<u>ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE</u>
<u>ENGINEERING AND TECHNOLOGY</u>

SHORT TERM LOAD FORECASTING BY USING ARTIFICIAL NEURAL NETWORKS

M.Sc. THESIS

Ali GHADIRIASL NOBARI

Department of Electrical Engineering

Electrical Engineering Programme

JUNE 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

SHORT TERM LOAD FORECASTING BY USING ARTIFICIAL NEURAL NETWORKS

M.Sc. THESIS

Ali GHADIRIASL NOBARI
(504171053)

Department of Electrical Engineering

Electrical Engineering Programme

Thesis Advisor: Prof. Dr. Belgin EMRE TÜRKAY

JUNE 2020

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY SİNİR AĞLARI KULLANILARAK KISA DÖNEMLİ YÜK TAHMİNİ

YÜKSEK LİSANS TEZİ

Ali GHADIRIASL NOBARI
(504171053)

Elektrik Mühendisliği Anabilim Dalı

Elektrik Mühendisliği Programı

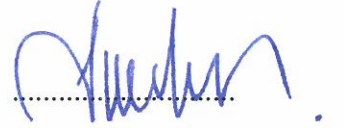**Tez Danışmanı: Prof. Dr. Belgin EMRE TÜRKAY**

HAZIRAN 2020

Ali Ghadiriasl Nobari, a M.Sc student of İTU Graduate School of Science Engineering and Technology 504171053 successfully defended the thesis entitled "SHORT TERM LOAD FORECASTING BY USING ARTIFICIAL NEURAL NETWORKS", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.
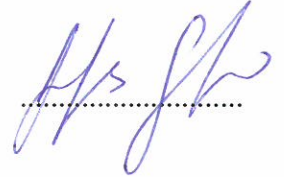
Thesis Advisor :   **Prof. Dr. Belgin EMRE TÜRKAY**
Istanbul Technical University

Jury Members :   **Assoc. Prof. Dr. Tahir ÇETİN AKINCI**
Istanbul Technical University

                              Asst. Prof. Dr. Hacer ÖZTURA
Yaşar University

Date of Submission : 15 June 2020
Date of Defense   : 20 July 2020

v

To my family,

FOREWORD

I owe my deepest gratitude to my thesis advisor Prof. Dr. Belgin Emre Türkay. Her door was always open whenever I had a question concerning my research and also she assisted me to access the required database. I wish to appreciate the Department of Electrical Engineering and all the instructors I had.

Finally, I must show my rather profound appreciation to my parents, my family, and especially my uncle, Rasoul Asghari, who has been my best friend since I was a child. No one has been more influential to me in the pursuit of this thesis than him.

This thesis would not have been achievable without them.

Thank you.

June 2020

Ali Ghadiriasl Nobari
(Electrical Engineer)

TABLE OF CONTENTS

ABBREVIATIONS

ANN	**:**	Artificial Neural Network
DIF	**:**	Diversity Factor
DF	**:**	Demand Factor
EA	**:**	Econometric analysis
EUA	**:**	End-use analysis
FF	**:**	Feed Forward Neural Network
FL	**:**	Fuzzy Logic
GA	**:**	Genetic Algorithm
LF	**:**	Load Factor
LR	**:**	Linear Regression
LTLF	**:**	Long- Term Load Forecast
MAE	**:**	Mean Absolute Error
MBE	**:**	Mean Bias Error
MSE	**:**	Mean Square Error
MTLF	**:**	Medium-Term Load Forecast
ReLU	**:**	Rectified Linear Unit
RNN	**:**	Recursive Neural Network
SDA	**:**	Similar-Day Approach
STLF	**:**	Short-Term Load Forecast
SVM	**:**	Support Vector Machine
TA	**:**	Trend Analysis
TS	**:**	Time Series
UF	**:**	Utilization Factor
UMN	**:**	University of Minnesota
VSTLF	**:**	Very-Short-Term Load Forecast

SYMBOLS

ReLU   **:** Rectified linear unit function

Tanh   **:** hyperbolic tangent function

σ   **:** Sigmiod function

LIST OF TABLES

LIST OF FIGURES

Page

SHORT TERM LOAD FORECASTING BY USING ARTIFICIAL NEURAL NETWORKS

SUMMARY

Load forecasting is considered as one of the most essential tools for managing power systems. It accommodates the system to maintain reliable power for consumers. Forecasting is the first step for investigating power system planning. Electrical energy injections and withdrawals/losses on the power system must be balanced at all times and the cost of generating electrical energy changes with demand. Electricity must be provided while it is being consumed because saving this valuable energy is not affordable. These most crucial characteristics of electricity has made the managing of power systems much more complicated. Since this topic is quite significant, numerous researchers have used different methods for achieving greater results.

All those methods could be classified into two groups, the Artificial Intelligence Approach, and the Statistical Approach. Artificial intelligence (AI) is an attempt to imitate the human brain functions of thinking and decision making. Artificial intelligence is highly practical in load forecasting for investigating the relationships between different features. Knowing the forthcoming consumption is extremely necessary for power companies. The short term forecasting could be used for anticipating the load ranging from an hour to several days.

In this thesis, a forecasting model was made by utilizing Artificial Neural Networks (ANN) and other methods. The ANN is a strong machine-learning method that processes the data with non-linear and complex functions. This algorithm is not so new, but it has become more popular these recent years. There are various types of ANN, each suitable for a particular case. Feed-Forward (FF) was applied as an ANN in this thesis.

The neural networks are suitable methods for both classification and regression cases which forecasting is a regression problem. The linear activation function is the best choice for the last layer.

Different combination of features was used to find out which one has the best results. The calendar features were used in the first study and the second one utilized the weather features. Finally, both of them were used in the last study. The errors of these three studies show that it is more beneficial to use all of the features in one model. The correlation methods were used for determining weather features. Some of the weather features like the cloudiness of the sky had such little correlation, so they were excluded from train data.

The main goal of this thesis is to forecast the load in the short term, but the model could also be used for longer periods. The model was prepared by using more than five years of data from Gebze city. The goal of this thesis is estimating the average load of the consuming amount. The weather data consists of information gathered from two different stations. It facilitated filling the missing values properly.

There are four steps in the load forecasting model. The first step is data preprocessing and the second step is defining the model. A large number of parameters and hyper-parameters must be tuned in defining the model. Not only these elements have a very crucial impact on accuracy, but they could also prevent the model from under-fitting and over fitting. The third and fourth steps are fitting the data into the model and validating the accuracy.

Three different validation functions were applied, RSQ, Mean absolute error (MAE), and Mean Squared Error (MSE). The RSQ is a number between zero and one; one being the best value for these criteria. The best value for MAE is zero. The MSE is a non-negative value that measures the quality of an estimator in which it must be a small value.

The thesis uses K-Nearest Neighbors, Linear Registration, and Decision Tree for comparing the results and finding the best algorithm. The best solution is the neural network algorithm result. The accuracy could be increased by using more data with more features and also by designing a better architecture for the neural network's model.

YAPAY SİNİR AĞLARI KULLANILARAK KISA DÖNEMLİ YÜK TAHMİNİ

ÖZET

Elektrik endüstrisi bir ülkenin altyapı endüstrilerinden biridir ve günümüz toplumlarının büyüme ve gelişmesinde çok önemli bir ayağıdır. Elektrik sektöründeki projeler bir yandan büyük yatırımlar ve uzun süreler gerektirdiği için diğer yandan mevcut teknoloji ile bu enerjiyi büyük ölçekte depolamak hala mümkün değildir. Bu nedenle elektrik talebini karşılayacak şekilde üretim planlaması yapılmalıdır. Bu nedenle yük tahmini, güç sistemlerinin geliştirilmesinde ve çalıştırılmasında önemli bir faktör olarak kabul edilir ve aslında karar vermeyi iyileştirmek için kullanılabilecek bir araçtır. Kaynak tahsisi sürecini tahmin etmek, elektrik şebekesinin geliştirilmesi için çok önemlidir.

Gelecekteki geliştirme planlamasında, bir yük tahmini güç sistemi çok önemlidir ve planlama çalışmalarının temelini oluşturur.Yük tahmin hatalarının miktarı özel bir önem taşır. Bu durumda karar verme sorunları, bir yandan sınırlı bütçeler ve bir yandan maliyetleri en aza indirme hedefi, diğer yandan da güç uzmanları ve mühendislerinin gelişmiş ve pahalı ekipman satın alma baskısı ile elektrik kullanımının kontrolsüz bir şekilde genişlemesi ile karşılaşıldığında daha da kötüleşir. Tahmin edilen yük, gerçek yükten daha az ise, güvenilirlik ve dolayısıyla hizmet kalitesi azalır ve bu, zorunlu kapatmalara bile yol açabilir. Bu, kendi başına sistem analistlerinin işini biraz zorlaştırır ve diğer yandan, bir sonraki sefer gerekli miktardan önce tahmin edilirse, çok fazla yatırım boşa harcanır ve finansal ihtiyaçlara yol açar. Gerçek şu ki, elektrik tüketimi sabit değildir ve her zaman çeşitli zamansal, çevresel, ekonomik ve dalgalanan parametrelerin doğrusal olmayan bir fonksiyonudur. Elektrik tüketimindeki değişiklikler, enerji şirketlerinin farklı zamanlarda güç sistemlerindeki enerjiyi daha iyi yönetmek için gereken bilgileri önceden tahmin etmesini gerektirmiştir.

Elektrik sektöründe güvenilirliğin ve verimliliğin artırılması, elektrik sektöründe maliyetlerin ve işlem maliyetlerinin düşürülmesi, tüketicilere daha fazla seçenek sunulması vb. Ekonomik, sosyal, politik teşvikler kaçınılmaz olarak elektrik sektörünün yönetimini dönüştürmüştür. Elektrik endüstrisinin yeniden yapılandırılması olarak anılır. Elektrik sektörünün yeniden yapılandırılmaya başlaması ve rekabetçi elektrik piyasalarına girmesi ile üretim, iletim ve dağıtım kısımları büyük ölçüde özel sektöre devredilmiştir.

Yük tahmini güç sistemlerini planlamak için en önemli araçlardan biridir. Sistemin tüketiciler için güvenilir güç sağlamasına yardımcı olur. Tahmin, güç sistemi planlamasının araştırılması için ilk adımdır. Elektrik sistemindeki elektrik enerjisi enjeksiyonları ve geri çekilmeleri kayıpları her zaman dengelenmeli ve elektrik enerjisi üretme maliyeti talebe göre değişmelidir. Elektrik tüketildiğinde sağlanmalıdır, çünkü bu enerjiden tasarruf etmek uygun değildir. Elektriğin bu önemli özellikleri, güç sistemlerinin yönetimini çok karmaşık hale getirdi. Bu konu çok

önemli olduğu için, birçok araştırmacı daha iyi sonuçlar elde etmek için farklı yöntemler kullandı.

