MARCH 2019 M.Sc in Industrial Engineering SELMA GÜLYEŞİL

UNIVERSITY OF GAZİANTEP

GRADUATE SCHOOL OF

NATURAL & APPLIED SCIENCES

JUTE CONSUMPTION FORECASTING: A CASE STUDY IN A CARPET MANUFACTURING

M.Sc. THESIS

IN

INDUSTRIAL ENGINEERING

SELMA GÜLYEŞİL

MARCH 2019

Jute Consumption Forecasting: A Case Study In A Carpet Manufacturing

M.Sc Thesis

in

Industrial Engineering

University of Gaziantep

Supervisor

Assoc. Prof. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU

by

Selma GÜLYEŞİL

March 2019

REPUBLIC OF TURKEY
UNIVERSITY OF GAZİANTEP
GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES
DEPARTMENT OF INDUSTRIAL ENGINEERING

Name of the thesis: Jute Consumption Forecasting: A Case Study In A Carpet Manufacturing

Name of the student: Selma GÜLYEŞİL

Exam Date: 01.03.2019

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Ahmet Necmeddin YAZICI

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Serap ULUSAM SEÇKİNER

Head of Department

This is to certify that we have read this thesis and that in our consensus opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU

Supervisor

Bu tez tarafımızca okunmuş, kapsam ve niteliği açısından bir Yüksek Lisans/~~Doktora~~ tezi olarak oy birliği/oy çokluğu ile kabul edilmiştir.

Examining Committee Members	Signature
Assoc. Prof. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU	....
Assoc. Prof. Dr. Alptekin DURMUŞOĞLU	....
Asst. Prof. Dr. Yunus EROĞLU	....

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

SELMA GÜLYEŞİL

ABSTRACT

Jute Consumption Forecasting: A Case Study In A Carpet Manufacturing

GÜLYEŞİL, Selma

M. Sc. Thesis, Industrial Engineering

Thesis Supervisor: Assoc. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU

March 2019

Pages 99

In today's increasingly competitive market conditions, it is necessary for the all companies to meet the demand of the customers in a timely and complete manner. For this reason, all resources (raw materials, semi-finished products, all energy sources, etc.) should be prepared and supplied at the right time and in sufficient quantities. This situation forces companies to make forecasts of raw materials to avoid possible lateness.

The most basic forecasting methods are statistical estimation methods. However, these statistical methods may be inadequate in the system with high uncertainties and high dynamism. At this point, the method of artificial neural network (ANN), which have been succesful and widespread in recent years, steps in. The studies show that the results obtained from ANNs may be much more successful than the statistical methods applied in different areas for estimation.

Jute is one of the most basic raw materials used in carpet production and it is important to estimate the jute consumption correctly in order to prevent possible production delays and to ensure customer satisfaction. In this study, jute consumption estimations were made by ANN method using data for the years between 2015-2017 of carpet company. In this thesis, a multiple linear regression (MLR) model was also established to evaluate the performance of ANN method. The results show that ANN is more successful than regression analysis considering the performance indicators' values (values of MSE- mean squared error- R^2 ve adjusted R^2).

Keywords: Artificial Neural Networks, Multiple Linear Regression, Raw Material Consumption, Jute Consumption

ÖZET

Jüt Tüketim Tahmini: Halı İmalatında Bir Vaka Çalışması

GÜLYEŞİL, Selma

Yüksek Lisans Tezi, Endüstri Mühendisliği

Tez Yöneticisi: Doç. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU

Mart 2019

99 Sayfa

Günümüzün artan rekabet koşullarında, tüm firmaların öne çıkabilmeleri için müşterilerin taleplerini tam zamanında ve eksiksiz olarak karşıyalabilmeleri gerekmektedir. Bu nedenle kullanılacak tüm kaynakların (hammadde, yarı ürün, tüm enerji kaynakları vs.) hazırlanması ve tedariği doğru zamanda ve yeterli miktarda yapılmalıdır. Bu durum firmaları olası gecikmeleri önlemek amacıyla tahminlerde bulunmaya zorlamıştır.

En temel tahminleme yöntemleri istatistiksel tahmin yöntemleridir. Ancak bu istatistiksel yöntemler yüksek belirsizlik ve yüksek dinamizme sahip sistemlerde yetersiz kalabilmektedir. Bu noktada son yılların başarılı ve yaygın kullanımına sahip yapay sinir ağları (YSA) yöntemi devreye girmektedir. Yapılan çalışmalar farklı alanlarda uygulanan tahmin çalışmalarında yapay sinir ağlarından elde edilen sonuçların istatistiksel yöntemlere göre çok daha başarılı olabildiğini göstermiştir.

Jüt, halı üretiminde kullanılan en temel hammaddelerden birisidir ve olası üretim gecikmelerini önlemek ve müşteri memnuniyetini sağlamak için jüt tüketimini doğru tahmin etmek önemlidir. Bu çalışmada bir halı firmasının 2015-2017 yılları arasına ait verileri kullanılarak yapay sinir ağları metodu ile jüt hammaddesi tüketimi tahminlemesi yapılmıştır. Bu tez çalışmasında, YSA metodunun performansını değerlendirmek için çoklu doğrusal regresyon modeli kurulmuştur. Elde edilen sonuçlar performans gösterge değerleri (ortalama kare hatası, R^2 ve düzeltilmiş R^2) göz önüne alındığında yapay sinir ağlarının regresyon analizine göre daha başarılı olduğunu göstermiştir.

Anahtar Kelimeler: Yapay Sinir Ağları, Çoklu Doğrusal Regresyon, Hammadde Tüketimi, Jüt Tüketimi

To my precious family…

ACKNOWLEDGEMENT

I would like to thank my thesis supervisor, Assoc. Prof. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU for her encouragement, advice, and guidance during this research. Her valuable support has been of great help in the completion of this work.

I am also grateful to the Department of Industrial Engineering for the knowledge and the experience that I gained during my graduate study.

I would also like to thank my dear parents and family members, my beloved husband Özkan and little son Ömer Asaf for 'giving meaning' to my life.

CONTENTS

ABSTRACT vi
ÖZET vii
ACKNOWLEDGEMENT ix
TABLES xiv
FIGURES xiv
ABBREVIATIONS xvi
CHAPTER 1. INTRODUCTION 1
CHAPTER 2. LITERATURE REVIEW 4
2.1 Information About Forecasting: 4
2.1.1 Types of Forecasting Methods: 4
2.1.1.1 Judgmental: 4
2.1.1.2 Casual Methods (Linear Regression): 4
2.1.1.3 Time-Series Methods: 4
2.2 Brief Explanation of ANN Method: 5
2.3 Studies About ANN in the Literature: 5
2.3.1 ANN Studies in Time-Series: 5
2.3.2 ANN Studies in the Field of Production: 8
2.3.3 ANN Studies in the Field of Electricity: 10
2.3.4 Studies on ANN in the Field of Economics: 12
2.3.5 ANN Studies in Other Fields: 13
CHAPTER 3. ARTIFICIAL INTELLIGENCE (AI) AND ARTIFICIAL NEURAL NETWORK (ANN) 16
3.1 ARTIFICIAL INTELLIGENCE (AI): 16
3.1.1 Historical Development of Artificial Intelligence 18
3.2 ARTIFICIAL NEURAL NETWORKS 20
3.2.1 Definitions of Artificial Neural Networks: 20
3.2.2 Historical Development of ANN: 21
3.2.3 Advantages of Artificial Neural Networks: 22
3.2.4 Structure of Artificial Neural Networks 24
3.2.4.1 The Human Brain (Biological Neural Networks): 24
3.2.4.2 Components of a Neuron: 25
3.2.4.3 The Basic Structure of Artificial Neural Networks: 26
3.2.4.3.1 Inputs: 26

3.2.4.3.2 Weights: 27
3.2.4.3.3 Propagation Function: 27
3.2.4.3.4 Activation Function: 27
3.2.4.3.5 Output: 30
3.2.4.4 The Difference of Neural Networks from Traditional Computing: 30
3.2.4.5 Design of Artificial Neural Networks: 31
3.2.4.6 Normalization: 31
3.2.4.7 Artificial Neural Network Topology Selection: 32
3.2.4.8 Learning Strategies: 34
3.2.4.8.1 Supervised Learning: 35
3.2.4.8.2 Unsupervised Learning: 35
3.2.4.8.3 Reinforcement Learning: 35
3.2.4.9 Learning Laws: 35
3.2.4.9.1 Hebb's Rule: 36
3.2.4.9.2 Hopfield Rule: 36
3.2.4.9.3 Delta Rule: 36
3.2.4.9.4 The Gradient Descent Rule: 36
3.2.4.9.5 Kohonen's Learning Law: 36
3.2.4.10 Number of Hidden Layer and Neuron Number in Each Layer: 37
3.2.4.11 Training and Testing of an ANN: 37
3.2.4.12 Stopping Criteria of an ANN: 38
3.2.4.13 Performance Indicators: 38
3.2.5 The Most Common Used Models in ANN 39
3.2.5.1 Perceptrons: 39
3.2.5.2 Multilayer Perceptrons (MLPs): 39
3.2.5.3 Learning Vector Quantization (LVQ): 43
3.2.5.4 Self-Organizing Map (SOM): 43
3.2.5.5 Adaptive Resonance Theory (ART): 43
3.2.5.6 Hopfield Networks: 43
3.2.5.7 Counterpropogation Networks: 44
3.2.5.8 Neocognitron Networks: 44
3.2.5.9 Boltzmann Machine: 44
3.2.5.10 Probabilistic Neural Network (PNN): 45

3.2.5.11 Elman Networks: 45
3.2.5.12 Radial Based Network (RBN): 45
CHAPTER 4. REGRESSION ANALYSIS 47
4.1 Simple Linear Regression (SLR): 47
4.2 Multiple Linear Regression (MLR): 50
4.3 Nonlinear Regression: 51
CHAPTER 5. TEXTILE INDUSTRY AND CARPET PRODUCTION 53
5.1 Weaving: 53
5.2 Finishing: 54
5.3 Bulked Continuous Flament (BCF): 54
5.3.1 BCF Production Line: 54
5.3.2 Bending Production Line: 54
5.3.3 Heat-Set: 55
5.3.4 Packaging: 55
5.4 Raw Materials: 55
5.4.1 Weft Yarns: 55
5.4.1.1 Jute: 55
5.4.1.2 Cotton Weft: 56
5.4.1.3 Chenille: 56
5.4.2 Warp Yarn: 56
5.4.3 Pile Yarns: 56
5.4.3.1 Acrylic Yarns: 56
5.4.3.2 Polype Yarns: 56
5.4.3.3 Viscose: 57
5.4.3.4 Spun Thread: 57
5.4.3.5 Bamboo: 57
5.5 Planning: 57
CHAPTER 6. COMPUTATIONAL EXPERIMENT 60
6.1 Data Analysis: 60
6.1.1 Data Collection: 60
6.1.1.1 Calculations for ‘Purchasing Price’ Variable: 61
6.1.1.2 Calculations for ‘Average Consumption in 1 m^2’ Variable: 62
6.1.2 Data Normalization: 64

6.2 Determination of Performance Indicators: 65
6.3 Forecasting with Artificial Neural Network Method: 66
6.3.1 Determination of Neuron Number in Hidden Layer: 66
6.3.2 Determination of Activation (Transfer) Function: 68
6.3.3 Determination of Training Function: 69
6.3.4 Determination of the Best Feasible ANN: 70
6.4 Forecasting with Multilinear Regression Method: 73
6.4.1 Checking for Regression Assumptions: 74
6.4.2 Equations and Results of MLR Method: 81
6.5 A Scenario with ANN Model: 84
7 CONCLUSION 87
REFERENCES 89
APPENDIX 94

TABLES

Table 3.1: Examples of propagation functions 28
Table 3.2: Examples of activation functions... 29
Table 3.3: Comparison of computing approaches.. 30
Table 3.4:ANN topologies according to their usage ares and purpose of use 33
Table 3.5: Inputs and output sets for XOR problem 41
Table 6.1: Variables definition table... 61
Table 6.2: Examples of data for calculations of 'purchasing price' variable.............. 62
Table 6.3: Examples of data for calculations of 'average consumption in 1 m2' variable 63
Table 6.4: The sample of the raw data used in application....................................... 64
Table 6.5: Normalized values of the data shown in Table 6.4 65
Table 6.6: Network characteristics of net_1 and net_2... 67
Table 6.7: Comparative MSE values of net_1 and net_2 networks........................... 68
Table 6.8: Network characteristics of net_1 and net_3... 69
Table 6.9: Comparative MSE values of net_1 and net_3... 69
Table 6.10: Network characteristics of net_1 and net_4... 70
Table 6.11: Comparative MSE values of net_1 and net_4... 70
Table 6.12: Values of performance indicators of net_1 72
Table 6.13: Weight values of each connection of net_1 at satisfied MSE value...... 73
Table 6.14: Correlation matrix for variables of jute consumption problem 74
Table 6.15: Calculated t values of Table 6.14... 75
Table 6.16: VIF values of each independent variable.. 76
Table 6.17: Analysis of Variance result of MLR.. 80
Table 6.18: Calculated t values of each regression coefficient................................. 81
Table 6.19: Regression equations of MLR 82
Table 6.20: Values of performance indicators of MLR 83
Table 6.21: Comparative results of ANN and MLR... 83
Table 7.1: Comparative results of ANN and MLR... 87
Table A.1: : Raw form of 246 Data Points Between The Years 2015 and 2017 for Jute Consumption Prediction Problem... 94

FIGURES

Figure 3.1: Block diagram representation of the neural system 24
Figure 3.2: Illustration of the biological neuron 25
Figure 3.3: A nonlinear model of a neuron 26
Figure 3.4: A sigmoid function for varying slope parameter α................................ 29
Figure 3.5: Examples of linearly nonseparable classes... 40
Figure 3.6: Graphical representation of Table 3.5 41
Figure 3.7:Architectural graph of an MLP with two hidden layers 42
Figure 3.8: Architectural graph of Hopfield network consisting of N=4 neurons..... 44
Figure 3.9: Architectural graph of an RBN 46
Figure 4.1: Simple Linear Regression Graph.. 50
Figure 4.2: MLR hyperplane for population.. 51
Figure 4.3: Graph of second-order nonlinear regression model 52
Figure 5.1:Workflow diagram of the company... 59
Figure 6.1: Network structure of net_1 67
Figure 6.2: Network structure of net_2 68
Figure 6.3: MSE graph of net_1.. 71
Figure 6.4: R graphs for each data group of net_1.. 72
Figure 6. 5: Residuals plot against fitted value 78
Figure 6.6: Normal probability plot of residuals.. 79

ABBREVIATIONS

ACO	Ant Colony Optimization
ADALINE	Adaptive Linear Neuron
ADE	Adaptive Differential Evolution
AI	Artificial Intelligence
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
APE	Absolute Percentage Error
APSOA	Adaptive Particle Swarm Optimization Algorithm
AR	Auto-regressive
ARIMA	Autoregressive Integrated Moving Average
ART	Adaptive Resonance Theory
BCF	Bulked Continuous Flament
BP	Back Propagation
BPN	Back Propagation Network
BPNNs	Back Propagation Neural Networks
DAN2	Dynamic Artificial Neural Network
DEA	Data Envelopment Analysis
FNN	Feed Forward Neural Network
GA	Genetic Algorithm
GARCH	Generalized Autoregressive Conditional Heteroscedasticity
GPS	General Problem Solver
GRA	Grey Relational Analysis
IEEE	Institute of Electrical and Electronics Engineers
LM	Levenberg-Marquardt
LMS	Least Mean Square
LVQ	Learning Vector Quantization
MAD	Mean Absolute Deviation
MAE	Mean Absolute Error
MAED	Model for Analysis and Energy Demand
MAPE	Mean Absolute Percentage Error
MDA	Multiple Discriminate Analysis
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression (Multilinear Regression)

MLTs	Manufacturing Lead Times
MRP	Material Requirement Planning
MSE	Mean Squared Error
NAV	Net Asset Value
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
RBN	Radial Based Network
RMS	Root Mean Square
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SA	Simulated Annealing
SANN	Seasonal Artificial Neural Network
SARIMA	Seasonal ARIMA
SKUs	Stock Keeping Units
SLR	Simple Linear Regression
SOM	Self-Organizing Map
SQP	Sequential Quadratic Programming
VIF	Variance Inflation Factor
VLSI	Very-Large-Scale-Integrated
WIP	Work-in-Process

CHAPTER 1. INTRODUCTION

Nowadays, people can shop not only from their own countries but also from other countries through developing technology. This made it necessary for existing firms not only to compete with competitors in their own countries, but also with rival companies in other countries. In the world-wide competitive market conditions, companies need to ensure customer satisfaction in order to survive. This can be achieved by providing timely products or services by accurately estimating the uncertain customer demands. Therefore, timely and complete supply of raw material supply has become important in all sectors. The carpet manufacturing sector is one of the production sectors where these conditions must be met.

The leading companies in the carpet sector are located in Gaziantep. These companies produce carpets for both domestic and abroad customers. There are advantages and disadvantages for companies that compete with each other in close locations. Some of the advantages are the rapid recognition of the developments in the competing company or quick access to qualified employees. However, having more than one manufacturer option for customers accelerates customer loss even in the smallest problem between customer and company about production or sales issues. This is one of the most important disadvantages. In order to increase customer satisfaction and to avoid possible production delays by adhering to the deadlines demanded by the customer, it is much more important that the raw materials to be used in production should be provided in timely and sufficient amount. For these reasons, raw material forecasting has become very important for carpet companies.

This thesis covers the subject of yarn consumption estimation. It is especially important for the procurement departments of the carpet manufacturing companies. As being the basic component of a carpet, carpet manufacturing requires huge amount of jute. In carpet production, the cost of jute is approximately 23 % of the total cost of yarns. In addition, purchasing prices of jute are in dollar terms. In this case, the fluctuating exchange rate affects the cost of jute. There are two factors that any typical carpet manufacturer considers. On one hand, they can keep small size of inventories of jute which may cause stopping of carpet production, on the other hand, excess inventories may result in a huge holding cost for the companies. With respect to inventory cost, they need to allocate capital and physical inventory locations for storing the jute itself. When we consider these limitations, a typical carpet manufacturer needs a good plan of jute procurement. Procurement plan is especially critical under some volatile business environments like hyperinflation or floating rate. Considering both limitations and unexpected business factors is increasing the decisions complexity on such procurement plan. This topic proposes an ANN based estimation method to ease this decision process. The method can be used to estimate jute consumption of carpet producers by taking some manufacturing, economical and logistical factors.

In this study, raw material consumption was estimated instead of sales forecast. The main reason for this is that there are too many different carpet models. Also, there are too many equivalent products. On the other hand, it is possible to work by grouping the products in the estimation of raw material consumption.

In fact, the problem behind this thesis was a problem in a carpet manufacturing company. Unfortunately, we could not make the right planning for one of the jute types and did not have enough stock in the past years. For this reason, we had to close several weaving machines at the same time and delay the deadline given to the customer. This problem showed us the importance of timely and sufficient supply of raw material.

For this study, data, (the amount of jutes consumed) which belongs to the years between 2015 and 2017, was taken from one of the carpet factories in Gaziantep. This factory has 39 weaving machines for manufacturing. Production planning is performed based on the orders of abroad customers. Detailed information about the company can be found in Chapter 5.

In this study, two methods are used in order to forecast raw material consumption. One of the methods is the Artificial Neural Network (ANN), that is the most popular and wide-ranging methods in recent years. The other method is one of the statistical methods of Multiple Linear Regression (MLR). For ANN method, the Neural Network Toolbox of MATLAB software was used. On the other hand, for MLP method, the Minitab software was used.

In this thesis, literature is reviewed in order to present previous studies and applications on ANN method. There are many different studies in distinct areas from banking to electricity consumption. But unfortunately, there is not much study on raw material consumption in the production area and this situation makes this thesis relatively important.

Throughout this study, figures that belong to MATLAB and Minitab software have been shown. In order to compare the performance of two methods, Mean Squared Error (MSE), R^2 and adjusted R^2 performance indicators have been used. And results show that ANN model has better performance (outperforms MLR model with respect to defined performance indicators).

In this thesis, the following outline is followed respectively;

- In Chapter 2, previous studies about ANN method from different science areas are mentioned.
- In Chapter 3, detail information about Artificial Intelligence (AI) and ANN is given.
- In Chapter 4, regression analysis is explained.
- In Chapter 5, information about the company is given.
- In Chapter 6, the case study is explained in detail (all computations, comparisons of two methods, obtained results and graphs).
- In Chapter 7, concludes the thesis and provides some suggestions for the future works.

CHAPTER 2. LITERATURE REVIEW

In this chapter, firstly, information about forecasting is given. Than, ANN is explained briefly. Advantages and disadvantages of the two methods (ANN and MLR) are also provided. After this part, studies from the literature about ANN method are given and grouped based on their areas.

2.1 Information About Forecasting:

Forecast can be defined as *'a prediction of future events used for planning purposes'* (Krajewski et al., 2013). On the other hand, planning is the process of taking managerial decisions in order to allocate limited resources to the best respond to the customer demand. Forecasting methods can be divided three main categories:

- Judgmental
- Casual
- Time-series methods.

2.1.1 Types of Forecasting Methods:

2.1.1.1 Judgmental:

When there is sufficient historical data, quantitative methods can be used for forecasting. But, in some situations judgmental methods are the only practical way for estimation. For example, if a new product is produced, or historical data is not enough, or technology is expected to change, judgment methods can be used (Krajewski et al., 2013).

2.1.1.2 Casual Methods (Linear Regression):

In situations when historical data is available and the relationship between dependent and independent variables can be defined, casual methods can be used for forecasting. Linear regression is one of the most known and commonly used casual forecasting methods. In linear regression, there are two types of variables. The first one is **dependent variable,** which is the variable to be forecasted. The second one is **independent variable,** which is assumed to affect dependent variable. Detailed information about regression analysis will be given in Chapter 4.

2.1.1.3 Time-Series Methods:

Time-series methods do not deal with independent variables like in the regression models. Time-series methods use historical data regarding only the forecasted variable. In these methods the main assumption is that the historical pattern of dependent variable will continue in the future (Krajewski et al., 2013).

2.2 Brief Explanation of ANN Method:

From the beginning of ANN studies, different definitions have been made by researchers. Two of these definitions are as follow:

- ANN is *'an approach to linear and nonlinear time series, is widely used to forecast the future. ANN provides linear and nonlinear modeling without the necessity of preliminary information and assumptions as to the relation between input and output variables. Therefore, ANN is more flexible and applicable than other methods.'* (Zhang et al., 1998).
- *'ANNs are universal and highly flexible function approximators first used in the fields of cognitive science and engineering.'* (Kaastra and Boyd, 1996).

First study about ANN method is the model of neurological networks studied by Warren McCulloch and Walter Pitts in 1943 (Kriesel, 2005). After this time, ANN methods have been used many researchers for different purposes and significant developments have been achieved. Detailed explanation on this subject will be given in Chapter 3.

2.3 Studies About ANN in the Literature:

In the literature, there are a lot of studies about ANN in different areas like economy, production, inventory level optimization, stock market business cycle prediction, electric load forecasting etc. In these section, some of these studies are grouped based on their working areas and information about them is given.

2.3.1 ANN Studies in Time-Series:

Zhang et. al. (2001) examined a simulation study of ANN for nonlinear time series forecasting. In this study, they examined the effects of three basic variables which were input nodes, hidden nodes, and sample size. It is concluded that the number of input nodes was more significant than the number of hidden nodes and a large sample size was important to ease the overfitting problem (Zhang et al., 2001).

Hwarng (2001) studied about ANN to understand better the modelling and forecasting ability of back propagation neural networks (BPNNs) on a special class of time series to improve performance. As a result of this study, when compared Box-Jenkins model BPNNs generally performed well (Hwarng, 2001).

