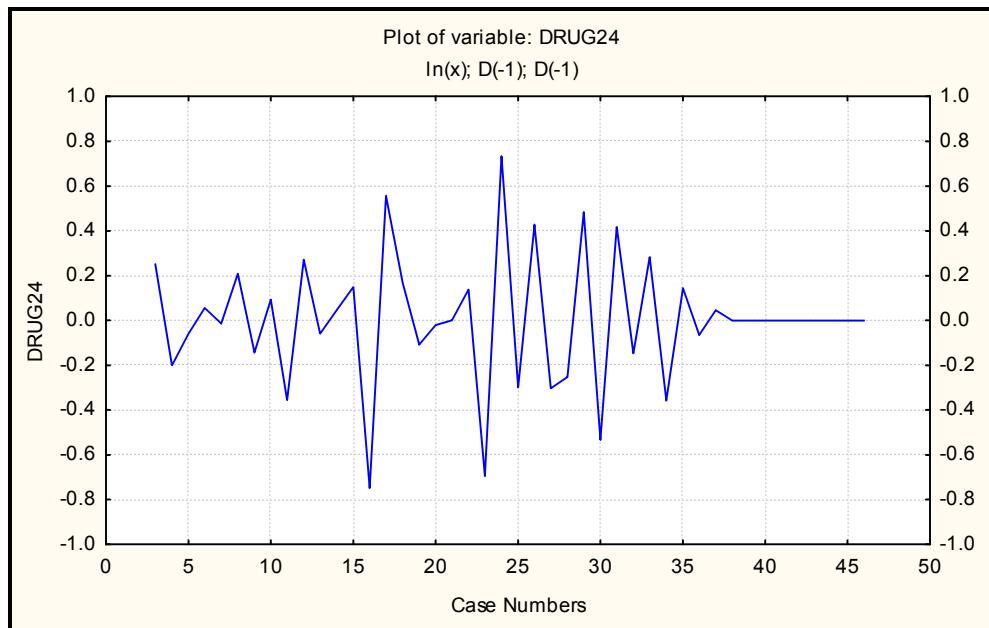


Figure 41: Time series plot of drug 24 sales data after differentiation and log transforming

- **Step 2: Identification**

At this phase (which is commonly called identification stage), we also require to choose how many autoregressive (p) and moving average (q) parameters are required to achieve an efficient but still parsimonious model of the procedure (parsimonious represents that it has the fewest parameters and maximum number of degrees of freedom among all models that adjust the data)

(STATISTICA 7, Electronic Manual). In practice, the numbers of the p or q parameters hardly ever require to be greater than 2. The main tools applied in the identification stage are plots of the series, correlograms of autocorrelation (ACF), and partial autocorrelation (PACF). The choice is not straightforward and in usual cases needs not only experience but also a proper deal of experimentation with substitute models (STATISTICA 7, Electronic Manual). Nevertheless, a majority of empirical time series patterns can be adequately estimated by means of one of the 5 fundamental models that can be recognized according to the form of the autocorrelogram (ACF) and partial autocorrelogram (PACF). The subsequent concise summary is derived from practical suggestions in literature reviews (McCleary and Hay, 1980; McDowall, McCleary, Meidinger, and Hay, 1980; Pankratz, 1983; Hoff, 1983; Vandaele, 1983; cited by STATISTICA 7, Electronic Manual):

1. One autoregressive (p) parameter: ACF - exponential decay; PACF - spike at lag 1, no correlation for other lags.
2. Two autoregressive (p) parameters: ACF - a sine-wave shape pattern or a set of exponential decays; PACF - spikes at lags 1 and 2, no correlation for other lags.
3. One moving average (q) parameter: ACF - spike at lag 1, no correlation for other lags; PACF - damps out exponentially.
4. Two moving average (q) parameters: ACF - spikes at lags 1 and 2, no correlation for other lags; PACF - a sine-wave shape pattern or a set of exponential decays.
5. One autoregressive (p) and one moving average (q) parameter: ACF - exponential decay starting at lag 1; PACF - exponential decay starting at lag 1.

We also have to decide about seasonality. Multiplicative seasonal ARIMA is a generalization and extension of the method mentioned in the earlier paragraphs to series in which a pattern recurs seasonally over time. Besides the non-seasonal parameters, seasonal parameters for a particular lag (determined in the identification phase) require to be approximated. Analogous to the simple ARIMA parameters, these are: seasonal autoregressive (ps), seasonal differencing (ds), and seasonal moving average parameters (qs). For instance, the model (0,1,2) (0,1,1) explains a model that consists of no autoregressive parameters, 2 regular moving average parameters and 1 seasonal moving average parameter, and these parameters were computed for

the series later than it was differenced once with lag 1, and once seasonally differenced. The seasonal lag applied in the seasonal parameters is typically established through the identification stage and must be clearly specified (STATISTICA 7, Electronic Manual). The common suggestions regarding the selection of parameters to be approximated (based on ACF and PACF) also employ in seasonal models. The major difference is that in seasonal series, ACF and PACF will demonstrate considerable coefficients at multiples of the seasonal lag (besides their general patterns indicating the non seasonal elements of the series) (STATISTICA 7, Electronic Manual).

Accordingly, we should now decide how many autoregressive (p) and moving average (q) parameters are necessary according to Figure 43 and 44 (autocorrelation and partial-autocorrelation functions). From the visual examination of plots in Figure 42 and 43 it is obvious that our case is similar to case number 2 above without seasonal parameters. However, all possible ARIMA models were identified for the data series. Up to this step, the best model for sales data of drug 24 was ARIMA (2, 2, 0). To be sure, of this step results, following steps should be followed.

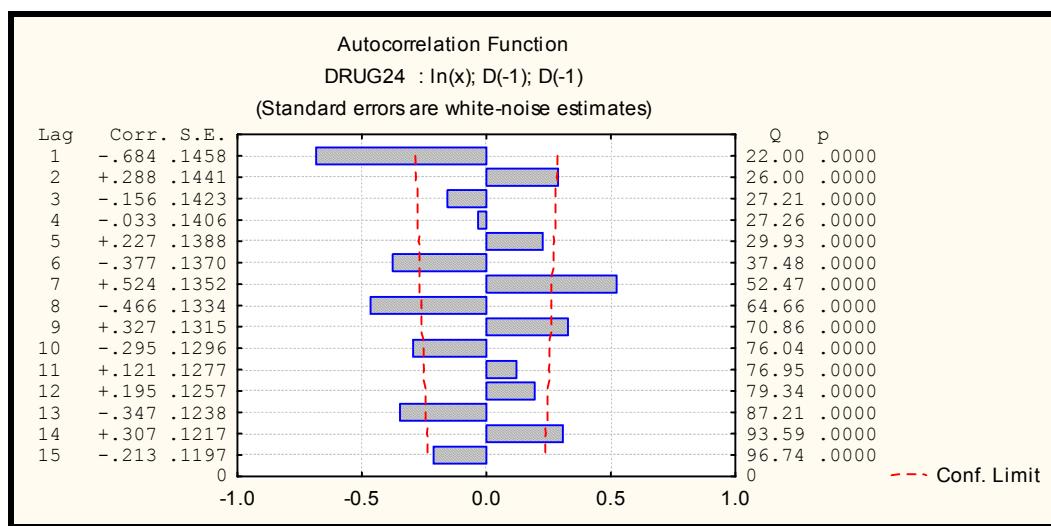


Figure 42: Auto-correlation Function of Drug 24 Sales Data

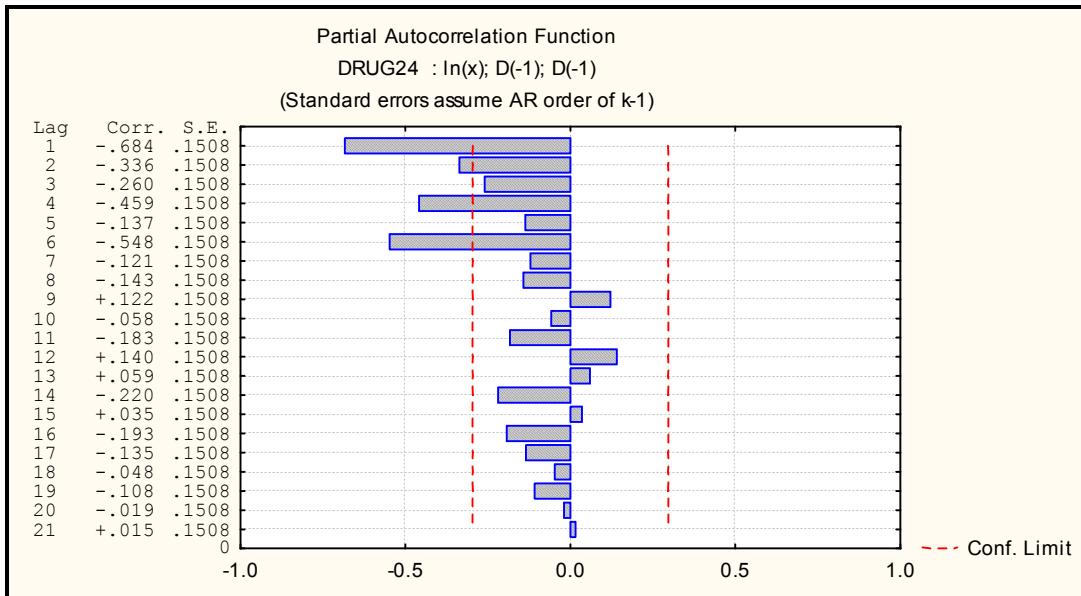


Figure 43: Partial Auto-correlation Function of Drug 24 Sales Data

- **Step 3: Estimation and Forecasting**

At this phase (Estimation), the parameters are approximated by applying function minimization procedures, so that the sum of squared (SS) residuals is minimized. The estimates of the parameters are applied in the last step (Forecasting) to compute new values of the series (further than those contained in the input data set) and confidence intervals for those forecasted values (STATISTICA 7, Electronic Manual). The estimation step is performed on transformed (differenced) data; before the forecasts are produced, the series is integrated (integration is the inverse of differencing) in order that the forecasts are indicated in values well matched with the input data. This automatic integration feature is demonstrated by the letter *I* in the name of the method (ARIMA = Auto-Regressive Integrated Moving Average) (STATISTICA 7, Electronic Manual). We used the exact maximum likelihood (Melard) technique according to Melard et. All, (2006) to compute the SS for the residuals as usually only a few iterations by means of the exact maximum likelihood technique are required to complete the parameter estimates. According to this method, we find parameters in STATISTICA and Table 10 below shows exactly the output of STATISTICA. From this table we can see that only p-value of P (1) is less than .001; thus we understand that only this parameter is significant and we do not need to add constant. Accordingly, the best model for our sales data is (1, 2, 0). The result of our forecasting for the test data is presented in Table 11. If we observe the values of residuals (the difference of

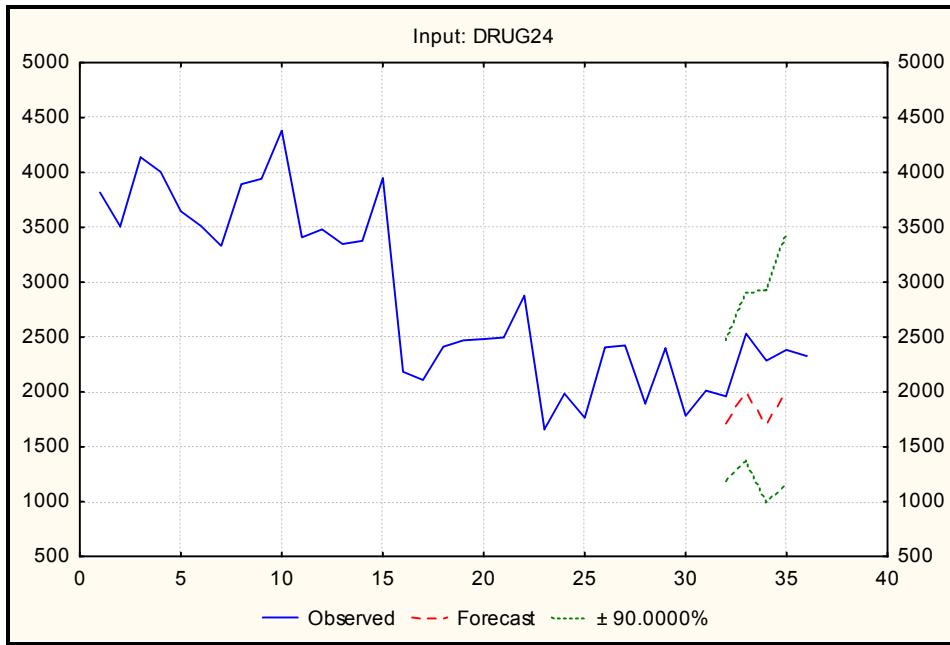
observed and forested values), MSE, and MAE, we see that although we chose the best ARIMA model, the result is not satisfactory enough. In next section, we are going to compare the result of this model with ANN and we will see that how better ANN is than ARIMA for our case. We can also observe Figure 45 to better see the forecasting result for test data. As it is clear, the red dashed line related to our forecasting and green dashed lines show the upper and lower limits according to confidence level of 90%.