Tüm bu yöntemler Yapay zeka yaklaşımı ve İstatistiksel yaklaşım olmak üzere iki gruba ayrılabilir. Yapay zeka (AI), insan beyni işlevini düşünme ve karar vermede taklit etmeye çalışmaktır. Yapay zeka, farklı özellikler arasındaki ilişkilerin araştırılmasında yük tahmininde çok pratiktir. Ertesi günün tüketimini bilmek güç şirketleri için çok yararlıdır. Kısa vadeli tahmin, yükü bir saatten birkaç güne kadar tahmin etmek için kullanılabilir.

Bu tez Yapay Sinir Ağları (YSA) ve diğer yöntemleri kullanarak bir tahmin modeli oluşturmuştur. YSA, verileri doğru olmayan ve karmaşık işlevlerle işleyen güçlü bir makine öğrenme yöntemidir. Bu algoritma yeni bir algoritma değildir, ancak son yıllarda popüler hale gelmektedir. Buna bağlı olarak nedenlerini şöyle açıklayabiliriz, birinci nedeni bilgisayarların işleme gücünün artmasıdır ve ikinci nedeni ise veri bankasının genişlemesidir.

Günümüzde elektrik ağının çoğu elemanları ölçme ve kayıt işlemlerini yapma özelliğine sahiptir. Tüm bunlar bizim için çok önemli veri tabanı hazırlamaktadır.

Her birinin özel bir durum için uygun olduğu çeşitli YSA türleri vardır. Bu tez YSA olarak kullanmak için İleri Beslemeyi (FF) kullanmıştır. Bir ileri beslemeli sinir ağı, birimler arasındaki bağlantıların bir döngü oluşturmadığı bir yapay sinir ağıdır. Bu nedenle, tekrar eden sinir ağlarından farklıdır.

İleri beslemeli sinir ağı, geliştirilen yapay sinir ağının ilk ve en basit türüdür. Bu ağda, bilgi, girdi düğümlerinden, gizli düğümler (varsa) aracılığıyla, sadece bir yönde ileri çıkış düğümlerine doğru gider. Ağda döngü yoktur.

Sinir ağları hem sınıflandırma hem de regresyon vakaları için uygun bir yöntemdir ve tahmin bir regresyon problemidir. Doğrusal etkinleştirme işlevi son katman için en iyi seçimdir.

Hangi kombinasyonun en iyi sonuçları verdiğini bulmak için farklı özelliklere sahip kombinasyonlar kullanıldı. İlk çalışmada takvim özellikleri, ikincisinde ise hava özellikleri kullanılmıştır. Son olarak, bu çalışmada her iki özellik de son çalışmada kullanılmıştır. Bu üç çalışmanın hatası, tüm özelliklerin bir modelde kullanılmasının daha iyi olduğunu göstermektedir. Hava özelliklerini seçmek için kullanılan korelasyon yöntemleri. Gökyüzünün bulanıklığı gibi bazı hava özellikleri arasında çok az korelasyon vardır, bu nedenle öğretim verilerinden çıkarılmıştır.

Bu tezin temel amacı kısa vadede yükün tahmin edilmesidir, ancak model daha uzun süre de kullanılabilir. Model, Gebze şehri için beş yıldan fazla veri kullanılarak eğitildi. Hedef değer ortalama yük tüketen miktardır. Hava durumu verileri iki farklı istasyondan gelen bilgilerden oluşur. Eksik değerlerin düzgün bir şekilde doldurulmasına yardımcı olur. Bu veriler Türkiye Meteoroloji Genel Müdürlüğü tarafından sağlanmaktadır.

Tahmin modelinin dört adımı vardır. Veri ön işleme, ikinci adımın modeli tanımladığı ilk adımdır. Modeli tanımlarken birçok parametre ve hiper parametre ayarlanmalıdır. Bu elemanların doğruluk üzerinde çok önemli bir etkisi vardır, aynı zamanda modelin yetersiz oturmasını ve aşırı takılmasını önleyebilirler. Üçüncü ve dördüncü adımlar, verileri modele uydurmak ve model doğruluğunu doğrulamaktır.

Bu makalede üç farklı doğrulama fonksiyonu kullanılmıştır: RSQ, Ortalama mutlak hata (MAE) ve Ortalama kare hatası (MSE). RSQ sıfır ile bir arasında bir sayıdır ve

bu miktar bir ne kadar yakın olursa iyidir. MAE için en iyi değer sıfırdır. MSE, bir tahmin edicinin kalitesini ölçen negatif olmayan bir değerdir. Daha iyi modeller küçük MSE değerlerine sahiptir.

Tez, sonuçları karşılaştırmak ve en iyi algoritmayı bulmak için diğer üç algoritmayı kullanır. Bu algoritmalar K Yakın Komşular, Doğrusal Kayıt ve Karar Ağacıdır. Sinir ağları en iyi sonuçları verir. Daha fazla özelliğe sahip daha fazla veri kullanılarak ve sinir ağının modeli için daha iyi mimari tasarlanarak doğruluk artırılabilir.

Karar Ağacı , ağaç yapısı şeklinde basit bir sınıflandırma algoritmasıdır. Data modelimizin dinamiklerine göre skorlama yöntemlerini kullanarak basit karar ağaçları oluşturabilir. Bu sayede giriş değerine göre çok hızlı sınıflandırma yaparak karar alabilir, hatta tahminlerde bulunabilir.

1. INTRODUCTION

Energy is the first characteristic of most of the world's debates in recent years that plays an essential role in the development programs of all countries. Different studies relate the energy consumption to the economic growth of the countries [1–3]. Most countries are developing extremely rapidly, and their energy demand is increasing regularly. The entire world is facing environmental issues, as fossil resources supply most of the energy. These resources are depleting at an alarming rate. Consequently, managing the energy resources and consumption properly and optimizing the production to minimize the costs is necessary [4].

Decision making for massive grids needs various real-time data from consumers, producers, and distribution systems. These real-time data are not available in traditional systems. So we need new, modern and efficient grids to obtain these sorts of data.

The researchers have established ways of efficient production, transmission, and distribution for many years. Khoi, M.Begvoic, and Damirin introduced the term smart grid in 1997 [5]. Although the smart grid does not possess a universal definition, several papers use various definitions. The smart grid is determined by the European Technology Platform (ETP) as "an electrical system that intelligently integrates the actions of all elements connected to it, to make the electricity secure and efficient" [6].

Electricity has some characteristics, which make a primary difference between electricity and other energy sources. It is hard to store economically, and the demand for electricity changes over time. Besides, production and consumption should be equal at any moment.

Load forecasting is one of the fundamental elements of power management systems in planning and operating. It contains accurate prognostication of electricity consumption over different time zones by using prior data. The central part of the data is related to previous load consumption, but other information like weather conditions also has a significant role.

1

The goal of this thesis is to perform short-term load forecasting for a small city called Gebze. This city is located in Kocaeli Province, Turkey.

After İzmit, Gebze is the largest city in the province. According to the latest reports, the population of Gebze is around 642,726. This paper studied Gebze consumption for five years from 1/10/2013 to 20/9/2019.

Various sets of data were used in this thesis as a feature. Most of these features are related to weather conditions like temperature, humidity, wind speed, and pressure that were provided by The Turkish State Meteorological Service. The data about the weather were gathered from two different weather stations, Kocaeli, and Gebze for five years. The consumption data for Gebze was taken every 15 minutes, but the weather were gathered every hour.

The benefit of load forecasting for a city like Gebze is that the operators acknowledge the average and peak value of load consumption so they would be able to design several programs to manage this value. Besides that, they could plan for future development more efficiently.

1.1 Purpose of Thesis

The first purpose of this thesis is to forecast the Gebze load curve with a shallow fault by using artificial neural network methods based on historical and weather data.

Since numerous features impact load consumption and most of them are correlated to each other, it is necessary to choose just essential features. Finding the optimum number of these features is quite significant. The training process could be incomplete if the number is insufficient, and the network could be very complicated if the number is significant. Hence the second purpose is to determine which feature has more impact on the model.

1.2 Literature Review

One of the crucial elements of the energy management systems in electrical power grids is load-forecasting. Unit commitment, decision-making, reducing the spinning reserve capacity, and planning the maintenance programs could be achieved by accurate load forecasting. Power system operators need to have a perspective about load consumption algorithms to operate the power system efficiently. The power

system must continuously follow the load demand rate, so all the parts of the power system need to have access to load forecasting information.

Generation companies utilize load forecasting methods to manage their resources to reach load demand. Transmission corporations utilize load forecasting methods to reduce congestion and overload by optimizing the transmission lines' power flow. Short-term load changes have a minimal effect on distribution systems, and short term load forecasting is not used by these utilities [7]. Load forecasting plays a vital role in having an optimal, secure, and reliable system. In the past, engineering strategies relied on charts and tables to predict future consumption. These strategies mainly considered meteoric and calendar information. These data are still being used in the recent engineering approaches, but with new methods.

Significant settlements depend on load forecast with lead times of minutes to years. According to different time zones, load forecasting has four main classifications, very-short-term, short-term, medium-term, and long-term load forecasting [8].

The first paper related to forecasting was published in 1918 [9, 10]. After that, lots of researches were published for load forecasting in recent years. These papers look at forecasting from different aspects. This thesis focused primarily on short-term load forecasting among them. These papers use different techniques to solve the forecasting problem. These techniques have two classifications, statical approach [11], and intelligence approach [12].

Different offline and online methods were reviewed by Abu-El-Magd and Sinha, which covered spectral decomposition, multiple regression approach, stochastic time series approach, and other methods [13]. This review also considers the advantages and disadvantages of each technique.

I. Moghram and S. Rahman compared five different techniques in their review paper [14]. They highlighted the different aspects of time series, exponential smoothing, multiple linear regression, state-space, and knowledge-based techniques. Algorithms implementing these forecasting techniques have been programmed and applied to the same database for direct comparison of these different techniques.

A. D. Papalexopoulos suggested a regression-based method to short term load forecasting [11]. The paper used this method to forecast the maximum hourly load for the next 24 hours. Temperature modeling by using heating and cooling degree

3

functions, weighted least square technique are some of the regression techniques that were used. The weather and load variables have a nonlinear relationship; this nonlinearity is modeled using a transformation technique in Haida and Muto's paper [15]. B. Krogh, E. S. de Llinas, and D. Lesser merged regression methods with the ARIMA model in their research [16].

Hippert et al. used the artificial neural network for reviewing short term load forecasting in more than 40 papers and researches in the 1990s. This paper revealed two main reasons for the possibilities of using ANN, which is over fitting and unsystematically carrying out. This paper also outlines designing a forecasting model in four steps:

1) Preprocessing

2) Architecture of the network

3) Performing the network

4) Validation.