Tseng et. al. (2002) proposed a hybrid model that combines the seasonal time series ARIMA (SARIMA) and the neural network back propagation (BP), called as (SARIMABP) to predict two different time series data – total production of Taiwan machinery industry and the soft drink time series. These two data set were used in four distinct methods – SARIMA, SARIMABP, differenced neural network, and deseasonalized neural network. In order to compare the performance of these models, the mean squared error (MSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE) were utilised. Results showed that proposed SARIMABP model had the lowest values of MSE, MAE, and MAPE. (Tseng et al., 2002).

Zhang and Qi (2005) examined the subject of how to effectively model time series with both seasonal and trend patterns. They studied the effectiveness of data preprocessing, which includes deseasonalization and detrending. Simulation and real data were examined and Box-Jenkins seasonal ARIMA model was used to compare. The results showed that neural networks were capable of capturing seasonal and trend variations effectively without any preprocessing of raw data and either detrending and deseasonalization could significantly reduce forecasting errors (Zhang and Qi, 2005).

Ghiassi et. al. (2005) proposed a dynamic ANN (DAN2) to predict time series events. In order to show the effectiveness of this model, they compared its performance with a group of benchmark time series events. MSE and MAD values were used as performance indicators and the results showed that DAN2 model outperformed both ANN and ARIMA models (Ghiassi et al., 2005).

Jain and Kumar (2007) proposed a hybrid neural network model for hydrologic time series prediction. In this study, they combined the conventional time series approach auto-regressive (AR) and ANN techniques to take advantage of both methods. The proposed method for time series forecasting was tested using the monthly streamflow data at Colorado River at Lees Ferry, USA. It is concluded that the proposed hybrid method porduced more accurate forecasts and was capable of capturing the nonlinear nature of the complex time series (Jain and Kumar, 2007).

Harpham and Dawson (2006) studied about time series forecasting with radial basis function (RBF). In this study, they stated that many studies that use RBF networks utilise one or two basis functions – generally Gaussian function. The aim of this study was to show the test set error variation among six determined basis functions. The tests were applied on three different data set – Mackey-Glass chaotic time series, Box-Jenkins furnace data, and flood prediction data sets for the Rivers Amber and Mole, UK. Two stage approach was used for RBF network training. For the first stage, k-means clustering algorithm was used and for the second stage, singular value decomposition was utilised. At the end of this study, they showed that basis function selection changed according to data set and assessment of all determined basis functions could be beneficial (Harpham and Dawson, 2006).

Hamzaçebi (2008) studied about improving artificial neural networks' performance in seasonal time series forecasting. He proposed a model called SANN (Seasonal Artificial Neural Network) that has better performance when the seasonality in time series is strong. According to the proposed SANN model; for monthly time series the number of input and output neurons should be 12, and for quarterly time series it should be 4 in order to improve forecasting performance. In this study, he used four real-world time series which are the airline passenger data set, the Taiwan machinery industry time series, the soft drink time series and the quarterly sales time series in order to test his SANN model. At the end of the study, researcher concluded that for seasonal time series forecasting SANN improved the ANN forecasting ability (Hamzacebi, 2008).

Hamzaçebi et. al. (2009) studied about multi-periodic time series forecasting. They compared direct and iterative artificial neural network approaches in mutli-periodic time series forecasting, and grey relational analysis (GRA) was used to compare the performance of these approaches. In iterative method, to predict the first subsequent period past observations were used. Then, the estimated value was used as an input. Until the last period was predicted, the process was repeated. In the direct forecast method, consecutive periods were predicted all at once. At the end of the study, it is concluded that the direct method was better than the iterative method (Hamzaçebi et al., 2009).

Khashei and Bijari (2011) studied about a time series forecasting by using a new hybrid model that combines ANN and ARIMA models. The purpose of this model was to reduce limitations of ANN model and obtain more accurate forecasting results. In this model, ARIMA was used to identify linear structure in data, and ANN model was used to capture the underlying data generating process and forecast the future. This new model was applied to three well-known real data sets and results showed that the proposed model outperformed each model used separately (Khashei and Bijari, 2011).

Sheikhan and Mohammadi (2013) studied about time series prediction based on load time series of The Institute of Electrical and Electronics Engineers (IEEE) Relability Test System. In this study, they used a multilayer feedforward neural network with partical swarm optimization (PSO) algorithm. The hybrid method that combines GA and ant colony optimization (ACO) algorithm was used in order to constitute small and suitable feature subset. In order to determine optimum neuron number of hidden layer, the PSO algorithm was used. It is concluded that the proposed hybrid model had lowest MAPE value than recent researches in this field. (Sheikhan and Mohammadi, 2013).

Wang et. al (2015) studied about time series forecasting and proposed a new approach that combines adaptive differential evolution (ADE) and back propagation neural network (BPNN), called ADE-BPNN. In this new approach, firstly ADE was utilised for searching the global initial connection weight and thresholds of BPNN. After this step, BPNN was used to profoundly research for the optimal weights and thresholds. Two real-life data sets were used for testing this hybrid approach and results showed that ADE-BP could enhance prediction accuracy with respect to basic BPNN, ARIMA and other hybrid models (Wang et al., 2015).

2.3.2 ANN Studies in the Field of Production:

Luxhoj et. al. (1996) studied about a hybrid econometric- neural network model to forecast total monthly sales. This model compounded the structurel properties of econometric models with the nonlinear pattern recognition features of neural networks to create a hybrid modeling. The model was tested with actual sales forecasting problem from a Danish company that produces consumer goods. The results showed that the proposed hybrid model had provided a 2.3 % reduction in the

MAPE value when compared with the current qualitative approach used by the company (Luxhøj et al., 1996).

Gaafar and Choueiki (2000) examined a neural network model for Material Requirement Planning (MRP) problem of lot-sizing. Different scenarios were used to evaluate the performance of the proposed model and common heuristics were utilised for comparison. It is concluded that developed ANN model was capable of solving lot-sizing problem with significant consistency and modest accuracy (Gaafar and Choueiki, 2000).

Partovi and Anandarajan (2002) studied about ANN model used for ABC classification of stock keeping units (SKUs) in a pharmaceutical company. Back propagation (BP) and genetic algorithm (GA) learning methods were used in the ANN model. Also, performance of the ANN was compared with multiple discriminate analysis (MDA) methods. It is concluded that ANN models had higher predictive accuracy than MDA (Partovi and Anandarajan, 2002).

Fonseca and Navaresse (2002) studied about ANN method for the traditional job-shop simulation approach. The established neural network architectures significantly estimated manufacturing lead times (MLTs) for orders simultaneously processed in a four-machine job shop. Also, the ANN based simulation was capable of capturing the underlying relationship between jobs' machine sequences (Fonseca and Navaresse, 2002).

Zhang (2003) studied about a hybrid methodology that combines both ARIMA and ANN models in order to take advantage of two models for linear and nonlinear modeling. In order to show the effectiveness of the proposed hybrid model, three data sets- the Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate data - are used. The sunspot data contained the annual number of sunspots from 1700 to 1987. The lynx series involved the lynx trapped number/year in the Mackenzie River district of Northern Canada. The last data set was the exchange rate between the British pound and US dollar. In order to compare the performance of hybrid model with ARIMA and ANN; MSE and Mean Absolute Deviation (MAD) performance indicators were used. It is concluded that the hybrid method could outperform each component model used in isolation (Zhang, 2003).

Co and Boosarawongse (2007) studied about forecasting Thailand's rice export. They investigated three models which were exponential smoothing, the Box-Jenkins model and ANN simulation in order to forecast rice exports from Thailand. For exponential smoothing, they used Holt-Winter method. At the beginning of this study, rice was grouped into 4 categories according to their export. They used MAE, MSE, MAPE, and root mean squared error (RMSE) performance indicators for comparing the performance of methods. It is concluded that ANN approach outperformed other two methods (Co and Boosarawongse, 2007).

Lin et. al. (2009) studied about ANN prediction method for optimization of work-in-process (WIP) inventory level for wafer fabrication. In this study, they developed an algorithm integrates ANN and sequential quadratic programming (SQP) method. At the end of this study, they stated that the proposed algorithm was an effective and systematic way to define optimal WIP level (Lin et al., 2009).

2.3.3 ANN Studies in the Field of Electricity:

Elkateb et. al. (1998) studied a medium-weather-dependent load forecasting using enhanced ANN / fuzzy neural network (FNN) and statistical techniques with monthly peak load demand data of Jeddah area. The average error value for ARIMA method was 11.7 %. Direct ANN and FNN methods showed poor performance. By introducing **'time index feature'**, that was assumed any value in the interval [0, 1.0] and increases by an equal step, performance of both ANN and FNN methods significantly enhanced (Elkateb et al., 1998).

Gareta et. al. (2006) studied about the neural network that can be used to predict short term hourly electricity pool prices. Extensive data sets were used to test neural network architecture and results showed that there was good aggrement between actual data and neural network results. Also, this method could be used to improve power plant generation capacity management (Gareta et al., 2006).

Azadeh et. al. (2007) studied about an integrated method that combines GA and ANN to predict electricity demand using stochastic procedures. The variables of price, value added, number of customers and consumption in historical data were selected as eceonomic indicators. While all values of parameters were tested simultaneously, the GA had been tuned for all its parameters and the best coefficients that had minimum error were identified. The estimated error values of the GA method were less than regression method. As a result of this study, the proposed integrated method outperformed time series approach (Azadeh et al., 2007a).

Azadeh et. al. (2007) studied about forecasting electrical consumption by integration of neural network, time series, and ANOVA. As the first step, data were preprocessed with time series using moving average method to enhance output. Then, ANN method was applied. After this step, in order to compare ANN, regression and actual data, ANOVA was applied. It is concluded that the ANN method had higher predictive accuracy (Azadeh et al., 2007b).

Hamzaçebi (2007) studied about the prediction of Turkey's net electricity consumption for four sectors that include industrial, residential, agriculture, and trasportation sectors. The purpose of this study was to show the importance of using alternative forecasting techniques. He used ANN method and Model for Analysis of Energy Demand (MAED) simulation technique. Absolute percentage error (APE) and MAPE were used to compare ANN and MAED results. As a result of this study, ANN method gave better results than MAED. But, researcher stated that this study did not show that ANN method was always better than MAED technique (Hamzaçebi, 2007).

Catalao et. al. (2007) studied about forecasting short-term electricity prices by using ANN method. They used three-layered feed forward neural network and Levenberg-Marquardt algorithm as a training algorithm. They evaluated the accuracy of ANN method by using the results from the electicity markets of mainland Spain and California. MAPE was selected as performance criterion and ANN results were compared with ARIMA technique and naïve procedure. In the naïve procedure, the forecasted prices for a given week were the actual prices of the previous week, that is, using no-change criterion. As a result of this study, ANN method outperformed both ARIMA and naïve procedure (Catalão et al., 2007).

Kheirkhah et. al. (2013) studied about estimation and prediction electricity demand by using ANN, principal component analysis (PCA), data envelopment analysis (DEA) and ANOVA methods. ANOVA was utilized to determine the best structure in the group that has been specified by DEA. DEA was utilized to compare the constructed ANN models according to their learning algorithm performance. PCA was utilized like input selection method, and proper time series model was selected from the linear (ARIMA) and nonlinear models. After that, Mcleod-Li test was used for determining the nonlinearity condition. After satisfying the nonlinearity condition, the requested nonlinear model was selected and compared with the selected ARIMA model, and the best time series model was preferred. Then, a new ANN-PCA-DEA-ANOVA algorithm was developed to estimate the time series. It is concluded that the developed algorithm ensured a correct solution to estimate electricity consumption (Kheirkhah et al., 2013).

2.3.4 Studies on ANN in the Field of Economics:

Kohzadi et al. (1996) studied about ANN in order to forecast commodity prices. In this study, data belonging to monthly live cattle and wheat prices from 1950 through 1990 were used. ANN method was compared autoregressive integrated moving average (ARIMA) method and results showed that mean absolute percentage error (MAPE) was lower for ANN method (Kohzadi et al., 1996).

Chiang et. al. (1996) used ANN method in order to forecast the end-of-year net asset value (NAV) of mutual funds. For the NAV data prediction, historical economic data were used. The results of ANN method were compared with linear and nonlinear regression analysis and results showed that ANN method outperformed regression models (Chiang et al., 1996).

Kaastra and Boyd (1996) studied about a practical introductry guide in order to design a neural network to predict economic time series data. They described the eight-step procedure to design a neural network forecasting model. These steps were; (1) variable selection, (2) data collection, (3) data preprocessing, (4) training, testing, and validation sets, (5) neural network paradigms, (6) evaluation criteria, (7) neural network training, and (8) implementation (Kaastra and Boyd, 1996).

Frances and Draisma (1997) studied a graphical method based on an ANN model to examine how and when seasonal patterns in macro economic time series change. This model was applied to quarterly Industrial Production in France and the Netherlands. The results of this study indicated that observations often changed in only one or a few seasons (Franses and Draisma, 1997).

Guresen et. al. (2011) studied about prediction of stock exhange rates by using three methods which were multi-layer perceptron (MLP), dynamic artificial neural network (DAN2), and a hybrid model that combines ANN and generalized autoregressive conditional heteroscedasticity (GARCH-ANN) in order to reduce deficiencies of MLP model. For evaluating these methods, NASDAQ Stock Exchange data were used. MSE and MAD were selected as performance indicators. It is concluded that, MLP model was better than other methods with a little difference (Guresen et al., 2011).

2.3.5 ANN Studies in Other Fields:

In this part, ANN studies that had not been gouped in the previous fields like traffic flow forecasting, runoff prediction of a watershed, or prediction of innovation performance are explained. Especially, the study of Efendigil et al. (2009) is about demand forecasting and the closest study to this thesis. But, there is a difference between these studies. Efendigil et. al. (2009) studied about demand forecasting, on the other hand this thesis is about raw material consumption prediction based on demand forecasting.

Markham and Rakes (1998) examined the robustness of simple linear regression and ANN with respect to different sample size and variance of the error. The root mean square (RMS) value was used as a performance indicator. It is concluded that sample size and variance variables were significant for the performance of regression analysis and ANN method. Both techniques outperformed with a large sample size and lower variance levels (Markham and Rakes, 1998).

Wang and Chien (2006) studied ANN method for a different purpose. Their study was about a forecasting model that predicts inovation performance using technical informational resources and clear innovation objects. In this study, they proposed a neural network approach that uses the Back-Propagation Network (BPN) to solve this problem and results compared with a regression model. It is concluded that BPN method was capable of predicting innovation performance and outperformed regression method (Wang and Chien, 2006).

Zhang et. al. (2007) studied a new hybrid algorithm that combines particle swarm optimization (PSO) algorithm and back propagation (BP) algorithm and this developed algorithm can be called PSO-BP. This new algorithm was developed to train the weights of feedforward neural network (FNN). This hybrid algorithm could use the powerful global searching ability of PSO algorithm and strong local searching capability of BP algorithm at the same time. In the PSO-BP algorithm, a heuristic approach was used for transition from particle swarm search to gradient descent search. It is concluded that, the developed hybrid PSO-BP algorithm outperformed adaptive particle swarm optimization algorithm (APSOA) and BP algorithm with respect to convergent speed and accuracy (Zhang et al., 2007).

Efendigil et. al (2009) studied about demand forecasting in a multi-level supply chain structure. The aim of this study was to develop a new forecasting model which was formed by artificial intelligence. This new developed model included the comparison of ANN and adaptive network-based fuzzy inference system (ANFIS). MAPE and MSE were used as performance indicators and results showed that the proposed model outperformed ANN structure (Efendigil et al., 2009).

Golmohammadi (2011) studied about a fuzzy multi-criteria decision making problem. A neural network was designed in order to learn the relationships among criteria and alternatives and then sort the choices. The developed model could use historical data and update these data for using in future decisions. In addition to this, this model could be used for very different kind of multi-criteria decision making problems (Golmohammadi, 2011).

Ren et. al. (2014) studied about wind speed prediction by using a new method that combines PSO-BP and Input parameter Selection, that is called IS-PSO-BP. The aim of this study was to show that comprehensive parameter selection could be useful to improve forecasting accuracy. In order to show prediction performance of the developed method, wind speed data of Jiuquan and 6-hourly wind speed data of Yumen, Gansu onf China from 2001 to 2006 were used. At the end of this study, they showed that this developed method outperforms the BP neural network and ARIMA model (Ren et al., 2014).

Moretti et. al. (2015) studied about urban traffic flow forecasting by using a hybrid model that combines ANN and simple statistical approach. The proposed approach was tested on three different classes of real streets. At the end of this study, it is showed that this new hybrid model outperformed the ANN and statistical method (Moretti et al., 2015).

Rather et. al (2015) studied about stock returns prediction by using a new hybrid model that combines two nonlinear models - autoregressive moving average and exponential smoothing- and a nonlinear model – recurrent neural network (RNN). In order to form optimal weights for developed model, an optimization model was generated and GA was used to solve this model. For improving forecasting accuracy, the developed hybrid model combined predictions from these three forecasting models. As a result of this study, the developed hybrid model outperformed RNN and produced satisfactory predictions (Rather et al., 2015).

Qui et. al. (2016) studied about stok market returns forecasting by using ANN. In this study, they proposed a new set of input variables for improving forecasting effectiveness. They predicted Japanese Nikkei 225 index by using back propagation (BP) learning algorithm with selected input variables. Also, GA and simulated annealing (SA) were used to improve ANN forecasting accuracy. It is concluded that the developed model that used GA and SA enhanced forecasting accuracy and outperformed BP learning algorithm (Qiu et al., 2016).

In this part of thesis, different studies about ANN method have been explained. It can be said that ANN method has a wide range application areas.

CHAPTER 3. ARTIFICIAL INTELLIGENCE (AI) AND ARTIFICIAL NEURAL NETWORK (ANN)

3.1 ARTIFICIAL INTELLIGENCE (AI):

In this part of the thesis, **‘artificial intelligence (AI)’** and **‘artificial neural network (ANN)’** approaches are explanied and detail information about them is provided.

Artificial intelligence (AI) is the latest field of study for engineering and also science. Studies about AI started after World War II. AI comprises many different kinds of study field in engineering and science disciplines such as proving mathematical formulas or diagnosing illnesses.

There are many different definitions about AI in the literature. Some of them are given below:

‘The study of mental faculties through the use of computational models.’ (CHARNIAK and McDermott, 1985).

‘The art of creating machines that perform functions that require intelligence when performed by people.’ (Kurzweil, 1990).

‘The study of how to make computers do things at which, at the moment, people are better.’(Rich and Knight, 1991)

‘Artificial intelligence (AI) is the part of computer science concerned with designing intelligent computer systems, that is, sytems that exhibit the characteristics we associate with intelligence in human behavior – understanding language, reasoning, solving problems, and so on.’ (Cohen and Feigenbaum, 2014).

As stated before, there are many activities in different subfields that AI is used. Some of these applications are below:

- **Problem solving:** One of the main achivement of AI is programs that can solve the puzzle and difficult problems. When solving difficult problems, AI programs use techniques called ‘search and problem reduction’ which divides hard problems into easier sub-problems. Another program that solves problem combines mathematical formulas symbolically and thus delivers great performance. This program is used by engineers and scientists (Cohen and Feigenbaum, 2014).

- **Logical reasoning:** Logical reasoning is one of the most continuously surveyed subfields of AI. This subarea is closely related to problem solving. Developed programs can prove claims by processing a database of facts. Special interest is given to problems of determining ways of focusing on just relevant truths in a large database and updating them when new data gains (Cohen and Feigenbaum, 2014).
- **Language:** At the beginning of AI studies, the field of language understanding also examined. Written programs perform several tasks such as translate sentences between two languages. Also these programs can attain information by reading textual material. (Cohen and Feigenbaum, 2014).
- **Speech recognition:** Different from translation between distinct languages, AI can be used for speech recognition. For example, in United Airlines automated speech recognition system can make a dialog with a traveler in order to book a flight (Russell et al., 2010).
- **Programming:** Programming is a significant area for AI studies on its own. Studies in this area are called 'automatic programming' and analyze systems that can write computer programs for their usage. Studies in this automatic programming field can result in semiautomated software-development systems and also AI programs that update their ways by updating their code (Cohen and Feigenbaum, 2014).
- **Logistic planning:** AI approach also can be used as a planning technique for logistic purposes. For example, in the Period of Persian Gulf crisis in 1991, Dynamic Analysis and Replanning Tool, called as DART, was used by U.S. government in order to make automated logistics planning and scheduling. In this transportation scheduling, there were 50,000 vehicles and people. By using AI planning techniques, a transportation plan was generated in hours (Russell et al., 2010).
- **Learning:** In the AI studies, there have been some attractive attempts comprising programs that learn from examples and from their own performance. Today's studies in the subfield of AI called 'artificial neural network' uses learning function very effectively. ANN systems learn

successfully from the examples and forecast with high accuracy about the future (Cohen and Feigenbaum, 2014).

- **Expertise:** The field of expert systems (i.e., knowledge engineering) is a new way for successful and practical applications of AI techniques. Generally, a user interacts with an expert system in a consultation dialogue. Today's expert systems accomplish consultation dialogs in high level performance (Cohen and Feigenbaum, 2014).
- **Robotics and vision:** Another subarea of AI studies comprises programs that manipulate robot equipments. Studies in this area comprise a wide range field from the movements of the robot's arm to the way of planning a sequence of actions to accomplish a robot's objective. Today, the millions of robots which are being used in industrial implementations are basic devices which are programmed to perform some basic iterative tasks (Cohen and Feigenbaum, 2014). For example, Roomba robotic vacuum cleaner which has been produced by the iRobot Corporation has been sold in huge amounts (Russell et al., 2010) From the vision perspective of AI research, most visual information can be processed actively. Developed programs can identify items and even small changes from one images to another like aerial reconnaissance.
- **Machine translation:** A computer program can translate between two different languages lik from English to Spain (Russell et al., 2010).

From start to today, the work area of AI is growing continuously and many AI research group in different countries like Japan, Germany, Canada, Britian, France contributes to these areas with new ideas and tools. Neverthless, much more people who want to work in these AI research groups are needed especially in universities.

3.1.1 Historical Development of Artificial Intelligence

Historical development of AI studies is sorted chronologically below:

- **Between 1943-1955:** The work of Warren McCulloch and Walter Pitts is the first study in the field of AI. They offered an artificial neurons model in which each neuron is characterized as being 'on' or 'off'. In this model, 'on' occurs in response to stimulation by an adequate number of neurons in neighbourhood. (It can be said that they defined the Boolean circuit model of the brain.) Also, they showed that some network of connected neurons can calculate any computable functions. In addition to these works, they proposed that properly defined nets can learn. In 1949, Donald Hebb showed an ordinary modifying rule to update the strenghts of connections between neurons. His rule is called **'Hebbian learning'** and is valid still today. In 1950, Marvin Minsky and Dean Edmonds built the first computer for the purpose of simulation of a neural network and this computer was called as 'SNARC'. This computer simulated a network with 40 neurons. And also in this year, one of the most important studies for the AI field was done by Alan Turing. Alan Turing proposed the **Turing Test,** genetic algorithms, and machine learning (Russell et al., 2010).