**Table 11: Parameter Estimation**

Input: DRUG24 : Transformations: ln(x); D(-1); D(-1), Model:(1,2,0), MS Residual= .04304						
	Param.	Asympt. Std.Err	Asympt. t	p	Lower	Upper
<b>Constant</b>	0.000921	0.014248	0.06461	0.948795	-0.02785	0.029695
<b>p(1)</b>	<b>-0.913700</b>	0.148895	-6.13652	<b>0.000000</b>	-1.21440	-0.613000
<b>p(2)</b>	-0.336469	0.148895	-2.25977	0.029214	-0.63717	-0.035769
Initial ss=3.7364, Final ss=1.7659, MS=.04307						

**Table 12: Four Months' Forecasting Results for the Original Data**

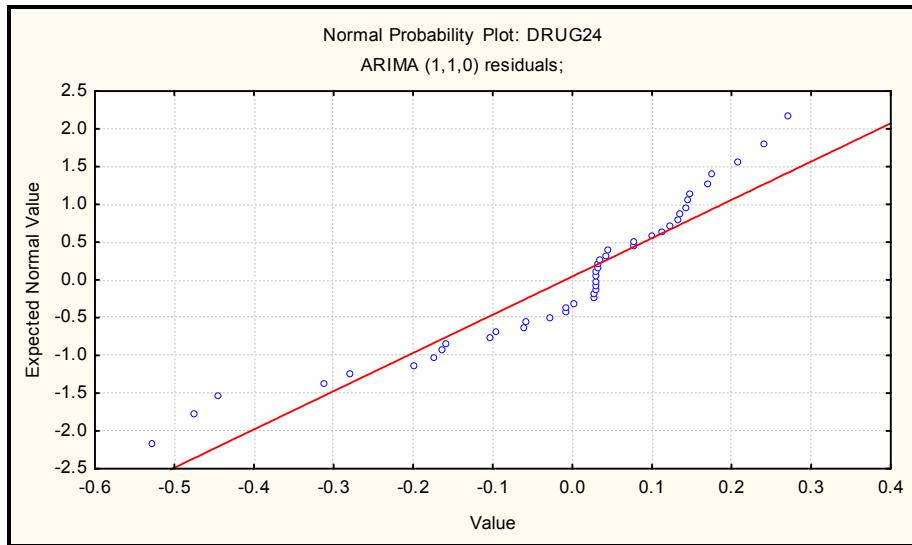
Input: DRUG24, Forecasts; Model:(1,2,0), MS Residual=.03312,						
	Forecast	Lower Limit	Upper Limit	Observed	Residual	MSE
33	2054.746	1429.366	2953.743	2531.000	476.2539	140793.3
34	1968.661	1356.720	2856.614	2286.000	317.3387	
35	2057.058	1206.779	3506.434	2384.000	326.9417	
36	1969.168	1147.046	3380.527	2328.000	358.8325	



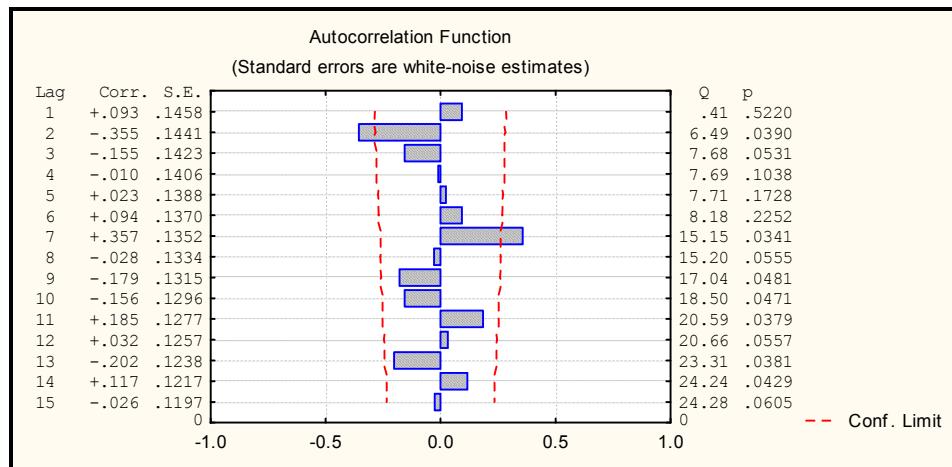
**Figure 44: Plot of Forecasting**

- **Step 4: Evaluation of the ARIMA Methodology (Diagnosis Checking)**

The simple and frequent measure of the reliability of the model is the precision of the forecasts it produces according to partial data, in order that the predictions can be compared with recognized (original) observations (STATISTICA 7, Electronic Manual). Nevertheless, a proper model should not only offer adequately precise forecasts, it should also be parsimonious and generate statistically independent residuals that include only noise and no systematic elements (e.g., the correlogram of residuals should not display any serial dependencies) (STATISTICA 7, Electronic Manual). A fine test of the model is to plot the residuals and examine them for any systematic trends, and to inspect the autocorrelogram of residuals (there ought to be no serial dependency among residuals) (STATISTICA 7, Electronic Manual). Hence, the aim of this step is to diagnose the residuals to check whether they are random and normally distributed, indicating a fine model (Box and Jenkins, 1976; Berthouex and Brown, 2002; (STATISTICA 7, Electronic Manual). Normal probability plot of the residuals in Figure 46 approves the normality of them (being distributed about an approximately straight line). Figure 47 and Figure 48 show that there is no serial dependency between residuals. This step again confirms the goodness of model obtained in previous step.



**Figure 45: Normal Probability Plot of Residuals for ARMA (1, 1, 0)**



**Figure 46: Autocorrelation Function of Residuals**

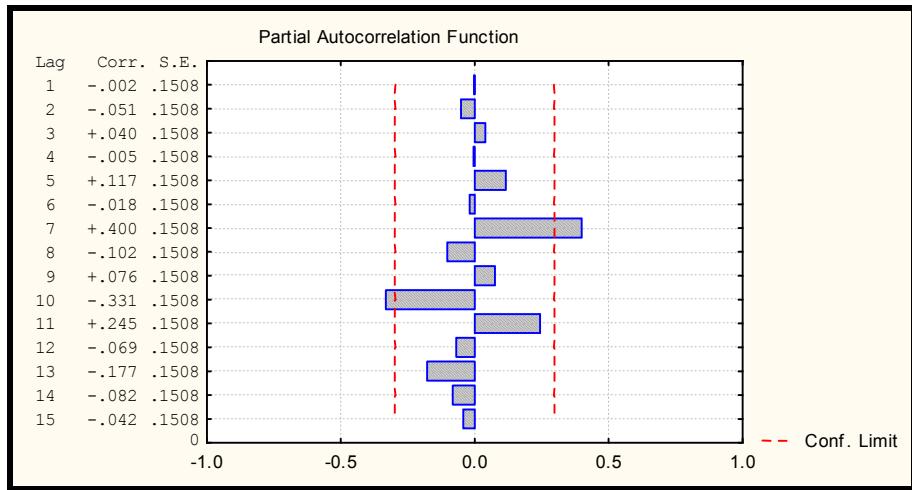


Figure 47: Partial Autocorrelation Function of Residuals

- **Limitations of ARIMA Methodology**

In addition to our poor result of ARIMA for sales forecasting above, we are bringing more theoretic reasons (according to the existing literatures) for rejection of this methodology in our research :

- 1- The ARIMA method is typically suitable for a time series that is stationary (that is its mean, variance, and autocorrelation should be roughly constant through time) (STATISTICA 7, Electronic Manual). Most of our sales data have not these characteristics.
- 2- It is suggested that there are minimum of 50 observations in the input data ) (STATISTICA 7, Electronic Manual), but we have only 36 records.
- 3- It is also supposed that the values of the approximated parameters are constant all through the series) (STATISTICA 7, Electronic Manual). It means that it only reveal linear relationships. As we found nonlinear behaviors in our data, this model is not proper for sales prediction.

Consequently, we could reject ARIMA (Box-Jenkins) methodology both empirically and theoretically.

#### 4.4.2. A Hybrid Neural Network Approach for Time Series Forecasting

“Recently, neural networks have emerged as an important tool for business forecasting” (Zhang, 2004). Subsequent to a review of the time series forecasting literature, we discovered

that neural networks acted at least as satisfactory as other methods in most of cases. However, in our case that we have nonlinear relationships and we have to put aside linear methods like ARIMA; it is the best decision to apply neural networks. Since our data exploratory phase revealed that there are both linear and nonlinear relationships among sales of drugs, we have to leave single methods like single neural networks. Neural networks encompass the capability of modeling both linear and nonlinear time series (Zhang, 2001; Zhang, 2003; Zhang, 2004). This ability is important because nonlinearities exist in our data like other real data series.

Accordingly, in this research we used a hybrid neural networks which means that we model linear parts with linear ANN and nonlinear parts with nonlinear ANN. In this section, we present a combined linear and nonlinear neural network approach for time series forecasting. “The idea is based on the fact that no single model is the best for any situation” (Zhang, 2004).

Due to the fact that we had access to only 36 records (which is not enough for time series prediction), we came to the new idea of grouping drugs according to their sales cross-correlation to increase our records and use group members’ sales records for each other. For example, if we wish to predict the sales of drug 24 which have connected to 19 other drugs, we can make use of its group members’ past sales information as well. However not all 19 drugs is used and the software do a feature selection and chose the most effective ones. In this section, one of the main tools for our prediction is Table 9 (we saw it before) because it presents the group members. This table helps us to obviously know that which drugs are useful for sales forecasting of other drugs. In general, “for building neural networks, there is no general theory that provides specifications for type of neural network, number of layers, number of nodes (at various layers), or learning algorithm” (Bansal et al., 1998). Without such an assumption, we have just common rules founded on studies by prior network designers (Bansal et al., 1998). Consequently, we must examine many neural networks before we find out the suitable neural network to resolve our crisis (Bansal et al., 1998). However, we will show that if make prediction by means of other drugs’ information we would have an innovative and successful ANN methodology.

According to Zhang, (2004), “the motivations for the combined (hybrid) method are three-fold. First, in many forecasting situations, it is often difficult to determine whether a time series under study is generated from a linear or nonlinear underlying process or whether one particular method is more effective than the other in out-of-sample forecasting”. Therefore, it is

hard for predictors to select the appropriate method for their unique condition at the beginning. Normally, various models are examined and the one with the most precise outcome is chosen (Zhang, 2004). Nevertheless, the ultimate chosen model is not certainly the best for future applications owing to numerous possible issues, such as sampling variation, model ambiguity, and structural alteration (Zhang, 2004). By merging diverse techniques, the issue of model selection can be relieved with small further attempt. Second, actual world time series are seldom fully linear or nonlinear. They frequently include both linear and nonlinear patterns (Zhang, 2004). Therefore, by hybrid models, intricate autocorrelation structures in the data can be modeled more precisely. Third, it is commonly approved in the forecasting surveys that no single technique is the finest in every circumstance (Zhang, 2004). “This is largely due to the fact that a real world problem is often complex in nature and any single model may not be able to capture different patterns equally well” (Zhang, 2004). “Therefore, combining different models can increase the chance to capture different patterns in the data and improve forecasting performance” (Zhang, 2004). Numerous empirical studies have previously recommended that by combining various models, forecasting precision can often be enhanced over the single models. Besides, the combined or hybrid model is typically more vigorous and is less possible to get considerably poorer outcomes than each individual model applied in isolation (Zhang, 2004).

Actually, our hybrid modeling includes three main steps: (1) fitting a linear neural network model to the time series under examination, (2) building a nonlinear neural network model derived from the residuals from the linear neural network model, and (3) combining the linear neural network forecasting and the nonlinear neural network outcome to form the ultimate forecast. By merging various models, we intend to take advantage of the single modeling ability of each unique model and advance forecasting performance significantly (Zhang, 2004). We are going to explain the model in more detail below.

- **The Hybrid Model**

“The hybrid model used in this study is relatively simple to understand and easy to implement in practice. We assume that a time series contains two basic components, a linear component  $L_t$  and a nonlinear component  $NL_t$ ” (Zhang, 2004). The model can be written as (Zhang, 2004):

$$Y_t = L_t + NL_t + a_t$$

Where  $Y_t$  is the time series observation at time period  $t$ ;  $L_t$  and  $NL_t$  are the linear and nonlinear formations of the time series, in that order, and  $a_t$  is the random error phrase (Zhang, 2004).

The fundamental design of this hybrid methodology is to allow linear ANN model the linear component and permit nonlinear ANN model the nonlinear element and afterward merge the outcomes from both linear and nonlinear models (Zhang, 2004). The model building procedure covers the subsequent three phases. First, a linear ANN is built founded on the sales records to approximate the linear part (Zhang, 2004). The residuals from the linear model are supposed to include several nonlinearities (since we have done a comprehensive exploratory analysis and see that in our sales data the relationships are mostly nonlinear; besides, it is widely accepted by researchers that an actual data set frequently includes roughly nonlinear relationships) (Zhang, 2004). Then, we employ ANNs to the residuals to approximate the nonlinear part. Lastly, we merge these two components together to predict (Zhang, 2004).