The Back Propagation (BP) model was used in the [17] for ANN modeling. Nevertheless, [18] used Radial Basis Function NN for forecasting. The paper shows that this method is more efficient than the BP model. Besides, the training time for the proposed model is somewhat shorter.

M. H. Choueiki and his coworkers introduced the Weighted Least Squares (WLS) technique and used it in training a neural network to solve the short-term load forecasting problem. Therefore, this paper recommended that the weighted least squares procedure be further studied by electric utilities that use neural networks to forecast their short-term load and experience large variabilities in their hourly marginal energy costs for 24 hours [19]. The data gathered from weather stations mostly consisted of false data. To overcome this issue, especially when the temperature error increases, [20] they used a multistage neural network. The [11] studied the sensitivity to the weather prediction in the regression-based technique. Some of the significant innovations in this paper are temperature modeling using heating and cooling degree functions, and including accurate holiday modeling by using binary variables.

The functional supervisory technique was used in 1996 in Taiwan by Chen et al. The paper divided consumers into industrial, commercial, and residential. The load was

considered with the relationships between load and temperature and the type of consumption.

l. Al-Fuhaid et al. included more than one weather data for forecasting. Before that, so many papers used just temperature as a weather variable. Nonetheless, this research utilized moisture and temperature and combined these two variables. In 1998 a Recurrent Neural Network (RNN) was offered by Vermaak and Botha. They used time alongside the other variables in their research.

After a while, researchers started to combine ANN with other methods like Fuzzy Logic and genetic algorithms. Papadakis et al. introduced a three-step fuzzy artificial neural network method. Likewise, Dash et al. used fuzzy logic but with the impact of the calendar variables. Kung et al. used genetic algorithms in the training process.

Some researchers also used other techniques to enhance the efficiency of ANNs. For instance, Chow and Leung applied time-series techniques. They used the Non-Linear Autoregressive Integrated method to predict the consumption of Hong Kong.

Drenza et al. introduced a new technique for selecting training data using the K nearest neighbor method.

Kandil et al. used a multilayer perceptron method. Multilayer perceptron employed forecasting by weather data and real load variables.

1.3 Outline of Thesis

This thesis includes five chapters in this format:

Chapter 1 explains the purpose of this research which is short load forecasting and investigates this topic in the previous researches.

Chapter 2 describes load forecasting, types of load forecasting, different methods, and different factors that impact forecasting.

Chapter 3 introduces Artificial Neural Networks for short term load forecasting. This chapter also discusses the architecture of the network.

Chapter 4 describes how important features were selected among other features, and also describes how artificial neural networks were implanted into data. Moreover, in the end, it discusses the results of this method.

Chapter 5 is the conclusion of the thesis by considering the results and tables.

2. LOAD FORECASTING IN POWER SYSTEMS

Electric load forecasting is the process used to anticipate future load, given historical load and weather information and current and forecasted weather information.

2.1 Power System Planning

The power system is extremely complicated and consists of different interconnected parts. This extensive system has many elements, like generators, stations, transmission lines, and transformers. The primary purpose of this system is to produce and distribute reliable energy to consumers.

The number and behavior of these consumers are changing fast, so power systems must be updated to be able to meet the user's needs. Developing the system to satisfy future needs is called power system planning [21]. The main concepts for these activities are:

• Quality

• Reliability

• Stability

 • Security

 • Economy

power system planning has essential considerations such as when and where the changes must be done. In any case, loads must be satisfied.

2.2 Load Forecasting

Forecasting is the first step for examining power system planning. Forecasting has two sides. The first side is demand forecasting, and the second side is energy forecasting. The capacity of power system elements is defined by demand forecasting. Furthermore, the second side of forecasting reveals the required type of generators. Some of the advantages of load forecasting are listed below:

- Forecasting helps the power system operators to have an accurate and complete plan for the future.
- Forecasting minimizes the economic risk for system operators.
- The forecasting maximizes the power generation by avoiding working under or over a generation.
- The best time for maintenance is learned by forecasting by studying the consumer's behavior pattern.
- Forecasting could be used to determine the best place for building new structures

2.3 Loads in Power System

Loads are the last part of the energy system. They consume electricity and convert it into another type of energy. According to the different behaviors in the alternating current, they have three classifications: resistive, capacitive, and inductive. Load curves are used to recognize the patterns in users' behaviors. Load curves are the graphical illustration of load in proper time sequences. The curves could be hourly, daily, monthly, seasonally, or annually.

Load curves have so much information about the power system, like maximum demand, average load, load factor, and variation of load [22].

For better investigating loads characteristics, many factors should be introduced. Demand Factor (DF), Load Factor (LF), Diversity Factor (DIF), and Utilization Factor (UF) are the most significant factors, which help for categorizing the loads.

$$DF = \frac{max.demand}{2aconnected\ loads} \qquad (2.1)$$

$$LF = \frac{avg.demand}{max.demand} \qquad (2.2)$$

$$DIF = \frac{sum\ of\ max.demands}{max.demand\ of\ power\ station} \qquad (2.3)$$

$$UF = \frac{max.demand\ on\ power\ station}{rated\ capacity\ of\ power\ station} \qquad (2.4)$$

According to these factors and characteristics, users of power systems categorized into four different groups:

2.3.1 Residential

Table 2.1 demonstrates the attributes for residential users. The biggest uses of electricity in the residential sector are air conditioning, lighting, space and water heating, and appliances and electronics.

Table 2.1 : Residential users' characteristics.

	DF	DIF	LF
Residential	70 -100 %	1.2 – 1.3	10 - 15 %

2.3.2 Agricultural

Table 2.2 displays the criteria for agricultural sectors.

Table 2.2 : Agricultural users' characteristics.

	DF	DIF	LF
Agricultural	90 -100 %	1 – 1.5	15 – 25 %

2.3.3 Commercial

The commercial sector includes government facilities, service-providing facilities and equipment, and other public and private organizations. Table 2.3 shows the attributes for this sector.

Table 2.3 : Commercial users' characteristics.

	DF	DIF	LF
Commercial	90 -100 %	1.1 – 1.2	25 – 30 %

2.3.4 Industrial

The attributes for these users are:

Small-unit: 0 – 20 kw

Medium-unit: 20 – 100 kw

Large-unit: more than 100 kw with DF = 70 -80 % and LF = 60 – 65 %

2.4 Factors Effecting The Load Forecasting

Various forecasting methods were described. Each method has an accuracy rate, but it is likely to change in different circumstances. Many factors can affect this accuracy

rate. The operators must be acquainted with them and consider their influence in calculations.

These factors vary from case to case, a special factor could have positive effects in one case, but it may not be valid for other cases. Therefore, before starting every project and choosing the best method, the engineers must consider all the factors.

These factors can be categorized into two groups. The first group consists of the factors from nature and obligatory circumstances, and the second group includes the decisions that operators make. The most significant factors are described below:

2.4.1 Economic factor

Currently, electricity consumption is the exact criterion for judging the economy of a country; the economy's growth should increase. More significant economies not only have more consumption rates but they also have different demand patterns. For instance, developed countries and undeveloped countries have different patterns for their daily load curves. The primary peak occurs between 11 a.m. and 4 p.m. in developed countries and after 6 p.m. in undeveloped ones.

The pricing system of electricity and is another critical issue. Higher prices force consumers to decrease their consumption. Some countries control the load curve by using pricing systems. These systems have different prices for electricity. These strategies control and reduce demand, so they are essential for short-term load forecasting.

2.4.2 Weather

The most significant independent variable is the weather. Weather conditions influence domestic and agriculture subscribers more than industrial ones. Because of weather variables, having different load curves for each season is a possibility. Power systems minimize the operation cost by using weather forecasting models in load forecasting.

The weather factors are listed below:

1. Temperature

2. Humidity

3. Rainfall

4. Wind speed index

5. Cloud cover index

2.4.3 Time

Time has the highest impact on the load curve which is periodic. The load curve holds patterns in each day, month, season, and year. These patterns vary for different places. The consumption rate in the 24-hour cycle is meager during the night-time. At 6:00 p.m., it starts to increase to its peak value. Weekends and the days before and after the weekends have different consuming algorithms. Months patterns are related to seasonal changes.

The total consumption in summer and winter is much higher than in spring and fall. Most of the people use air conditioners in these two seasons.

2.4.4 Random

Humans are unpredictable creatures that no science is able to anticipate their whole behaviours and it even twists more when the case is related to a population. Every engineering project has its undesired issues.

Random factors are divided into two main groups. Changes in consumer consumption cause the first group, and the second one is related to faults in the power system.

2.5 Time Periods

Fig 2.1 depicts the different time horizons for forecasting
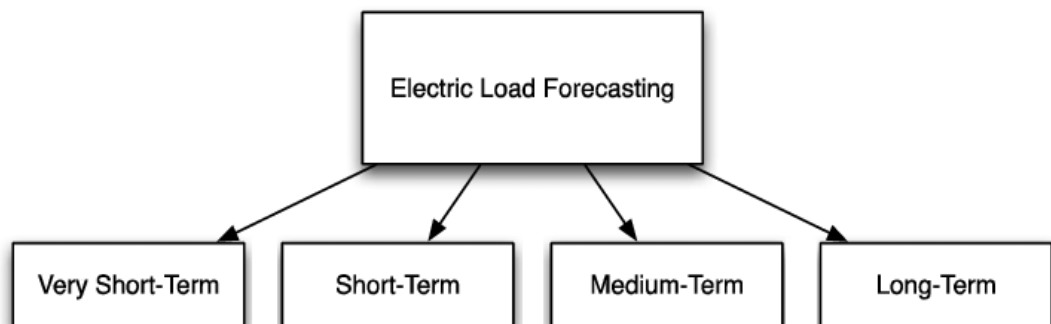
Figure 2.1: Time Horizons For Forecasting [23].

2.5.1 Very short term load forecast

The timing for Very Short Term Load Forecast (VSTLF) starts from several seconds to several minutes. VSTLF is used in economic dispatch and load frequency control and predicts the load from thirty seconds to thirty minutes ahead.

2.5.2 Short-term load forecast

Power system operators use Short Term Load Forecasting (STLF) to overcome the overloading and increasing the reliability of the power system. The timing for STLF starts from an hour to a month. STLF is very crucial for operators, and the information gained from STLF is extremely valuable. The short term is the primary interest of this paper, so the main methods for STLF and main elements that affect STLF are shortly discussed in the next part.

2.5.3 Medium-term load forecast

This type of forecasting varies between one month to one year. Most of the generator units use a Medium Term Load Forecast (MTLF) to predict the amount of needed fuel, but on the bigger scale, all power system operators use MTLF to anticipate the future expansion, and material purchases.

2.5.4 Long- term load forecast

Long- Term Load Forecast (LTLF) anticipates load for more than one year. Long-term forecasting is used for preparing the needs of the station for the future, from fuel to workers. LTLF helps operators to make decisions for long periods.