- **In 1956:** The **'artificial intelligence'** approach was suggested in Dartmouth College (Russell et al., 2010).
- **Between 1952-1969:** At the beginning of 1952, Arthur Samuel developed a series of programs which quickly learned to play the game better than its builder (Poole and Mackworth, 2010). Thus, he could disprove the idea that the computers can do only thing what you say. General Problem Solver (GPS) was designed from the beginning to imitate human problem-solving procedures. The GPS was the first program to concretize the approach of 'thinking humanly'. In this times, some of the first AI programs were produced at IBM (Russell et al., 2010). In 1959, Herbert Gelernter built the Geometry Theorem Prover that could demonstrate theorem which was quite tricky for many students of mathematics. In 1958, John McCarthy defined the high-level programming language called **'Lips'** that was to become favourite programming language of AI studies along 30 years. In 1962, Bernard Widrow enhanced Hebb's learning methods and called his networks as **'adalines'.** In 1963, the SAINT program developed by James Slagle could solve closed-form calculus integration problems. In 1968, the ANALOGY program developed by Tom Evans could solve geometric analogy problems in IQ tests (Russell et al., 2010). In 1967, STUDENT program developed by Daniel Bobrow was able to solve algebra story problems (Poole and Mackworth, 2010).
- **Between 1966-1979:** In these times, AI studies met with some computational difficulties. The first difficulty is about machine translation from one language to another. Another difficulty is solving larger problems by early AI

programs. At the end of this situations, AI studies came to a stopping point. Also, in these years, knowledge-intensive systems arose. The DENDRAL program written for solving the problem inferring molecular structure from the information supplied by a spectrometer is the first knowledge-intensive system (Poole and Mackworth, 2010).

- **In 1980-2005:** In 1980, AI became an industry. The first accomplished expert system, R1, started to its operations at the Digital Equipment Corporation. In the mid-1980s, AI studies started to become popular again. In 1998, AI systems became most common in Web-based applications. In addition to this, AI Technologies used for many Internet tools like search enginees, and Web site aggregators. Between the years 2000 and 2005, robot toys were released to the market (Russell et al., 2010).

As it can be understood from this part, AI approach has been used in many different areas since the emergence of this approach and continues to attract the attention of many researchers.

3.2 ARTIFICIAL NEURAL NETWORKS

3.2.1 Definitions of Artificial Neural Networks:

In the literature, different definitons are made for artificial neural networks (ANNs). Some of them are as follows:

- *'An ANN is an information processing system that has certain characteristics in common with biological neural networks. ANN has been developed as generalizations of mathematical models of human cognition or neural biology*' (Fausett and Fausett, 1994).
- *'A neural network is a massively parallel distributed processor made up of simple processing units that has a natural propensity for storing experiential knowledge and making it available for use'* (Haykin and Haykin, 2009).
- *'ANNs are universal and highly flexible function approximators first used in the fields of cognitive science and engineering.'* (Kaastra and Boyd, 1996).

In summary, ANNs are a mathematical approach that is inspired by the human brain's ability to learn and produce new information and is used in different fields and for distinct purposes, especially in the solutions of nonlinear problems.

3.2.2 Historical Development of ANN:

It can be said that the first computers that were produced could do only arithmetic calculations. Today's computers can also learn cases and decide about them according to environmental factors. ANN is one of the scientific fields that can contribute to this development.

Historical development of ANN is explained in chronological order as follow:

- **In 1943:** The first model of neurological networks using simple binary threshold functions were introduced by Warren McCulloch Walter Pitts (Mehrotra et al., 1997).
- **In 1949:** The classical **Hebbian Rule,** which is the basis for almost all neural learning procedures, was formulated by Donald Hebb. The Hebbian Rule states that when two neurons are active simultaneously, the connections between them is strengthened (Kriesel, 2005).
- **In 1954:** Learning filter was invented by Gabor. This approach utilises gradient descent method in order to find 'optimal weights' which minimize MSE value between actual output and generated output produced by net (Mehrotra et al., 1997).
- **In 1957-1958:** The first prospering neurocomputer, that was called Mark I perceptron, was developed by Frank Rosenblatt, Charles Wightman, and their cooperator (Kriesel, 2005).
- **In 1960:** The **ADALINE (ADAptive LInear NEuron),** a quick and certain adaptive learning system was introduced by Bernard Widrow and Marcian E. Hoff. The ADALINE was the first commercially widespread used neural network (Kriesel, 2005).
- **In 1963:** 'Perceptron Convergence Theorem' was proofed by Novikoff (Mehrotra et al., 1997).
- **In 1969:** A certain mathematical analysis of a perceptron was produced by Marvin Misky and Seymour Papert. This analysis indicated that some significant problems like XOR problems cannot be solved by using a perceptron model. This search caused a significant decrease in both ANN studies and research funds to this area for a while (Kriesel, 2005).

As a result of the study of Misky and Papert, ANN studies lost their popularity and only some of the researchers continued their investigations. In addition to this, there was no any conference or other events in the scientific field about ANN. The studies after that time are expressed below:

- **In 1972:** A model of **linear associator (associative memory)** was introduced by Teuvo Kohonen (Fausett and Fausett, 1994).
- **Between 1976-1980:** Stephen Grossberg analyzed several neural networks mathematically. Also, **Adaptive Resonance Theory (ART)** was introduced at the end of the research of him and Gail Carpenter (Kriesel, 2005).
- **In 1982: Self-organizing future maps (SOMs)** – also called as Kohonen maps – was introduced by Teuvo Kohonen (Fausett and Fausett, 1994).
- **In 1983: Neocognitron** neural model, that was capable of identifying handwritten characteristics, was introduced by Fukushima, Miyake, and Ito (Kriesel, 2005).
- **In 1985: Hopfield nets** was introduced in an article by John Hopfield. This net offered acceptable solutions for the Travelling Salesman problem (Fausett and Fausett, 1994)
- **In 1988:** A model of Radial Basis Functions neural networks (RBF) was introduced by Broomhead and Lowe. In the article, they stated that this neural network can be an alternative to the multilayer perceptron neural networks.

3.2.3 Advantages of Artificial Neural Networks:

It can be said that an artifcial neural network takes its power from two properties: the first one is its structure. An ANN shows a parallel structure and these characteristics give it speed when solving problems. The second property is its ability to learn and generalize. The useful characteristics and capabilities of ANN can be sorted as follows:

- **Nonlinearity:** Most of the daily life problems are very complex and nonlinear. While traditional models are often used for linear problems, an

ANN can be used for both linear and nonlinear problems (Haykin and Haykin, 2009).

- **Input-Output Mapping:** A supervised learning, that comprises an alteration of the weights between neurons in an ANN by using a set of labeled training examples, is a popular learning approach. Each example involves a unique input signal and a corresponding target (desired) signal. The network is represented with a random example from the set, and the weights between the neurons are updated in order to minimize the difference between the desired and actual response. The network training is iterated by using examples from the set until reaching a steady state of the network. In the steady state situation of a network, there are no important changes in the connection weights between neurons. In this way, network learns by using examples by constructing input-output mapping (Haykin and Haykin, 2009).
- **Adaptivity:** A neural network has an inbuilt capability to adapt its weights between neurons when a change occurs in its enviroment. Especially, a neural network can be easily retrained to cope with some changes in the surrounding environment. Also, when a neural network work in a nonstationary environment, it may be designed to alter its weights between neurons in real time (Haykin and Haykin, 2009).
- **Evidential Response:** As part of pattern classification, a neural network can be designed to supply information for both which specific pattern to be selected and the confidence in the decision. The confidence in the decision can increase the classification performance of the neural network by rejecting ambiguous patterns (Haykin and Haykin, 2009).
- **Contextual Information:** Knowledge is presented by the activation state and structure of an artificial neural network. All neurons in the network affect each other with their global activity. As a result, the term ‘contextual information’ is the object of the neural network (Haykin and Haykin, 2009).
- **Fault tolerance:** A neural network consists of a number of cells connected and distributed in parallel. The information that enters the network is distributed over all connections. The ineffectiveness of some cells in the network training does not affect considerably the network in order to produce correct information. As a result, it can be said that artificial neural networks

have more fault tolarance when compared to traditional methods (Haykin and Haykin, 2009).

- **VLSI (Very-Large-Scale-Integrated) Implementability:** The structure of a neural network is massively in parallel form and this feature makes it fast for computation of complex problems. Also, this characteristics allow an artificial neural network to be used in VLSI (very-large-scale-integrated) technology. One of the most important features of a VLSI technology is capturing completely complex behaviour in a hierarchical form (Haykin and Haykin, 2009).
- **Uniformity of Analysis and Design:** Essentially, neural networks are universal information processors. This means that the same notation is used in all areas of neural networks applications. It can be understood from different properties of neural networks (Haykin and Haykin, 2009). These properties are:
  - ✓ In all neural networks, neurons are the common ingredient.
  - ✓ With this commonality, theories and learning algorithms can be shared among different neural network applications. Also, with error-free integration of modules, modular networks can be installed.

3.2.4 Structure of Artificial Neural Networks

3.2.4.1 The Human Brain (Biological Neural Networks):

Biological neural networks are capable of processing high and complex events. It is estimated that there are roughly 10 billion neurons in the human cortex and approximately 60 trillion synapses between them. An ANN aims to provide this ability for computers. Therefore, it is more useful to recognize the biological nerve cell before moving on to the subject of ANN (Haykin and Haykin, 2009).

The human neural system can be summarized as a three-stage system as showed in the block diagram of Figure 3.1:

Figure 3.1: Block diagram representation of the neural system

A brain that is showed by the neural net is the centre of the system. This part perpetually receives information, detects it, and makes proper decisions. In the block diagram, there are two sets of arrows. The arrows showed from left to right specify the forward transmission of information-bearing signals throughout the system. Those pointing from right to left indicate the presence of feedback in the nervous system. The receptors are responsible for converting stimuli from the external environment or the human body into electrical impulses that transmit information to the brain. On the other hand, the effectors are responsible for transmitting electrical impulses that are generated by the brain into noticeable responses (Haykin and Haykin, 2009).

The biological neural system is a center that takes information, interprets and produces appropriate decisions. The basic elements of the neural system are called neurons.

3.2.4.2 Components of a Neuron:

A biological nerve cell consists of dendrites, cell body, axons, and synapses. The biological nerve cell is shown in Figure 3.2:

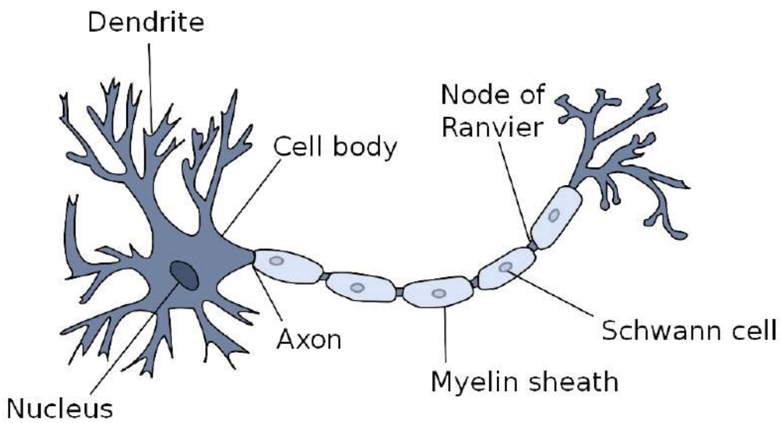

Figure 3.2: Illustration of the biological neuron (Kriesel,2005)

Synapses are one of the basic structural and functional units which are responsible for supplying interactions between neurons. Dentrits are structures extending from the cell nucleus called soma and these structures look like tree branchs. The task of the dentrits is receiving electrical signals and transmitting into the nucleus. Soma is the nucleus of the cell and accumulates all of the signals that have been taken from dentrits. When the accumulated signals exceed a specific value (threshold value), then the soma activates an electrical pulse to the other neurons. The electrical pulse to other neurons is sent by axons (Kriesel, 2005).

3.2.4.3 The Basic Structure of Artificial Neural Networks:

Neural networks have nerve cells, such as biological neural networks. Artificial neural networks are the modeling of biological neural networks. Artificial neural networks consist of a number of interconnected neurons and these neurons receive information and produce an output with various functions. In Figure 3.3, the model of a neuron is shown. In this figure, all basic components of a neuron can be seen. These 5 basic components are as follows:

- Inputs
- Weights
- Propagation function
- Activation function
- Output

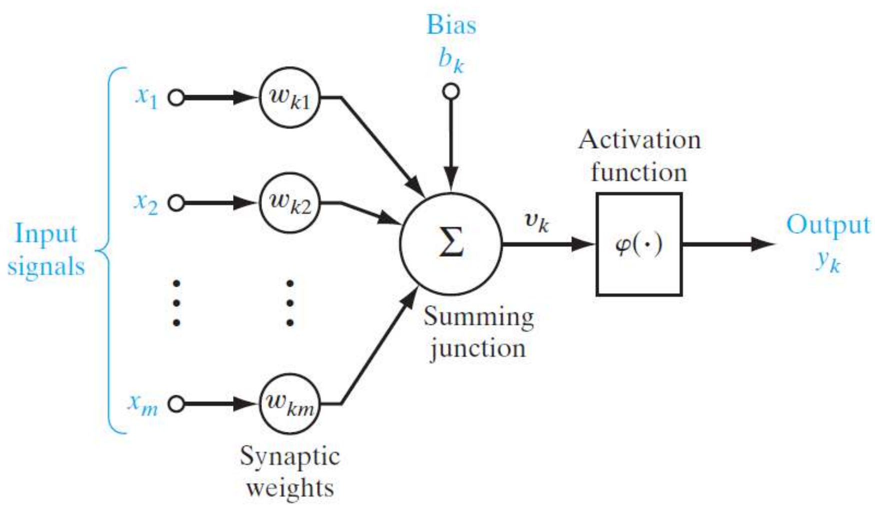

Figure 3.3: A nonlinear model of a neuron (Haykin,2009)

All these basic components of a neuron can be explained as follows:

3.2.4.3.1 Inputs:

Inputs are information which are taken from the environment, other cells or itself by an artificial nerve cells and output is produced for these inputs.

3.2.4.3.2 Weights:

The weights (w) indicate the effect of data on an artificial nerve cell. The fact that the weights are large or small does not mean that they are significant or unimportant. The zero value of a weight can be the most important event for that network. A plus (+) or minus (-) weight indicates that the effect is positive or negative. Weights can be changeable or fixed value.

3.2.4.3.3 Propagation Function:

This function calculates the net input to a cell. Different functions are used for this purpose. The most common one is to find weighted sum function. In this function, each incoming input value is multiplied by its own weight. Thus, the net input to the network is found. The formula of the weighted sum function is shown in Equation (3.1) (Haykin, 2009):

$$Net = \sum_{i} X_j W_{kj} \qquad (3.1)$$

In this formula, X_j shows input signals, and W_{kj} is the respective weights of neuron k. However, it is not always necessary to use this formula in an ANN. Different formulas have been used as the propagation function in the literature. In Table 3.1, five different propagation functions and their explanations are given (Öztemel, 2012). It can be seen from Table 3.1, calculation methods of each function is different. Therefore, the selected propagation function can affect the performance of ANN model. There is no defined rule for choosing the propagation function, therefore the most appropriate function can be found by trial and error.

3.2.4.3.4 Activation Function:

This function processes the net input coming to the cell and determines the output that the cell will generate for that input. Like in the propagation function, different formulations can be used for activation function. Some of the models like multilayer perceptrons require that this function must be a function that can be derived. The most suitable activation function for a problem can be found by trial and error by the designer of the network. For multilayer perceptron models, a sigmoid function is generally used for activation function. The formula for the sigmoid function is given in Equation (3.2) (Haykin, 2009):

$$\varphi(v) = \frac{1}{1 + \exp(-\alpha v)} \qquad (3.2)$$

In this formula, φ(υ) shows the activation function for the induced local field υ, and α is the slope parameter of the sigmoid function. By changing the α parameter, we can obtain distinct sigmoid functions as shown in Figure 3.4:

Table 3.1: Examples of propagation functions

Net	Statement
Multiplication $Net = \prod_i X_j W_{kj}$	The weight values are multiplied by the inputs and t calculated values are multiplied with each other to calculate the net input.
Maximum $Net = Max\ (X_j W_{kj)}\ i = 1,2,\ldots,N$	After the inputs are multiplied by their weight values, the maximum value is considered the net input of the network.
Minimum $Net = Min\ (X_j W_{kj)}\ i = 1,2,\ldots,N$	After the inputs are multiplied by their weight values, the minimum value is considered the net input of the network.
Majority $Net = \sum_i sgn(\ X_j\ W_{kj)}$	Firstly, each input is multiplied by its own weight and the numbers of positive and negative ones are found. Then, great value is considered the net input of the network.
Cumulative Total $Net = Net(previous) + \sum X_j\ W_{kj}$	Each input is multiplied by its own weight and all multiplications are summed. After that, this value is added to previous information and the net input of the networks is found.

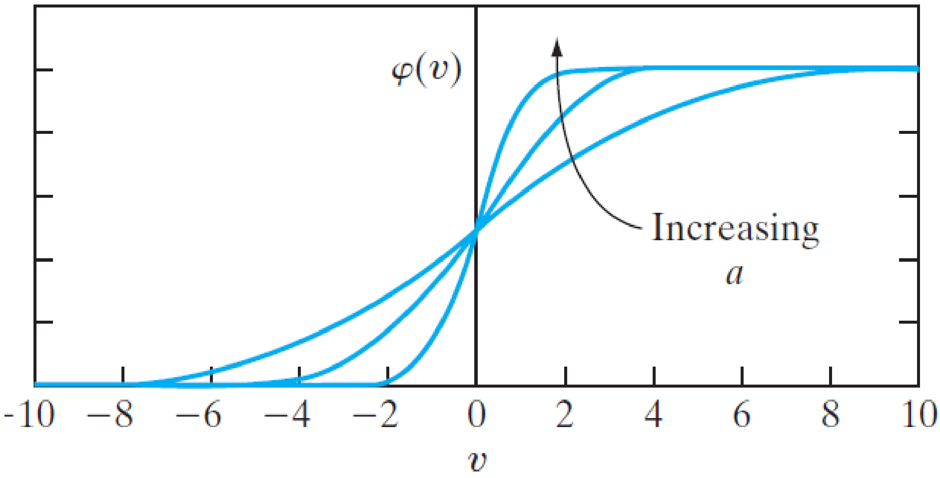

Figure 3.4: A sigmoid function for varying slope parameter α (Haykin,2009)

Some of the activation functions are given in Table 3.2. In Table 3.2, *Net* shows the net input value coming to the neuron. Step function is often used in single-layer nets. On the other hand, hyperbolic tangent function is often preferred when the output in a range of (-1,1) (Fausett and Fausett, 1994).

Table 3.2: Examples of activation functions

Activation Functions	Statement
Lineer function $F\ (Net) = Net$	The output of the cell is accepted as is the input form.
Step function $F(Net) = \begin{cases} 1, if\ Net > threshold \\ 0, if\ Net \leq threshold \end{cases}$	The output of the cell takes a value of 0 or 1 depending on whether the incoming net input value is below or above a specified threshold value.
Sinus function $F\ (Net) = \sin(Net)$	This function is used in cases where the events taught to be learned are distributed according to the sinus function.
Threshold function $F\ (Net) = \begin{cases} 0, if\ Net < 0 \\ 1, if\ Net \geq 0 \end{cases}$	In this function, the output of a neuron is 1, if the Net value is higher than 0. Otherwise, it takes the value of 0.
Hyperbolic tangent function $F\ (Net) = \dfrac{(e^{Net} - e^{-Net})}{(e^{Net} + e^{-Net})}$	The net value is calculated by passing the tangent function.

3.2.4.3.5 Output:

This is the output value determined by the activation function. The output is sent to the environment or to another cell. The output is the solution of the problem for an ANN.

3.2.4.4 The Difference of Neural Networks from Traditional Computing:

When compared traditional computing methods, a new and different way is offered by neural networks in order to analyze data, and identify a pattern within that data. The problems which can be well qualified like balancing checkbooks or keeping tabs of inventory can be solved by using traditional computing methods. Experts systems can be called the fifth generation of computing and these systems are a widening form of traditional computing methods. On the other hand, artificial neural networks can be called the sixth generation of computing and this method proposes a different approach for solving problems. The problems in areas like machine learning, vision, and pattern recognition can be solved by using artificial intelligence (Anderson and McNeill, 1992). The basic differences between traditional computing (including expert systems) and ANN are given in Table 3.3. For example, while an expert system can ben used for inventory operations, digital communication or accounting, an ANN model can be used for sensor processing, text recognition or speech recognition. As stated in Chapter 3, for instance, in United Airlines automated speech recognition system can make a dialog with a traveler in order to book a flight (Russell et al., 2010).

Table 3.3: Comparison of computing approaches

Characteristics	**Traditional Computing (Including Expert Systems)**	**Artificial Neural Networks**
Processing style	Sequential	Parallel
Functions	*Logically (left brained via) *Rules *Concepts *Calculations	*Gestault (right brained) via *İmages *Pictures *Controls
Learning method	By rules (didactically)	By example (socratically)
Applications	*Accounting, *Word processing, *Math, *Inventory, *Digital comunication	*Sensor processing, *Speech recognition, *Pattern recognition, *Text recognition

3.2.4.5 Design of Artificial Neural Networks:

The network needs to be well designed for an ANN to be successful. The success of the model depends on the design of the model.

When a neural network is built, some parameters and variables which define the network should be determined. One important decision for building a neural network is finding proper architecture, the number of layers, and the number of nodes for each layer. Other decisions include the determination of the learning strategy, the activation functions, the training algorithm, normalization methods, and performance measures. Unfortunately, there is no rule for determining the best selections for these parameters. Generally, the most feasible ANN structure is found by trial and error (Golmohammadi, 2011).

3.2.4.6 Normalization:

All data in an ANN must be in one range and in a similar fashion with each other. If all data are not in one range and not proportional to each other, that is, if the difference between the data is too large or to small, the resulting net inputs will be out of proportion. For this reason, the data must be scaled at a certain interval in order to avoid computational errors. Usually, the range of [0,1] is preferred for scaling. Data normalization is implemented before the training process begins. The most common normalization formulas are given in Equation (3.3), (3.4), (3.5), and (3.6). In linear transformation to range [0,1], the maximum datum is transformed to 1, and the minimum datum is transformed to 0. All other data points are transformed to (0,1) interval. In linear transformation to range *(a,b),* the maximum datum is transformed to b, and the minimum datum is tranformed to a. In simple normalization, all data points is divided by the maximum datum. In statistical normalization method, mean and standard deviation of all data are used for normalization.

$$x_n = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \qquad (3.3)\, Linear\ transformation\ to\ range\ (0,1)$$

$$x_n = (b-a)\frac{x_0 - x_{min}}{x_{max} - x_{min}} + a \qquad (3.4)\, Linear\ transformation\ to\ range\ (a,b)$$

$$x_n = \frac{x_0}{x_{max}} \qquad (3.5) \textit{Simple normalization}$$

$$x_n = \frac{x_0 - \bar{x}}{s} \qquad (3.6) \textit{Statistical normalization}$$

In this equations, x_n and x_0 show the normalized and original data respectively. In addition to this, x_{min}, x_{max}, x_n, $\bar{x}$, and s are the minimum, maximum, normalized data, mean, and standard deviation, respectively.

3.2.4.7 Artificial Neural Network Topology Selection:

It is necessary to determine the neural network topology correctly in order to solve the problem successfully with this method. In the selection of an ANN topology, input neuron number, layer number of the net, neuron number in each layer, and output neuron number are determined. The choice of ANN topology depends on the type of problem. In Table 3.4, some of the neural network types in different categories are shown according to their usage areas and purpose of use. (Anderson and McNeill, 1994). Table 3.4 shows more common network topologies. Some of the networks that have been grouped more than one network type can be used to solve for different problems.