If we name the estimated element from the linear model  $L^e_t$ , and the estimated nonlinear part  $NL^e_t$ ; subsequently, the combined prediction  $Y^e_t$  will be (Zhang, 2004):

$$Y^e_t = L^e_t + NL^e_t$$

- **Step 1: Data Selection**

As the methodology that we offer is the same for all kinds of drugs, there is no need to examine it with all 217 drugs. Therefore, we selected 21 drugs from different categories (with different degrees) and build Table 9 (that was shown before) that explicitly presents that each drug with its group members. For instance, we want to predict sales of drug 71, by using this table we know exactly that in addition to past sales records of drug 71, records of which drugs should be added to our input variables. Thus, by observing this table we notice that our input variables should be drug 40, 61, 71, 95, 96, 103, 114, 121, 122, 130, 174, and 216. However, for making a comparison, we built hybrid neural network model with two different approaches: 1) Just using each drug's own past records, 2) Using each drug's own past records and its group members' past records as well.

- **Step 2: Split Data into Train and Test Set**

Neural network model building is performed through the standard cross-validation method, that is the model parameters are approximated with the training data and then the validation sample is applied to choose the final model. By means of the training data, we approximated numerous diverse neural network models to superior acquire the concealed behavior of the time series (Zhang, 2004). Thus, for every data set, we keep the last 4 months' data as the holdout sample (test data) for forecasting assessment and the rest of the data (in-sample) are applied in model selection and estimation. The holdout sample is not applied in the model building procedure and it signifies a set of concealed future observations for testing the suggested model efficacy as well as for model comparisons (Zhang, 2004). In our neural network modeling, the in-sample data are separated into two parts of a training sample and a validation sample. The validation sample includes the last 4-month observations whereas the training sample includes all earlier observations. The training sample is applied to approximate the parameters for any explicit model building. The validation set is subsequently applied to choose the finest model and afterward this estimated model is employed in out-of-sample prediction (Zhang, 2004).

- **Step 3: Model Building (Hybrid Neural Networks)**

- **Building Linear ANN:** In this phase, for all 21 drugs and for our two approaches, we first fitted a linear neural network model to the time series under study and find the residuals.
- **Building Nonlinear ANN:** In this part, we applied ANNs to the residuals (that contain some nonlinearity) from linear model to estimate the probable nonlinear components. For this step, we tried different types of network such as, Linear, PNN or GRNN, Radial Basis Function and One, Two and Three- Layer Perceptron. In the experiment, we chose number of hidden units 0-9 for radial basis function and 0-26 for MLP and examined all of them to find the best architecture. We defined 1-6 range for steps (number of time steps used as inputs) since it is mentioned in literatures that 6 steps is enough. For activation functions, we tried linear and sigmoid logistic function [ $\text{logistic}(x) = f(x) = 1 / (1 + e^{-x})$ ] that is an S-shaped (sigmoid) curve, with output in the range (0, 1). And finally, for training, many different learning algorithms like BP, KM, KN, CG, PI, ... were examined automatically.

➤ **Sensitivity Analysis:** “A technique called sensitivity analysis can be used to get an idea of how opaque models work” (Berry and Linoff, 2004). Sensitivity signifies the relative significance of the inputs (past records of group members) to the outcome of the network (Berry and Linoff, 2004). In other words, it is a kind of feature selection. Sensitivity analysis employs the test set to find out how sensitive the output of the network is to every input (Berry and Linoff, 2004). In many problem domains, a range of input variables are available that can be used to train a neural network, but it is not clear which of them are most useful, or indeed are needed at all. The problem is further complicated when there are interdependencies or correlations between some of the input variables. Therefore, by means of feature selection approach with sensitivity based techniques STATISTICA software recognized and selected input variables that contribute considerably to the performance of networks, and removed others. The fundamental phases are as below (Berry and Linoff, 2004):

1. “Find the average value for each input. We can think of this average value as the center of the test set” (Berry and Linoff, 2004).
2. “Measure the output of the network when all inputs are at their average value” (Berry and Linoff, 2004).
3. “Measure the output of the network when each input is modified, one at a time, to be at its minimum and maximum values” (Berry and Linoff, 2004).

We bring a type of sensitive analysis with STATISTICA for drug 71 in Table 12. For all drugs, the same analysis was done in STATISTICA.

**Table 13: Sensitivity Analysis of Drug 71 for the best 5 models of 20 tried models**

Sensitivity Analysis												
	Drug40	Drug61	Drug71	Drug95	Drug96	Drug103	Drug114	Drug121	Drug122	Drug130	Drug174	Drug216
<b>Ratio. 1</b>	1.079342	1.012156	5344.8	0.98894	1.00161	1.06680	1.007003	1.694962	1.015267	1.00123	0.99346	1.079536
<b>Rank. 1</b>	4.000000	7.000000	1.0	12.000000	9.00000	5.00000	8.000000	2.000000	6.000000	10.000000	11.000000	3.000000
<b>Ratio. 2</b>			6229.3			1.15946		2.172442				
<b>Rank. 2</b>			1.0			3.00000		2.000000				
<b>Ratio. 3</b>					0.99501			1.224900	1.094701			0.929170
<b>Rank. 3</b>					3.00000			1.000000	2.000000			4.000000
<b>Ratio. 4</b>	0.947780	0.984084	622417.0	0.75529	0.89520	0.88606	0.936070	1.017130	0.971316		0.95644	0.971989
<b>Rank. 4</b>	7.000000	3.000000	1.0	11.000000	9.00000	10.000000	8.000000	2.000000	5.000000		6.000000	4.000000
<b>Ratio. 5</b>	0.968394	1.000040	571684.8	0.79229	0.90084	0.93463	0.964804	1.059068	0.991069		0.98005	1.000744
<b>Rank. 5</b>	7.000000	4.000000	1.0	11.000000	10.000000	9.000000	8.000000	2.000000	5.000000		6.000000	3.000000

#### Step 4: Combining both Models

Finally, we combined these two parts together to make a forecast. Actually, we tried above steps with 25 different neural network models with different architectures and learning algorithms by use of STATISTICA software. Actually, we wanted software to show us best 5 or 7 models for each drug and the best one of these models is extracted and presented here. In fact, we made prediction on our test data for four months After finding the best model, prediction results of drugs 24 and 71 are extracted and presented in Tables 13,14,15, and 16 below. The prediction results of all other 19 drugs are presented in Appendix 1.

#### 4.4.3. Model Evaluation

After that prediction models have been created and tested, their performances were analyzed. In this study, two forecast error measures, Mean Squared Error (MSE), and Mean Absolute Error (MAE) were employed for model evaluation and model comparison. In fact, the

performance of a model do not evaluated on the training set, because it is used in generating the model, so will exaggerate the model's exactness (Berry and Linoff, 2004). The model's precision is always calculated on a test set that is extracted from the identical population as the training, but has not been applied by any means to generate the model (Berry and Linoff, 2004). Thus, in this study, as explained above, the dataset were divided in to training and test set (4 months for test, 4 months for validation and the rest for training). Train set was used for model building and test set for measure the model's accuracy. Results were returned when the model run on the test set and were shown in a summary table for prediction results and a statistics table for accuracy measures. Table 13, 14, 15, and Table 16 show us the best fitted model, training algorithms that were used, predicted values in comparison with observed values, residuals which are difference between real and predicted values,  $R^2$ ,  $|R|$ ,  $(MSE = \sum_i (d_i - y_i)^2 / n)$ , and finally

$(MAE = \sum_i |d_i - y_i| / n)$  for drug 24 and drug 71. As it is obvious from these tables, the

prediction performance was improved greatly when we used partners' past records as well. For instance, if we observe the prediction result of drug 24 and drug 71 below, we see that both MAE and MSE are lower for prediction with partners' records than prediction with their own records. Table 13, which is related to sales prediction of drug 24 with its own records, shows that **MSE= 4094.8**, and **MAE= 51.5**. However, Table 14 which is related to sales prediction of drug 24 with its partners' records, shows that **MSE= 1973.5**, and **MAE= 33** that indicates a great improvement. By comparing these result with results of ARIMA model in previous section (**MSE= 140793.3**, and **MAE= 369.8417**) we can see the difference between our new methodology and ARIMA. Accordingly, we proved that our methodology outperformed both ARIMA methodology and ANN sales forecasting without use of partner's records considerably. The prediction results of all other drugs (19 drugs) are presented in Appendix 1. If we compare accuracy measures (MAE and MSE) of both ANN approaches for all those 19 drugs as well (except for drug 31 and 86 that do not have any group member), we see that in 16 cases of them, the prediction performance was improved greatly when we used partners' past records.

**Table 14: Forecasting Results and Accuracy Measures of Test Data for Drug 24 (with Its Own Past Records)**

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
MLP s6-1:6:0 <sup>1</sup>	BP100,CG47b	2531.000	2267.458	-263.542	69454	263.542	18695	90.568
		2286.000	2318.301	32.301	1043	32.301		
		2384.000	2318.552	-65.448	4283	65.448		
		2328.000	2328.98	0.980	1	0.980		

**Table 15: Forecasting Results and Accuracy Measures of Test Data for Drug 24 (with Its Partner's Past Records)**

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
MLP s6-16:3:0	BP100,CG20,CG5b	2531.000	2531.000	0.0000	0.0	0.0000	442.2	10.5143
		2286.000	2243.943	-42.0573	1768.8	42.0573		
		2384.000	2384.000	0.0000	0.0	0.0000		
		2328.000	2328.000	0.0000	0.0	0.0000		

**Table 16: Forecasting Results and Accuracy Measures of Test Data for Drug 71 (with Its Own Past Records)**

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
RBF s6-1:7:0	BP100,CG47b	7.000	9.138	2.138	4.6	2.1	4094.8	51.5
		-44.000	-78.462	-34.462	1187.6	34.5		
		22.000	-42.093	-64.093	4107.9	64.1		
		8.000	-97.257	-105.257	11079.1	105.3		

**Table 17: Forecasting Results and Accuracy Measures of Test Data for Drug 71 (with Its Partner's Past Records)**

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
RBF s6-11:2:0	BP100,CG20,CG5b	7.000	-64.036	-71.036	5046.1	71.0	1973.5	33.0
		-44.000	-44.741	-0.741	0.5	0.7		
		22.000	-30.868	-52.868	2795.1	52.9		
		8.000	15.238	7.238	52.4	7.2		

<sup>1</sup> See Appendix 2 for the definition of these codes.

To conclude, this section demonstrated that by grouping drugs, using records of group members as input variables, and combining both linear and nonlinear models, forecasting performance can be significantly improved for two reasons: 1) Grouping our data, and using group members' past sales records solved the problem of not having enough past sales records for each drug, 2) The hybrid approach examined in this research conquer the restriction of a pure linear or nonlinear modeling methods whereas simultaneously took advantage of their single modeling ability to reach various patterns in the data (Zhang, 2004). More detailed conclusions about the obtained results are presented in the following chapter.

## **4.5. Summary**

To summarize, in this chapter first we prepared our raw data to make them useful for our objective. Next, we did an exploratory analysis in order to better understand the nature of our data. Then, we did a comprehensive network based analysis in order to find clique-sets, group members and visualize the network of drugs. Afterwards, we built sales forecasting models with three different approaches: 1) ARIMA methodology for time series forecasting, 2) hybrid neural network approach for time Series forecasting by means of each drug's past records, and 3) hybrid neural network approach for time Series forecasting by means of each drug's past records and its group members' past records. Then, we presented model evaluation and results. Due to our outcomes, we noticed that our new methodology (number 3 above) is the best approach, and the weakest one was ARIMA methodology.

# **CHAPTER 5**

## **CONCLUSION**

### ***5. Conclusion***

*In this chapter, a brief review of the entire research, also conclusions are mentioned. Subsequently, contributions and managerial implications and limitations of this research are presented. At the end, suggestions for further research related to this study are offered.*

#### **5.1. Conclusion**

As a matter of fact, most pharmaceutical distribution companies in Iran using simple traditional (statistical) techniques or their managers' intuition and experience to predict future

sales and determine inventory levels for different drugs. Accordingly, they suffer from lack of precise sales forecasting methods. That is why the objective of this thesis has been defined as accurate sales forecasting for a pharmaceutical distribution company (Pakhsh Hejrat Co.) using past sales data assembled by the company in order to minimize costs of excessive inventory also prevents losing potential customers because of drug shortage.

After defining the problem and our goals, to reach our objectives, different studies related to time series forecasting were reviewed in chapter 2. Based on literature, we realized that there are different methods for sales forecasting. Among many existing methods, two different models that have been frequently used for time series prediction are: 1) Linear models: ARIMA modeling, and 2) Nonlinear models: neural networks. By reviewing related studies, we found out most researchers concluded that neural networks outperform linear models in that neural networks can model any type of nonlinear relationships, and they are non-parametric in nature. However, we decided to examine both methods to become assured that applying neural networks is the best approach in this research. After reviewing various techniques and choosing appropriate ones, we went through analysis and modeling phase.