2.6 Forecasting Techniques

Forecasting is a very complicated process. Numerous methods are used for achieving higher accuracy for each different case. These methods are based on some fundamental techniques. The forecasting techniques is widely divided into three main groups.

1. Correlation techniques

2. Extrapolation techniques

3. Combination of correlation and extrapolation techniques

2.7 Short Term Forecasting Methods

The mentioned kinds of load forecasting need different prediction methods. In other words, every condition needs to have its requirements. Furthermore, it is essential to choose the proper method for each circumstance. The econometric approach and end-use methods are two main ways that could be used for long and medium-term forecasting. Nevertheless, operators have more options for short and concise term forecasting. Lots of different methods were used over the years, from mathematical approaches to computer base approaches.

Short term load forecasting could be classified into two categories: statistics methods and artificial intelligence methods. Nowadays, researchers found a third group called hybrid. The hybrid methods use more than one method and combine them to reach better efficiency. The essential approaches are as follow.

2.7.1 Linear regression

Linear Regression (LR) is a simple and basic tool for predictive analysis which considers two conditions:

1. Whether the result is accurately predicted by input variables

2. Which input variables have more influence on the accuracy of outputs

These predictions are applied to find the connection between the dependent and independent variables like in the line equation.

Forecasting is our primary interest in using linear regression. LR can be used in forecasting future values and the influence of changes. In various cases, there is more than one variable that influences the output. Multilevel Linear Regression (MLR) is utilized more than other kinds of linear regressions. MLR outlines how one variable which is linearly influenced by many predictor variables can predict electrical load at a defined time by using weather and other variables. Correlation analysis is used to select these variables, which have more influence on the load curve [24,25].

2.7.2 Similar-day approach

One of the approved methods for short term load forecasting is the Similar-Day Approach (SDA). At first, some of the more essential features are chosen, features like Weather,

date, the day of the week, and season. These features have a higher impact on forecasting outcomes.

In the next step, a day with similar features is selected and then a prediction is made on the load for a particular day. Sometimes, this search results in more than one day, so choosing the right day among these days could reduce the total error [26].

2.7.3 Time series

Time series (TS) is one of the oldest and most practical methods in load forecasting. Every data arranged according to the time could be included in the time series. They could have equal or unequal time intervals. Time series forecasts the future load pattern by utilizing previous load data, and ignores other features like temperature. This method has three main disadvantages:

1. This method is not very accurate.

2. This method needs a big database, and most problems do not have significant historical data.

3. This method could not adapt to the changes in the circumstances

The forecasting process utilizes the linear filter to process the white noise with zero mean and constant variance as an input. Forecasting methods could be classified according to this filter.

The autoregressive outlines the output according to previous data and a noise signal. Unlike autoregressive, the forecasting output is calculated according to current and previous data and a noise signal in the moving-average method. The noise is composed of the fault which occurs in the forecasting process.

The autoregressive moving-average method uses two previous methods to determine the output. Then the load curve with both Autoregressive (AR) and Moving-Average (MA) could be predicted.

These three methods are part of the stationary process. The stationary process has a constant mean and variance. Nevertheless, processes do not always have that system. A different method is needed for non-stationary situations. Some of the suitable methods for these circumstances is Autoregressive Integrated Moving Average (ARIMA), Autoregressive Moving Average with Exogenous Variables (ARMAX) and Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX).

2.7.4 Artificial intelligence methods

Artificial Intelligence (AI) is an attempt to imitate the human brain function in thinking and decision making. Artificial intelligence is useful in load forecasting during investigating the relationships between different features. The artificial intelligence has two main classifications: strong and weak. The weak AI is designed to do one particular task, and it is also being used in so many cases like assistants on smart phones. On the other hand, strong AI is designed to solve more complicated problems. Strong AI can be utilized in the case that it could make a decision without human intervenes like driverless cars.

The artificial intelligence has different techniques, and each technique has its specialties. The most important ones are:

1. Genetic Algorithm (GA)

2. Support Vector machine (SVM)

3. Fuzzy Logic (FL)

4. Expert Systems (ES)

5. Artificial Neural Network (ANN)

2.7.4.1 Expert system

Expert systems have been widely used since they were developed in 1965. They were used in chemistry initially, and now they are beneficial not only in engineering fields but also in financial problems.

This method is a sort of computer program, which solves complicated problems by utilizing artificial intelligence. The artificial intelligence applies some rules based on IF-THEN, on the database to forecast. Most of these rules are constant in the forecasting process, but some of them must change frequently.

2.7.4.2 Support vector machine method

Supported vector machine is one of the machine learning methods used in both registration and classification problems. Such a machine can separate the new data according to trained data. The separating process is straightforward when the data could solve linearly. Nevertheless, in some cases, different categories are mixed, and there is no way of separating. In this situation, SMV uses other dimensions to change

its point of view. Some of these inseparable data could be separated simply by using one or more dimensions. The support vector machine changes the nonlinearity into linearity by using the kernel functions.

2.7.4.3 Genetic algorithm method

The genetic algorithm is a kind of optimization method which is inspired by the theory of natural selection. The genetic algorithm is efficient in electrical load forecasting and other modeling problems.

Choosing the most proper individuals is the first step of natural selection. The next generation consists of children of chosen individuals. The attribute of parents is inherited by these children. They have higher chances to survive with better attributes. Finally, after various repetitions, all of the members have the most desirable attributes.

2.7.4.4 Fuzzy logic

The Boolean logic is the central concept of Fuzzy logic (FL). It has three principal parts, linguistic variables, rules, and sets. The linguistic variables are inputs of the system. These inputs are transformed into output by using sets of rules. Fuzzy logic does not need complicated mathematical equations because it consists of linguistic variables, so they are understandable. FL could adapt to new conditions by adding or eliminating a rule.

2.7.4.5 Artificial neural network

The artificial neural network is a network of neurals that can solve problems. The process has two parts. In the linear part, each node's value multiple into each weight and with the bias. The second part is nonlinear.

The result of the first part is transformed into a nonlinear value with the help of the activation function.

2.8 Medium and Long-term Term Forecasting Methods

The need for medium and long term load forecasting is growing by the increase of the use of renewable energies. The future needs more accurate and secure grids. Short term forecasting methods were mostly used in medium and long term forecasting with sufficient accuracy.

Medium and long term method needs more amount of data and training. There are three main methods for the medium and long term:

1. Trend Analysis (TA)
2. End-Use Analysis (EUA)
3. Econometric Analysis (EA)
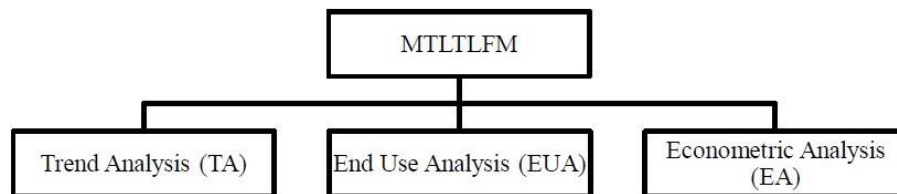
Figure 2.2: Long and Medium Forecasting [27].

2.8.1 Trend analysis

TA uses the consumption of the past year to forecast consumption for next year. The basic idea for TA is that possible future developments are predictable by changes in the past.

Using this method has many advantages; some of them are:

1. TA has a very simple procedure.

2. TA is very quick.

And these are some important shortcomings, besides those advantages:

1. TA relies on the old demand curves, so it could be inaccurate.

2. TA method does not innclude the important factors which have an impact on consumption.

2.8.2 End-use analysis

Generally, the people who are using a finished commodity were called end-users. The consumption of all commodities is related to these people, and electricity is not an exception, and this is the main idea of EUA. Data from the past is used to find out how much electricity each consumer used in each device. The predicted consumption for each residential user is equal to the amount of consumption for each device multiplied by the number of devices. The result is an estimation of the electricity demand in the future for a residential area. EUA could be inaccurate because this method assumed that the consumers' behavior is constant over the forthcoming years. Furthermore, this

assumption could not be valid for extensive periods. The second cause of inaccuracy is that this method uses current data. The same approach could also be used for commercial and industrial users.

These are some of the essential advantages of EUA:

1. Less historical data is required.

2. Shows how much each type of users, consume.

2.8.3 Econometric analysis

The econometric analysis makes use of both previous methods. Complicated mathematical equations show the relationships between demand and the factors that impact it. These equations were used to find the value of essential factors. So the accuracy of the model depends on the accuracy of these factors. The main advantages of EA are:

1. Provides the reasons of demand changes

2. Forecasts for each type of user individually

3. ARTIFICIAL NEURAL NETWORK

The Neural Network (NN) is a powerful machine-learning method that processes the data with non-linear and complex functions. This machine-learning algorithm is not new, but it has become popular in recent years. The lack of variable data and the inadequate power of the computers are two critical barriers. The simple networks have three layers. The first one and the last one is related to inputs and outputs, and the middle layer is used for calculation. For more complicated problems, the number of layers can be increased.

The human brain's functionality inspired the concept of neural networks. The brain works non-linearly to solve the problems and make decisions. The next chapter explains the NN thoroughly.

The neural networks have many advantages to other algorithms. The most important items are listed below:

- Neural networks predict the future load without planning a system model

- They accept a large ratio of inputs

- They are resistant against noise

- They have adequate performance in the nonlinear cases

- They do not need any mathematical equations for modeling the systems

- They can handle multiple tasks without losing the efficiency

- They can achieve better results with fewer data

No method is without disadvantages and in this case the neural network is not an exception. These are some of the disadvantages of ANN:

- The neural networks are running with numerical data, so the non-numerical data must be transformed into numerical

- The deep networks are needed to process with parallel processing capability

- They need to be trained before each process, and this is time-consuming.

- They are not adaptable to circumstances

- The training process could be time-consuming

3.1 Benefits of Neural Networks

The neural network is not a new method, but nowadays, other deep learning methods have become very popular. Fig. 3.1 shows the performance according to data; the performance could be anything like training and accuracy.
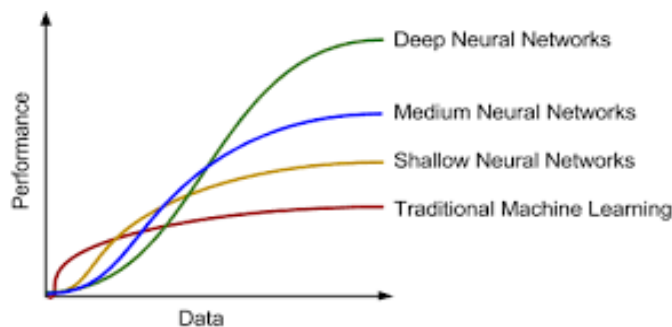
Figure 3.1: Performance VS. Data [28].