In Table 3.4, 17 different networks were grouped according to their purpose of use in the second column. The most known networks that are LVQ, counterpropagation, PNN, Hopfield, Boltzmann Machine, ART, and SOM will be explained in detail in the later section. Other networks are explained briefly as follows:

- **Backpropagation Algorithm:** The backpropagation algorithm is a generalized form of least mean squared algorithm. In least mean squared algorithm, weights of connections are changed in order to minimize MSE value between the actual output and produced output by the network. Backpropagation algorithm uses supervised learning (Mehrotra et al., 1997).
- **Delta Bar Delta:** In delta bar delta algorithm, each connection weights have its own learning rate and these learning rates can change with time while training process continue. There are two heuristics in order to determine the proper changes in the learning rates for each weight. If the weight

modification is in the same direction (increasing or decreasing) during a lot of steps, the learning rate of the weight is increased. Otherwise, the learning rate of this weight is decreased (Fausett and Fausett, 1994).

Table 3.4: ANN topologies according to their usage ares and purpose of use

Network Type	**Networks**	**Use of Network**
Prediction	*Back-propagation *Delta bar delta *Extended delta bar delta *Directed Random Search *Higher order Neural Networks *Self-organizing Map into Back-Propagation	Use input values to predict some output (e.g. pick the best stocks in the stock market, predict the weather, identify people with cancer risks)
Classification	*Learning vector quantization *Counter-propagation *Probabilistic Neural Network	Use input values in order to determine classification (e.g. is the input the letter A, is the blob of the video data a plane and what kind of plane is it)
Data association	*Hopfield *Boltzmann Machine *Hamming network *Bidirectional associative memory *Spatio-temporal pattern recognition	Like classification but it also recognizes data that contains errors (e.g. not only identify the characters that were scanned but also identify when the scanner was not working properly)
Data conceptualization	*Adaptive resonance network *Self-organizing map	Analyze the inputs so that grouping relationships can be inferred (e.g. extract from a data base the names of those most likely to buy a particular product)
Data filtering	*Recirculation	Smooth an input signal (e.g. take the noise out of a telephone signal)

- **Extended Delta Bar Delta:** Extended delta bar delta algorithm was developed by Ali Minai and Ron Williams. They improved the delta bar delta algorithm by changing the learning rate increase with the help of exponential decay and also by adding the momentum coefficient back in. Thus, they limitted the learning rate and momentum coefficient (Anderson and McNeill, 1992).
- **Directed Random Search:** The directed random search architecture is different from previous architectures. The previous network architectures use

the calculus based learning rules. On the other hand, the weights in the directed search architecture are modified randomly. This random search paradigm is very fast and easy to use when compared to other network architectures (Anderson and McNeill, 1992).

- **Higher Order Neural Networks:** The Higher Order Networks was developed by Giles and Maxwell (1987). In this type of network, there is no hidden layer, but there are nodes which are called ***'higher order processing units'.*** These nodes are used to compute complex computations (Mehrotra et al., 1997).
- **Self-Organizing Map into Back-Propagation:** This is a hybrid network that uses self-organizing map in order to conceptually distinguish the data before using in the back-propagation manner. The self-organizing map into back-propagation uses unsupervised learning (Anderson and McNeill, 1992).
- **Hamming Network:** The Hamming network was developed by Richard Lippman. This network is also developed form of Hopfield network. The Hamming network performs a classifier relying on the least error for binary input vectors. For defining this error, Hamming distance is used (Anderson and McNeill, 1992).
- **Bidirectional Associative Memory:** The Bidirectional Associative Memory was developed by Kosko in 1988 (Fausett and Fausett, 1994). In this type of network, the number of input neurons are the same with the number of output neurons. The first aim of the Kosko's study was optical processing.
- **Spatio-temporal Pattern Recognition:** This network was the result of the Grossberg's work. The basic aim of this study was to clarify certain cognitive processes in order to recognize changing sequences of events with time (Anderson and McNeill, 1992).
- **Recirculation:** Recirculation networks were developed by Geoffrey Hinton and James McClelland as an alternative for BPNs. These networks use unsupervised learning. Different from back-propagation, in this network, data is processed in one direction (Anderson and McNeill, 1992).

3.2.4.8 Learning Strategies:

For a neural network, there are different learning strategies. The learning system and the learning algorithm that will be used for learning change according to these strategies. Generally, there are three different learning strategies for an ANN. These are;

- Supervised learning
- Unsupervised learning
- Reinforcement learning

3.2.4.8.1 Supervised Learning:

It is also called learning with a teacher. In this strategy, we can think of the teacher who has knowledge about the environment, and knowledge represents the set input-output examples. The teacher gives samples to the system as an input-output set. The task of the system is to map inputs to outputs determined by the teacher. In this way, the relationships between the inputs and outputs of the event are learned. The multilayer perceptron is an example of networks that use this strategy (Haykin and Haykin, 2009).

3.2.4.8.2 Unsupervised Learning:

There is no teacher in this strategy that helps the system to learn. Only input values are shown to the system. The relations between the parameters in the samples are expected to be learned by the system. Generally, this strategy is used for clustering problems (Mehrotra et al., 1997). Adaptive Resonance Theory (ART) networks are examples of systems that use this strategy. The detailed information about ART network will be given in the topic of 'The Most Common Used Models in ANN'.

3.2.4.8.3 Reinforcement Learning:

In this type of strategy, a logical or real value (or signal) is sent to the network, after completion of each input set. This logical or real value indicates whether the output produced is right or wrong. It can be said that reinforcement learning strategy should be more effective than unsupervised learning. Because, the network takes a specific signal for the solution of the problem. The Learning Vector Quantization (LVQ) network is an example of systems that use this strategy (Kriesel, 2005).

3.2.4.9 Learning Laws:

In learning systems such as ANN, no matter which learning strategy is used, the learning process is performed according to some rules. There are different learning rules used in learning systems. Some of them are as follows:

3.2.4.9.1 Hebb's Rule:

Hebb's rule is the oldest learning rule. It forms the basis of other learning rules. This rule was developed by Donald Hebb in 1949. This rule can be explained as follows: If a neuron takes an input from another neuron and both neurons are active (mathematically have the same sign), the weight between these neurons is strengthened (Fausett and Fausett, 1994).

3.2.4.9.2 Hopfield Rule:

This rule is similar to the Hebb's rule, but there is a difference. In this rule, strengthening or attenuation of neuron connections is carried out with the help of the learning coefficient. If the expected output and inputs are both active or passive, the weight values are increased by the learning coefficient. Otherwise, weight values are reduced by the learning coefficient (Anderson and McNeill, 1992).

3.2.4.9.3 Delta Rule:

This rule can also be called the Widrow-Hoff Learning Rule and the Least Mean Square (LMS) Rule. Delta rule is a developed form of the Hebb's rule. This rule is based on the principle of perpetually changing the weight values of the neuron connections to reduce the difference (the delta) between the desired output and the actual output. It is aimed to minimize the mean squared error of the network. It is significant that the input data set should be well randomized in using of Delta rule. If the training set is well ordered, then the network may not converge to the desired accuracy. In this situation, the network is not capable of learning the problem (Anderson and McNeill, 1992).

3.2.4.9.4 The Gradient Descent Rule:

This rule is the same as with the Delta rule. As in the Delta rule, also this rule uses the derivative of the transfer function in order to change the delta error before using in the connection weights. However, an additive proportional constant engaged to the learning rate is used for final changing factor acting upon the weight. This rule is widely used, however, it converges stability point slowly (Anderson and McNeill, 1992)

3.2.4.9.5 Kohonen's Learning Law:

This rule was developed by Teuvo Kohonen. According to this rule, neurons compete each other to change their weight. The neuron which produced the largest output is the winning neuron and the connections weights are changed. This means that the winning neuron becomes stronger against the neighbour neurons. Both the winning neuron and their neghbours are allowed to change the connections weights (Anderson and McNeill, 1992).

3.2.4.10 Number of Hidden Layer and Neuron Number in Each Layer:

The number of layer in neural networks and the number of neurons in each layer affect the performance of the network. As the number of layers and the number of neurons in each layer increase, the level of learning increases and the time for finding the solution of the network is prolonged. However, if the number of layers and the number of neurons in each layer are very small, then the network cannot learn and the resulting output will be less sensitive. There is no definitive rule or method to specify the number of layers and neurons. Usually, the trial and error method is used and the most appropriate network is selected by comparing the performances of the resulting networks.

3.2.4.11 Training and Testing of an ANN:

In an ANN, weights of connections between neurons are initially assigned randomly. The training of an ANN is the process of changing the randomly assigned weight values in order to produce the desired output from the input data set. The aim of learning is to determine the most appropriate weight values for the network and to produce the correct output for the problem. Modifying weight values is done according to the learning rules. When the ANN finds the correct weight values as a result of trials, then this network gains the generalization ability for other samples of the problem. But in some situations, the network attaches to a solution and it is not possible to improve the performance of the network. Therefore, it is necessary to accept an error tolerance value, ε, for the solutions of the problem. At any point below the error tolerance value, it can be said that this network has gained the generalization ability and this solution can be acceptable. Acceptable solutions below tolerance value are called ***local solutions*** (Öztemel, 2012).

The best way to show network learning is to draw an error graph. The graph is plotted throughout the iteration number. After a certain iteration, the error does not decrease further. This means that the network has stopped learning and no better results can be found. If the results are not satisfactory, some properties like network topology, activation function or number of samples can be changed to find new solutions.

After the training process, testing process is started in order to measure whether the network have learn. For testing process, the samples that are unseen by the network through training process are used. During testing, weight values of connections are not changed. Test samples are shown to the network. The network uses the weight values determined at the end of the training process to produce output values for unseen samples. The degree of accuracy of the obtained outputs gives information about the performance of the network. The better results show the better performance for the training of network (Öztemel, 2012).

3.2.4.12 Stopping Criteria of an ANN:

In an ANN, the point of stopping affects the network performance. Therefore, it is necessary to avoid excessive training of the network. Two methods are used to determine the point where network training should be stopped:

- Error value below the specified level
- Completion of the specified number of iteration

3.2.4.13 Performance Indicators:

Performance indicators are used to show how accurate the predictions of an ANN (i.e, learning ability of the networks). There are multiple performance indicators. MSE, RMSE, and MAPE are some of the most commonly used performance indicators and their equations are given in Equations (3.7), (3.8), and (3.9) respectively (Özşahin, 2013):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(t_i - td_i)^2 \qquad (3.7)$$

$$RMSE = \frac{1}{N}\left(\sum_{i=1}^{N}\left[\left|\frac{t_i - td_i}{t_i}\right|\right]\right) \times 100 \qquad (3.8)$$

$$MAPE = 1 - \frac{\sum_{i=1}^{N}(t_i - td_i)^2}{\sum_{i=1}^{N}(t_i - \bar{t})^2} \qquad (3.9)$$

In these equations, t_i, td_i, and N represent the network output, actual output, and a total number of data respectively.

3.2.5 The Most Common Used Models in ANN

In an ANN, the selected topology of ANN, propagation and activation functions of neurons, learning strategy and selected learning law defines the model of this network. Many network models have been developed. The most common used models for an ANN as follow (Öztemel, 2012):

- Perceptrons
- Multilayer perceptrons
- Learning Vector Quantization (LVQ)
- Self-Organizing Map (SOM)
- Adaptive Resonanse Theory (ART)
- Hopfield networks
- Counter-propagation networks
- Neocognitron networks
- Boltzmann machine
- Prababilistic neural network (PNN)
- Elman networks
- Radial Based Network (RBN)

3.2.5.1 Perceptrons:

Perceptron was first developed by Rosenblatt in 1958 in order to pattern classification. Perceptron is based on the principle that one neuron takes more than one input and produces one output. Threshold function value is used in order to calculate output value (Fausett and Fausett, 1994)

3.2.5.2 Multilayer Perceptrons (MLPs):

The multilayer perceptron was developed in order to find solutions of practical problems that cannot be solved by single layer perceptrons (Rosenblat’s perceptron).

XOR problem means *exclusive-OR* problem, these problems cannot be solved by single layer perceptrons. At this point, MLPs were developed in order to find solutions to XOR problems. Before continue with MLPs, XOR problem should be explained.

In order to explain XOR problem, firstly the term of 'linear separability' should be explained. If we have two-dimensional inputs and there is a line which separates all samples between two classes, then a proper perceptron can be derived from this separating line's equation. These classificication problems can be called ***'linearly separable'.*** (Mehrotra et al., 1997)**.** This separating line has the folllowing Equation (3.10):

$$w_0 + w_1 x_1 + w_2 x_2 = 0 \qquad (3.10)$$

In Equation (3.10), w_0, w_1, and w_2 shows the connection weights from inputs *1*, x_1, and x_2, respectively.

On the contrary, if there is no line to linearly separate the samples, than any simple perceptron cannot solve these classification problem. This is the most basic limitation of simple perceptrons. Most- real life classificcation problems are not linearly separable, therefore a simple perceptron cannot be used to solve these problems. Figure 3.5 shows some examples of linearly nonseparable classes. One of them is XOR problem. In Figure 3.5, crosses are samples which are belong to one class, and circles shows samples that belong to other class. As it can be seen from Figure 3.5, there cannot be drawn a line to linearly separate these classes (Mehrotra et al., 1997).

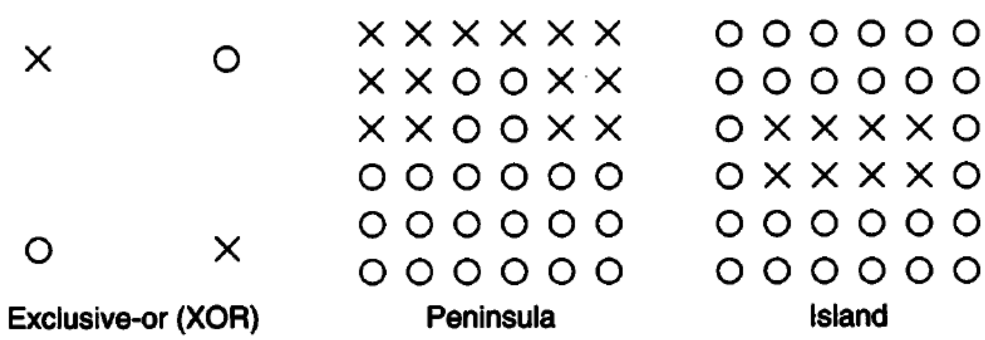

Figure 3.5: Examples of linearly nonseparable classes

The XOR problem can be explained with the following two-input and single-output sets as in Table 3.5.

Table 3.5: Inputs and output sets for XOR problem

Input (x1,x2)	**Output (t)**
(1,1)	-1
(1,-1)	+1
(-1,1)	+1
(-1,-1)	-1

In Table 3.5, for example, for $x_1 = 1$ and $x_2 = 1$, the desired response is (-1). Figure 3.6. shows the graphical illustration of Table 3.5. It can be seen from the Figure 3.6 that there cannot be drawn a line to linearly separate the (-) and (+) outputs. Therefore, MLPs are developed to solve these problems.

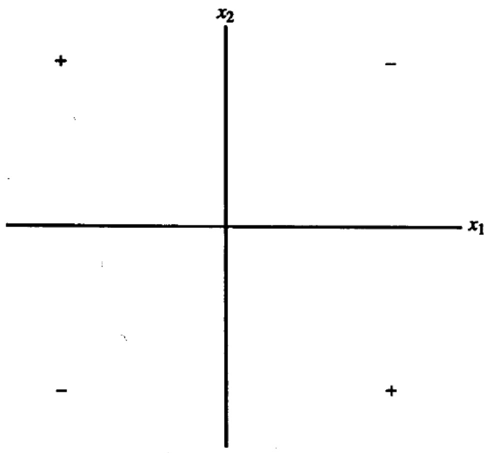

Figure 3.6: Graphical representation of Table 3.5

In the multilayer perceptrons (MLPs) there are one or more layers that are called ***hidden layers*** between the input and output nodes. MLPs uses supervised learning algorithm and also MLPs are a developed form of Delta learning rule. In the developed form of Delta learning rule, there are two stage;

- The first stage is a *forward phase*. In this phase, weight values of connections are fixed and according to input-output set, the output of the network is calculated.

- The second stage is a *backward phase*. In this phase, an error value is calculated by comparing the output produced by the network and the actual output. The aim is to minimize calculated error value. For this reason, this error value is propagated through the value of the weight in the backward direction (i.e., in the direction from output neurons to input neurons).

In Figure 3.7, an architectural representation of MLP with two hidden layers are shown:

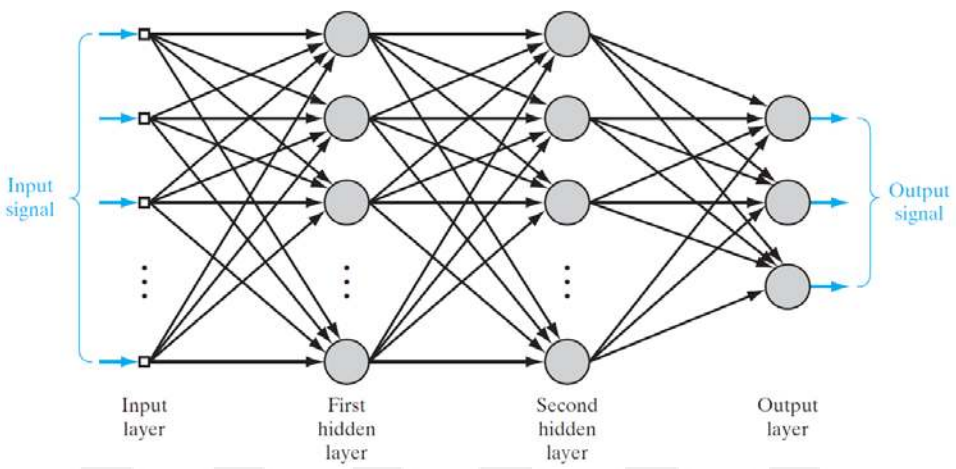

Figure 3.7:Architectural graph of an MLP with two hidden layers (Haykin,2009)

Working steps of an MLP are as follows:

- Data related to the problem are collected in order to find a solution.
- The network topology is determined.
- Learning parameters of the network are specified.
- Inital random weight values are assigned.
- Selected values are shown to the network.
- Forward phase calculations are made in order to find the output of the network.
- Actual output and desired output (the output produced by the network) are compared to calculate the error value.
- Backward calculations are made to modify weight values between connections.

These steps are repeated until the network completes its learning, that is, the difference between the actual and desired output belows a specific value.

3.2.5.3 Learning Vector Quantization (LVQ):

Learning vector quantization (LVQ) was developed by Kohonen in 1984. It is a pattern classification method (Fausett and Fausett, 1994). In LVQ model, there are one input layer, one Kohonen layer, and one output layer. In output layer, the number of neurons are the same as with the number of distinct categories of classes. The Kohonen layer has a number of neurons that are grouped for each of these classes. It can be said that LVQ is used in order to map *n-dimensional* vector into an *m-dimensional* vector (Anderson and McNeill, 1992).

3.2.5.4 Self-Organizing Map (SOM):

SOM network was developed by Kohonen. The main properties of this type of network are that it uses unsupervised learning. The most basic aim of the SOM is to transform an incoming signal pattern of random dimension into discrete map. This discrete map is one or two dimensional. SOM performs this transformation process in a topologically ordered manner (Haykin and Haykin, 2009).

3.2.5.5 Adaptive Resonance Theory (ART):

ART network was developed by Stephen Grossberg and Gail Carpenter (Kriesel, 2005). ART networks use unsupervised learning strategy and it forms the basis for other networks like counterpropagation and bi-directional associative memory networks. ART networks are used for clustering problems (Fausett and Fausett, 1994). The aim of ART networks is to categorize patterns into classes. But the most important feature of ART is to permit the user for controlling the degree of similarity among members in the same cluster by using the predefined constant which is called ‘vigilance parameter’ (Mehrotra et al., 1997).

3.2.5.6 Hopfield Networks:

Hopfield network is a network with one layer and feedback. All of the neurons are both input and output simultaneously. This means that Hopfield networks form a *multiple-loop feedback system*. The number of feedback loops and the neuron numbers is the same. Essentially, the output of each neuron is fedback to the other neurons with a time-delay element (Haykin,2009). An architectural representation of a Hopfield network is shown in Figure 3.8:

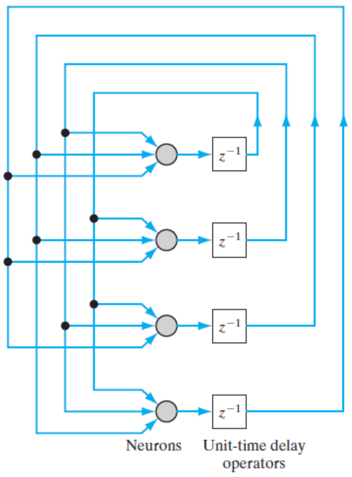

Figure 3.8: Architectural graph of Hopfield network consisting of N=4 neurons

There are two different types of Hopfield networks:

- Discrete Hopfield Network: These networks are used as associative-memory.
- Continuous Hopfield Network: These networks are used for the solution of combinatorial optimization problems.

3.2.5.7 Counterpropogation Networks:

Counterpropagation network is one of the multilayer feedforward neural networks and developed by Hecht-Nielsen (1987). This type of network is a hybrid neural network model that combines Kohonen and Grossberg's algorithm. Counterpropagation networks can be used for function approximation and data compression (Mehrotra et al., 1997).

3.2.5.8 Neocognitron Networks:

Neocognitron networks are developed by Fukushima, Miyake, and Ito. Neocognitron networks are extended form of cognitron networks and it is developed in order to recognize handwritten characters (Kriesel, 2005). Also, this type of network has been used for visual pattern recognition (Mehrotra et al., 1997).

3.2.5.9 Boltzmann Machine:

The Boltzmann machine is a stochastic binary machine that composes of stochastic neurons. For a stochastic neuron, there are two possible states probabilistically. These two states can be represented as 1 for the 'on' state and 0 for the 'off' state. Also in the Boltzmann machine, there are symmetric synaptic connections between neurons. This characteristic distinguishes Boltzmann machine from others (Haykin,2009).

3.2.5.10 Probabilistic Neural Network (PNN):

PNN was developed by Donal Specht. This network type is used for pattern classification pproblems with a statistical approach called Bayesian classifiers (Anderson and McNeill, 1992).

3.2.5.11 Elman Networks:

Elman networks use the same learning rule with MLPs. In an Elman network; there are 4 different neuron type as input, context, hidden, and output neuron. The input and output neurons are in a relationship with the environment. Input neurons take information from the external world and transmit to hidden neurons. Output neurons transmit the output produced by the network to the environment. Context neurons are used to recall previous activity values of hidden neurons. These neurons are responsible for transmitting the values of the previous iteration to the next iteration as an input. In Elman networks, there is one context neuron for each hidden neuron (Kriesel, 2005). Elman networks are successfully applied for modeling linear systems with first degree. These networks are capable of processing time-dependent events and transmit the results of previous times to the next times. Because of these caharacteristics, Elman networks are effectively used for speech recognition and voice recognition problems.

3.2.5.12 Radial Based Network (RBN):

RBN was developed in 1988. Like in MLPs, there are three layers in an RBN as an input layer, hidden and an output layer. But, an RBN uses radial-based activation function and nonlinear clustering analysis between the input and hidden layers. Architectural representation of an RBN is shown in Figure 3.7. In Figure 3.9, there is one single computational unit and Gaussian function is used as the radial-basis function in the hidden layer. In Gaussian function, σ_j shows the *width* measure of the *j*th Gaussian function with center x_j (Haykin and Haykin, 2009).