The first step of modeling was data collection. Actually, we gathered sales data from Pakhsh Hejrat Co. that has got near 1200 kinds of drugs which are bought from different manufacturers and sold to 18 provinces of our country. In fact, Pakhsh Hejrat Co. provided us with monthly sales information from December 2004 to December 2007, so the total number of records was 36 months for each pharmaceutical product. Gathered data were in the two text files and three excel files. For data analysis and modeling purpose all these five files were combined and integrated into one Excel file. As a form of preprocessing, data that were only an overhead and were not helpful to us in any way, also drugs that had missing values were deleted from table. Thus, we extracted 217 kinds of drugs that were sold in all 36 months and we changed the drugs' code to 1-217 for more simplicity. In this way, after data cleaning, there is a great data reduction in dataset from nearly 14381 rows to nearly 220 rows. In addition, kinds of drug reduced from more than 1200 to 217. Since for time series prediction we should have at least 50 past recodes, these 36 months of past sales records were not enough. Accordingly, in next step, we came to the new idea of finding groups of drugs that have similar sales behavior or have high cross-correlation to make prediction by means of past records of group members. On this

account, we firstly got cross-correlation matrix from sums of sales in all months for all drugs. Then, we built different incident matrixes for various values of correlation threshold ( $\theta = 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85$ , and  $0.9$ ) separately. We selected  $\theta = .65$  owing to the fact that it gave us appropriate number of cliques with the best combination.

Next, we built incident matrixes which contains of 0, 1 values. We gave value 0 if the correlation between two drugs was less than selected  $\theta = 0.65$  and gave 1 if the cross-correlation was greater than  $\theta = 0.65$ . Actually, this matrix was the basis of the grouping phase. After data preprocessing, in order to better understanding the nature of our data, we did a thorough exploratory analysis. By using STATISTICA software, we drew the sales plots and surface plots of our time series data for different drugs. We also computed some descriptive measures in this section. From our exploratory analysis, we concluded that most drugs have various natures and characteristics; also, relationships among drugs are both linear and nonlinear, but mostly nonlinear. Accordingly, we found that it would better to put linear models aside and make use of nonlinear or hybrid models for our sales forecasting. In next phase (network based analysis), we found clique sets, degree of each drug, group members, and we visualized the networks of drugs. In this section, we carried out all our analysis by means of UCINET and NetDraw softwares that are useful for network analysis. The output of this section helped us greatly in model building phase as we found the drugs, which could help each other in sales forecasting.

In prediction phase, three different methods such as ARIMA modeling and two approaches of neural networks (building hybrid ANNs by using each drug's own past records, building hybrid ANNs by using each drug's own past records and its group members' past records as well) were performed and compared. In all three methods, train set was applied for model building and test set for evaluate the model's accuracy. Thus, 32 months were used as training set in ARIMA methodology and as training and validation set for ANNs.

In ARIMA modeling, four stages of transforming, identification, estimation and diagnosis were performed for drug 24 by STATISTICA software. In this part, we fitted the best model and made prediction for 4 months and found the values of residuals (errors), MSE, and MAE according to our test data. Then, in our neural network approach, we used STATISTICA software and we applied hybrid neural networks in order that we modeled linear components

with linear ANNs and nonlinear components with nonlinear ANNs. Our hybrid modeling contained three steps (Zhang, 2004): (1) fitting a linear neural network model to the time series under examination, (2) generating a nonlinear neural network model according to the residuals from the linear neural network model, and (3) combining the linear neural network outcome and the nonlinear neural network outcome to form the ultimate forecast (Zhang, 2004). We have examined the above steps with two different approaches. In our first approach, we used each drug's own past records as the input. In our second approach, we inputted each drug's own past records and its group members' past records as well. In both approaches, we considered last four months as test data, and the rest (32 months) as validation and training sets. Subsequently, we made sales prediction for four months of 21 various drugs (with different degrees) and we found the values of R (errors), MSE, and MAE according to our test data. The results of proposed three methods for drug 24 (as an example) are as follows:

- 1) ARIMA modeling for drug 24: MSE= 140793.3, and MAE= 369.8417,
- 2) Building hybrid neural networks by inputting past records of drug 24: MSE= 4094.8, and MAE= 51.5,
- 3) Building hybrid neural networks by inputting drug's 24 past records and its group members' past records as well: MSE= 1973.5, and MAE= 33.

By comparing the prediction result of three methods for drug 24, we observed that ARIMA methodology had the poorest result. In addition, the best results were related to the third approach (building hybrid neural networks by using partners' past records). We also compared accuracy measures (MAE and MSE) of both neural network approaches for 19 kinds of drugs; we see that in 16 cases the prediction performance was improved greatly when we used partners' past records as well. Our final results showed that our proposed methodology (building hybrid ANNs by using each drug's own past records and its group members' past records as well) outperformed the other ones.

## **5.2. Contributions**

If we classify different contributions of researches into three categories of theoretical contribution, methodological contribution and empirical contribution, we can say that our research has methodological and empirical contributions. These two contributions are described below:

### **5.2.1. Methodological Contribution**

To make sales prediction for a specific item, all previous researches that were related to time series sales prediction, used only past sale data of that item. However, since we suffered from lack of data (we had only sales records of 36 months) in this research, we introduced a new method of grouping drugs in order to make use of group members' past sales data for each other in our sales prediction (model building) and increase the accuracy of the sales prediction. In this way, we found that which drugs could help each other in sales forecasting; thus, we significantly increased our past sales data for each drug. Accordingly, we built ANNs by using each drug's own past records and its group members' past records as well. Our introduced scheme outperformed two previously known methods of 1) ARIMA modeling, and 2) building ANNs by just using each drug's own past records.

### **5.2.2. Empirical Contribution**

We can consider the empirical contribution of our research from two different points of view:

Firstly, we have chosen a real case (problem) and a real company (Pakhsh Hejrat Company) with its real sales data in this research, we have compared our sales prediction with unseen real data of this company, and we have got acceptable results. It is the first time in pharmaceutical industry in Iran, that this kind of non-heuristic sales prediction method is proposed and conducted.

Secondly, although we had access to sales records of only Pakhsh Hejrat Co., we have interviewed and discussed with eight managers and experts of four main and famous pharmaceutical distribution companies that stand for more than 70% of total sales of drugs to drug stores and hospitals in Iran and. All of them were attracted by both our results and our new

method of sales prediction, and they wanted us to offer our method for them after finishing this research. Accordingly, we achieved empirical supports for our proposed methodology.

### **5.3. Managerial Implications**

The findings of this research have important implications for a large number of companies, specifically distribution companies and more specifically pharmaceutical distribution companies. Actually, every company who needs to establish its policy upon future sales may use these findings.

Actually, great deals of past sales information are available to many companies and organizations. This data can be a rich source of knowledge, if only properly used. This can be very beneficial for the companies using data mining to extract knowledge and useful information from this available source of data. Distributors, who deal with customer demand, by mining these data can optimized both their buying and selling policies according to their clients' needs and reduce the inventory costs. Thus, one of the managerial implications of this research is to inform managers about the advantages and importance of data mining in their strategic planning. As stated before, in this research we aimed to mine past sales data in order to make prediction for the future sales. Accordingly, in this research we applied one of the data mining techniques (ANNs) that is used for prediction to estimate the future sales of drugs using past sales data to advance sales and inventory management also decision making at pharmaceutical distribution companies. In this way, on every customer demand surely there will be enough, not so many not so few, products to satisfy the customer and meanwhile not to cause extra inventory costs.

In order to increase each drug's sales records and improve the accuracy of our sale prediction, we carried out a comprehensive analysis for network of drugs and we summarized and visualized the sales of drugs. Although our purpose in drugs' network identification phase was to obtain useful information about the behavior of the drugs' sales and find drugs whose sales have high cross-correlation with each other, this phase (specially visualizing the networks of drugs and clique sets) would contribute a great deal to distribution companies. For instance, according to the groups or clusters of drugs, they may allocate a specific visitor and sales or marketing planning to each group without consideration of drugs' characteristics. They may also

do a basket analysis in order to make packages of drugs that have similar sales behavior. This packaging would help them both in warehousing and in transportation. It also would decrease the cycle of receiving to delivering the orders. In addition, some drugs that act critical and pivotal roles like drug 24 would be highlighted by this method.

Actually, we talked with experts and sales managers of some leading pharmaceutical distribution companies and they were unanimous that this network-based analysis would help them in their planning and decision-makings. They were eager to investigate that why these drugs have similar sales behavior. They indicated that there might be logic behind the groups, which we have found, and most probably, they are not located at the same groups arbitrary.

After discussing with managers and experts, we were also assured that all of them suffered from lack of precise and valid sales prediction method. As a matter of fact, all distribution companies wish to keep their inventory level as low as possible (not too much and too few) in order to decrease the related costs, also not to face shortage of drugs and lose their commitments accordingly. Actually, pharmaceutical distributors buy products from manufacturers and often pay them immediately, but they sell products and receive the related money gradually. Thus, this gap causes them undesired expenses. Monthly and precise sale prediction would shorten or even eliminate this gap. In addition, by an accurate sale prediction, these companies can increase their return on investment or the profitability of their investment in order that they would not keep excessive products, which most probably would not be sold, and they would be expired. Obviously, the overall ROI for a company is occasionally applied as a method to score how well a corporation is directed. Thus, it is a competitive advantage to be able to increase it by means of accurate sales prediction methods. Furthermore, distributors need to have clear relationships with their suppliers. In Iran, distributors have annual meeting with their suppliers in order to obtain a mutual agreement on amount of products that should be produced by manufacturers and sold by distributors. They also revise their decisions every three months. However, their planning is purely based on traditional methods and experience or intuition of the managers. Our new sales prediction method would help them to better decide for both production and selling of drugs. In other words, our proposed sales forecasting method would help 1) manufacturers to plan for the material procurement and production, and 2) distributors to place reasonable amount of order for each drug; thus, have a better budget management or control. As a consequence of above

mentioned matters, we can conclude that managers who can predict the future sales based on what has occurred in the past can clearly have tremendous advantages over someone who cannot.

Additionally, many researches highlighted the managerial effects of accurate sales prediction for companies. We are bringing some of them below:

- As Yip et al. (1997) mentioned, “sales forecasting is a valuable management tool for determining the future magnitudes, timing, and possible effects of uncontrollable influences over a company's future success. It is an estimate of the expected demand for a company's products, and it is an indispensable element of management planning for many major company activities, e.g. marketing, production, finance, research and development, personnel, capital investment determination, purchasing, inventory and warehousing” (Yip et al., 1997).
- Owing to the tough competition that exists nowadays, most companies are in a permanent effort for raising their profits and decreasing their costs. Precise sales forecasting is definitely an economical way to meet the abovementioned objectives, as this causes superior customer service, decreased lost sales and product returns (Doganis et al., 2006). Especially for the pharmaceutical industry, successful sales forecasting systems can be very beneficial, due to the short shelf-life of many pharmaceutical products and the importance of the product quality which is closely related to the human health (Doganis et al., 2006).
- An effective forecasting system can decrease inventories, boost profits and attain greater adaptability to alterations. Specifically, sales forecasting is very significant, as its result is employed by numerous functions in a distribution company (Mentzer and Bienstock, 1998; cited by Doganis et al., 2006): “The sales department requires a good knowledge of the sales volume of each product, as it is charged with the job of organizing the sales force” (Mentzer and Bienstock, 1998; cited by Doganis et al., 2006). Purchasing department needs an accurate forecast of future sales in order to plan for buying sufficient (not much more and not far less than it is needed) products from manufacturers. Marketing requires a vision of the future market to plan its actions and evaluate the effect of alters in the marketing strategy on sales amount. Lastly, logistics also requires precise sales forecasts of various horizon lengths: a long-term prediction to set up and expand logistics infrastructure and a short-term prediction to describe precise logistics requires (Mentzer and Bienstock, 1998; cited by

Doganis et al., 2006). Pharmaceutical distribution companies like food companies are more related with sales prediction according to their particular characteristics, such as the short shelf life of their products, the requirement to keep high product quality and the ambiguity and variations in customer demands. Since products can just be sold for a short period, both lack and surplus of goods can cause loss of profits for the organization.

In summary, managers should recognize the advantages and significance of sales prediction by means of data mining techniques in their strategic planning since winning application of data mining methods could be a considerable payoff for the companies (Lee and Siau, 2001). They should use past data, which is stored in databases, to extract knowledge and useful information about the future. In other words, they ought to map the past to forecast the future.