As it is evident, the performance for different methods is the same; when the amount of data is skimp. The curve of traditional methods becomes stable after a while. It means that these methods could not achieve better performances for bigger databases. Nevertheless, the performance of neural networks is related to the amount of data, so it is reasonable to use NNs for huge problems.

The neural networks have two conditions for achieving high performance:

1. They need adequate data for training

2. They need to be able to train large amounts of data

Also, neural networks use three tools to enhance performance: data, computation, and algorithms.

3.2 Kinds of Neural Networks

Neural networks have diverse kinds of variables that are used to justify NN for the primary purpose. Number of layers, number of cells in each layer, number of the hidden layers, and the activation functions are all part of these justifications. Because of these different choices, NN has different sorts of architecture. They are categorized into four main groups, which will be mentioned in the following part.

3.2.1 Feed-forward neural network

All the neural networks consist of layers and nodes in each layer and the way they are connected is of great importance.

The Feed-Forward (FF) connection is one of the most common methods, which became functional in the 1950s. All the nodes are connected to other nodes in the feed-forward method, and the data continues only in one direction, from input to output nodes without any back-loop.

Fig 3.2 shows the FF network with one hidden layer and two nodes in the input and hidden layer.
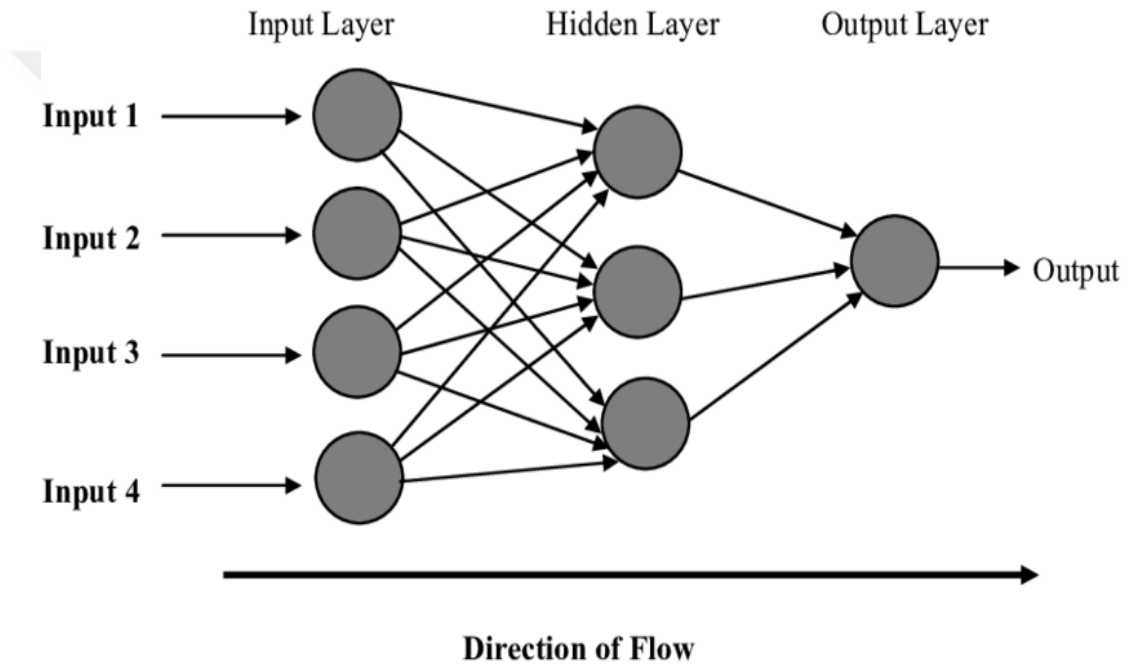
Figure 3.2: Fead Forward Neural Network [29].

3.2.2 Radial basis function neural network

This structure has two layers. The range of any point from the center is the main plan in the radial basis function. Functions are linked with the features in the hidden layer. The obtained output is used for calculating the output of the next step. The power systems are becoming more prominent each day.

The complexity of power systems increases the risk of blackouts. This structure helps systems to restore as soon as possible.

3.2.3 Multilayer Perceptron

Multilayer perceptron neural networks have more than two fully connected layers, and they are used when the Data is not linearly separated. The activation functions are non-linear in this NN.

The multilayer perceptron structure is mostly used in translation and speech recognition programs. Fig 3.3 depicts the Multilayer perceptron neural networks.
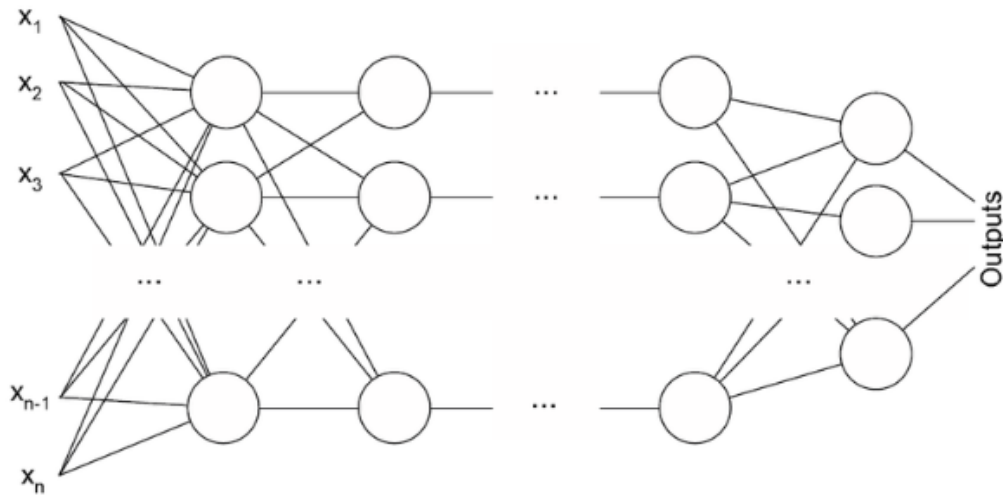
Figure 3.3: Multi-Layer Neural Network [30].

3.2.4 Convolutional neural network

Convolutional networks have at least one convolutional layer. In some cases, convolutional layers are completely interconnected but they can also be pooled. This neural network is mostly used in language processing and video or photo recognition programs because the convolutional layer helps the network solve more complicated problems with fewer parameters. The convolutional network is also very useful in signal and image processing.

3.2.5 Recurrent neural network

The output of every layer in Recurrent Neural Network is saved in memory then it is fed back to the input. The format is the same as the feed-forward neural network in the first layer. The next layers use this memory in calculating. In other words, the network works like feed-forward as usual but saves the necessary information. If the forecast is incorrect, the network detects the fault and operates towards making the correct

forecast through the backpropagation. Recurrent Neural Network is commonly used in the text-to-speech conversion.

3.2.6 Modular neural network

This structure consists of different independent networks. These networks work separately during the calculation. The complex problems could be solved faster when they are processed Integrated. Each network in a modular neural network system solves a special problem individually. Fig 3.4 depicts the modular neural network.
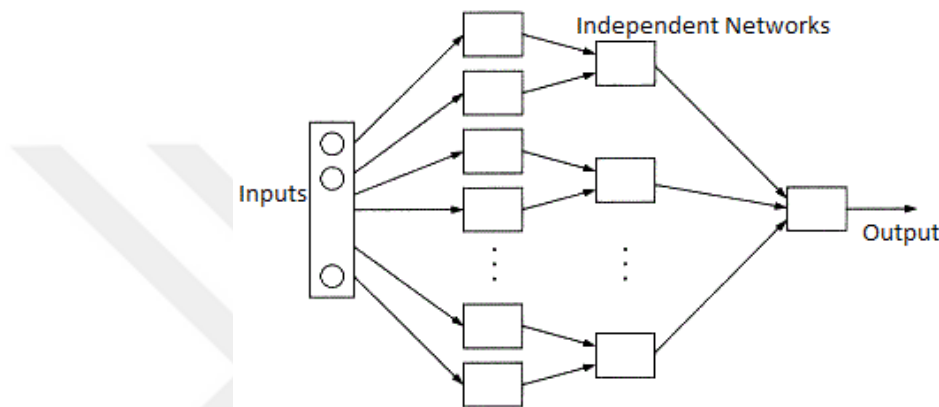
Figure 3.4: Modular Neural Network [31].

3.2.7 Sequence-to-sequence models

This models consists of two RNN. Sequence to sequence has an encoder and decoder for processing the input and output. These units could have equal or unequal parameters. It is suitable for the cases where input and output have different lengths, such as machine translation systems.

3.2.8 Recursive neural network

The second structure, which was introduced in 1990, is Recursive Neural Networks (RNN). These non-linear models were widely used in machine-learning problems. RNN can handle deeply related issues and solve both supervised and unsupervised data, and they are very efficient in tree-shape models and hierarchical ones.

Increasing the layers expands the network's complexity, and handling this complexity is one of the largest problems in NN. The back-propagation algorithm, which is the most used in RNN, could solve this problem. The back-propagation can be used in regression and classification.

3.3 Neural Networks for Forecasting

Neural networks have been used in the short term load forecasting since the 18th century. These networks imitate human brain functionality in programming languages. Neurons are the main component in neural networks. The artificial interconnection of these neurons makes the network crucial for calculations. Neurons and their mathematical relationships are determined in the figure.

As the figure shows, the neural networks have three critical features in their relationship: inputs, weights, and bias. The output is formed by the combination of each input with related weights plus bias. Hence the output could be calculated in this form:

$$A_1 = w_1 * x + b_1 \tag{3.1}$$

$$y_1 = g(A_1) \tag{3.2}$$

3.4 Parts of Neural Networks

Each neuron can make simple decisions and can feed them to other neurons, organized in interconnected layers. The neural network can emulate almost any function, and answer practically any question, given enough training samples and computing power. A "shallow" neural network has only three layers of neurons.

Neurons send the result of the decisions, which they make, to other neurons. The interconnected networks of these neurons could solve any problem. The lack of training data and computer processing power are the only barriers to solving the problems. A typical neural network consists of three kinds of layer, which will be explained below:

3.4.1 Input layer

The input layer, which is the first layer of the network, receives real values input data sets. Each of these real values is sent to one particular neuron. Then the input layer passes the information to the next layer.

3.4.2 Hidden layer

The hidden layer takes the input layer's data and delivers them to the output layer, so this layer is only connected to other layers and has no connection from the outer side.

The hidden layer could be more than one layer; the deeper networks have more than one hidden layer. Having several hidden layers could enhance the whole network's power, but having more layers is equal to more complexity, so finding the optimum number of the layers of every problem is very critical. Extracting existence patterns is the primary purpose of hidden layers.

3.4.3 Output layer

This layer provides the output information after several calculations for the user. The output layer fully processes the data which is received from the hidden part.