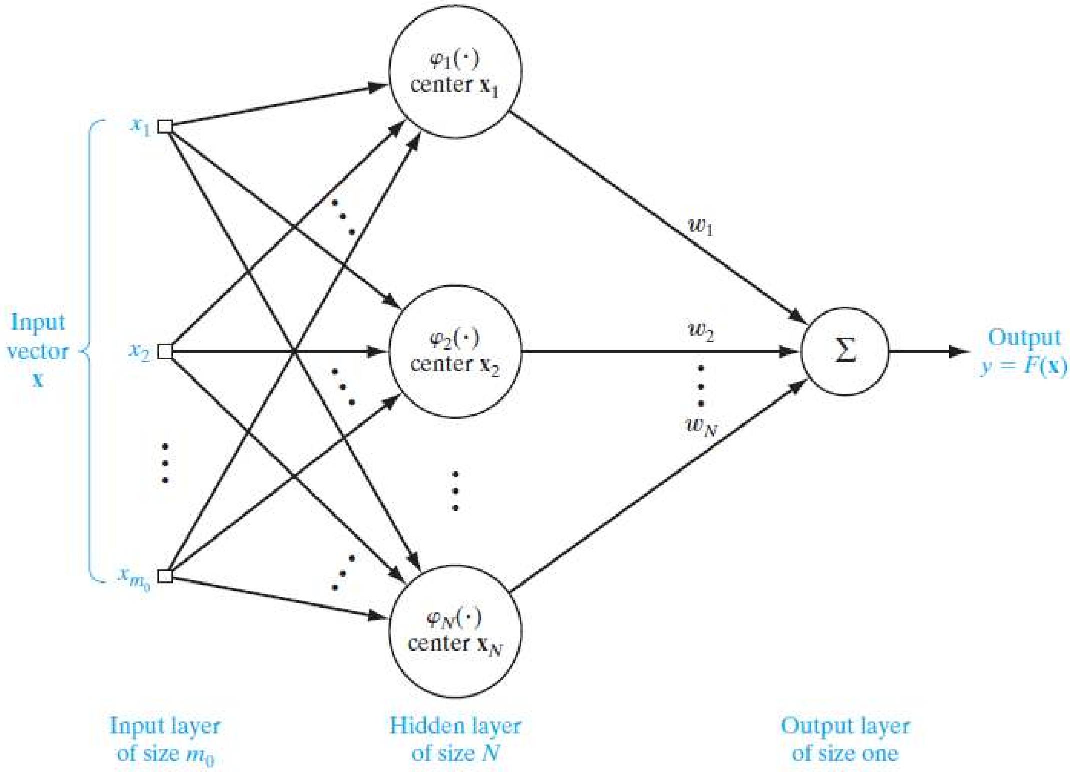

Figure 3.9: Architectural graph of an RBN (Haykin,2009)

In this part of the thesis, AI and ANN have been explained in detail. As stated before, there are many different application areas of ANN. Also, many different ANN models are available. According to usage areas and purpose of use, one of these ANN model can be selected for solution of a problem.

CHAPTER 4. REGRESSION ANALYSIS

At each stage of life, when there is more than one option, one has to decide by choosing one of them. In this situation, decision makers can evaluate the relationship between two or more choices. Regression analysis is one of the statistical techniques for analyzing the relationship among variables. There are 3 types of regression analysis:

- Simple Linear Regression
- Multiple Linear Regression
- Nonlinear Regression

4.1 Simple Linear Regression (SLR):

Simple linear regression analysis is a statistical technique that is used to explain the linear relationship between one independent variable and one dependent variable. A ***dependent variable*** is a variable whose variation needs to be explained. An ***independent variable*** is a variable that clarifies variation of the dependent variable. Simple linear regression model can be shown as Equation (4.1) (Groebner, 2011):

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (4.1)$$

Where;

y = value of the dependent variable

x = value of the independent variable

β_0 = Population's y intercept

β_1 = slope of the population regression line

ε = random error term

Coefficient β_1 shows the average change in y variable according to one unit of change in x variable. This coefficient can take the value of zero, positive, or negative according to the relationship between dependent and independent variables. β_0 coefficient shows the average value of y, in the case of x is zero.

When generating a regression model by using a set of data, there may be too many possible regression lines for these data points. In this situation, a criterion should be set in order to select the best line. This criterion is ‘least squares criterion’. ***Least squares criterion*** is defined as *‘the criterion for determing a regression line that minimizes the sum of the squared prediction errors’*. In the best regression line, the sum of the prediction errors is zero (Groebner,2011). In other words, the aim of simple linear regression is to get the value of coefficients β_0 and β_1 so that the sum of the squared deviations of the actual data is minimized. Simple linear regression graph can be used for this reason (Krajewski et al., 2013).

The simple linear regression model has 4 assumptions (2011) These are as follows:

- **Assumption 1:** ε is called error term and the first assumption is about independency of ε values. Each individual values of ε are statistically independent. For each of *x* level, these error terms show a random sample from the population of possible error term values.
- **Assumption 2:** For any *x*-value, possible error terms must be distributed normally.
- **Assumption 3:** For every *x*-values, the distribution of possible error terms have equal variances.
- **Assumption 4:** For every defined values of the *x*-variable, the means of the *y*-variable can be attached with a straight line. This line is called ***population regression model.***

Before explain the checking ways for these assumptions, the term of ‘residuals’ should be defined. ***Residual*** can be defined as *‘the difference between the actual value of the dependent variable and the value predicted by the regression model’* (2011).

The following methods are used for checking assumptions (Groebner,2011):

- **Checking for Linearity (Assumption 4):** A residuals plot is helpful for determining a linear function is proper for the regression model or not. In residuals plot, residuals are on the vertical axis and the independent variable is on the horizontal axis. If the residuals do not show a systematic pattern, it can be said that the linear model is proper.
- **Checking for Having Equal Variances (Assumption 3):** A plot of residuals is also helpful for detecting the residuals have a constant variance or

not. As explained in checking for linearity assumption, residuals should not show a systematic pattern to ensure having equal variance assumption. If multiple linear regression model is developed, the plot of residuals against the fitted values can be helpful for determining this assumption. In this case, if the residuals scattered like cone-shaped, it can be said that this model do not satisfy having equal variance condition.

- **Checking for Independency of Residuals (Assumption 1):** A residuals plot against time can be helpful for detecting the independency of residuals, if the data used in the regression model are gathered through a time. In order to ensure the condition of independency of residuals, the residuals should be scattered randomly around the mean of zero through time.
- **Checking for Normally Distribution of Residuals (Assumption 2):** There are different ways for checking the normally distribution of residuals. The first method is to generate the frequency histogram of residuals in order to detect general shape is normal or not. The other method for checking this assumption is the plot of standardized residuals. The last method is to form a probability plot of residuals. The more linear the line, the more normally distribution of residuals.

When checking regression assumptions, it should be noted that the normality assumption should be checked at the end. Because, if a regression model cannot meet the requirements of having a constant variance or the independence of the residuals, these situations may make the residuals appear to be not normal.

Regression analysis is used two primary objectives: description and prediction. For example; to describe the factors which affect the customer demand for a product, regression analysis can be used by market researchers. In another example; to predict future demand for a product, regression analysis can be helpful.

A graphical example for a simple linear regression analysis is shown in Figure 4.1. Figure 4.1 demonstrates the simple linear regression line and scatter plot of the data points for simple linear regression analysis between y and x_1. The linear regression line passes the points such that the sum of squared errors is minimized.

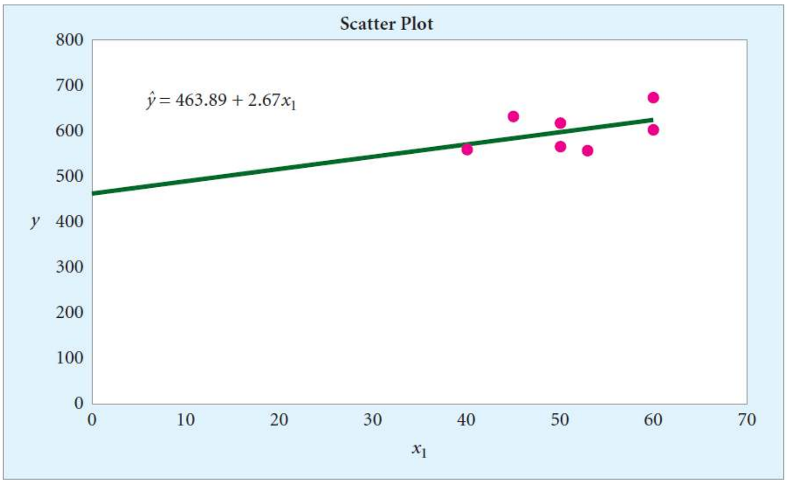

Figure 4.1: Simple Linear Regression Graph

4.2 Multiple Linear Regression (MLR):

Multiple linear regression analysis is a statistical technique that is used to explain the linear relationship between one dependent variable and more than one independent variables. Multiple linear regression model can be shown as Equation (4.2) (Groebner,2011):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots. + \beta_k x_k + \varepsilon \qquad (4.2)$$

Where;

β_0 = Population's regression constant

β_j = Population's regression coefficient for each variable x_j = 1,2,…..,k

k = Number of independent variables

ε = Model error

The primary difference between the simple linear regression and MLR is that simple linear regression forms a straight line in two-dimensional space, but MLR constitutes a hyperplane through multidimensional space. Each coefficient in the MLR model shown in Eqution (4.2) represents distinct slope. An example for MLR hyperplane is shown in Figure 4.2 (2011):

In Figure 4.2, the regression equation constitutes a hyperplane. This hyperplane forms from data such that sum of squared error is minimized. This is identical with least square criterion.

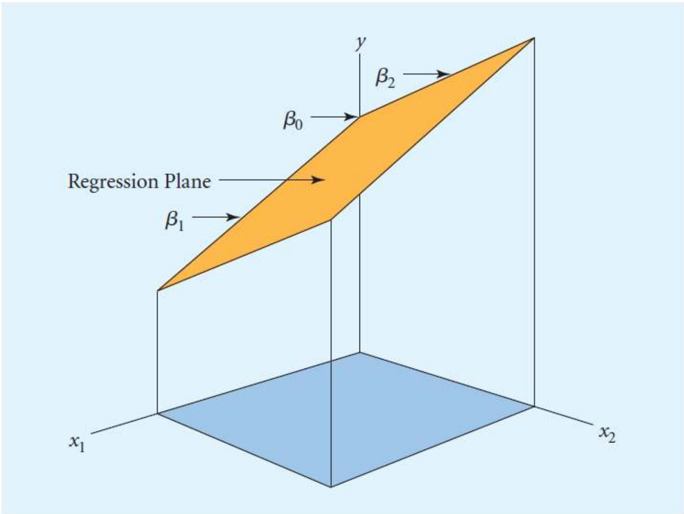

Figure 4.2: MLR hyperplane for population

4.3 Nonlinear Regression:

In many instances, there is linear relationship between two variables. On the other hand, in some situations there is curvilinear relatioship between two variables. For example, electricity demand has increased exponential rate according to the population growth in some regions.

If there is curvilinear relationship between variables, terms that will produce 'curves' should be incorporated into the MLP model. For generating these curves, included terms should have an exponent larger than 1. If a model has such terms, it can be called ***polynomial model.*** In Equation (4.3), most used equation for polynomial model with one independent variable is shown (Groebner,2011):

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \varepsilon \qquad (4.3)$$

Where;

β_0 = Population regression's constant

β_j = Population's regression coefficient for variable x^j ; j = 1,2,3,…,p

p = Order (degree) of the polynomial

ε = Model error

The largest exponent of the independent variable specified the model’s degree. The graph of second-order regression model can be shown in Figure 4.3 (Groebner,2011):

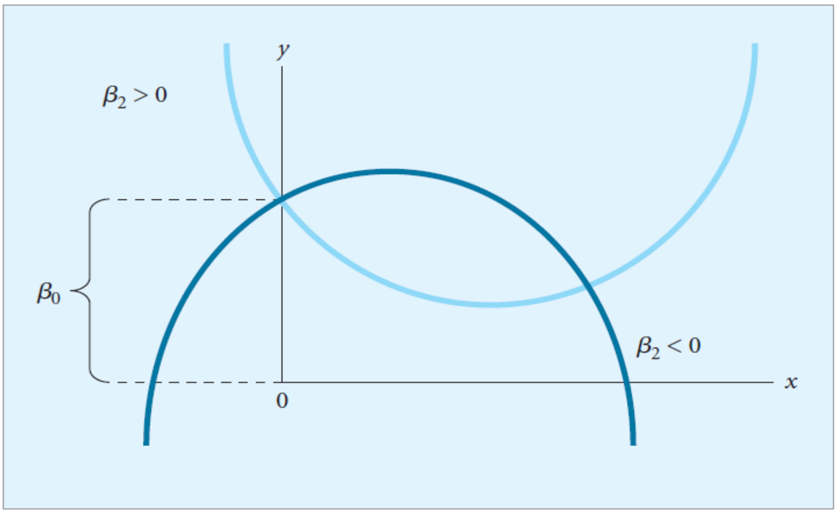

Figure 4.3: Graph of second-order nonlinear regression model

In Figure 4.3, it can be seen that a second-order polynomial model generates a parabola. For $\beta_2 > 0$, the parabola opens upward. On the other hand, for $\beta_2 < 0$, the parabola opens downward.

In this part of thesis, the subject of regression analysis and the most basic terms about this topic has been explained. In order to use one of the regression models for the solution of a problem, firstly assumptions should be checked. If the model is not appropriate, some corrective actions can be taken.

CHAPTER 5. TEXTILE INDUSTRY AND CARPET PRODUCTION

In this part of the thesis, information about the company will be given. The company's three main production areas, raw materials and planning process will be explained.

This company is one of the leading companies in domestic and foreign carpet market and is in the top 500 rankings of Turkey’s 500 largest industrial corporations (ISO500). It operates in a 50.000 m^2 closed area in Gaziantep Organized Industrial Zone. It has a wide range of customers both in Turkey and abroad with its well-assoted products.

There are three main processes in the company:

- Weaving
- Finishing
- Bulked Continuous Flament (BCF) Production

5.1 Weaving:

There are 39 weaving machines in this company. These machines differ according to the number of reed. There are 5 different reed numbers of machines including 16, 32, 42, 48 and 100 reed. The annual production capacity of the weaving section is approximately 10 million m^2 as area rug. The **reed number** of a weaving loom means the number of tie knots of 10 centimeters in width of the carpet produced by that machine. The **weft number** is the number of tie knots per 10 centimeters in length. For example; a carpet with 42*94 model means that this carpet is weaved in a machine with 42 reeds. The reed number of a machine can be changed by a technician. This process is done very rarely. The frequency (density) of a carpet increases as the number of wefts and reeds increase. The production capacity of each machine varies according to the model of the produced carpet at that time. For example; a machine with 16 reeds produces carpets faster and in large quantity. However, the production of a machine with 100 reeds is slower.

There are 3 basic yarn types used in the weaving section:

- Weft yarn

- Warp yarn
- Pile yarn

Jute that is the subject of this thesis is in the weft yarn category.

5.2 Finishing:

The finishing department is one of the production area where finishing processes of the woven carpets are done. In this department, the carpets are first passed through an area called **chemical finish** and subject to heat and adhesive application. The purpose of the adhesive is to prevent the slipping of the carpet and provide better adhesion of the carpet yarns to each other. After this section, carpets are completely processed according to the customer demands. For example, some customers want cardboard pipe inside the carpet in order to prevent destroying of the carpet shape. In another example; some customers want 3 different labels on the back of the carpet, while some customers want a single label. This section works completely according to customer demands. Carpets completed in finishing are sent to the shipping department.

5.3 Bulked Continuous Flament (BCF):

BCF is an abbreviation of the term Bulked Continuous Flament. In this section, polype carpets are produced from polypropylene raw materials. BCF yarn is produced by melting the granulated polypropylene raw material with the help of the extruder and converting it into a yarn. The BCF production steps are described below:

5.3.1 BCF Production Line:

It is the first stage of yarn production. The raw material (polypropylene), masterbatch and spinfinish are mixed in a certain ratio to make a yarn. Single color, double color and three color yarn can be produced.

5.3.2 Bending Production Line:

It is the next stage after the BCF yarn production. . In this section, the yarn is twisted as a single layer, or double layer. With the help of bending process, the yarn is strengthened and the point effect is provided.

5.3.3 Heat-Set:

After bending production line, the yarns are subjected to heat-set process. In this section, the yarn is heat treated to obtain high strength and volume. At the same time, the point effect of the yarn is significantly increased. Yarns are visible on the carpet. In addition, the yarn is made more resistant to the external environment.

5.3.4 Packaging:

After heat-set process, yarns are conveyed to the bagging section. These yarns are put into sack or package according to the customer order and made ready for shipment.

5.4 Raw Materials:

The company has a wide variety of raw materials depending on the production department. However, since the jute raw material is used in the weaving section, 3 different types of yarn used in carpet weaving will be explained.

5.4.1 Weft Yarns:

Jute, cotton weft and chenille are used as weft yarns.

5.4.1.1 Jute:

Jute is the most produced botanical fiber in the world after cotton. 80 % of the world production is covered by India, Pakistan, and Bangladesh. A large part of the jute is used as a sack, cover, rope twine and weft yarn for carpet production.

Different jute types are used in the production of different carpet models. 11 different types of jute are used in this company. These jute types are 10/1 LBS, 11/2 LBS, 12/2 LBS, 13/1 LBS, 13/2 LBS, 16/1 LBS, 16/2 LBS, 18/1 LBS, 20/1 LBS, 20/2 LBS, and 22/1 LBS. LBS is one of the weight units and it can be used for jute raw material. For example; 11/2 in LBS code; The 11 is the number of jute type and 2 is the number of yarn layer. The yarn layer is the number of flament numbers in the jute.

A large amount of jute raw material is used in the orders of foreign customers. Jute raw material is purchased from abroad. Jute is taken in large quantities and is delivered by sea shipment. That is, it takes approximately 6 months to come to the company after the order has been placed. Therefore, accurate estimation of jute order quantities is very important in order not to delay the deadline given to the customer and not to increase the cost of inventory in the excess supply of raw materials. For this reason, it was aimed to solve this problem in this study.

5.4.1.2 Cotton Weft:

Cotton is one of the oldest fibers used in textiles. This plant grows in a more humid and hot climate.

Cotton weft is mostly used in domestic customer orders. It is much more is preferred in Turkey, because it is much more white when compared to jute colour. The company purchases cotton weft from domestic suppliers.

5.4.1.3 Chenille:

Chenille is also one of the types of weft yarn. However, it is not necessary to use chenille on every carpet. Chenille is used to give a pattern to the carpet. Since the raw material of chenille is not used too much, the stock quantity is low and the supply is provided by domestic companies.

5.4.2 Warp Yarn:

The yarn in the longitudinal direction of a woven carpet is called a warp yarn. Warp yarn constitutes the carpet floor with jute.

5.4.3 Pile Yarns:

The most important raw materials that expand the product range in this company are pile yarn types. Types of pile yarn used in the company are as follows:

- Acrylic
- Polype
- Viscose
- Spun thread
- Bamboo

5.4.3.1 Acrylic Yarns:

Acrylic yarns are short fiber yarns. Most of the acrylic yarns are used for orders of domestic customers. The acrylic yarn is ordered to domestic suppliers.

5.4.3.2 Polype Yarns:

BCF yarns are manufactured at the company's own BCF production site. Therefore, yarn requirement is determined and production planning of the BCF section is made according to the incoming order. Generally, BCF yarn stock is not available, because it is manufactured according to the order.

5.4.3.3 Viscose:

Viscose yarn is one of the raw materials supplied from abroad. It is usually supplied from India by sea shipment. Some domestic companies also produce viscose yarns, but the domestic purchase price is higher than abroad. Therefore, the stock is kept for the viscose yarn.

5.4.3.4 Spun Thread:

Spun thread brighter than other yarn types. For this reason, it is mostly used to give an aesthetic appearance to the carpet. This yarn type is obtained from domestic suppliers.

5.4.3.5 Bamboo:

Bamboo yarn is a softer and thin than other yarns. Although it is thin, it is resistant to wear. Bamboo carpets are sold to domestic and foreign customers. This yarn is purchased from foreign suppliers.

As the most basic raw material used in the carpet production is yarn, the problems encountered in the supply phase of this raw material cause lateness in production and sometimes change the whole production schedule. In order to avoid these problems, some precautions are taken. If the yarn is supplied domestically, more than one supplier company is involved and agreements are made. Thus, raw materials that cannot be supplied from one manufacturer are immediately ordered from other suppliers. If the yarn is supplied from abroad; the excess stock for this yarn is kept available. However, inventory costs may increase in this case. In cases where the raw material received from abroad is not sufficient, the procurement stage from domestic suppliers starts. Domestic purchase price is higher than the purchase price from abroad. In this situation, the domestic purchase cost increases. For all of these reasons, it is much more important that the raw materials which are procured from abroad should be supplied at the right time and in sufficient quantities.

5.5 Planning:

In this section, finally the planning procedure of the company is explained.

The planning phase begins with the arrival of the order from the customers. First, incoming orders are assigned to the respective machines according to the reed numbers of machines. The criteria to be considered during this assignment are as follows:

- Is the raw material required for the order available and sufficient?
- If the raw material is not available and is supplied from another supplier, is the order quantity and due date determined?
- If the BCF yarn produced in the company is inadequate, is this yarn included in the BCF production plan to provide the required quantity?
- Is the workload distribution on the machines suitable for the due date requested by the customer?
- When setting the machine to which the order is to be produced, is the setup between different orders set to minimum time?

All of these stages are reviewed one by one and the plan created by the planning chief is approved by the planning manager. The customer is informed about the deadline of his order. If a problem occurs at any stage of the specified steps, the entire plan is reviewed and if necessary the order is assigned to another machine.

Figure 5.1 shows the general workflow diagram of the company. In Figure 5.1, information flows and feedback points between departments are shown.

In this part of the thesis, information about company, carpet production processes, and raw materials used in carpet production are given. In the next section, the problem of jute consumption and the computations of ANN and MLR methods will be explained.

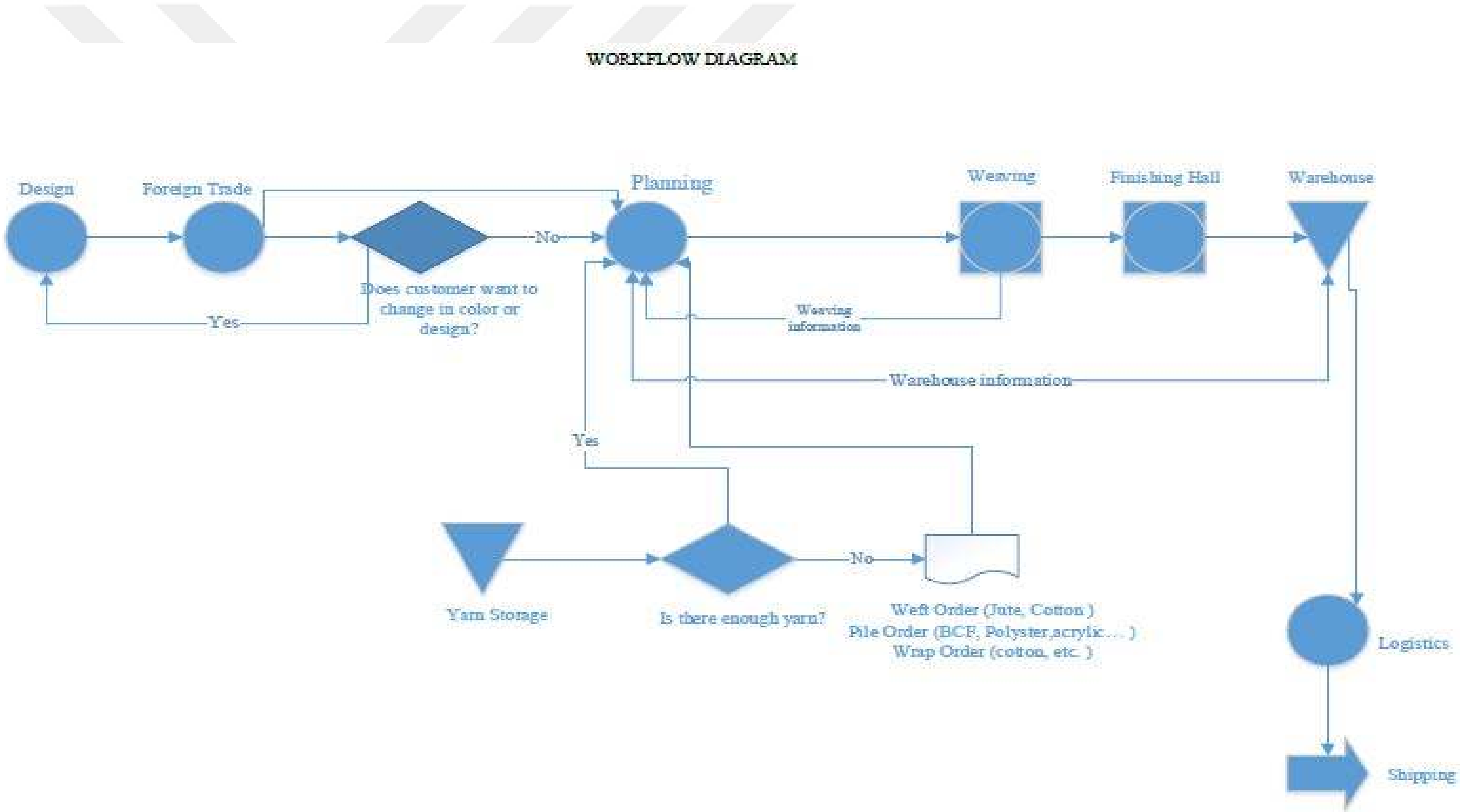

Figure 5.1:Workflow diagram of the company

CHAPTER 6. COMPUTATIONAL EXPERIMENT

In this thesis, the usage amount of jute yarn for the future period has been estimated. For this purpose, 246 monthly based data points (actual data) that include 6 independent variables and 1 dependent variable were used for 11 different types of jute. These data belong to years between 2015 and 2017 (Gülyeşil et al., 2018).