## **5.4. Research Limitations**

This study, like all other researches, has some limitations as follows:

- Actually, one of the important limitations of this research was data collection. As we know for data mining purpose secondary data is used; thus, we had to convince managers to provide us with the desired data. Since sales data are confidential information of each company, they are usually very unwilling to offer these kinds of data. Thus, it was really challenging to gather the required data.
- Our target company offered us 36 months of their sales records. However, for time series prediction we needed to have at least 50 records. Therefore, we had to add a new phase (network based analysis and grouping) to our research in order to make use of group members' data for each other and increase the sales information of each drug accordingly.
- We had not access to some critical data, which would be beneficial for our prediction, of the mentioned company. For instance, they did not give us information about probable offered promotions and discounts<sup>2</sup>, or even the number of their sales representatives in particular periods. In fact, this industry is really special and unusual industry as most policies and prices are set by the government in this industry. Hence, manufacturers and distributors

---

<sup>2</sup>. Usually discounts are not monetary.

cannot act upon freely. As an illustration, in this industry, promotion and giving discount are prohibited by the Ministry of health. Thus, the related information was confidential for these companies and they were unwilling to give us these data. Hence, these sudden and unknown variables may decrease the accuracy of our prediction.

- There are some specific drugs (that are mostly imported from foreign countries) that the government gives subsidies on them. Since these drugs are bought and sold in special conditions and not routinely, we cannot consider their sales records as time series. Thus, Pakhsh Hejrat Co. did not offer them to us and we did not consider them in our data set.

## 5.5. Further Research

- In this research, we summed up the sales of all provinces for each drug and considered a central depot, so we did not consider the regions. As a further research, it is proposed to cluster regions in order to see which regions or provinces have similar sales behavior. It means that we can find cross-correlation among different regions or provinces on order to cluster them. It is desirable for Pakhsh Hejrat Company like all other distributors to know where they should locate their depots or to know the optimized place and even the required space for each depot. For instance, this company realized that their depot in Ghazvin is not required any more, but they still do not know that from which depot they should give service to this city. By accurate clustering of the regions according to their sales behavior, we can understand that which depot should give service to which cities.
- In this research, we had monthly sales records of only three years. We recommend researches to collect sales records of at least five years or even to collect daily sales records in that if we have more records as our inputs, we will obtain more accurate results. Thus, there would be no need to group the products in order to make use of group member's records.
- In this research, we had access only to sales records of a pharmaceutical distribution company. However, it is suggested for further research to examine the proposed methodology in other industries such as food industry and make sales prediction for food distribution companies.
- As mentioned before, we did not have access to the information of external or environmental variables such as promotions, discounts, competitors' past sales records and so on.

Accordingly, we did not consider these variables in our modeling and sales prediction. For a further research, it is recommended to persuade distribution companies to offer the related data in order to obtain more accurate results.

- As a further study, researchers can investigate the groups or cliques of drugs, which were presented in chapter 4, in order to find that why specific drugs are located in the same groups according to their sales cross-correlations. It means that they can discover some features of these drugs (other than their healing characteristics) that cause them to be located in special groups or cliques.
- Future research may answer this question: Is the hybrid model used in this study still effective if the linear and nonlinear components are not separable?

# REFERENCES

- Aburto, L., and Weber, R., (2007), “Improved supply chain management based on hybrid demand forecasts”, *Applied Soft Computing*, 7:1, 136–144.
- Adya, M. and Collopy, F., (1998), “How effective are neural networks at forecasting and prediction? A review and evaluation”. *Journal of Forecasting*, 17, PP. 481–495.
- Agrawal, D., Schorling, C., (1996), “Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model”, *Journal of Retailing*, 72:4, 383-408.
- Agrawal, R. and Srikant, R. (1995) “Mining sequential patterns”, *In Proceeding of the 11th Int. Conf. on Data Engineering*, March 1995, Taipei, Taiwan, Published by IEEE Computer Society Press.
- Aiken, M. and Bsat, M., (1999), “Forecasting market trends with neural networks”, *Information Systems Management*, 16:4, 42–48.
- Ainscough, T. L., and Aronson, J. E., (1999), “An empirical investigation and comparison of neural networks and regression for scanner data analysis”, *Journal of Retailing and Consumer Services*, 6:4, 205–217.
- Albadvi, A., (2007), “Research Method in Marketing and E-commerce”, *A Master Course in TMU University*, Iran.
- Alon, I.,Qi, M., and Sadowski, R. J., (2001), “Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional methods”, *Journal of Retailing and Consumer Services*, 8:3, 147–156.
- Ansuj, A. P.,Camargo, M. E.,Ragharaman, R., and Petry, D. G., (1996), “Sales forecasting using time series and neural networks”, *Computers and Industry Engineering*, 31: (1/2), 421–424.
- Aslanargun, A., Mammadov, M., Yazici, B., Yolacan, S., (2007), “Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting”, *Journal of Statistical Computation and Simulation*, 77:1, 29–53.
- Atiya, A. F., (2001), Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on Neural Networks*, 12:4, 929–935.
- Appelqvist, P., Lehtonen, J-M., and Kokkonen, J., (2004), “Modeling in product and supply chain design: literature survey and case study”, *Journal of Manufacturing Technology Management*, 15:7, 675–686.
- Ayad, A.M. (2000) “*A New Algorithm for Incremental Mining of Constrained Association Rules*”, A thesis submitted to the Department of Computer Science and Automatic

Control in partial fulfillment of the requirements for the degree of Master of Science, Alexandria University - Faculty of Engineering.

Balaguer, E., Palomares, A., Soria, E. and Martin-Guerrero, J. D., (2008), "Predicting service request in support centers based on nonlinear dynamics, ARMA modeling and neural networks", *Expert Systems with Applications*, 34: 1, 665–672.

Bansal, K., Gupta, A., Vadhavkar, S., (1998), "Neural Networks Based Forecasting Techniques for Inventory Control Applications", *International Journal of Agile Manufacturing*, 2: 1. (Also Presented in: (Bansal, K., Vadhavkar, S., Gupta, A (1998). Neural networks based forecasting techniques for inventory control applications. *Data Mining and Knowledge Discovery*, 2, 97–102.)

Bansal, K. Vadhavkar, S. and Gupta, A., (1998), "Neural Networks Based Data Mining Applications For Medical Inventory Problems", *Data Mining and Knowledge Discovery*, 2:1, 97-102.

Berry, M. J. A. and Linoff, G. S., (2004), *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management*, (2nd edn), Indiana, Wiley Publishing, Inc.

Berthouex, P. M., and Brown, L. C., (2002), *Statistics for environmental engineers*, New York, Lewis Publishers.

Bishop, M., (1995), *Neural Networks for Pattern Recognition*, Oxford, UK: Oxford University Press.

Bigus, J. P., (1996), *Data Mining with Neural Networks: solving business problems from Application development to decision support*, Printed in the United States of America, The McGraw-Hill Companies, Inc.

Boginski, V., Sergiy Butenko, S., and Pardalos, P.M., (2005), "Statistical analysis of financial networks", *Computational Statistics and Data Analysis*, 48, 431 – 443.

Boginski, V., Sergiy Butenko, S., and Pardalos, P.M., (2006), "Mining market data: A network approach", *Computers and Operations Research*, 3, 3171–3184.

Bowerman, B. L., and O'Connell, R. T., (1986), *Time Series Forecasting: Unified Concepts and Computer Implementation*, Boston, MA, USA, PWS Publishing Co., Translated by Reza Shiva, (1th edn) in 1996, Iran.

Box, G.E., Jenkins, G.M., (1976), *Time Series Analysis*, San Francisco, CA. Holden-Day.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time series analysis: Forecasting and control* (3rd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Brockwell, P. J. and Davis, R. A., (2002), *Introduction to time series and forecasting*, New York, Springer- Verlag,

Burgess, K., Singh, P.J. and Koroglu, R., (2006), "Supply chain management: a structured literature review and implications for future research", *International Journal of Operations and Production Management*, 26: 7,703-729.

Buxey, G., (2006), "Reconstructing inventory management theory", *International Journal of Operations and Production Management*, 26: 9, 996-1012.

Cabena, P. P. Hadjinian, R. Stadler, J.Verhees, and A. Zansi, (1997), *Discovering Data Mining: From Concept to Implementation*, Englewood Cliffs, NJ: Prentice-Hall,

Callen, J. L.,Kwan, C. Y.,Yip, C. Y., and Yuan, Y., (1996), "Neural network forecasting of quarterly accounting earnings", *International Journal of Forecasting*, 12, 475–482.

Carboneau, R., Lafraimboise, K., and Vahidov, R., (2008), "Application of machine learning techniques for supply chain demand forecasting", *European Journal of Operational Research*, 184: 3, 1140–1154.

Chakraborty, K., Mehrotra K., Mohan C.K., (1992), "Forecasting the behavior of multivariate time series using neural networks", *Neural Networks*, 5, 961–970.

Cooper and Schindler, (2003), *Business Research Methods* (8<sup>th</sup> edn), USA, McGraw-Hill, Inc.

Chang, P.C. and Wang, Y.W., (2006), Fuzzy Delphi and back-propagation model for sales forecasting in PCB industry, *Expert Systems with Applications*, 30: 4, 715–726.

Charytoniuk, W.,Box, E. D.,Lee, W.-J.,Chen, M.-S.,Kotas, P., and Olinda, P. V., (2000), Neural-network-based demand forecasting in a deregulated environment. *IEEE Transaction on Industry Applications*, 36:3, 893–898.

Chen, X.,Racine, J., and Swanson, N. R., (2001), "Semi-parametric ARX neural-network models with an application to forecasting inflation:", *IEEE Transactions on Neural Networks*, 12: 4, 674–683.

Chen, F.L., and Ou, T.Y., (2008), Gray relation analysis and multilayer functional link network sales forecasting model for perishable food in convenience store, *Expert Systems with Applications*, 36: 3, 7054-7063.

Cheng, Y.H. Liao, H.W., and Chen Y.S., (2006), "Implementation of a Back-Propagation Neural Network for Demand Forecasting in a Supply Chain- A Practical Case Study, presented in International Conference on Service Operation and Logistics, and Informatics, SQLI 06.IEEE, 1036-1041.

Christopher, M., (1992), *Logistics: The Strategic Issues*, London, Chapman and Hall.

Church, K.B., and Curram, S.P., (1996), "Forecasting consumers' expenditure: A comparison between econometric and neural network models", *International Journal of Forecasting*, 12, 255–267.

Cooper, M.C., Lambert, D.M. and Pagh, J.D., (1997), "Supply chain management: more than a new name for logistics", *The International Journal of Logistics Management*, 8:1, 1-13.

Cooper, D. R. and Schindler, P.S., (2003), *Business Research Methods*, (8<sup>th</sup> edn), USA, McGraw-Hill, Inc.

Council of Logistics Management (CLM), (1998), *What It's All About*, 4: 6, Oak Brook, IL.

Cox, J.F. III, Blackstone, J.H. Jr, and Spencer, M.S., (1998), *APICS Dictionary* (9th edn), the APICS Educational and Research Foundation, Falls Church, VA.

Crone, S.F., (2003), "Prediction of white noise time series using artificial neural networks and asymmetric cost functions", *Proceedings of the International Joint Conference on Neural Networks*, 20-24 July 2003, IEEE, 4: 20-24, 2460 – 2465.

Darbeyay, G. A. and Slama, M., (2000), "Forecasting the Short-term Demand for Electricity: Do neural networks stand a better chance?", *International Journal of Forecasting*, 16, 71 – 83.

Dasgupta, C.G., Dispensa, G.S. and Ghose, S., (1994), "Comparing the predictive performance of a neural network model with some traditional market response models. International Journal of Forecasting", 10:2, 235-244.

De Carvalho, M. C. M., Dougherty, M. S., Fowkes, A. S., and Wardman, M. R., (1998), "Forecasting travel demand: A comparison of logic and artificial neural network methods" *Journal of Operational Research Society*, 49, 717–722.

De Gooijer, J.G., and Hyndman, R.J., (2006), "25 years of time series forecasting", *International Journal of Forecasting*, 22:3, 443–473.

Dhar, V. and Chou, D., (2001), "A comparison of nonlinear methods for predicting earnings surprises and returns", *IEEE, Transactions on Neural Networks*, 12: 4, 907–921.

Dhond, A., Gupta, A., and Vadhavkar, S., (2000), "Data Mining Techniques for Optimizing Inventories for Electronic Commerce", *In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, Pages: 480 – 486, Boston, Massachusetts, United States.

Di'az-Robles, L.A., Ortega, J.C., Fu, J.S., Reed, G.D., Chow, J.C., Watson, J.G., and Moncada-Herrera, J.A., (2008), "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile", *Atmospheric Environment*, 42:35, 8331–8340.

Diez, R.P. and Fernández J.P., (2005), *Forecasting the Price of Energy in Spain's Electricity Production Market*, IX Congreso de Ingeniería de Organización Gijón, 8 y 9 de septiembre de 2005.

Doganis, P., Alexandridis, A., Patrinos, P., and Sarimveis, H., (2006), "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing", *Journal of Food Engineering*, 75:2, 196–204.

Dong, X. and Wen, G., (2006), An Improved Neural Networks Prediction Model and Its Application in Supply Chain, *Nature and Science*, 4: 3, 23-27.

Elkateb, M. M., K. Solaiman, and Al-Turki, Y., (1998). "A comparative study of medium-weather-dependent load forecasting using enhanced artificial/fuzzy neural network and statistical techniques", *Neurocomputing*, 23:1-3, 3–13.