3.5 Activation Function

Most of the natural phenomenon could be classified through the perspective of linearity into two classes, and this means how Causes and Effects are related to each other. This mentioned relationship is the central concept to understand activation functions. The relationship is linear in simple problems, but it is not in most cases. The main purpose of activation functions is to fill this gap. If the neural network does not use the activation function, it would be like simple regression.

The inputs are multipled by weights linearly. After adding the bias, the activation function runs on the output. The activation functions have different kinds, and each kind has different mathematical operations. The essential kinds will be listed below.

3.5.1 Step function

The output is one for positive inputs and zero in other cases. The step function is helpful in binary classification problems.

3.5.2 Sigmoid function

As Fig. 3.5 depicts, the output of this function is between 0 and 1, and it is used for classification. The desired output for binary classification is zero or one, it would be one when the output of the function is more than .5, and it would be zero when the output is less than 0.5. The mathematical equation for sigmoid, which relates inputs to outputs is:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \tag{3.3}$$
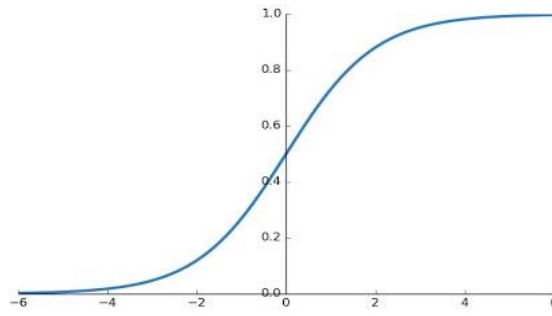
Figure 3.5: The Sigmoid Function [32].

3.5.3 Tanh function

The Tanh function is the same as sigmoid function but in the range of [-1,1]. Both functions could be obtained from one another, but the Tanh function is better than sigmoid. The mathematical relation is:

$$\tanh(x) = 2\sigma(2x) - 1 \tag{3.4}$$

3.5.4 ReLU function

The rectified linear unit *(*ReLU) is used more than other functions in neural networks. Although the mathematical relation is straightforward, the rectified linear unit's learning rate is quite high. The result of ReLU is in the range of [0, inf).

$$f(x) = \max(0, x) \tag{3.5}$$

3.5.5 Soft-max function

This non-linear function is a kind of sigmoid function that is practical in classifications. The main difference between soft-max and sigmoid function is that soft-max is widely used for multi-classification tasks, but sigmoid is used for binary classifications.

3.6 Logistic regression

Logistic regression is a straightforward and small neural network. This algorithm classifies inputs into discrete outputs by using the sigmoid function. The output could be two or more classes. The main difference between linear and logistic regression is that linear regression outputs are continuous but logistic regressions are not. The logistic regression has three main groups.

1. Binary logistic regression (Pass/Fail)

2. Multi logistic regression (Cats, Dogs, Sheep)

3. Ordinal logistic regression (Low, Medium, High)

The sigmoid function is used for mapping predicted values to probabilities. The output value of this function is in the range of (0,1). The sigmoid function:

$$\text{The predicted value} = \sigma\,(w*x + b) \tag{3.6}$$

3.7 Making Decisions

The output of the sigmoid is an S-shape curve between 0 and 1. The threshold value (t) needs to transform this range into the discrete classes. The values more than the threshold are one class, and the other class values are lower than the threshold value.

$$p \geq t,\ \text{class} = 1 \tag{3.7}$$

$$p < t,\ \text{class} = 0 \tag{3.8}$$

Prediction function is made by using the threshold value and function. This function determines the probability of input values in the right class. Generally, this class is shown in 1.

3.8 Lost and Cost Functions

The algorithms need a tool for estimating their accuracy. Loss functions are considered as these tools, which help machines to learn. The large numbers for loss function mean that the method is not accurate, and the output values are far from actual values. The main goal is to decrease this number. The loss function considers each member of the data individually.

On the other hand, the cost function considers all data. The algorithms need a tool for estimating their accuracy. Loss functions are considered as these tools, which help machines to learn. The large numbers for loss function mean that the method is not accurate, and the output values are far from actual values. The main goal is to decrease this number. The loss function considers each member of the data individually. On the other hand, the cost function considers all data.

There are different types of loss functions, and each problem needs a particular function. Loss functions are categorized into two Classification and Regression losses according to their learning tasks.

3.9 Regression Loss

These approaches are the three most common loss functions for Machine Learning Regression.

3.9.1 Mean square error

It measures the difference between real values and outputs. Because it squares this difference, the number shows the distance without the direction and has a significant penalty for inaccurate values. This loss function has a simple equation for computing gradients.

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^{\wedge}2}{n} \tag{3.9}$$

3.9.2 Mean absolute error

MAE measures the absolute difference between real values and outputs. The MAE does not contain the direction and shows the distance. MAE has a simple equation, but the process of calculating the gradients is not very easy.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{3.10}$$

3.9.3 Mean bias error

The equation of mean bias error is the same as MSE, without using the absolute value. This function could be less accurate because positive and negative values could eliminate their impact. Accordingly, this method must be used very carefully. MBE is less common than other functions

$$MBE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n} \tag{3.11}$$

3.10 Classification Loss

When the output of the model is consists of discrete values, the model is a classification model. The most used approaches listed in the below.

3.10.1 Multi class SVM loss

The multiclass SVM loss is not a differentiable function which is used in SVM. The score of the correct classification must be more significant than other scores, plus one.

Working with the convex curve of this function is more common than other curves.

$$SVM\ LOSS = \sum_{j \neq s_i}^{n} max\ (0, s_i - s_{yi} + 1) \tag{3.12}$$

3.10.2 Cross entropy loss

Cross entropy loss is the most used function for classification cases. The more considerable distances between the output values and actual values, the larger the CEL. The equation has two parts. The second part is eliminated when the actual value is one, and the first part is eliminated when the actual value is zero. This function largely penalizes the wrong outputs.

$$Cross\ Entropy\ Loss = -(y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i)) \tag{3.13}$$

3.11 Gradient Descent

The gradient descent is the most used learning algorithm in Machine Learning, and is a way to minimize the cost function, and this optimization tool has two main steps, the first step is the initialization of the related variables. These variables in the NN are weights and biases. These initialization values could be zero or very small random values. The second step is the Correction of these variables. The second step must be irritated as long as the output is in the desired range. The function corrects the direction in each irritation.

1) Initializing the weight and bias
2) Repeating the

$$w = w - \alpha \frac{dJ}{d\omega}\ , \quad b = b - \alpha \frac{dJ}{d\omega} \tag{3.14}$$

Where,

$$J(w, b) = \frac{1}{m} \sum_{n=1}^{m} y^{(i)} \log(y^{*(i)}) + (1 - y^{(i)})\log(1 - y^{*(i)}) \tag{3.15}$$

And w and b are weight and bias for NN.

The learning rate is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated. The learning rate may be the most significant when configuring the neural networks. The learning rate determines the amount of change in each step. Choosing a proper learning rate is a fundamental issue that manages the speed of the algorithm. Large learning rates increase the minimum point's speed, but significant learning rates may cause overshooting. Choosing a low learning rate increases reliability and also calculation

time, so the learning rate must be chosen at the optimum point. Fig 3.6 shows the relationship between gradient descent and the learning rate
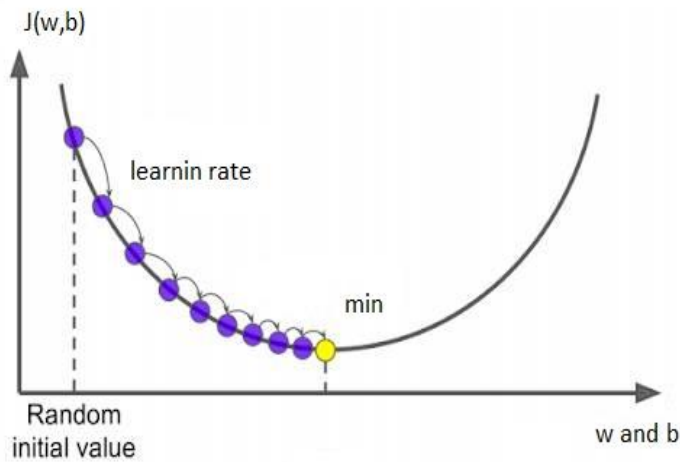
Figure 3.6: The gradient descent and learning rate [33].

3.12 Multilayer Perceptron

Multilayer Perceptron networks have two or more hidden layers with input and output layers. Using one hidden layer is always sufficient in the linear programs, but the network needs more than one in nonlinear cases to achieve high efficiency.

3.13 Different Kinds of Data in Machine Learning

The Machine Learning Algorithms can be sorted into four divisions according to their goal:

1. Supervised learning

2. Unsupervised learning

3. Semi-supervised learning

4. Reinforcement learning

3.13.1 Supervised learning

In supervised learning, each input data maps to a particular output. Supervised learning learns the mapping function between the training data and the output and uses it to predict the output for unseen data. Backpropagation is one of the most important methods which is used in neural networks.

3.13.2 Unsupervised learning

Both input and output variables have a label in the supervised learning. It means that every input has a specified value, and the network could learn the pattern by investigating the relationship between output and input. On the other hand, data in unsupervised learning includes of only inputs. K-means clustering and association rules are two mostly used algorithms for unsupervised learning. This kind of learning is classified into clustering cases.

The neural network considers these inputs as random information, and the weights in the neural network could be changed according to these inputs.

3.13.3 Semi-supervised learning

As mentioned before, in supervised learning, all inputs have output, but this does not happen in the real world. Most of the time, information is not complete. Semi-supervised learning has output, but not for all input; it is a combination of both learning methods. It conveys that only some of the inputs are mapped into the output.

3.13.4 Reinforcement learning

The last type of machine learning is reinforcement learning, which is widely used in robotics, automation, and autonomous vehicles.

Reinforcement learning is used by different computer programs to achieve the best solution and to choose the best path in each case. The training data in supervised learning has the output, and the algorithm uses these answer keys in the training process. However, reinforcement learning does not have these answer keys, so there is no choice for algorithm for learning except experience. Reinforcement learning has three main approaches:

1. Value-based

2. Policy-based

3. Model-based

3.14 Backpropagation Algorithm

One of the essential ANN parts is the Backpropagation (BP) algorithm, which is used as an algorithm for training. This method was launched in 1960 and, after that, became

more and more popular. Artificial neural networks consist of numerous nodes in their layers. Each node has two values, the first one is weight, and the second one is biased. The total accuracy of the network is related to finding proper values for these two elements. Finding accurate values is a very intricate process and needs lots of training. The BP method makes this process possible.