In this section, firstly, data analysis is explained in two part that include data collection and data normalization. Then forecasting with ANN method is explained. After that, steps of MLR method are expressed. Finally, comparisons of two methods is represented.

6.1 Data Analysis:

6.1.1 Data Collection:

As stated before, the subject of this study is to predict the consumption amount of 11 different jute types for the future period. For this purpose, 6 input variable were selected because of their relation with the total jute consumption (Gülyeşil et al., 2018). These variables and their statements are as given in Table 6.1.

In Table 6.1, the variables of 'jute type', 'production %' and ' average consumption in 1 m^2' were taken from weaving department. 'Purchasing price' variable was taken from purchase department. 'Amount of stock' variable was taken from planning department. Finally 'exchange rate' variable was taken from the official web address of the Central Bank of the Republic of Turkey.

After the data were taken from the related departments, some calculations were made with the variable of 'purchasing price' and 'average consumption in 1 m^2' and the obtained data were used for prediction purposes. These calculations are explained in follow:

Table 6.1: Variables definition table

Variable Symbol	**Variable Name**	**Statement**
X_1	Jute Type	There are 11 types of jute to be used in different carpet models. The main differences between different types of jute are the thickness and the number of yarn layers.
X_2 ($)	Purchasing Price	The purchase price of jute of different quality (different thickness and number of layers) is different. In addition, a jute type of the same quality can be obtained from different suppliers. In this case, the weighted average of purchase prices based on order quantity was used.
X_3	Exchange Rate	As the purchasing prices are in $ terms, the exchange rate was used. This data was taken from the official web address of the Central Bank of the Republic of Turkey.
X_4 (tons)	Amount of Stock	The monthly stock of each type of jute varies.
X_5	Production %	Different types of jute are used in the production of different carpet models. The produced m^2 of these carpets are kept separately in the ERP system used by the company. The 'production %' shows the percentage of the stated jute in the total amount.
X_6 (kg)	Average Consumption in 1 m^2	In different carpet models, even if the same type of jute is used, the amount of consumption can be different. While the average consumption was taken in this data type, weighted usage was calculated on the basis of m^2 produced.
Y (tons)	Amount of Consumption on the Basis of Jute Type	The amount of jute consumption based on forecasted input variables values for the future period

6.1.1.1 Calculations for 'Purchasing Price' Variable:

The purchase price of jute of different quality (different thickness and number of layers) is different. In addition, a jute type of the same quality can be obtained from different suppliers. In this case, the weighted average of purchase prices based on order quantity was used. For example; assume that the jute type 3 was supplied from two different companies with different prices and quantities as in the Table 6.2:

Table 6.2: Examples of data for calculations of 'purchasing price' variable

Jute Type	Suppliers	Purchasing Price ($)	Purchasing Amount (tons)
3	A	1.55	26
3	B	1.45	52

According to Table 6.2, weighted purchasing price was calculated with Equation (6.1):

$$\frac{\sum(purchasing\ price * purchasing\ amount)}{\sum amount} \qquad (6.1)$$

The weighted purchase price result of Table 11 is;

$$\frac{\sum(1.55 * 26) + (1.45 * 52)}{\sum(26 + 52)} = \$1.48$$

These calculations were done for each monthly data.

If the jute type specified in any month has not been purchased, the ‘purchase price’ was indicated as zero (0) for this jute type. Also if the jute type specified in any month has not been purchased, the ‘purchasing amount’ was indicated as zero (0) for this jute type.

6.1.1.2 Calculations for ‘Average Consumption in 1 m^2’ Variable:

In different carpet models, even if the same type of jute is used, the amount of consumption can be different. While the average consumption was taken in this data type, weighted usage was calculated on the basis of m^2 produced. For example; assume that the jute type 1 was used in three different carpet models with different consumption amount. Also, the produced m^2 of these carpet models varied as in Table 6.3:

Table 6.3: Examples of data for calculations of 'average consumption in 1 m2' variable

Jute Type	Carpet Model	Average consumption for 1 m^2	Produced m^2
1	A1	0.524813	1200
1	A2	0.471622	900
1	A3	0.601521	1300

According to Table 6.3, weighted usage was calculated with Equation (6.2):

$$\frac{\sum(average\ consumption * produced\ m2)}{\sum(produced\ m2)} \qquad (6.2)$$

The weighted consumption result of Table 12:

$$\frac{\sum(0.524813 * 1200) + (0.471622 * 900) + (0.601521 * 1300)}{\sum(1200 + 900 + 1300)} = 0.540063\ kg$$

These calculations were done for each monthly data and each jute type.

A sample data that has been used is given in Table 6.4. In this table, data that has belong to different months has been shown. Because of this reason, 'exchange rate' for each line is different. For example, in Table 6.4, in the first row; the input variables belong to 2015 March and the output variable belongs to 2015 April. In this first row, explanations of each column are as follow:

- 'Purchasing price' indicates that the purchase price for jute type 1 in 2015 March.
- 'Amount of stock' indicates that the total stock amount of that time in 2015 March.
- 'Production %' shows the total production of all carpet models which the jute type 1 was used as a raw material in total production in 2015 March.
- 'Average consumption in 1 m^2' shows the average consumption of jute type 1 based on produced m^2 of carpet models which used jute type 1 as raw material.
- 'Purchasing amount' shows the purchasing quantity of jute type 1 in 2015 April.

Table 6.4: The sample of the raw data used in application

Jute Type	**Purchasing Price ($)**	**Exchange Rate**	**Amount of Stock (tons)**	**Production %**	**Average Consumption in 1 m^2**	**Purchasing Amount (tons)**
1	1.27	2.58	21.787	4.317	0.52429	26
3	1.176	2.65	40.339	32.783	0.77986	130
4	1.03	3.00	20.945	5.982	0.67708	26
8	1.125	3.26	94.144	5.884	0.71875	104
11	1.11	3.48	45.457	2.177	0.72569	78

6.1.2 Data Normalization:

Since the types of data used ($, ton, kg) are different, they must be subjected to normalization before being used in the software. In addition, normalization of the data makes the ANN method much more meaningful to be used in the solution of nonlinear problems. Min_Max normalization method was used to normalize the data. In this method, Min is the smallest value in data, Max represents the highest value in the data. With the Min_Max method, the data is reduced to a range of [0,1]. The formula used for normalization is shown in Equation (6.3):

$$X' = \frac{Xi - Xmin}{Xmax - Xmin} \qquad (6.3)$$

An example of raw data is shown in Table 6.4. Table 6.5 shows the values after the normalization of the data in Table 6.4. For example; in Table 6.4, the amount of stock for jute type 4 is 20.945 tons. The max value in this data is 303.956 tons. The min value in this data is 0 that means there was no stock for one of the jute type in any month. By using the Equation (6.3), the normalized value of this 20.945 tons is calculated as follow;

$$X' = \frac{20.945 - 0}{303.956 - 0} = 0.068907$$

In Table 6.5, the first input variable (jute type) was not normalized because this is categorical variable and it does not need normalization process.

Table 6.5: Normalized values of the data shown in Table 6.4

Jute Type	Purchasing Price ($)	Exchange Rate	Amount of Stock (tons)	Production %	Average Consumption in 1 m^2	Purchasing Amount (tons)
1	0.875862	0.100508	0.071678	0.113772	0.52429	0.2
3	0.811034	0.150757	0.132713	0.863983	0.77986	1
4	0.710344	0.431149	0.068907	0.157653	0.67708	0.2
8	0.775862	0.634719	0.309729	0.155070	0.71875	0.8
11	0.765517	0.807794	0.149551	0.057374	0.72569	0.6

6.2 Determination of Performance Indicators:

In this thesis, two methods were used which are ANN and MLR. In order to compare the performance of these methods, three performance indicators were selected as MSE, R^2 and adjusted R^2.

MSE means Mean Squared Error and this value shows the difference between the actual data and computed data from the model. Small MSE value indicates that the constructed model and actual data has overall good fit.

R^2 demonstrates the value of the coefficient of determination. This term can be defined as *'the portion of the total variation in the dependent variable that is explained by its relationship with the independent variable'* (Groebner, 2011). R^2 can take a value on the range of [0,1]. As R^2 goes to the 1, this indicates a perfect linear relationship between variables.

Adjusted R^2 is a percentage that explaines the variation in the dependent variable in MLR analysis. But this value also considers the relationship between the sample size and the number of independent variables in the MLR model. Adjusted R^2 value can be computed with Equation (6.4) (Groebner, 2011):

$$R - sq(adj) = R_A^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k-1}\right) \qquad (6.4)$$

Where;

n = sample size

k = number of independent variables

R^2 = coefficient of determination

6.3 Forecasting with Artificial Neural Network Method:

In this study, the ANN method was used for the estimation of the amount of raw material 'jute'. Feed forward and back propagation multilayer ANN model was preferred for the solution of this problem. In a feed forward neural network, the direction is towards from input layer to output layer. Most widely used ANNs in forecasting problems are multilayer perceptron (MLPs) (Hamzacebi, 2008). MATLAB software was used for the training and testing of the networks created for this purpose (Gülyeşil et al., 2018).

As the first stage of the solution, data were partitioned as training, validation, and testing sets. The corresponding share in total is as follow. These are default values of Neural Network Toolbox of MATLAB software partitions:

- Training = 70 %
- Validation = 15 %
- Testing = 15 %

The number of neurons, activation function and training algorithm parameters in the hidden layer constituting the structure of the ANN affect the performance of the resulting network. For this reason, models with different network topologies and learning parameters were established and experiments were made and the obtained results were compared by using MSE values of each network. The most suitable network for the problem has been found by trial and error.

6.3.1 Determination of Neuron Number in Hidden Layer:

In order to measure the effect of the number of neurons in a hidden layer on the performance of the neural network, two networks were used, and MSE values of them were compared. The first network was called as *net_1* with 5 neurons, and the second network was called as *net_2* with 10 neurons. Properties of these networks are given in Table 6.6.

Table 6.6: Network characteristics of net_1 and net_2

Network No	**Net_1**	**Net_2**
Network Type	Feed forward–back propagation	Feed forward–back propagation
Training Function	TRAINLM	TRAINLM
Adap. Learning Function	LEARNGDM	LEARNGDM
Number of Hidden Layers	1	1
Number of Neurons	5	10
Transfer Function	TANSIG	TANSIG
Epoch	500	500

In Table 6.6, TRAINLM was selected as training function. TRAINLM is a network training function that updates bias and weight values according to Levenberg-Marquardt optimization. It is often the fastest backpropagation training algorithm and is the most used supervised algorithm, but it use more memory when compared to other algorithms. LEARNGDM was selected as learning function. LEARNGDM is the gradient descent with momentum weight and bias learning function. Learning occurs according to LEARNGDM's learning parameters which are learning rate and momentum constant. There was one hidden layer for both nets. In net_1, there was 5 neuron in hidden layer and in net_2, there was 10 neuron in hidden layer. TANSIG was selected as transfer function. TANSIG is a hyperbolic tangent sigmoid transfer function. Transfer functions calculate a layer's output from its net input. Finally, epoch was 500. An epoch is a measure of the number of times all of the training vectors are used once to update the weights. For batch training, all of the training samples pass through the learning algorithm simultaneously in one epoch before weights are updated.

Also, network structures of net_1 and net_2 networks are as shown in Figure 6.1 and Figure 6.2 respectively:

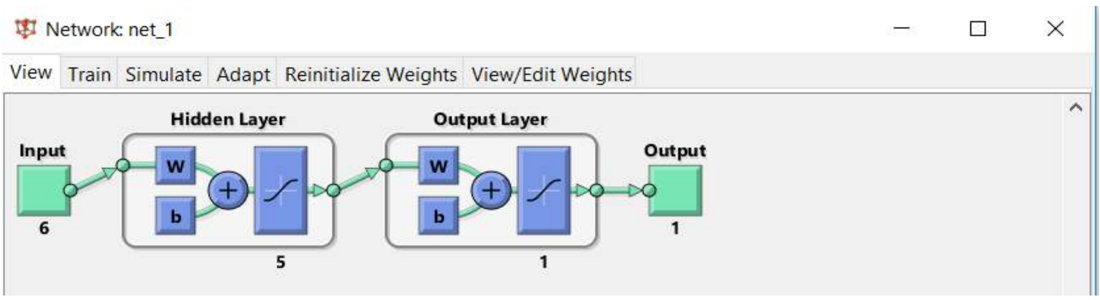

Figure 6.1: Network structure of net_1 (MATLAB software)

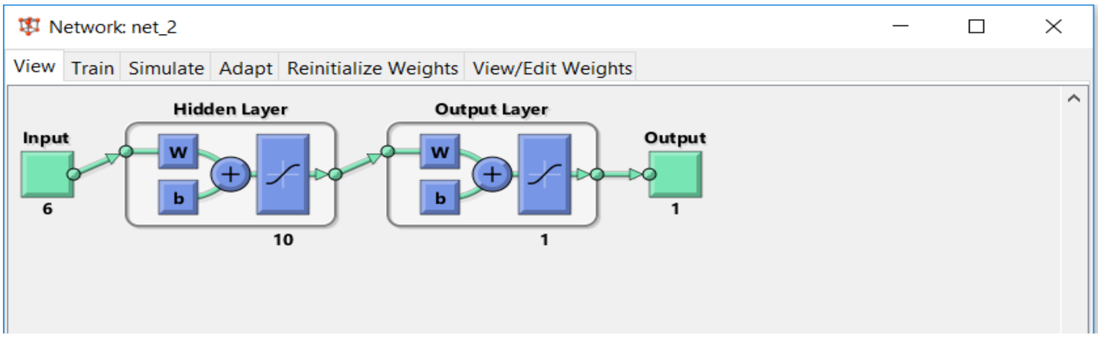

Figure 6.2: Network structure of net_2 (MATLAB software)

Net_1 and net_2 networks were run six times. After six running, the MSE values did not vary significantly. The obtained MSE values are shown in Table 6.7:

Table 6.7: Comparative MSE values of net_1 and net_2 networks

Training No	**MSE Values for Net_1**	**MSE Values for Net_2**
Train 1	0.442	0.0688
Train 2	0.0232	0.0208
Train 3	0.0133	0.0194
Train 4	0.0140	0.0190
Train 5	0.0116 (*)	0.0183
Train 6	0.0140	0.0183

As it can be seen from the Table 6.7, the smallest MSE value obtained with the net_1 network is 0.0116 (5.training result) (with (*)). On the other hand, the smallest MSE value obtained with the net_2 network is 0.0183. Because the small MSE value shows that the performance of the network is better, the net_1 network with 5 neurons is more suitable for solving this problem.

6.3.2 Determination of Activation (Transfer) Function:

In order to measure the effect of the activation function on the performance of the network, a new network was formed called *net_3* with LOGSIG activation function. MSE value of this network was compared with the net_1 network that was formed in the previous section. Network properties of net_1 and net_3 are shown in Table 6.8:

In Table 6.8, all variables were the same with Table 6.6 except the transfer function of net_3. In net_3, LOGSIG was selected as a transfer function. LOGSIG means logaritmic sigmoid transfer function. Transfer functions calculate a layer's output from its net input.

Table 6.8: Network characteristics of net_1 and net_3

Network No	**Net_1**	**Net_3**
Network Type	Feed forward–back propagation	Feed forward–back propagation
Training Function	TRAINLM	TRAINLM
Adap. Learning Function	LEARNGDM	LEARNGDM
Number of Hidden Layers	1	1
Number of Neurons	5	5
Transfer Function	TANSIG	LOGSIG
Epoch	500	500

Net_1 and net_3 networks were run six times. The obtained MSE values are shown in Table 6.9:

Table 6. 9: Comparative MSE values of net_1 and net_3

Training No	**MSE Values for Net_1**	**MSE Values for Net_3**
Train 1	0.442	0.139
Train 2	0.0232	0.0150
Train 3	0.0133	0.0135
Train 4	0.0140	0.0120
Train 5	0.0116 (**)	0.0145
Train 6	0.0140	0.0156

As it can be seen from the Table 6.9, the smallest MSE value obtained with the net_1 network is 0.0116 (5.training result) (with (**)). On the other hand, the smallest MSE value obtained with the net_3 network is 0.0120. Because the small MSE value shows that the performance of the network is better, the net_1 network with TANSIG activation function is more suitable for solving this problem.

6.3.3 Determination of Training Function:

For evaluating the effect of selected training function on the performance of the network, *net_1* network with 5 hidden neurons and TRAINLM (Levenberg_Marquardt training algorithm) was compared with a *net_4* network with 5 neurons in the hidden layer and TRAINBR (Bayesian Regularization training algorithm). Network characteristics of net_1 and net_4 are shown in Table 6.10:

Table 6.10: Network characteristics of net_1 and net_4

Network No	Net_1	Net_4
Network Type	Feed forward–back propagation	Feed forward–back propagation
Training Function	TRAINLM	TRAINBR
Adap. Learning Function	LEARNGDM	LEARNGDM
Number of Hidden Layers	1	1
Number of Neurons	5	5
Transfer Function	TANSIG	TANSIG
Epoch	500	500

In Table 6.10, the different factor was the training function of net_4. TRAINBR was selected as a training function for net_4. TRAINBR is a network training function that updates the weight and bias values according to Levenberg-Marquardt optimization. It minimizes a combination of squared errors and weights, and then determines the correct combination so as to produce a network that generalizes well. The process is called **Bayesian regularization.**

Net_1 and net_4 networks were run six times. The obtained MSE values are shown in Table 6.11.

As it can be understood from the Table 6.11, each network were run six times, and the smallest MSE value of net_1 is 0.0116 (with (***)) and 0,0178 for net_4. As a result, the net_1 network is the most suitable ANN design among the alternatives for the solution of this problem.

Table 6.11: Comparative MSE values of net_1 and net_4

Training Number	MSE Values for Net_1	MSE Values for Net_4
Train 1	0.442	0.630
Train 2	0.0232	0.0249
Train 3	0.0133	0.0232
Train 4	0.0140	0.0192
Train 5	0.0116 (***)	0.0196
Train 6	0.0140	0.0178

6.3.4 Determination of the Best Feasible ANN:

The most appropriate ANN design for the defined problem in this thesis has been found as net_1. One of the disadvantages of an ANN is that there is no certain rule or method for determining important parameters that effect the performance of the network such as the number of neurons in hidden layer, activation, and training function. Therefore, these parameters are defined with trial and error (Öztemel, 2012). For this reason, as a result of trial and error, net_1 with the smallest MSE value is selected. After deciding the use of net_1, the training process was continued and finally the solution was completed with the most appropriate MSE and R values.

MSE values for each data groups (training, validation, and testing) of net_1 is shown in Figure 6.3. In Figure 6.3, MSE value reach to steady-state position at 500 epochs for training, validation, and testing steps. The best validation performance 0.0170 and reached at epoch 0.

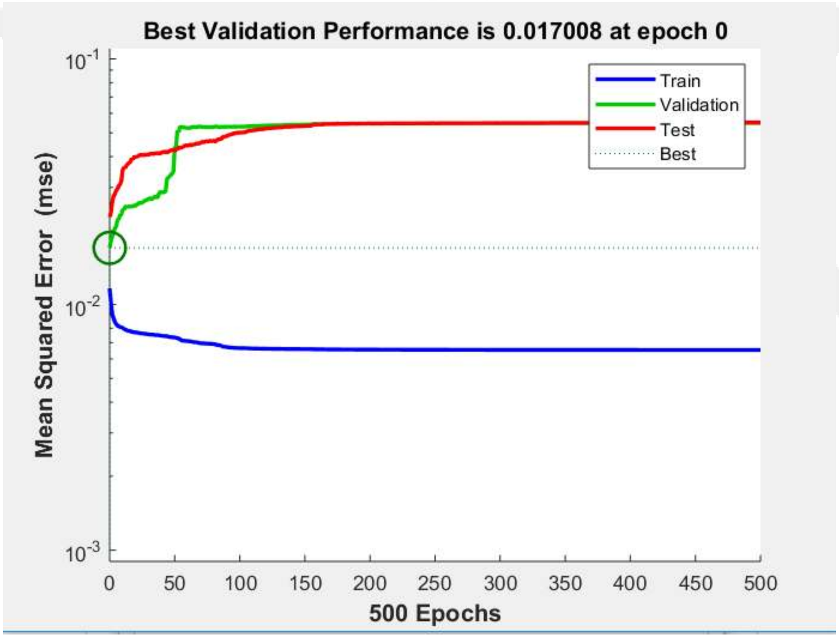

Figure 6.3: MSE graph of net_1 (MATLAB software)

Before explaining regression line graphs, the definition of R should be explained. R means correlation coefficient and can be defined as *'the direction and strenght of the relationship between the independent variable and the dependent variable.'*(Krajewski et al., 2013). The correlation coefficient can take any value between -1 and +1. As the R value is closer to +1, the fitting of the data points with the regression line increases.

Figure 6.4 shows the regression line graphs for each data groups. In Figure 6.4, each circle shows the data points. As each data points converges to the regression line, the least square error value decreases. In addition to this, according to the regression graph, we can see the harmony between target and data that is used. Also, the R value that is bigger than 0.90 shows that the prediction has a high accuracy rate.