Ellram, L. and Cooper, M., (1993), "Characteristics of supply chain management and the implications for purchasing and logistics strategy", *International Journal of Logistics Management*, 4: 2, 1-10.

Fausett, L., (1994), *Fundamentals of neural networks*, Englewood, NJ: Prentice Hall.

Fayyad, U. and Uthurusamy, R., (1996), "Data mining and knowledge discovery in databases", *Communications of the ACM*, New York, NY, USA, 39:11, 24-26.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., (1996a), "From data mining to knowledge discovery: an overview", In Fayyad, U., Shapiro, P. G., Smyth, P. and Uthurusamy, R. (eds.), *Advances in Knowledge Discovery and Data*, Palo Alto, Cambridge, Massachusetts: CA: AAAI/MIT Press, 1996, PP 1-34, also in *AI Magazine*, 17:3, 186-193.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., (1996b), From Data Mining to Knowledge Discovery in Databases, *American Association for Artificial Intelligence, AI MAGAZINE*, 17, (3). 37-54.

Foster, B., Collopy, F., Ungar, L., (1992), "Neural network forecasting of short, noisy time series", *Computers and Chemical Engineering*, 16: 12, 293-297.

Fu, Y., (1997), "Data Mining: Tasks, Techniques and Applications", *Potentials, IEEE*, 16: 4, 18–20.

Ganesan, R., Boone, T., and Stenger, A.J., (2001), "The impact of inventory and flow planning parameters on supply chain performance: An exploratory study", *International Journal of Production Economics*, 71: 1, 111-118.

Ganti, V., Gehrke, J. and Ramakrishnan, R., (1999), "Mining Very Large Databases", *IEEE Computer*, 32: 8, 68-75.

Gruca, T.S., Bruce, R.K., and Furr Petersen, E.A., (1999), "Mining Sales Data using a Neural Network Model of Market Response", *SIGKDD Explorations*, 1: 1, 39-43.

Gupta, A., Maranas, C.D., and McDonald, C.M., (2000), "Mid-term supply chain planning under demand uncertainty: customer demand satisfaction and inventory management", *Computers and Chemical Engineering*, 24: 12, 2613–2621.

Hanna, M., (2004), "Data mining in the E-learning Domain", *Journal of Campus-Wide Information Systems*, 21: 1, 29 – 34.

Han, J. and Kamber, M., (2006), *Data Mining: Concepts and Techniques*, San Francisco, U.S.A, Morgan Kaufman Publishers.

Hand, D. J., (1999), "Statistics and Data Mining: Intersecting Disciplines", *ACM SIGKDD Explorations Newsletter*, 1: 1, 16 – 19.

Haykin, S., (1999), *Neural network: a comprehensive foundation*. (2nd edn), Upper Saddle River, NJ: Prentice-Hall.

He, Y., Li,F., Song, Z., and Zhang, G.,( 2002), "Neural Networks Technology for Inventory Management", *Computer Engineering and Application*, 15, 182-184.

Heravi, S., Osborn, D. R., and Birchenhall, C. R., (2004), "Linear versusneural network forecasting for European industrial production series", *International Journal of Forecasting*, 20, 435 –446.

Herrera, J. L., (1999), *Time Series Prediction Using Inductive Reasoning Techniques*, PHD Thesis, Instituto de Organizaci.on y Control de Sistemas Industriales.

Hill, T., Marquez, L., O'Connor, M., Remus, W., (1994), "Artificial neural network models for forecasting and decision making" *International Journal of Forecasting*, 10, 1–15.

Hill, T., Marquez, L., O'Connor, M., Remus, W., (1996), "Neural networks for time series forecasts", *Manage. Sci.*, 42: 7, 1082–1092.

Hippert, H. S., Pedreira, C. E., and Souza, R. C., (2001), "Neural networks for short-term load forecasting: A review and evaluation", *IEEE Transactions on Power Systems*, 16:1, 44–55.

Holme, I.M. and Solvand, B.K., (1991), *Forskningsmetodik. Om Kvalitativa och Kvantitativa Metoder*, Lund, Sverige: Studentlitterature.

Houlihan, J.B. (1987), "International supply chain management", *International Journal of Physical Distribution and Materials Management*, 17: 2, 51-66.

Ho, T.B., (nd), "Knowldge Discovery and Data Mining Techniques and Practice", Unesco Course (online), (cited October 2006).

Ho, T.B., (2002), "Introduction to Knowledge Discovery and Data Mining", Available at: [http://www.netnam.vn/unescocourse/knowlegde/know\\_frm.htm](http://www.netnam.vn/unescocourse/knowlegde/know_frm.htm).

Hossein Javaheri, S., (2007), *Response Modeling in Direct Marketing – a Data Mining Based Approach for Target Selection*, Master Thesis for a Joint Program in Marketing and E-commerce of Tarbiat Modares University and Lulea University of Technology.

Hugos, M., (2006), *Essentials of Supply Chain Management*, (2nd edn), Hoboken, New Jersey, Published by John Wiley and Sons, Inc.

Jain, A. K., Mao, J., and Mohiuddin, K., (1996), “Artificial Neural Networks: A Tutorial”, *IEEE Computer*, 29:3, 31-44.

Jain, A., and Kumar, A., (2007), “Hybrid neural network models for hydrologic time series forecasting”, *Applied Soft Computing*, 7:2, 585–592.

Johnson, J.C. and Wood, D.F., (1996), *Contemporary Logistics*, (6th edn), Upper Saddle River, NJ: Prentice-Hall.

Kantardzic, M., (2003), “*Data Mining: Concepts, Models, Methods, and Algorithms*” John Wiley and Sons, Inc., edited by ff.

Kathawala, Y.K. and Abdou, K., (2003), “Supply chain evaluation in the service industry: a framework development compared to manufacturing”, *Managerial Auditing Journal*, Vol. 18 No. 2, pp. 140-9.

Kennedy, R. L., Lee, Y. Roy, B. V. Reed, C. D. and Lippman, R. P., (1997) ‘Solving Data Mining Problems Through Pattern Recognition’, Prentice Hall Professional Technical Reference, New Jersey.

Khashei, M. and Bijari, M., (2010). “An artificial neural network (p,d,q) model for timeseries forecasting”, *Expert Systems with Applications*, 37:1, 479-489.

Kong, J.H.L. and Martin, G.P.M.D., (1995), “A back-propagation neural network for sales forecasting”, *IEEE International Conference on Neural Networks*, 2, 1007-1011.

Kumar, A., Rao, V. R., and Soni, H., (1995), “An empirical comparison of neural network and logistic regression models”, *Marketing Letters*, 6:4, 251–263.

Kuo, R. J., and Cohen, P. H. (1998), “Intelligent tool wear estimation system through artificial neural networks and fuzzy modeling”, *AI in Engineering*, 12, 229–242.

Kuo, R.J. and Xue, K.C., (1998), “An intelligent sales forecasting system through integration of artificial neural network and fuzzy neural network”, *Computers in Industry*, 37: 1, 1–15.

Kuo, R. J., (2001), “A sales forecasting system based on fuzzy neural network with initial weights generated by genetic algorithm”, *European Journal of operation research*, 129, 496–571.

Kuo, R.J., Wu , P. and Wang, C.P., (2002), “An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination”, *Neural Networks* , 15: 7, 909–925.

Kuo, R.J., Hu, T.L., and Chen, Z.Y., (2009), “Application of Radial Basis Function Neural Network for Sales Forecasting” *International Asia Conference on Informatics in Control, Automation and Robotics, IEEE Computer Society. Washington, DC, USA*: 325–328.

Krycha, K.A., Wagner, U., (1999), “Applications of artificial neural networks in management science: a survey”, *Journal of Retailing and Consumer Services* 6, 185-203.

Langley, P., Simon, H., (1995), “Applications of Machine Learning and Rule Induction”, *Communications of the ACM*”, 38: 11, 54 – 64.

Larose, D. T., (2005), *Discovering Knowledge in Data: An Introduction to Data Mining*, Hoboken, New Jersey, Published by John Wiley and Sons, Inc.

Law, R., (2000), “Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting”, *Tourism Management*, 21, 331–340.

Lee, S. J. and Siau, K., (2001), “A review of data mining techniques”, *Industrial Management and Data System*, 101:1, 41-46.

Lummus, R.R. and Alber, K.L., (1997), “Supply Chain Management: Balancing the Supply Chain with Customer Demand”, Falls Church, VA., The Educational and Resource Foundation of APICS.

Lummus, R.R. and Vokurka, R.J., (1999), “Defining supply chain management: a historical perspective and practical guidelines”, *Industrial Management and Data systems*, 99:1, 11-17.

Lummus, R.R., Krumwiede, D. W., Robert J. Vokurka, R.J., ( 2001), “The relationship of logistics to supply chain management: developing a common industry definition”, *Industrial Management & Data Systems*, 101:8, 426-431.

Luxhoj, J. T., Riis, J. O., and Stensballe, B., (1996), “A hybrid econometric-neural network modeling approach for sales forecasting”, *International Journal of Production Economics*, 43, 175–192.

Makridakis, S., and Hibon, M., (1997), “ARMA Models and the Box-Jenkins Methodology”, *Journal of Forecasting*, 16, 147-163.

Makridakis, S. G. Wheelwright S. C. and Hyndman R. J., (1998), *Forecasting: methods and applications*, New York: Wiley.

Malhotra, N., K., (2006), *Marketing Research, An applied Orientation*, (5<sup>th</sup> edn), New Dehli, Printice Hall of India.

Malhotra, N. and Birks, D., (2003), *Marketing Research, An Applied Approach*, Harlow, Prentice Hall.

Malhotra, N., K and Peterson, M., (2006), *Business Research Methods, A decision-Making Approach*, New Jersey, Pearson Prentice Hall.

Manchanda, P., Wittink, D.R., Ching, A., Cleanthous, P, Ding, M., Dong, X.J., Leeflang, P.S.H., Misra, S., Mizik, N., Narayanan, S., Steenburgh, T., Wieringa, J.E., Wosinska, M., and Xie, Y., (2004), “Understanding Firm, Physician and Consumer Choice Behavior in the Pharmaceutical Industry”, Part of the Sixth Invitational Choice Symposium during June 4–8, hosted by the University of Colorado and held at Estes Park, Published by Puneet Manchanda , Also in: *Marketing Letters*, 16:3/4, 293–308, 2005.

Masters, T., (1994), *Signal and image processing with neural network*, John Wiley and Sons Inc.

McKee, T. E. and Greenstein, M., (2000), Predicting Bankruptcy Using Recursive Partitioning and Realistically Proportioned Data Set. *Journal of Forecasting*, 19, 219–230.

McKenzie, E. (1984), “General exponential smoothing and the equivalent ARMA process” *Journal of Forecasting*, 3, 333-344.

Melard, M., Roy, R., and Saidi, A., (2006), “Exact maximum likelihood estimation of structured or unit root multivariate time series models”, *Computational Statistics and Data Analysis*, Vol. 50, No. 11, PP. 2958-2986.

Mentzer, J.T., DeWitt, W., Keebler, J.S., Min, S., Nix, N.W., Smith, C.D. and Zacharia, Z.G., (2001), “Defining supply chain management”, *Journal of Business Logistics*, 22: 2, 1-25.

Mitra, S., and Acharya, T., (2003), *Data Mining Multimedia, Soft Computing, and Bioinformatics*, New Jersey, Hoboken, Published by John Wiley and Sons, Inc.

Moshiri, S., and Cameron, N., (2000), “Neural Network Versus Econometric Models in Forecasting Inflation”, *Journal of Forecasting*, 19, 201-217.

Nag, A. K. and Mitra, A., (2002), “Forecasting Daily Foreign Exchange Rates Using Genetically Optimized Neural Networks”, *Journal of Forecasting*, 21: 7, 501–512.

Nam, K. and Schaefer, T., (1995), “Forecasting international airline passenger traffic using neural networks”, *Logistics and Transportation Review*, 31: 3, 239–251.

Nelson, M., Hill, T., Remus, B., O'Connor, M., (1994), “Can neural networks be applied to time series forecasting learn seasonal patterns: an empirical investigation”, *Proceedings of the 27 Annual Hawaii International Conference on System Sciences*, 649-655.

Ntungo, C., and Boyd, M., (1998), “Commodity futures trading performance using neural network models versus ARIMA models”, *The Journal of Futures Markets*, 18:8, 965–983.

Outsourcing Data Entry Services, (2007), “Importance of Data Mining in Today’s Business”, Available at: <http://outsourcingdataentry.blogspot.com/2007/07/importance-of-data-mining-in-todays.html>

Ohsuga S., (2005), “Difference between Data Mining and Knowledge Discovery- A View to Discovery from Knowledge-Processing”, *International Conference on Granular Computing*, Beijing, China, 25-27 July, *IEEE*, 1, 7- 12.

Partovi, F.Y., and Anandarajan, M., (2002), “Classifying inventory using an artificial neural network approach”, *Computers and Industrial Engineering*, 41, 389-404.

Pazzani M. J., (2000), “Knowledge discovery from data”, *IEEE INTELLIGENT SYSTEMS*, 15: 2, 10-13.