Backpropagation works in ANN problems in the supervised class, according to Widrow-Hoff rules. Backpropagation method has five steps:

1. All weights and bias in different layers are randomly initialized.

2. The output is calculated according to these random values.

3. The error is calculated for this situation by considering the output from step 2 and the desired output.

4. The algorithm uses this error for changing the weight and bias values.

5. The algorithm repeats these four steps until a minimum error is found.

The flowchart deprived in Fig.3.3 is described in the ten steps:

1. Initializing the weights
2. determining the product of hidden layer by using the activation function.

$$y = f(w*x + b) \tag{3.16}$$

Where,

x is the input

y is the output

w is the weight

f is the activation function

3. determining the product of the output layer by using the hidden layer's outcome

$$yk = f(w*y + b) \tag{3.17}$$

4. Determining the backpropagation fault
5. Calculating the weight correction
6. Generateing a fault by adding the delta for every hidden neuron
7. Calculating the weight correction
8. Calculating the weights by using weight correction
9. Repeating step two
10. Process is completed

The backpropagation algorithm is depicted in Fig 3.7
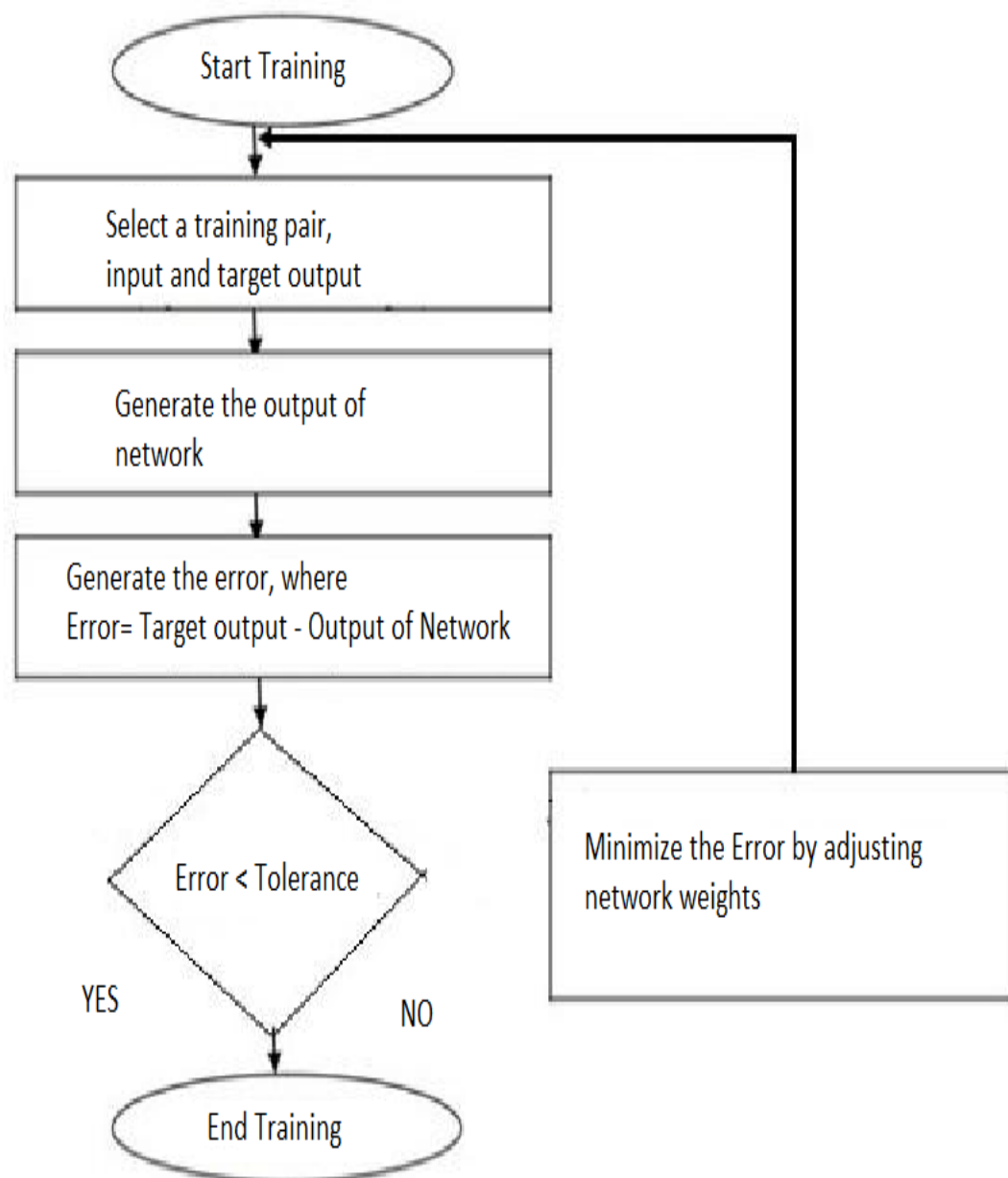
Figure 3.7: The backpropagation algorithm [34].

4. METHODOLOGY AND RESULTS

This chapter exclusively discusses implementing the neural network and other machine learning algorithms like Decision tree for data and a case study for this thesis. This paper has three parts.

The first part is about preparing and preprocessing the data. This part discusses different features related to each other and also how some features are eliminated from the process. The second part is using the NN and other models for STLF. This part shows the procedure of forecasting. And finally, the last part is about results.

4.1 Data

Data is one of the significant elements for having an accurate model, so it is very crucial to investigate it very precisely. Data that have used in this paper could consider in two parts, each part gathered from different sources and has unique characteristics. The first part has the Consumption's information, and the second part is related to weather information. The first part has reliable information, but the weather data has many useless values.

4.1.1 Load data

The load data consists of data of Gebze and has three main features, minimum, average, and maximum load Consumption. This information was gathered for 15 minutes intervals from 1/10/2013 to 20/9/2019. Each day has 96 samples in the dataset. Fig. 4.1 is a part of this information, which depicts the first 27 samples from the dataset.

This thesis considers the average load consumption as a target value, and all other values are features. The format of the date and hour in the table was achieved by using Python libraries. The table could not show the relationship between the consumption

and the data variables. Hence new features were added to the charts to find these relations.

These new features are season, month, and days of the week. By using these new features plotting and analyzing covariance between calendar variables has become quite simple. Weekdays shows the name of each day.

Table 4.1: The Consuming data.

Date	Avr [W]	Min [W]	Max [W]
10/1/13 3:00:00 AM	81101.516	85843.813	77710.344
10/1/13 3:15:00 AM	80353.93	86267.813	77003.242
10/1/13 3:30:00 AM	82334.68	90006.633	78021.633
10/1/13 3:45:00 AM	82048.313	88869.531	77371.016
10/1/13 4:00:00 AM	79951.32	88522.852	75210.492
10/1/13 4:15:00 AM	78183.852	83476.023	74081.945
10/1/13 4:30:00 AM	83486.195	92971.039	78194.469
10/1/13 4:45:00 AM	83394.656	93697.313	79177.344
10/1/13 5:00:00 AM	82997.336	91215.813	78858.734
10/1/13 5:15:00 AM	84283.695	123175.063	78157.352
10/1/13 5:30:00 AM	84270.508	101983.125	76399.602
10/1/13 5:45:00 AM	85372.93	95162.258	76806.773
10/1/13 6:00:00 AM	83746.398	91986.625	78273.578
10/1/13 6:15:00 AM	82440.336	92455.578	74894.93
10/1/13 6:30:00 AM	86992.891	98222.016	75547.031
10/1/13 6:45:00 AM	78221.844	95982.484	62184.578
10/1/13 7:00:00 AM	87213.742	99004.234	73667.352
10/1/13 7:15:00 AM	82113.32	117480.563	71285.711
10/1/13 7:30:00 AM	87424.672	102927.578	73082.617
10/1/13 7:45:00 AM	105107.086	143460.453	78926.375
10/1/13 8:00:00 AM	122419.484	166717.5	100552.031
10/1/13 8:15:00 AM	144472.578	187540.375	122549.453
10/1/13 8:30:00 AM	144079.719	185473.406	132365.453
10/1/13 8:45:00 AM	150533.813	211380.438	130793.18
10/1/13 9:00:00 AM	169180.047	208552.656	150392.188
10/1/13 9:15:00 AM	178015.969	218363.531	151730.031

The NN could use only number variables, so this feature interprets the name of the day for users. All the other new features are in number. Plotting the target value according to calendar values is the best way to find the load patterns between these values.

Table 4.2: The Consuming data after preprocessing.

Date	Month	Season	Week_days	Day	Avr	Min	Max
2013-10-01 03:00:00	10	4	Tuesday	1	81101.52	85843.81	77710.344
2013-10-01 03:15:00	10	4	Tuesday	1	80353.93	86267.81	77003.242
2013-10-01 03:30:00	10	4	Tuesday	1	82334.68	90006.63	78021.633
2013-10-01 03:45:00	10	4	Tuesday	1	82048.31	88869.53	77371.016
2013-10-01 04:00:00	10	4	Tuesday	1	79951.32	88522.85	75210.492
2013-10-01 04:15:00	10	4	Tuesday	1	78183.85	83476.02	74081.945
2013-10-01 04:30:00	10	4	Tuesday	1	83486.2	92971.04	78194.469
2013-10-01 04:45:00	10	4	Tuesday	1	83394.66	93697.31	79177.344