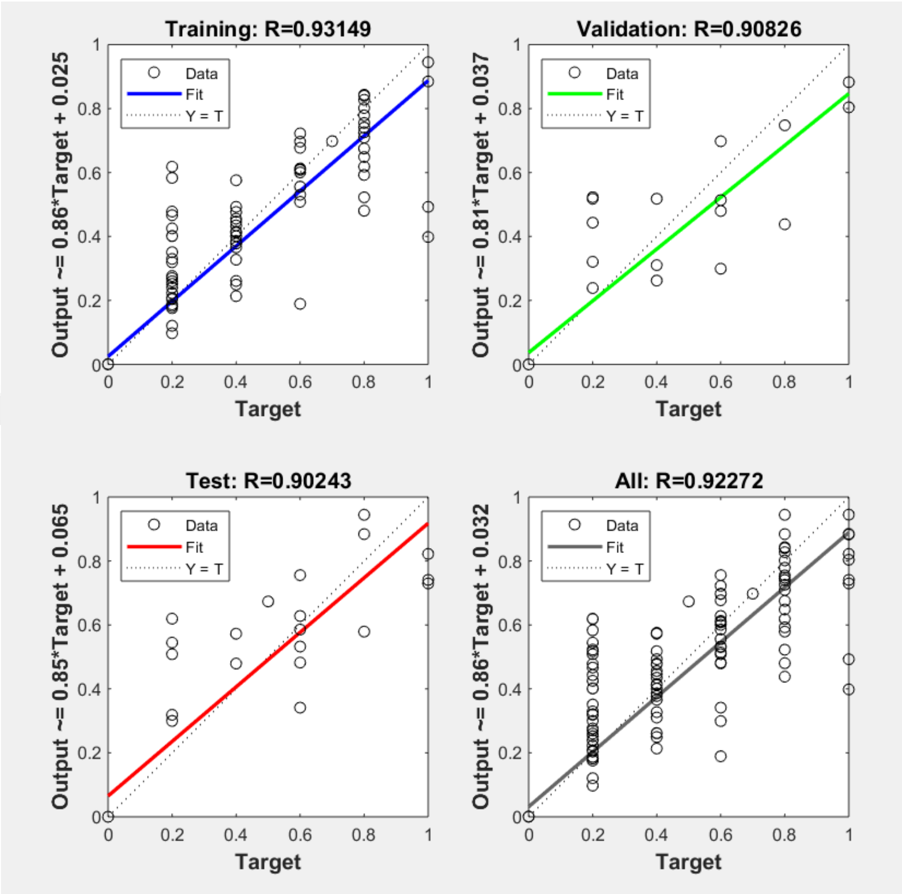

Figure 6.4: R graphs for each data group of net_1 (MATLAB software)

MLR method is used in order to evaluate the performance of the ANN method. The performance indicators used to compare two methods are MSE, R^2, and adjusted R^2 values. As a result; MSE, R^2 and adjusted R^2 values for net_1 are given in Table 6.12:

Table 6.12: Values of performance indicators of net_1

Performance indicator	Result of Net_1
MSE	0.0116
R^2	0.8464
Adjusted R^2	0.8712

In a feedforward and backpropagation multilayer neural network, the weights of each connection are tried to improve the performance of the network by updating weights. The current weight values at the time the net_1 network reach the minimum MSE values among our trials are shown in Table 6.13:

Table 6.13: Weight values of each connection of net_1 at satisfied MSE value

	N1	**N2**	**N3**	**N4**	**N5**
I1	7.64	-0.42	0.94	5.86	0.43
I2	17.31	1.16	3.52	15.58	-3.55
I3	-6.78	-1.14	-0.76	-2.84	1.47
I4	1.35	-3.68	0.54	2.81	4.43
I5	0.31	-5.04	0.89	12.21	6.01
I6	7.34	5.07	2.78	-0.65	-6.74
B_H	-15.50	-8.11	-3.54	8.07	11.30
W_O	-0.68	5.58	1.51	6.58	5.24
B_O	-5.99	-5.99	-5.99	-5.99	-5.99

In Table 6.13, the notations which are I_x, N_x, B_H, W_O, B_O are explained as follows:

I_x = the number of input variable (x = 1,2,3,4,5,6)

N_j = the j^{th} neuron in the hidden layer (j = 1,2,3,4,5)

B_{Hj} = bias value for j^{th} neuron in the hidden layer (j = 1,2,3,4,5)

W_{jO} = weight value from j^{th} neuron in the hidden layer to output neuron (j = 1,2,3,4,5)

B_O = bias value for output neuron

6.4 Forecasting with Multilinear Regression Method:

In this part of thesis, MLR method for the jute consumption prediction problem was explained. Firstly, regression assumptions were checked. After all conditions were satisfied, MLR method was applied to this problem.

In forecasting with multiple linear regression method (MLR), Minitab software was used. The same 246 data were used for MLP study and results of performance indicators were calculated.

6.4.1 Checking for Regression Assumptions:

In order to apply regression analysis to any problem, firstly regression assumptions should be checked. These assumptions are as follows:

The first assumption that should be checked is that linear relationship between independent and dependent variables. In order to check this assumption, correlation matrix and t-test can be used. Table 6.14 shows the correlation matrix for the variables of the jute consumption problem.

Table 6.14: Correlation matrix for variables of jute consumption problem

	Jute Type	**Purchasing Price**	**Exchange Rate**	**Amount of Stock**	**Prod. %**	**Avg. Consump. in 1 m^2**
Purchasing Price	-0.269					
Exchange Rate	0.143	0.167				
Amount of Stock	-0.182	0.175	0.127			
Prod. %	-0.154	0.253	-0.170	0.146		
Avg. Consump. in 1 m2	0.160	0.199	-0.171	0.195	0.348	
Purchasing Amount	-0.164	0.803	0.151	0.215	0.404	0.192

After finding correlation matrix, by applying Equation (6.5) the t values of these variables are found and t-test is applied (Groebner, 2011):

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \qquad (6.5)$$

Where;

r = correlation coefficient

n = number of sample size

By applying Equation (6.5), the calculated t values of Table 6.14 are shown in Table 6.15:

Table 6.15: Calculated t values of Table 6.14

	Jute Type	**Purchasing Price**	**Exchange Rate**	**Amount of Stock**	**Prod. %**	**Avg. Consump. in 1 m^2**
Purchasing Price	-4.362					
Exchange Rate	2.256	2.645				
Amount of Stock	-2.891	2.776	7.874			
Prod. %	-2.434	4.084	-2.694	2.305		
Avg. Consump. in 1 m2	2.531	3.171	-2.711	3.105	5.798	
Purchasing Amount	-2.596	21.046	2.386	3.438	6.898	3.055

Other necessary variables to perform t-test are given as follows:

Criticial $\alpha = 0.05$

Degrees of freedom = n – 2 = 244

The null and alternative hypotheses are given below:

$H_0 : \rho = 0$

$H_A : \rho \neq 0$

For α is 0.05, the t-table value is approximately ± 1.96. For t-values which are greater than 1.96 or less than 1.96, this coefficient is defined as significant and there is linear relationship between these variables. All t-values that are shown in Table 6.15 are greater than 1.96 or less than 1.96, therefore H_0 should be rejected and it can be said that the linear relationship assumption is satisfied.

The other assumption that should be satisfied is the multicollinearity constraint. In order to apply the MLR method for any problem, the independent variables in the model should satisfy multicollinearity constraint. Multicollinarity can be defined as; '*Multicollinearity means a high correlation between two independent variables such that the two variables contribute redundant information to the model.'* (2011). When a high correlation occurs between independent variables, this situation can affect the regression results in the negative direction (Groebner, 2011).

One of the methods to measure multicollinearity is known as the ***variance inflation factor (VIF).*** VIF value shows the increment amount of variation of a predicted regression coefficient if there is a relationship between independent variables.

When multicollinearity increases, the VIF value increases. The formulation of VIF value is given in Equation (6.6) (Groebner, 2011):

$$VIF = \frac{1}{\left(1 - R_j^2\right)} \qquad (6.6)$$

Where;

R_j^2 = Coefficient of determination when the *j*th independent variable is regressed against the remaining *k-1* independent variables.

Generally, it can be said that if VIF < 5 for a certain independent variable, there is no multicollinearity for this variable (Groebner, 2011).

In this study, 246 data points were used in the Minitab software and obtained VIF results are given in Table 6.16.

Table 6.16: VIF values of each independent variable (Minitab software)

Variable Name	VIF value
Purchasing price	1.26
Rate	1.12
Amount of stock	1.27
Production %	3.16
Avg. consumption in 1 m2 carpet	1.44
Jute type 1	1.32
Jute type 2	2.11
Jute type 3	2.37
Jute type 4	1.85
Jute type 5	1.97
Jute type 6	2.15
Jute type 7	2.02
Jute type 8	2.02
Jute type 9	2.03
Jute type 10	2.40
Jute type 11	2.15

As it can be seen from Table 6.16, all of the VIF values for independent variables are less than 5. Therefore, it is concluded that there is no multicollinearity among all x variables and ensure necessary condition.

In addition to multicollinearity, the presence of autocorrelation between residuals should be checked. There should be no autocorrelation between residuals. The Durbin-Watson statistic can be used to check autocorrelation. For Durbin-Watson test, the following hypotheses are used.

$H_0 : \rho = 0$

$H_A : \rho \neq 0$

If there is positive autocorrelation between residuals, H_0 is rejected.

In the Durbin-Watson table, there are two critical values, d_L and d_U.

If calculated Durbin-Watson value (d value) is smaller d_L, H_0 is rejected and it is concluded that there is positive autocorrelation between residuals. If d value is greater than d_U, H_0 is not rejected and it is concluded that there is no positive autocorrelation between residuals. If d value is between d_L and d_U, the test is inconclusive (Groebner, 2011).

Other necessary variables to perform Durbin-Watson test are given as follows:

Criticial $\alpha = 0.05$

$n = 246$

$k = 6$

According to these values, the Durbin-Watson table value for d_L is 1.623 and for d_U is 1.735. The calculated Durbin-Watson value for the jute consumption prediction problem is 1.95984 that is taken from Minitab software. In conclusion, it can be said that H_0 is not rejected and there is no positive autocorrelation between residuals.

The other constraint is homocedasticity. This term means the same variance. If the residuals of a regression model has the same variance for all values of the independent variables, it can be said that this model satisfies the homocedasticity assumption. In order to check this assumption, residuals plot against fitted values on the horizontal axis can be used. In Figure 6.5, the plot in which the residuals against the fitted values is shown. In this figure, the residuals plot is not cone-shaped, which means that the assumption of equal variance is satisfied.

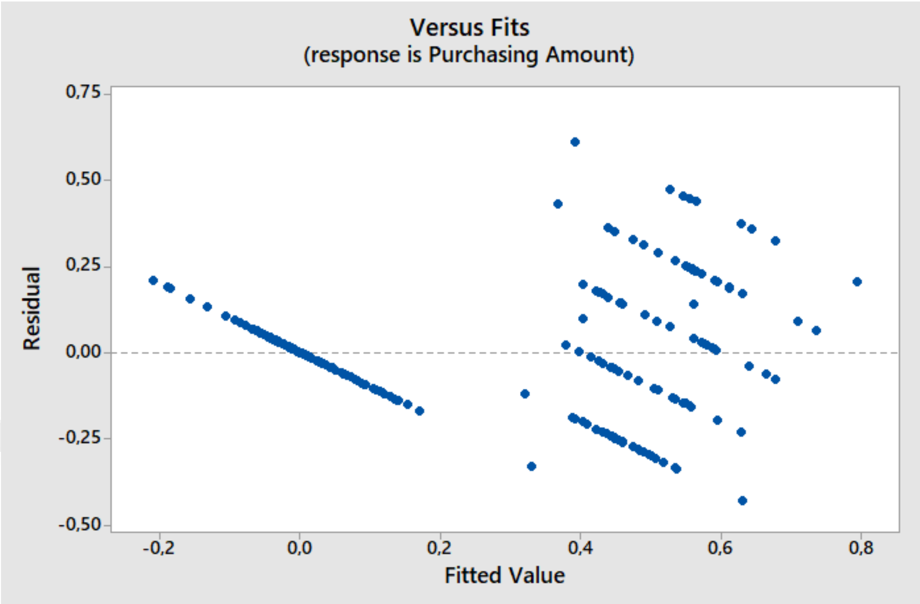

Figure 6.5: Residuals plot against fitted value (Minitab software)

After this assumptions are satisfied, the MLR model for the jute consumption problem is established and the following assumptions are checked.

The first assumption is the normality distribution of residuals. There are more than one ways to check this assumption. Normal probability plot of residuals is one of them. Figure 6.6 shows the normal probability plot of residuals for the jute consumption prediction problem. In Figure 6.6, the line of distribution of residuals is close to linear. Therefore, it can be said that this normality assumption of residuals is satisfied.

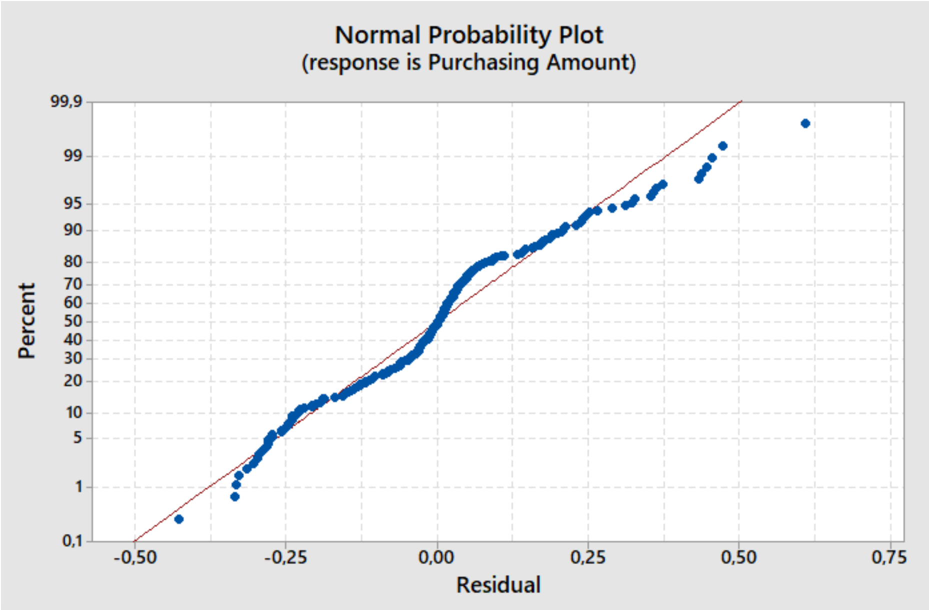

Figure 6.6: Normal probability plot of residuals (Minitab software)

The other assumption that should be checked is the significance of overall regression model. The F-test is a method for testing whether the regression model explains a certain proportion of the variation in the y, dependent variable. The formulation of F-test is given in Equation (6.7) (Groebner, 2011):

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} \qquad (6.7)$$

Where;

SSR = sum of squares regression

SSE = sum of squares error

n = sample size

k = number of independent variables

degrees of freedom = D1 = k, and D2 = (n-k-1)

The result of the analysis of variance is given in Table 6.17. In Table 6.17, the F-value is important in order to check the significance the overall regression model.

Table 6.17: Analysis of Variance result of MLR (Minitab software)

Source	**DF**	**Adj SS**	**Adj MS**	**F-value**	**P-value**
Regression	15	16.8491	1.1233	39.73	0.000
Error	230	6.5029	0.0283		
Total	245	23.3520			

The null and alternative hypotheses are given below:

$H_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$

H_A = At least one $\beta_i \neq 0$

If the null hypothesis is true and all the slope coefficients of independent variables are equal to zero, the overall regression model is not appropriate for prediction.

Other necessary variables to perform F-test are given as follows:

Criticial $\alpha = 0.05$

$k = 6$

$n - k - 1 = 239$ degrees of freedom

F-table value is equal to approximately 2.14.

$F = 39.73 > 2.14$

Therefore, H_0 is rejected and it can be concluded that the overall regression model explains a significant proportion of the variation.

Finally, the last assumption is the significance of each regression coefficient. In the previous assumption, it is concluded that the overall model is significant. But, this is not the evidence for the significance of all the variables.

In order to check the significance of each regression coeffcients, t-test can be used. In this test, following hypotheses are used:

$H_0 : \beta_j = 0$

$H_A : \beta_j \neq 0$ for all j

Other necessary variables to perform t-test are given as follows:

Criticial α = 0.05

n – k – 1 = 239 degrees of freedom

According to these values, t-table value is approximately 1.97.The calculated t value that are less than 1.97 or greater than 1.97 satisfy the significance of regression coefficient assumption. The calculated t values which are taken from Minitab software are shown in Table 6.18:

Table 6.18: Calculated t values of each regression coefficient (Minitab software)

Variable Name	**t Value**
Purchasing price	20.72
Rate	-2.71
Amount of stock	2.70
Production %	3.90
Avg. consumption in 1 m2 carpet	-3.15
Jute type 1	-3.61
Jute type 2	3.50
Jute type 3	2.98
Jute type 4	3.21
Jute type 5	2.85
Jute type 6	3.67
Jute type 7	2.64
Jute type 8	3.13
Jute type 9	3.83
Jute type 10	2.96
Jute type 11	3.71

According to Table 6.18, it can be concluded that H_0 should be rejected and all regression coefficients are significant.

6.4.2 Equations and Results of MLR Method:

All regression assumptions were checked and finally it could be said that using the MLR method is proper for the jute consumption prediction prediction.

The regression equations are given in Table 6.19. As it can be seen, each equation changes according to the type of jute.

Table 6.19: Regression equations of MLR (Minitab software)

Jute Type	Regression Equations of MLR Method
Jute Type 1	***Consumption Amount*** = *-0.2080 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 2	***Consumption Amount*** = *-0.0129 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 3	***Consumption Amount*** = *-0.0823 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 4	***Consumption Amount*** = *-0.0472 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 5	***Consumption Amount*** = *-0.0576 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 6	***Consumption Amount*** = *-0.0097 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 7	***Consumption Amount*** = *-0.0722 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 8	***Consumption Amount*** = *-0.0441 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 9	***Consumption Amount*** = *-0.0072 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 10	***Consumption Amount*** = *-0.0762 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*
Jute Type 11	***Consumption Amount*** = *-0.0217 + 0.6411 (Purchasing Price) -0.0843 (Rate) +0.0470 (Amount of stock) + 0.428 (Production %) – 0.0103 (Average consumption in 1m^2)*

One of the performance indicators that is used to compare two models is MSE. The MSE value is calculated in MATLAB software but it is not for Minitab. Therefore, MSE value is calculated by hand with using following Equation (6.8):

$$MSE = \frac{\sum (y_t - F_t)^2}{n} \qquad (6.8)$$

Where;

y_t = actual value of data t

F_t = forecasted value of data t

n = sample size

As a result; MSE, R^2 and adjusted R^2 values for MLR method are as given in Table 6.20.

Table 6.20: Values of performance indicators of MLR

Performance indicator	Regression results
MSE	0.1289
R^2	0.7215
Adjusted R^2	0.7034

In Table 6.20, the MSE value of MLR method is at satisfactory level. But R^2 and adjusted R^2 values are lower than 0.80, and this means that the prediction accuracy of MLR method is not satisfactory for this problem.

Relative results of ANN and MLR models according to MSE, R^2 and adjusted R^2 is given in Table 6.21.

Table 6.21: Comparative results of ANN and MLR

Performance indicator	Results of the ANN	Results of the MLR
MSE	0.0116	0.1289
R^2	0.8464	0.7215
Adjusted R^2	0.8712	0.7034

In Table 6.21, MSE value of ANN method is less than the half of the MLR method. Also, R^2 and adjusted R^2 values of ANN method higher than MLR method. This means that the outputs of ANN method is more closer to actual values.

6.5 A Scenario with ANN Model:

The following scenario is created in order to show the effectiveness of ANN model from the cost saving point of view. The given cost values are the current prices that was taken from purchasing department.

Assume that the company have purchased 26 tons of jute for jute type 1 based on experiential judgements. Actually, it will need 48 tons jute and the ANN model has computed 49 tons of jute type 1. In this situation, six weaving looms that use the jute type 1 at the same time should stop for one shift (8-hours). Also, the remaining raw material need (22 tons) should be supplied from domestic supplier.

The current cost values about jute type 1are as follows:

- Purchasing price for 1 ton from the abroad supplier is $ 1.175.
- Purchasing price for 1 ton from domestic supplier is $1.275.
- The cost of one weaving looms for stopping the production is $1.000/day. This data was calculated by weaving department and they took into consideration the number of employees for one weaving loom, the monthly salary of these employees, energy and raw material cost etc. while calculate this cost value.
- Holding cost is approximately 5% of the total purchasing cost. This data was calculated by accounting department and they took into consideration the land value of the warehouse area, the salary of the workers in the warehouse, the cost of the energy resources spent in the warehouse, and the average interest rate considered when purchasing the yarn etc. while calculate this cost value.

By using these cost values, we can make the following calculations:

Calculations for Judgmental Method:

In this method, 26 tons of jute type 1 are supplied from abroad supplier, 22 tons are supplied from domestic supplier. Also, six machines are stopped in one shift.

The cost of stopping production:

This value indicates the cost value that should be incurred when the company should stop production due to lack of jute raw material. In this calculation, firstly stopping duration of 6 machines is calculated in hours and then converted to day time. Then, the cost value of stopping production is calculated.

$$(8\ hours) * (6\ machines) = 48\ hours$$

$$\frac{(48\ hours)}{(24\ hours)} = 2\ days$$

$$(\$1.000) * 2 = \$2.000$$

For 48 hours, the cost of stopping production is \$2.000.

Total cost of jute from abroad supplier:

This value represents the purchase cost of 26 tons of jute from the foreign supplier.

$$(26\ tons) * (\$1.175) = \$30.550$$

Total cost of jute from domestic supplier:

This value indicates the purchase cost of 22 tons of jute from the domestic supplier. This 22 tons jute is required to complete the orders of customers.

$$(22\ tons) * (\$1.275) = \$28.050$$

Holding cost:

This value shows the holding cost for total 48 tons of jute.

$$[(\$30.550) + (\$28.050)] * 0.05 = \$2.930$$

Total cost of Case 1:

The total cost of Case 1 is the sum of stopping cost, purchasing cost for 26 tons of jute from foreign supplier, purchasing cost for 22 tons of jute from domestic supplier, and total holding cost for 48 tons of jute.

$$(\$2.000) + (\$\ 30.550) + (\$28.050) + (\$2.930) = \$\mathbf{63.530}$$

Calculations for ANN Method:

In this method, 49 tons of jute type 1 are supplied from abroad supplier. In this situation, none of machines are stopped for production.

Total cost of jute from abroad supplier:

This value shows the purchase cost for 49 tons of jute from foreign supplier.

$(49 \text{ tons}) * (\$1.175) = \$57.575$

Holding cost:

This value shows the holding cost for total 49 tons of jute.

$(\$57.575) * 0.05 = \$2878$

Total cost of Case 2:

The total cost of Case 2 is the sum of purchasing cost for 49 tons of jute from foreign supplier, and holding cost for 49 tons of jute.

$(\$57.575) + (\$2878) = \$\mathbf{60.453}$

Cost of saving of ANN Method is:

$(\$63.530) - (\$60.453) = \$3.077$

$(\$3.077) * (5.45)(exchange\ rate) = 16.749\ TL$

As it can be seen from the above calculations, by using the ANN model the company can make profit 16.749 TL.

7 CONCLUSION

In today's competitive market conditions, the firms operating in the production and service sectors must meet the quantities demanded by the customers in a timely and complete manner. For this purpose, all the resources to be used must be provided on time and in sufficient quantity. In this case, most accurate sales and raw material consumption predictions should be done by companies.

Several methods have been used for prediction studies. The advantages and usage areas of each method are different. Traditional statistical methods are sufficient for linear problems. However, most business problems are nonliner and traditional methods are inadequate.

Artificial intelligence and applications of this discipline have increasing popularity and application areas in today's world. Artificial neural networks, which are one of the sub-branches of artificial intelligence, are used in many areas ranging from demand estimation to banking sector, health field applications to public sector and the obtained results are satisfactory.

In this study, the future consumption amount of jute raw material is estimated. For this purpose, 246 data were used. An ANN with 6 input and one hidden layer with 10 neurons was established. In addition, the MLR model was used and the results were compared with ANN results. MSE, R^2 and adjusted R^2 were used as performance indicators. Comparative results of these two methods are shown in Table 7.1:

Table 7.1: Comparative results of ANN and MLR

Performance indicator	**Results of the ANN**	**Results of the MLR**
MSE	0.0116	0.1289
R^2	0.8464	0.7215
Adjusted R^2	0.8712	0.7034

In Table 7.1, MSE value of ANN method is less than the half of the MLR method. Also, R^2 and adjusted R^2 values of ANN method higher than MLR method. This means that the outputs of ANN method is more closer to actual values.