Frawley, W. J., Piatetsky- Shapiro, G. and Matheus, C. J., (1992), *Knowledge Discovery in Databases*, AI Magazine, 13:3, 57-70.

Piatetsky-Shapiro, G., (2000), “Knowledge Discovery in Databases: 10 years after”, *SIGKDD Explorations*, 1: 2, 59-61, Available from: <http://doi.acm.org/10.1145/846183.846197>.

Prybutok, V.R., Yi, J., Mitchell, D., (2000), “Comparison of neural network models with ARIMA and regression models for prediction of Houston’s daily maximum ozone concentrations”, *European Journal of Operational Research*, 122:1, 31–40.

Pyle, D., (1999), *Data Preparation for Data Mining*, San Francisco, CA, Morgan Kaufmann Publishers.

Qi, M., (1996), “Financial applications of artificial neural networks” In: Maddala, G.S., Rao, C.R. (Eds.), *Handbook of Statistics, Statistical Methods in Finance*, 4, Amsterdam, North-Holland, Elsevier Science Publishers, pp. 529-552.

Qi, M., and Maddala, G.S., (1999), “Economic factors and the stock market: a new perspective”, *Journal of Forecasting*, 18: 3, 151-166.

Qi, M., (2001), “Predicting US recessions with leading indicators via neural network models” *International Journal of Forecasting*, 17: 3, 383–401.

Raghunathan, S., (1999), “Inter-organizational collaborative forecasting and replenishment systems and supply chain implications”, *Decision Sciences*, 30: 4, 1053-71.

Read, B.J., (1999), “Data Mining and Science? Knowledge discovery in science as opposed to business”, *12th ERCIM Workshop on Database Research*.

Routroy, S., and Kodali, R., (2005), “Differential evolution algorithm for supply chain inventory planning”, *Journal of Manufacturing Technology Management*, 16: 1, 7-17.

Rosenblatt, F., (1958), “The perceptron: A probabilistic model for information storage and organization in the brain”, *Psychological Review*, 65: 6, 386-408.

Rosenblatt, F., (1962), *Principles of Neurodynamics, Perceptrons and the Theory of Brain Mechanisms*, Washington, D.C.: Spartan Books.

Saunders, M., Lewis, P. and Thornhill, A., (2000), *Research Method for Business Students*, England, Pearson Education Limited.

Shanmugasundaram, J., Nagendra Prasad, M. V., Vadhavkar, S., Gupta, A., (2002) “Use of Recurrent Neural Networks for Strategic Data Mining of Sales Information”, MIT, Sloan School of Management, Working Paper, No. 4347-02.

Sivanandam, S. N., and Deepa, S. N., (2006) - *Introduction To Neural Net With Matlab 6.0*, New Delhi, Tata McGraw-Hill Publishing Company Limited.

Snyder, R., (2002), “Forecasting sales of slow and fast moving inventories”, *European Journal of Operational Research*, 140: 3, 684–699.

SPSS White Paper, (1999), “*Data mining: an introduction Clementine – Working with health care*”.

Stanfield, J., (2002), “Agents in Supply Chain Management”, *Agent Link News*, 9, 11-12.

STATISTICA , (2006), *STATISTICA statistical software, Version 7, Electronic Manual*. StatSoft Inc.

Supply-Chain Council, (1997), November 20, available from: <http://www.supply-chain.com/info/faq.html>

Svensson, G., (2002), “Supply chain management: the re-integration of marketing issues in logistics theory and practice”, *European Business Review*, 14: 6, 426-436.

Symeonidis, A.L., Nikolaidou, V. and Mitkas, P.A., (2006), Exploiting Data Mining techniques for improving the efficiency of a Supply Chain Management agent, *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, (WI-IAT 2006 Workshops).

Tang, Z., Almeida, C., and Fishwick, P. A., (1991), “Times series forecasting using neural networks vs. Box–Jenkins methodology” *Simulations Councils*, 303–310.

Thiesing, F.M.; Middelberg, U.; Vornberger, O., (1995), Short term prediction of sales in supermarkets, *IEEE International Conference on Neural Networks*, 2, 1028 – 1031.

Thiesing, F.M. and Vornberber, O., (1997), Sales forecasting using neural networks, *in: Proceedings International Conference in Neural Networks*, Houston, 4, 2125–2128. Available from: <http://citeseer.ist.psu.edu/43652.html>

Tkacz, G. (2001), ” Neural network forecasting of Canadian GDP growth”, *International Journal of Forecasting*, 17, 57–69.

Thomassey, S., Happiette, M., and Castelain, J.M., (2005), “A global forecasting support system adapted to textile distribution”, *Int. J. Production Economics*, 96, 81–95.

Tseng, F.-M., Yu, H.-C., and Tzeng, G.-H., (2001), “Applied hybrid grey model to forecast seasonal time series”, *Technological Forecasting and Social Change*, 67, 291–302.

Ture, M. and Kurt, I., (2006), “Comparison of four different time series methods to forecast hepatitis A virus infection”, *Expert Systems with Applications*, 31: 1, 41–46.

Tsay, R.S., (2005), *Analysis of Financial Time series*, (2<sup>nd</sup> edn), Newyork, Wiley.

Two Crows (1999), Introduction to Data Mining and Knowledge Discovery, (3<sup>rd</sup> edn), U.S.A., by Two Crows Corporation.

Vellido, A., Lisboa, P. J. G., and Vaughan, J., (1999), “Neural networks in business: A survey of applications (1992–1998)”, *Expert Systems with Applications*, 17, 51–70.

Vincenti, G., Hammell II, R.J., and Trajkovski, G., (2005), Scouting for Imprecise Temporal Associations to Support Effectiveness of Drugs during Clinical Trials, *NAFIPS – 2005 Annual Meeting of the North American Fuzzy Information Processing Society*.

Voges, K.E., and Pope, N.K.L1., (2000), An Overview of Data Mining Techniques from an Adaptive Systems perspective, *ANZMAC 2000 Visionary Marketing for the 21st Century: Facing the Challenge*.

Wagner, B.A. and Alderdice, A.D.G., (2006), “Managing the distribution channel: the case of Scot Trout and Salmon”, *Supply Chain Management: An International Journal*, 11: 2, 104–107.

Wang, J.-H. and Leu, J.-Y., (1996), Stock market trend prediction using ARIMA-based neural networks, *IEEE International Conference on Neural Networks*, 2160–2165.

Weigend, A. S. and Gershenfeld, N. A., (1993), *Time Series Prediction: Forecasting the Future and Understanding the Past*, USA, Addison-Wesley.

West, P.M., Brockett, P.L., Golden, L., (1997), “A comparative analysis of neural networks and statistical methods for predicting consumer choice”, *Marketing Science*, 16: 4, 370-391.

Widrow, B., Rumelhart, D., and Lehr, M. A., (1994), “Neural networks: Applications in industry, business and science”, *Communications of the ACM*, 37: 3, 93–105.

Witten, I. H., and Frank, E, (2005), *Data Mining: Practical Machine Learning Tools and Techniques*, (2nd edn), San Francisco, CA, The Morgan Kaufmann Series in Data Management Systems.

Yang, Z. R., (1999), “Probabilistic neural networks in bankruptcy prediction, *Journal of Business Research*, 44, 67–74.

Yip, D.H.F., Hines, E.L. and Yu, W.W.H., (1997), Application of Artificial Neural Networks in Sales Forecasting, *IEEE International Conference on Neural Networks*, Houston, TX, USA., 4, 2121-2124,

Yule, G., (1927), “On a method of investigating periodicity in disturbed series with special reference to Wolfer’s sunspot numbers”, *Philosophical Transactions of the Royal Society of London*, 226, 267–298.

Zhang, G. P., Patuwo, E. P., and Hu, M. Y., (1998), “Forecasting with artificial neural networks: The state of the art”, *International Journal of Forecasting*, 14, 35–62.

Zhang, G. P. (2001), “An investigation of neural networks for linear time-series forecasting”, *Computers and Operations Research*, 28, 1183–1202.

Zhang, G. P. and Qi, M., (2002), “Predicting consumer retail sales using neural networks. In Smith and Gupta (Eds.)”, *Neural Networks in Business: Techniques and Applications* (pp. 26–40). Hershey, PA: Idea Group Publishing.

Zhang, G. P., (2003), “Time series forecasting using a hybrid ARIMA and neural network model”, *Neurocomputing*, 50, 159–175.

Zhang, G. P., (2004), *Neural Networks in Business Forecasting*, USA, Idea Group Publishing (IGP).

Zikmund, W., (2003), *Essential of Marketing Research*, USA, Thompson South-Western.

# APPENDIX 1

## Tables of Prediction Results with Neural Networks

### Forecasting Results and Accuracy Measures of Test Data for Drug 84 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>MLP</b> s6- 1:7:0	BP100, CG20	8154.000	6252.959	1901.04	3613956.6	1901.0	968377.6	618.0
		6035.000	6542.361	507.36	257415.5	507.4		
		6462.000	6485.883	23.88	570.4	23.9		
		6401.000	6440.598	39.60	1568.0	39.6		

### Forecasting Results and Accuracy Measures of Test Data for Drug 84 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>RBF</b> s3- 3:3:0	KM,KN,PI	8154.000	8288.959	134.96	18213.9	135.0	686472.1	540.1
		6035.000	7643.888	1608.89	2588519.1	1608.9		
		6462.000	6508.458	46.46	2158.3	46.5		
		6401.000	6771.131	370.13	136996.9	370.1		

### Forecasting Results and Accuracy Measures of Test Data for Drug 12 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>GRNN</b> s3- 1:16:2	SS	15199.00	15290.46	91.48	8369.1	91.5	1624244.3	1032.1
		13955.00	15349.30	1394.30	1944076.9	1394.3		
		17354.00	15307.20	-2046.80	4189377.9	2046.8		
		14888.00	15483.95	595.95	355153.2	595.9		

### Forecasting Results and Accuracy Measures of Test Data for Drug 12 (with its partner's Past Records)

<b>Best Fitted Model</b>	<b>Training Algorithm</b>	<b>Observed Test Data</b>	<b>Predicted Test Data</b>	<b>Residual (R)</b>	<b>RSQ (R^2)</b>	<b>ABSE (  R  )</b>	<b>MSE</b>	<b>MAE</b>
<b>GRNN s4- 9:16:2</b>	SS	15199.00	15061.64	-137.36	18867.6	137.4	18390.4	125.0
		13955.00	13918.70	-36.30	1317.8	36.3		
		17354.00	17200.52	-153.48	23555.0	153.5		
		14888.00	15060.69	172.69	29821.3	172.7		

### Forecasting Results and Accuracy Measures of Test Data for Drug 22 (with its Own Past Records)

<b>Best Fitted Model</b>	<b>Training Algorithm</b>	<b>Observed Test Data</b>	<b>Predicted Test Data</b>	<b>Residual (R)</b>	<b>RSQ (R^2)</b>	<b>ABSE (  R  )</b>	<b>MSE</b>	<b>MAE</b>
<b>RBF s7- 1:3:0</b>	KM,KN,PI	3711.000	3768.814	57.81	3342.5	57.8	291633.2	353.6
		3295.000	3213.187	-81.81	6693.3	81.8		
		3634.000	3856.539	222.54	49523.5	222.5		
		4663.000	3610.872	-1052.13	1106973.7	1052.1		

### Forecasting Results and Accuracy Measures of Test Data for Drug 22 (with its partner's Past Records)

<b>Best Fitted Model</b>	<b>Training Algorithm</b>	<b>Observed Test Data</b>	<b>Predicted Test Data</b>	<b>Residual (R)</b>	<b>RSQ (R^2)</b>	<b>ABSE (  R  )</b>	<b>MSE</b>	<b>MAE</b>
<b>MLP S6- 7:7:0</b>	BP100, CG20	3711.000	3284.735	-426.26	181701.8	426.3	51155.5	171.5
		3295.000	3385.750	90.75	8235.5	90.7		
		3634.000	3732.279	98.28	9658.8	98.3		
		4663.000	4733.894	70.89	5026.0	70.9		

### Forecasting Results and Accuracy Measures of Test Data for Drug 30 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
RBF s6 1:1:0	KM,KN,PI	14834.00	14641.69	-192.3	36982.2	192.3	247727 6.6	1186.2
		11746.00	14656.01	2910.0	8468133.3	2910.0		
		15453.00	14797.71	-655.3	429401.0	655.3		
		14097.00	15084.21	987.2	974590.0	987.2		

### Forecasting Results and Accuracy Measures of Test Data for Drug 30 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
Linear s12 14:0:0	PI	14834.00	15210.00	376.00	141376.0	376.0	2032221.6	1197.7
		11746.00	14213.84	2467.84	6090245.2	2467.8		
		15453.00	16396.00	943.00	889249.0	943.0		
		14097.00	15101.00	1004.00	1008016.0	1004.0		

### Forecasting Results and Accuracy Measures of Test Data for Drug 31 (Drug 31 Has Not Got any Group Member)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
GRNN s12 1:16:2	SS	6774.000	6481.676	-292.32	85453.6	292.3	343539.5	426.6
		7012.000	5896.736	-1115.26	1243813.4	1115.3		
		6322.000	6161.395	-160.60	25793.8	160.6		
		6113.000	6251.192	138.19	19097.1	138.2		

### Forecasting Results and Accuracy Measures of Test Data for Drug 37 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
RBF s6 1:2:0	KM,KN, PI	6774.000	7838.45	11084.5	122865138.2	11084.5	404633 39.4	5444.4
		7012.000	5645.09	3926.1	15414206.0	3926.1		
		6322.000	6600.38	3965.4	15724276.3	3965.4		
		6113.000	5997.74	2801.7	7849737.0	2801.7		

### Forecasting Results and Accuracy Measures of Test Data for Drug 37 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
GRN N s7 7:16:2	SS	-3246.00	-3241.05	4.95	24.5	4.95	87854.6	150.0
		1719.00	2311.78	592.8	351388.3	592.8		
		2635.00	2632.67	-2.3	5.4	2.3		
		3196.00	3195.95	-0.1	0.0	0.1		

### Forecasting Results and Accuracy Measures of Test Data for Drug 51 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
RBF S6- 1:2:0	KM,KN, PI	5754.00	5872.64	118.6	14074.7	118.6	679480.7	601.6
		7376.00	5863.71	-1512.3	2287032.6	1512.3		
		5325.00	5953.75	628.7	395324.8	628.7		
		6285.00	6138.40	-146.6	7849737.0	146.6		

### Forecasting Results and Accuracy Measures of Test Data for Drug 51 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>RBF s6-2:4:0</b>	KM,KN,PI	5754.00	5403.75	-350.3	122678.5	350.3	547755.7	633.9
		7376.00	6123.49	-1252.5	1568791.4	1252.5		
		5325.00	5970.95	646.0	417256.9	646.0		
		6285.00	5998.13	-286.9	82296.0	286.9		

### Forecasting Results and Accuracy Measures of Test Data for Drug 61 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>RBF s6 1:2:0</b>	KM,K N,PI	4547.00	4712.91	165.9	27525.7	165.9	4351902.6	1517.1
		467.00	3243.40	2776.4	7708420.7	2776.4		
		402.00	3511.89	3109.9	9671401.2	3109.9		
		277.00	260.79	-16.2	262.9	16.2		

### Forecasting Results and Accuracy Measures of Test Data for Drug 61 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>RBF s6 5:7:0</b>	KM,KN, PI	4547.00	5332.03	785.0	616267.9	785.0	2272363.4	866.9
		467.00	-2021.33	-2488.3	6191810.3	2488.3		
		402.00	307.07	-94.9	9012.1	94.9		
		277.00	376.38	99.4	9877.0	99.4		

### Forecasting Results and Accuracy Measures of Test Data for Drug 66 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
GRNN s7- 1:16:2	SS	-114.00	-114.000	0.0	0.0	0.0	519776.2	508.0
		10.00	1110.734	1100.7	1211614.5	1100.7		
		4.00	935.392	931.4	867490.1	931.4		
		3.00	3.000	0.0	0.0	0.0		

### Forecasting Results and Accuracy Measures of Test Data for Drug 66 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
GRNN s4- 4:16:2	SS	4547.00	38.101	152.1	23134.7	152.1	13299.3	103.2
		467.00	80.277	70.3	4938.8	70.3		
		402.00	158.367	154.4	23829.2	154.4		
		277.00	38.981	36.0	1294.7	36.0		

### Forecasting Results and Accuracy Measures of Test Data for Drug 78 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
MLP s6 1:8:0	BP50b	15425.00	15481.63	56.6	3206.7	56.6	5421172.9	1638.2
		13615.00	15869.07	2254.1	5080827.2	2254.1		
		16545.00	16373.62	-171.4	29369.5	171.4		
		17753.00	13682.22	-4070.8	16571288.4	4070.8		

### Forecasting Results and Accuracy Measures of Test Data for Drug 78 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
GRNN s2- 6:16:2	SS	15425.00	16261.88	836.9	700376.4	836.9	2339520.9	1169.0
		13615.00	16403.58	2788.6	7776178.0	2788.6		
		16545.00	16664.41	119.4	14257.8	119.4		
		17753.00	16821.73	-931.3	867271.2	931.3		

### Forecasting Results and Accuracy Measures of Test Data for Drug 86 (Drug 86 Has Not Got any Group Member)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
GRNN s3- 1:16:2	SS	15425.00	42778.69	14443.7	208620108	14443.7	52155027.1	3610.9
		13615.00	35156.00	0.0	.6	0.0		
		16545.00	29544.00	0.0	0.0	0.0		
		17753.00	41364.00	0.0	0.0	0.0		

### Forecasting Results and Accuracy Measures of Test Data for Drug 95 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
RBF s6 1:4:0	KM,KN, PI	5529.00	4315.42	-1213.58	1472777.0	1213.6	1564831.5	955.3
		1061.00	3221.27	2160.27	4666757.7	2160.3		
		1536.00	1859.03	323.03	104345.6	323.0		
		1442.00	1566.28	124.28	15445.7	124.3		

### Forecasting Results and Accuracy Measures of Test Data for Drug 95 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>RBF s6-13:4:0</b>	KM,KN, PI	5529.00	5839.00	310.00	96100.0	310.0	143032.1	362.5
		1061.00	1365.00	305.00	93025.0	305.0		
		1536.00	1786.00	286.00	81796.0	286.0		
		1442.00	893.18	-548.82	301207.2	548.8		

### Forecasting Results and Accuracy Measures of Test Data for Drug 103 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>RBF s6 1:5:0</b>	KM,KN, PI	5.000	475.556	470.56	221422.8	470.56	109700.6	316.7
		23.000	327.586	304.59	92772.9	304.59		
		5.000	294.440	289.44	83775.8	289.44		
		39.000	241.067	202.07	40831.0	202.07		

### Forecasting Results and Accuracy Measures of Test Data for Drug 103 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>RBF s6 10:2:0</b>	KM,KN, PI	5.000	281.795	276.80	76615.6	276.8	56819.2	237.2
		23.000	255.736	232.74	54166.0	232.7		
		5.000	223.229	218.23	47624.0	218.2		
		39.000	-182.069	-221.07	48871.3	221.1		

### Forecasting Results and Accuracy Measures of Test Data for Drug 124 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
GRNN 1:16:2	SS	622.0000	454.7388	-167.261	27976.3	167.3	7468.1	59.4
		529.0000	503.1756	-25.824	666.9	25.8		
		527.0000	493.8358	-33.164	1099.9	33.2		
		504.0000	492.6265	-11.374	129.4	11.4		

### Forecasting Results and Accuracy Measures of Test Data for Drug 124 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
RBF S6- 2:2:0	KM,KN, PI	622.0000	564.9799	-57.020	3251.3	57.0	861.1	20.1
		529.0000	521.6795	-7.321	53.6	7.321		
		527.0000	521.4329	-5.567	31.0	5.6		
		504.0000	493.5914	-10.409	108.3	10.4		

### Forecasting Results and Accuracy Measures of Test Data for Drug 127 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
MLP s6- 1:8:0	BP50b, CG 4b	389.000	552.134	163.13	26612.7	163.1	34114.3	142.7
		554.000	481.701	-72.30	5227.1	72.3		
		297.000	309.093	12.093	146.2406	12.093		
		669.000	345.780	-323.22	104471.3	323.2		

### Forecasting Results and Accuracy Measures of Test Data for Drug 127 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE ( R )	MSE	MAE
MLP s6- 11:7:0	BP100,C G20,CG5 b	389.000	735.107	346.11	119789.9	346.1	34091.5	136.0
		554.000	458.395	-95.60	9140.2	95.6		
		297.000	314.973	17.97	323.0	18.0		
		669.000	584.661	-84.34	7113.0	84.3		

### Forecasting Results and Accuracy Measures of Test Data for Drug 132 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE ( R )	MSE	MAE
GRN N s6- 1:16:2	SS	7884.00	6307.08	-1576.92	2486682.3 0.2	1576.92	621768.9	399.4
		5179.00	5178.58	-0.42	0.1	0.42		
		6726.00	6725.68	-0.32	393.1	0.32		
		6380.00	6399.83	19.83		19.83		

### Forecasting Results and Accuracy Measures of Test Data for Drug 132 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE ( R )	MSE	MAE
GRN N s2- 10:16: 2	SS	7884.00	7928.29	44.29	1961.7	44.3	512.8	14.5
		5179.00	5186.08	7.08	50.1	7.1		
		6726.00	6725.71	-0.29	0.1	0.3		
		6380.00	6386.27	6.27	39.3	6.3		

### Forecasting Results and Accuracy Measures of Test Data for Drug 167 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>RBF s6-1:9:0</b>	KM,KN, PI	13069.00	18521.23	5452.2	29726804.2	5452.2	20348171.4	4013.2
		16830.00	16362.65	-467.4	218420.3	467.4		
		29686.00	24387.04	-5299.0	28078984.2	5299.0		
		41069.00	36234.90	-4834.1	23368477.1	4834.1		

### Forecasting Results and Accuracy Measures of Test Data for Drug 167 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>Linear s6-12:0:0</b>	PI	13069.00	19173.74	-6104.7	37267792.2	6104.7	10173585.6	2223.2
		16830.00	17211.00	-381.0	145161.0	381.0		
		29686.00	30451.00	-765.0	585225.0	765.0		
		41069.00	42711.00	-1642.0	2696164.0	1642.0		

### Forecasting Results and Accuracy Measures of Test Data for Drug 168 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
<b>GRN N s12-1:16:2</b>	SS	13686.00	19255.38	5569.4	31017984.2	5569.4	71612145.9	7270.3
		20363.00	23780.05	3417.0	11676220.1	3417.0		
		38176.00	32703.39	-5472.6	29949441.3	5472.6		
		61742.00	47119.93	-14622.1	213804937.8	14622.1		

### Forecasting Results and Accuracy Measures of Test Data for Drug 168 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
GRNN s6- 8:16:2	SS	13686.00	19311.08	5625.1	31641518.8	5625.1	13221096.4	3267.2
		20363.00	24054.90	3691.9	13630147.6	3691.9		
		38176.00	35763.92	-2412.1	5818123.0	2412.1		
		61742.00	60402.37	-1339.6	1794595.9	1339.6		

### Forecasting Results and Accuracy Measures of Test Data for Drug 174 (with its Own Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
MLP s6 1:8:0	BP100,C G4b	89929.0	89991.8	62.8	3943.7	62.8	62575380.0	5803.3
		69885.0	83007.5	13122.5	172199281.4	13122.5		
		97403.0	88659.5	-8743.5	76448120.8	8743.5		
		98593.0	99877.6	1284.6	1650174.1	1284.6		

### Forecasting Results and Accuracy Measures of Test Data for Drug 174 (with its partner's Past Records)

Best Fitted Model	Training Algorithm	Observed Test Data	Predicted Test Data	Residual (R)	RSQ (R^2)	ABSE (  R  )	MSE	MAE
Linea rs3 17:1:0	PI	89929.0	90922.9	993.9	987926.9	993.9	274513013.3	12195.9
		69885.0	100698.0	30813.0	949441066.1	30813.0		
		97403.0	87588.0	-9815.0	96333475.8	9815.0		
		98593.0	105754.7	7161.7	51289584.5	7161.7		

## APPENDIX 2

### Definitions of Codes in Result Tables of Forecasting

Our result tables contain a concise description of the type of networks and training algorithms used to train and optimize the network. For example, 1:6:0 means that the network has 1 input, 6 nodes in first hidden layer, and zero node in second hidden layer. The code CG47b indicates that the Conjugate Gradient Descent algorithm was used, the best network discovered and selected while running, and this network was found on the 47th epoch.

The codes are:

<b>BP</b>	Back Propagation
<b>CG</b>	Conjugate Gradient Descent
<b>QN</b>	Quasi-Newton
<b>LM</b>	Levenberg-Marquardt
<b>QP</b>	Quick Propagation
<b>DD</b>	Delta-Bar-Delta
<b>SS</b>	(sub)Sample
<b>KM</b>	K-Means (Center Assignment)
<b>EX</b>	Explicit (Deviation Assignment)
<b>IS</b>	Isotropic (Deviation Assignment)
<b>KN</b>	K-Nearest Neighbor (Deviation Assignment)
<b>PI</b>	Pseudo-Invert (Linear Least Squares Optimization)
<b>KO</b>	Kohonen (Center Assignment)
<b>PN</b>	Probabilistic Neural Network Training
<b>GR</b>	Generalized Regression Neural Network Training
<b>PC</b>	Principal Components Analysis

The terminal codes are:

<b>b</b>	Best Network (the network with lowest selection error in the run was restored)
<b>s</b>	Stopping Condition (the training run was stopped before the total number of epochs elapsed as a stopping condition was fulfilled)
<b>c</b>	Converged (the algorithm stopped early because it had converged; that is, reached and detected a local or global minimum. Note that only some algorithms can detect stoppage in a local minimum, and that this is an advantage not a disadvantage!)