As a result, it is concluded that the results of the ANN method are much closer to actual values when compared MLR and therefore ANN seems to be more appropriate for dynamic real world conditions. In this study, the ANN method has been shown to be a reliable method in estimating jute consumption. In addition, by using this method customer demands can be estimated roughly and thereby they can be met in a timely manner. In this respect, customer satisfaction can be increased. In addition to this, the company can make profit by using the ANN prediction model according to the calculations in the scenario stated in the section 6.5. If the company faces this situation four times in a year, annual saving will be $ 12.308. Also, this calculations are done only for one jute type. The ANN prediction model can be used for all raw materials in the company in order to effectitvaly manage the supply chain. As a future work, this method can be extended to include several macro level parameters like the interest rate. The proposed approach can be applied to other manufacturing sectors for forecasting the possible raw material consumption. In addition, the predictive results can be improved by combining the ANN method with some heuristic models.

REFERENCES

Anderson, D. ve McNeill, G.,1992. Artificial Neural Networks Technology, Rome Laboratory, ELIN: A011, New York.

Azadeh, A., Ghaderi, S.F., Tarverdian, S., and Saberi, M. (2007a). Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. Applied Mathematics and Computation *186*, 1731–1741.

Azadeh, A., Ghaderi, S.F., and Sohrabkhani, S. (2007b). Forecasting electrical consumption by integration of Neural Network, time series and ANOVA. Applied Mathematics and Computation *186*, 1753–1761.

Catalão, J.P.S., Mariano, S.J.P.S., Mendes, V.M.F., and Ferreira, L.A.F.M. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. Electric Power Systems Research *77*, 1297–1304.

CHARNIAK, E.A., and McDermott, D.V. (1985). Introduction to Artificial Intelligence (Addison-Wesley Pub.Company).

Chiang, W.-C., Urban, T.L., and Baldridge, G.W. (1996). A neural network approach to mutual fund net asset value forecasting. Omega *24*, 205–215.

Co, H.C., and Boosarawongse, R. (2007). Forecasting Thailand’s rice export: Statistical techniques vs. artificial neural networks. Computers & Industrial Engineering *53*, 610–627.

Cohen, P.R., and Feigenbaum, E.A. (2014). The Handbook of Artificial Intelligence (Butterworth-Heinemann).

Durmuşoğlu, Z., and Gülyeşil, S. (2018). Jute Yarn Consumption Prediction by Artificial Neural Network and Multilinear Regression. Natural and Engineering Sciences, *3,* 43-53

Efendigil, T., Önüt, S., and Kahraman, C. (2009). A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis. Expert Systems with Applications *36*, 6697–6707.

Elkateb, M.M., Solaiman, K., 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*, 3–13.

Fausett, L.V., and Fausett, L. (1994). Fundamentals of Neural Networks: Architectures, Algorithms, and Applications (Prentice-Hall).

Fonseca, D.J., and Navaresse, D. (2002). Artificial neural networks for job shop simulation. Advanced Engineering Informatics *16*, 241–246.

Franses, P.H., and Draisma, G. (1997). Recognizing changing seasonal patterns using artificial neural networks. Journal of Econometrics *81*, 273–280.

Gaafar, L.K., and Choueiki, M.H. (2000). A neural network model for solving the lot-sizing problem. Omega *28*, 175–184.

Gareta, R., Romeo, L.M., and Gil, A. (2006). Forecasting of electricity prices with neural networks. Energy Conversion and Management *47*, 1770–1778.

Ghiassi, M., Saidane, H., and Zimbra, D.K. (2005). A dynamic artificial neural network model for forecasting time series events. International Journal of Forecasting *21*, 341–362.

Golmohammadi, D. (2011). Neural network application for fuzzy multi-criteria decision making problems. International Journal of Production Economics *131*, 490–504.

Guresen, E., Kayakutlu, G., and Daim, T.U. (2011). Using artificial neural network models in stock market index prediction. Expert Systems with Applications *38*, 10389–10397.

Hamzaçebi, C. (2007). Forecasting of Turkey's net electricity energy consumption on sectoral bases. Energy Policy *35*, 2009–2016.

Hamzacebi, C. (2008). Improving artificial neural networks' performance in seasonal time series forecasting. Information Sciences *178*, 4550–4559.

Hamzaçebi, C., Akay, D., and Kutay, F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. Expert Systems with Applications *36*, 3839–3844.

Harpham, C., and Dawson, C.W. (2006). The effect of different basis functions on a radial basis function network for time series prediction: A comparative study. Neurocomputing *69*, 2161–2170.

Haykin, S., and Haykin, S.S. (2009). Neural Networks and Learning Machines (Prentice Hall).

Hwarng, H.B. (2001). Insights into neural-network forecasting of time series corresponding to ARMA(p,q) structures. Omega *29*, 273–289.

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

Kaastra, I., and Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. Neurocomputing *10*, 215–236.

Khashei, M., and Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing *11*, 2664–2675.

Kheirkhah, A., Azadeh, A., Saberi, M., Azaron, A., and Shakouri, H. (2013). Improved estimation of electricity demand function by using of artificial neural network, principal component analysis and data envelopment analysis. Computers & Industrial Engineering *64*, 425–441.

Kohzadi, N., Boyd, M.S., Kermanshahi, B., and Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. Neurocomputing *10*, 169–181.

Krajewski, L.J., Ritzman, L.P., and Malhotra, M.K. (2013). Operations Management: Processes and Supply Chains (Pearson).

Kurzweil, R. (1990). The age of intelligent machines (MIT Press).

Kriesel, R. (2005). A Brief Introduction to Neural Networks

Lin, Y.-H., Shie, J.-R., and Tsai, C.-H. (2009). Using an artificial neural network prediction model to optimize work-in-process inventory level for wafer fabrication. Expert Systems with Applications *36*, 3421–3427.

Luxhøj, 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.

Markham, I.S., and Rakes, T.R. (1998). The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. Computers & Operations Research *25*, 251–263.

Mehrotra, K., Mohan, C.K., and Ranka, S. (1997). Elements of Artificial Neural Networks (MIT Press).

Moretti, F., Pizzuti, S., Panzieri, S., and Annunziato, M. (2015). Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling. Neurocomputing *167*, 3–7.

Öztemel, E. (2012). Yapay sinir ağları.

Partovi, F.Y., and Anandarajan, M. (2002). Classifying inventory using an arti®cial neural network approach. Industrial Engineering 16.

Poole, D.L., and Mackworth, A.K. (2010). Artificial Intelligence: Foundations of Computational Agents (Cambridge University Press).

Qiu, M., Song, Y., and Akagi, F. (2016). Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. Chaos, Solitons & Fractals *85*, 1–7.

Rather, A.M., Agarwal, A., and Sastry, V.N. (2015). Recurrent neural network and a hybrid model for prediction of stock returns. Expert Systems with Applications *42*, 3234–3241.

Ren, C., An, N., Wang, J., Li, L., Hu, B., and Shang, D. (2014). Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting. Knowledge-Based Systems *56*, 226–239.

Rich, E., and Knight, K. (1991). Artificial Intelligence (McGraw-Hill).

Russell, S.J., Russell, S.J., and Norvig, P. (2010). Artificial Intelligence: A Modern Approach (Prentice Hall).

Sheikhan, M., and Mohammadi, N. (2013). Time series prediction using PSO-optimized neural network and hybrid feature selection algorithm for IEEE load data. Neural Computing and Applications *23*, 1185–1194.

Tseng, F.-M., Yu, H.-C., and Tzeng, G.-H. (2002). Combining neural network model with seasonal time series ARIMA model. Technological Forecasting and Social Change *69*, 71–87.

Wang, T.-Y., and Chien, S.-C. (2006). Forecasting innovation performance via neural networks—a case of Taiwanese manufacturing industry. Technovation *26*, 635–643.

Wang, L., Zeng, Y., and Chen, T. (2015). Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. Expert Systems with Applications *42*, 855–863.

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

Zhang, G.P., and Qi, M. (2005). Neural network forecasting for seasonal and trend time series. European Journal of Operational Research *160*, 501–514.

Zhang, G., Eddy Patuwo, B., and Y. Hu, M. (1998). Forecasting with artificial neural networks: International Journal of Forecasting *14*, 35–62.

Zhang, G.P., Patuwo, B.E., and Hu, M.Y. (2001). A simulation study of arti"cial neural networks for nonlinear time-series forecasting. Operations Research 16.

Zhang, J.-R., Zhang, J., Lok, T.-M., and Lyu, M.R. (2007). A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. Applied Mathematics and Computation *185*, 1026–1037.

(2011). Business statistics: a decision-making approach (Upper Saddle River, N.J: Prentice Hall/Pearson).

APPENDIX

APPENDIX A. ALL DATA USED IN THIS THESIS

Table A.1: : Raw form of 246 Data Points Between The Years 2015 and 2017 for Jute Consumption Prediction Problem

Jute Type	Purchasing Price	Exchange Rate	Amount of Stock (tons)	Production %	Average Consumption in 1 m2	Purchasing Amount (tons)
1	1.38	2.4552	15.639	4.228	0.52313	26
2	1.18	2.4552	35.225	15.123	0.80209	78
3	1.15	2.4552	89.413	20.474	0.77556	78
4	0	2.4552	41.217	0.529	0.75833	0
5	1.125	2.4552	5.414	8.556	0.73667	52
6	1.07	2.4552	20.669	1.961	0.77778	26
7	1.06	2.4552	44.225	21.499	0.68254	78
8	1.06	2.4552	15.690	5.430	0.7375	26
9	1.06	2.4552	18.509	2.242	0.79167	52
10	1.045	2.4552	1.277	16.570	0.7963	104
11	0	2.4552	0	3.388	0.61111	0
1	1.27	2.5838	21.787	4.317	0.52429	26
2	1.181	2.5838	61.215	18.352	0.79827	104
3	1.146	2.5838	48.156	20.205	0.77667	130
4	0	2.5838	23.182	0.	0	0
5	1.107	2.5838	47.874	9.801	0.7463	52
6	0	2.5838	32.822	1.639	0.77778	0
7	1.073	2.5838	16.885	16.182	0.73016	78
8	1.05	2.5838	20.883	5.858	0.74	26
9	0	2.5838	19.747	0.967	0.78646	0
10	1.048	2.5838	28.726	16.664	0.85031	104
11	0	2.5838	0	6.015	0.61111	0
1	0	2.6481	33.157	3.158	0.50416	0
2	1.206	2.6481	71.628	19.815	0.78572	104
3	1.176	2.6481	40.339	32.783	0.77986	130
4	1.03	2.6481	18.346	4.403	0.63421	26
5	1.115	2.6481	37.887	6.039	0,74479	26
6	0	2.6481	36.801	0	0	0
7	0	2.6481	43.597	3.070	0.75556	0
8	1.08	2.6481	33.940	3.611	0.74	26
9	0	2.6481	36.552	3.925	0.73958	0
10	1.063	2.6481	65.546	18.772	0.8559	104
11	0	2.6481	0	4.424	0.61111	0
1	0	2.6461	28.000	2.874	0.51665	0
2	0	2.6461	56.317	12.812	0.80209	0
3	0	2.6461	28.268	37.944	0.75595	0
4	0	2.6461	13.491	6.581	0.6771	0
5	0	2.6461	35.320	7.585	0.75833	0

6	0	2.6461	28.841	0	0	0
7	0	2.6461	35.828	5.435	0.73333	0
8	0	2.6461	24.848	2.442	0.74375	0
9	0	2.6461	32.564	3.329	0.74769	0
10	0	2.6461	10.487	17.612	0.81481	0
11	0	2.6461	0	3.386	0.76389	0
1	1.1	2.6946	54.880	17.779	0.52314	52
2	1.0425	2.6946	204.905	9.221	0.84983	104
4	0.97	2.6946	17.162	5.589	0.66806	52
5	0.973	2.6946	55.571	6.467	0.7463	78
6	0	2.6946	80.715	0	0	0
7	0	2.6946	107.213	1.081	0.77037	0
8	0	2.6946	42.517	9.821	0.74167	0
9	0	2.6946	87.365	2.635	0.83333	0
10	0	2.6946	143.326	22.021	0.82778	0
11	0	2.6946	0	6.055	0.61111	0
1	0	2.8455	11.850	7.151	0.51665	0
2	1.0425	2.8455	130.228	12.218	0.85556	104
4	0	2.8455	16.201	9.446	0.66626	0
5	0.973	2.8455	43.697	4.519	0.76435	52
6	0	2.8455	28.609	1.106	0.77778	0
11	0	2.8455	0	8.778	0.76389	0
1	1.22	3.0027	21.266	12.172	0.51288	26
2	0	3.0027	80.639	9.886	0.85301	0
3	0	3.0027	14.341	21.430	0.79167	0
4	0	3.0027	10.175	0.117	0.60486	0
5	0	3.0027	23.435	6.046	0.75833	0
6	0	3.0027	23.585	0	0	0
7	0	3.0027	24.702	9.936	0.74921	0
8	0	3.0027	26.229	6.530	0.77083	0
9	0	3.0027	18.433	1.078	0.75	0
11	0	3.0027	0	7.727	0.61111	0
1	1.22	2.9295	18.293	9.364	0.54629	65
2	0	2.9295	165.759	12.479	0.84601	0
4	1.16	2.9295	9.000	5.524	0.6771	78
5	1.14	2.9295	68.560	12.626	0.73486	91
6	0	2.9295	75.696	0.440	0.75556	0
7	0	2.9295	101.460	12.413	0.70833	0
8	0	2.9295	44.048	2.422	0.75938	0
9	0	2.9295	80.893	0	0	0
10	0.9725	2.9295	116.960	20.220	0.82407	78
11	0	2.9295	0	3.881	0.61111	0
1	1.266	2.8712	41.340	9.251	0.51805	26
2	1.15	2.8712	161.422	6.468	0.84028	52
3	1.169	2.8712	142.991	16.807	0.77632	104
4	0	2.8712	49.311	7.883	0.6613	0

5	1.14	2.8712	11.952	15.651	0.74028	78
6	0	2.8712	75.262	0.777	0.7	0
7	0	2.8712	184.135	6.948	0.74603	0
8	1.0675	2.8712	27.386	4.222	0.74375	52
9	0	2.8712	79.553	3.017	0.75	0
10	1.0533	2.8712	148.823	20.882	0.85648	104
11	0	2.8712	0	8.094	0.61111	0
1	1.25	2.9172	15.389	8.693	0.57715	26
2	1.15	2.9172	169.461	9.106	0.80209	78
3	1.125	2.9172	90.973	21.156	0.78177	26
5	1.11	2.9172	0	9.689	0.7475	52
6	0	2.9172	71.927	1.820	0.70556	0
7	0	2.9172	151.782	3.017	0.74222	0
8	1.0675	2.9172	13.209	6.002	0.75	26
9	0	2.9172	66.524	1.927	0.75	0
11	0	2.9172	0	7.697	0.61111	0
1	1.19	3.0069	15.645	9.928	0.54306	130
2	1.0833	3.0069	188.744	12.759	0.83264	52
3	0	3.0069	72.002	22.546	0.77208	0
4	1.03	3.0069	20.945	5.982	0.67708	26
5	1.04	3.0069	83.243	16.253	0.74931	130
6	0	3.0069	62.637	0.229	0.87778	0
7	0	3.0069	149.297	3.327	0.70556	0
8	0	3.0069	40.380	3.866	0.74167	0
9	0	3.0069	113.727	0	0	0
10	0	3.0069	25.929	16.643	0.76389	0
11	0	3.0069	0	8.467	0.61111	0
1	1.2036	2.9406	48.231	9.523	0.52604	26
2	1.089	2.9406	115.999	9.443	0.85556	104
3	1.04	2.9406	76.733	13.992	0.7494	78
4	0	2.9406	45.234	7.159	0.67708	0
5	1.0676	2.9406	95.266	14.682	0.74931	52
6	0.993	2.9406	56.047	2.355	1	78
8	0.98	2.9406	18.170	6.649	0.7575	78
9	0	2.9406	108.551	0	0	0
10	0.99	2.9406	141.765	17.064	0.75397	26
11	0	2.9406	0	1.479	0.61111	0
1	1.1	2.8917	5.196	6.610	0.51476	26
2	0	2.8917	50.696	5.006	0.84028	0
4	1.135	2.8917	41.056	7.552	0.66956	104
5	1.02	2.8917	11.899	16.604	0.74673	130
6	0.993	2.8917	37.940	2.051	0.85	26
7	0.993	2.8917	39.166	9.809	0.66667	104
8	0.98	2.8917	5.184	7.466	0.75938	104
9	0	2.8917	108.551	0.705	0.75	0
11	0	2.8917	0	4.822	0.61111	0

1	1.25	2.8347	9.271	12.132	0.51476	52
2	0	2.8347	92.142	10.977	0.81481	0
3	1.12	2.8347	47.650	14.308	0.76667	104
4	1.135	2.8347	32.835	8.835	0.69063	130
5	1.086	2.8347	45.909	9.625	0.77278	26
6	0	2.8347	23.505	0	0	0
7	0	2.8347	0	10.080	0.75309	0
8	1.025	2.8347	0	4.764	0.75938	52
9	1.09	2.8347	78.327	4.058	0.79167	26
10	1.0387	2.8347	98.559	19.860	0.75926	104
11	0	2.8347	0	5.361	0.61111	0
1	1.225	2.9265	65.741	11.676	0.55258	78
3	1.15	2.9265	20.763	13.970	0.76042	52
4	0	2.9265	0	12.998	0.66806	0
5	1.125	2.9265	58.457	19.290	0.77865	130
6	0	2.9265	15.260	0.049	1	0
7	0	2.9265	22.592	8.712	0.69815	0
8	0	2.9265	5.283	0.262	0.75	0
9	0	2.9265	42.699	3.013	0.83333	0
10	0	2.9265	128.634	17.927	0.8	0
11	0	2.9265	0	5.927	0.61111	0
1	0	2.9169	73.043	8.905	0.5571	0
2	0	2.9169	114.289	10.778	0.84028	0
3	0	2.9169	10.628	13.392	0.74902	0
4	0	2.9169	5.168	7.080	0.65451	0
5	0	2.9169	83.634	9.508	0.75231	0
6	0	2.9169	43.866	0	0	0
7	0	2.9169	31.902	16.217	0.6679	0
8	0	2.9169	68.377	2.300	0.75	0
9	0	2.9169	19.390	5.149	0.80556	0
10	0	2.9169	23.410	22.005	0.78611	0
11	0	2.9169	0	4.666	0.76389	0
1	0	2.9575	83.222	12.432	0.50955	0
2	0	2.9575	101.227	8.681	0.82755	0
3	0	2.9575	78.000	10.544	0.74444	0
4	0	2.9575	22.410	8.466	0.67708	0
5	0	2.9575	125.093	14.194	0.74028	0
6	0	2.9575	121.866	1.265	0.7	0
7	0	2.9575	122.336	8.358	0.65833	0
8	0	2.9575	79.312	0.932	0.74167	0
9	0	2.9575	19.694	5.198	0.80556	0
10	0	2.9575	32.410	21.774	0.82222	0
11	0	2.9575	0	8.156	0.61111	0
1	1.23	2.9628	63.848	7.441	0.53299	26
2	1.068	2.9628	68.431	12.013	0.825	130
3	0	2.9628	272.664	12.732	0.76009	0

4	0	2.9628	104.861	8.895	0.65	0
5	0	2.9628	205.706	11.510	0.76556	0
6	0	2.9628	119.264	0.436	1	0
7	0.965	2.9628	197.806	7.208	0.7037	52
8	0	2.9628	179.052	6.514	0.74375	0
9	0.95167	2.9628	9.937	2.616	0.83333	78
11	0	2.9628	0	4.360	0.61111	0
1	0	2.9601	98.164	11.080	0.53299	0
2	0	2.9601	0	4.286	0.83073	0
3	1.09	2.9601	303.956	15.418	0.75897	78
4	0	2.9601	134.732	5.466	0.63947	0
5	0	2.9601	253.822	16.647	0.75231	0
6	0	2.9601	110.102	1.278	1	0
7	0	2.9601	162.546	7.838	0.7037	0
8	0	2.9601	145.854	4.656	0.75	0
9	0	2.9601	5.215	7.911	0.725	0
10	0	2.9601	71.648	22.949	0.84722	0
11	0	2.9601	8.813	2.471	0.61111	0
1	1.28	3.0679	55.469	5.127	0.53194	52
2	1.15	3.0679	71.500	5.260	0.85301	26
3	1.09	3.0679	226.884	21.535	0.74755	104
4	1.14	3.0679	82.321	3.189	0.67708	78
5	1.1016	3.0679	185.971	16.153	0.75317	78
6	0	3.0679	110.102	0	0	0
7	1.055	3.0679	124.646	12.726	0.69444	52
8	0	3.0679	124.026	1.382	0.75	0
9	0	3.0679	10.471	8.076	0.80729	0
10	1.04	3.0679	14.433	25.826	0.78704	52
11	0	3.0679	27.524	0.726	0.61111	0
1	1.28	3.2674	43.973	5.486	0.52604	104
2	0	3.2674	31.105	8.574	0.85301	0
3	0	3.2674	97.700	8.311	0.76296	0
4	1.14	3.2674	31.285	4.188	0.67708	52
5	1.1016	3.2674	37.212	15.132	0.76607	52
6	0	3.2674	96.668	0	0	0
7	1.055	3.2674	28.648	13.932	0.64815	26
8	1.125	3.2674	94.144	5.884	0.71875	104
9	0	3.2674	3.956	14.145	0.79861	0
11	0	3.2674	53.924	5.236	0.61111	0
1	0	3.4889	38.542	5.865	0.56481	0
2	0	3.4889	0	8.494	0.80972	0
3	0	3.4889	35.231	12.725	0.77396	0
4	0	3.4889	37.411	4.808	0.65774	0
5	0	3.4889	0	17.026	0.7794	0
6	0	3.4889	53.092	0	0	0
7	0	3.4889	16.270	12.995	0.67619	0

8	0	3.4889	19.497	4.757	0.72625	0
9	0	3.4889	2.593	6.508	0.84635	0
11	1.11	3.4889	45.457	2.177	0.72569	78
7	1.2	3.7348	155.075	12.748	0.72222	104
11	1.11	3.7348	73.883	5.137	0.61111	26
1	1.38	3.6723	68.170	7.249	0.52431	26
4	1.235	3.6723	75.799	4.474	0.71094	26
7	1.2	3.6723	97.466	12.418	0.68611	78
9	1.205	3.6723	112.427	9.142	0.79861	104
11	1.15	3.6723	88.367	1.546	0.61111	52
1	1.42	3.6659	39.165	4.211	0.51736	26
8	1.225	3.6659	9.656	3.867	0.74375	78
9	1.205	3.6659	51.085	11.446	0.79861	78
11	1.195	3.6659	71.400	2.203	0.61111	78
1	1.45	3.6538	115.903	6.806	0.51042	26
7	1.19	3.6538	112.546	13.129	0.66667	26
8	1.225	3.6538	70.192	1.818	0.745	26
11	1.2	3.6538	115.820	1.155	0.72569	26
1	1.38	3.5638	78.998	6.857	0.5248	52
4	1.2	3.5638	80.360	5.949	0.66625	26
6	1.18	3.5638	16.972	0	0	26
7	1.19	3.5638	40.799	5.436	0.70952	52
10	1.1714	3.5638	16.870	27.861	0.8179	130
11	1.14	3.5638	7.960	7.482	0.61111	104