Fig. 4.3, Fig. 4.4, and Fig. 4.5 show the correlation between month, season, and weekdays with the target value, the average Consumption. These figures are demonstrating the relation between consumption and other features.
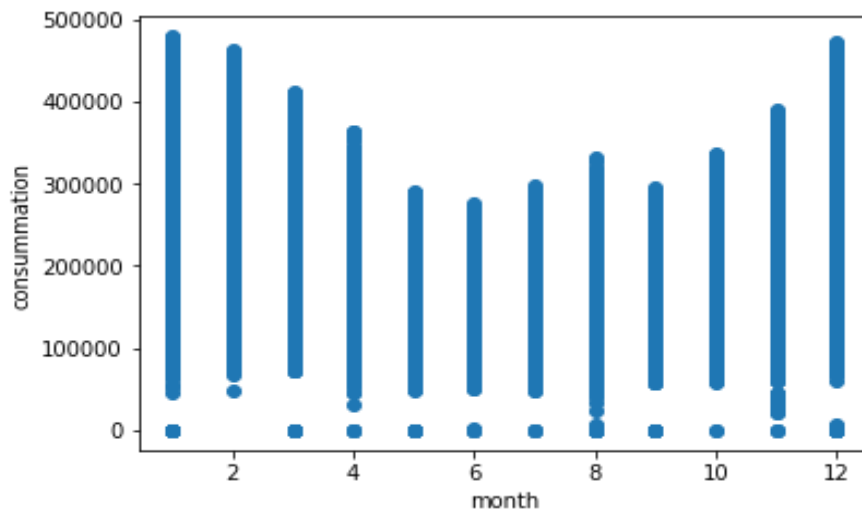
Figure 4.1: The consumption according to month.
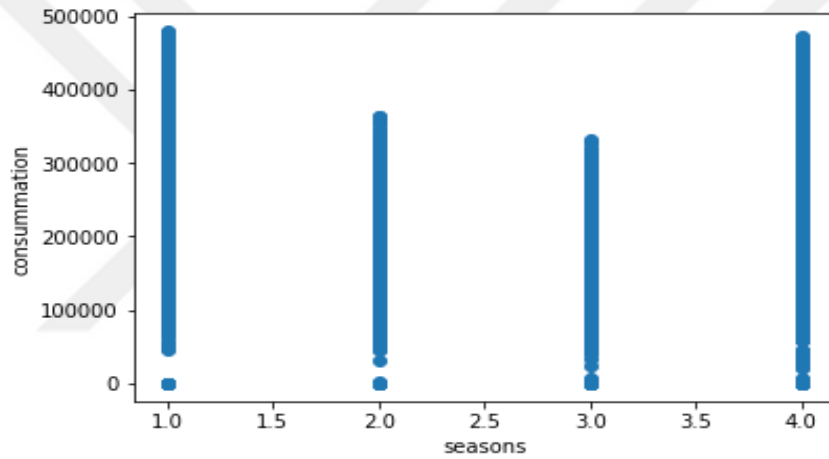
Figure 4.2: The consumption according to season.
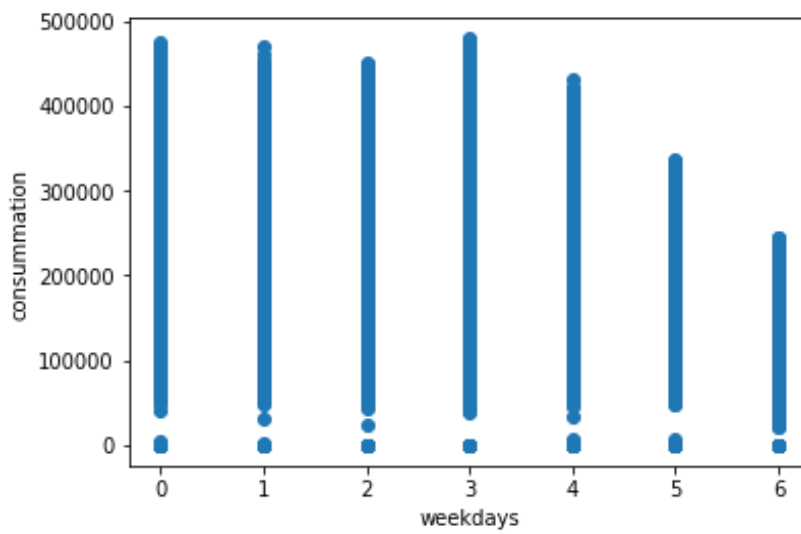
Figure 4.3: The consumption according to weekdays.

4.1.2 Weather data

The Meteorology Genel Müdürlüğü maintains the weather data, which comes from two different stations and were gathered in the monthly format, so it was transformed into a suitable format in the first place. The weather data consists of humidity, temperature, maximum temperature, minimum temperature, pressure, and wind speed.

The data also has other features that are not useable for this thesis. Some of them are not related directly to load Consumption. Furthermore, some others have many "Nan" values like cloudiness amount, sunbath intensity, and wind direction. The extra features were omitted in the first place.

The weather data were gathered at hourly intervals, but the target values are in 15 minute intervals.

Table 4.3: The Weather data.

| Date	Pressure	Temp	Min_Temp	Max_Temp	Humidity	Speed
2013-01-01 00:00:00	1014	7.4	7	7.4	83	1.3
2013-01-01 01:00:00	1013.7	7	7	7.1	85	1.5
2013-01-01 02:00:00	1013.3	7.1	7	7.1	84	1.3
2013-01-01 03:00:00	1013.2	7	6.8	7	85	1.2
2013-01-01 04:00:00	1013.1	6.8	6.7	6.9	87	1.4
2013-01-01 05:00:00	1013.2	6.8	6.8	6.9	86	1.2
2013-01-01 06:00:00	1013.5	6.8	6.8	7	85	1
2013-01-01 07:00:00	1013.6	7	7	7.2	83	1
2013-01-01 08:00:00	1013.8	7.2	7.2	7.6	79	1.3
2013-01-01 09:00:00	1013.6	7.6	7.5	7.6	77	1.5
2013-01-01 10:00:00	1013.2	7.5	7.5	7.9	77	1.2
2013-01-01 11:00:00	1012.5	7.8	7.7	7.9	76	1
2013-01-01 12:00:00	1012.6	7.8	7.6	7.8	76	1.4
2013-01-01 13:00:00	1012.8	7.8	7.8	7.9	76	1.5
2013-01-01 14:00:00	1012.6	7.9	7.8	7.9	77	1.6
2013-01-01 15:00:00	1012.7	7.9	7.8	7.9	77	1.6

There are some solutions to this problem. The consumption data has four values for each hour. The first solution uses the maximum value between these four; the second one uses the minimum value. However, the best method is equating from these four

values. This thesis utilizes an averaging method. Because using the average value increases the total accuracy.

4.2 Preprocessing

All of the features must be in a reasonable range. The network finds the relationship between variables in the training process. The network has the most significant impact of larger values, so if one of the features has a tremendous value, the network will considers it as a wrong considerable impact on this feature. It is essential to normalize input data before using them in the training process. In the ANN, the input data is separated into two groups. These two groups have a different number of members. The bigger group is the training group, and the smaller one is the test group. However, some of the data in these groups are not ideal for the training and testing process. These incorrect data have two origins, quality, and correlation. Dealing with these two groups, and solving their problems could enhance the training process.

4.2.1 Quality evaluation

The data quality has a huge impact on model accuracy. Data were gathered from different sources in an extended period, so reliability is a significant factor. The data was taken from various weather stations in load forecasting problems, and some of these stations had more reliable data. Data errors could be caused by human errors in measuring part or by devices.

4.2.1.1 Missing values

Missing values are prevalent in the weather station data. There are many of these missing values in the used data in this paper. Missing values are caused during the collection. There are two standard solutions, which this paper used each of them in different cases.

1. Approximate missing values

If the number of missing values is in a reasonable range, missing values could be filled by mode, mean, or median value of near values.

2. Omitting the whole row with the missing value

This is the easiest method, and sometimes really efficient. If data has numerous features, the feature with the missing value could be omitted.

4.2.1.2 Irrelevant values

Irrelevant values are the values that are not in their field. For example, in the weather condition data, all the data in the chart could be numerical. The only way to manage this data is by eliminating it.

4.2.1.3 Repetitive data

These are the information that are duplicated in the database. They must be removed before starting the training process.

4.2.2 Correlation

Correlation is a tool to calculate how much two variables are related to each other. The number that shows this ratio is the correlation coefficient. This number, which shows the direction and degree of relation, is in the range of [-1,1].

Zero is the minimum correlation coefficient, and both 1 and -1 are the strongest correlations but in the opposite direction. This number is obtained by calculating the standard deviation and covariance between input and output. The equation explains this calculation:

$$r(x,y) = COV(x,y) / SxSy \qquad (4.1)$$

Where;

COV(x,y) is the covariance,

Sx is the standard deviation of x,

Sy is the standard deviations of y, and

r(x,y) is the correlation

Fig. 4.7 shows the part of the data after preprocessing. All data are gathered together and normalized

	pressure	temprature	min_temperature	max_temprature	humidity	speed	month	season	day_of_week_in_number	avg	min	max
0	0.004653	0.004989	0.005047	0.004917	0.004057	0.002920	0.006152	0.006633		0.001305	0.002265	0.002223	0.002168
1	0.004652	0.004963	0.005102	0.005483	0.004178	0.003451	0.006152	0.006633		0.001305	0.002259	0.002272	0.002144
2	0.004653	0.005543	0.005617	0.005921	0.003452	0.002655	0.006152	0.006633		0.001305	0.002342	0.002606	0.002169
3	0.004659	0.006071	0.006079	0.006153	0.003028	0.005044	0.006152	0.006633		0.001305	0.002304	0.002398	0.002034
4	0.004661	0.005886	0.005970	0.005741	0.003633	0.013009	0.006152	0.006633		0.001305	0.002515	0.002932	0.002076

Figure 4.4: The final data.

41

4.3 Using Neural Network

This thesis has three case studies for forecasting. The first one uses just calendar variables for forecasting the load, and the second was only used in weather variables. At first, some of the features were chosen with more impact on the load curve in both of the cases.

The third case uses all of the variables to gather. The accuracy of the model is calculated in each case and compared with others to find the best solution.

Most of the deep learning methods have available functions in different software. These functions are beneficial for commercial uses. These functions consider some of the parameters constant, so in this thesis those functions are not used. All of the functions were rewritten for this thesis. The leading software that was used is Python. Different libraries were imported to Python.

The best result for neural networks is when the model used 12 different features. The neural network used has one input layer, four hidden layers, and one output layer. All of the layers except the output layer used ReLU function as an activation function. The activation function for the last layer is linear. This is the best choice for regression problems. The other hyper-parameters were calculated by using tuning methods. The layers have 128, 255, 255, 255, 255, 1, nodes. Table 4.1 reveals the three scoring numbers for each case study.

Table 4.4: Scoring for neural networks.

Number of features	MAE	MSE	RSQ
Calendar features	0.0008458	1.13e-06	0.62579
Weather features	0.0004181	2.55e-07	0.91549
12 features	0.0003469	1.95e-07	0.98530

The results of forecasting for 24 hours are in Table 4.2. This figure also has the original values and error amount in percent. These numbers visualized in Fig. 4.8. The axis-x indicates 24 hours based time.

Table 4.5: Forecasting resaults by using neural networks.

Hour	Avr	Predict	Error (%)
0	0.002557	0.002030	20.609161
1	0.002469	0.002590	4.911772
2	0.002345	0.002520	7.460068
3	0.002349	0.002549	8.520803
4	0.00269	0.002941	9.321570
5	0.003174	0.003470	9.332144
6	0.005141	0.005027	2.216693
7	0.008532	0.008732	2.338513
8	0.00965	0.009654	0.045352
9	0.009974	0.009950	0.238616
10	0.00912	0.009134	0.151728
11	0.00905	0.009450	4.425156
12	0.009052	0.009052	0.002207
13	0.008967	0.008930	0.413829
14	0.007967	0.007900	0.835677
15	0.007347	0.007347	0.002873
16	0.007605	0.007605	0.005902
17	0.007118	0.007172	0.752039
18	0.005988	0.005941	0.790597
19	0.00471	0.004747	0.790960
20	0.004077	0.004002	1.845438
21	0.003909	0.003919	0.247940
22	0.003603	0.003631	0.790448
23	0.003335	0.003347	0.362164
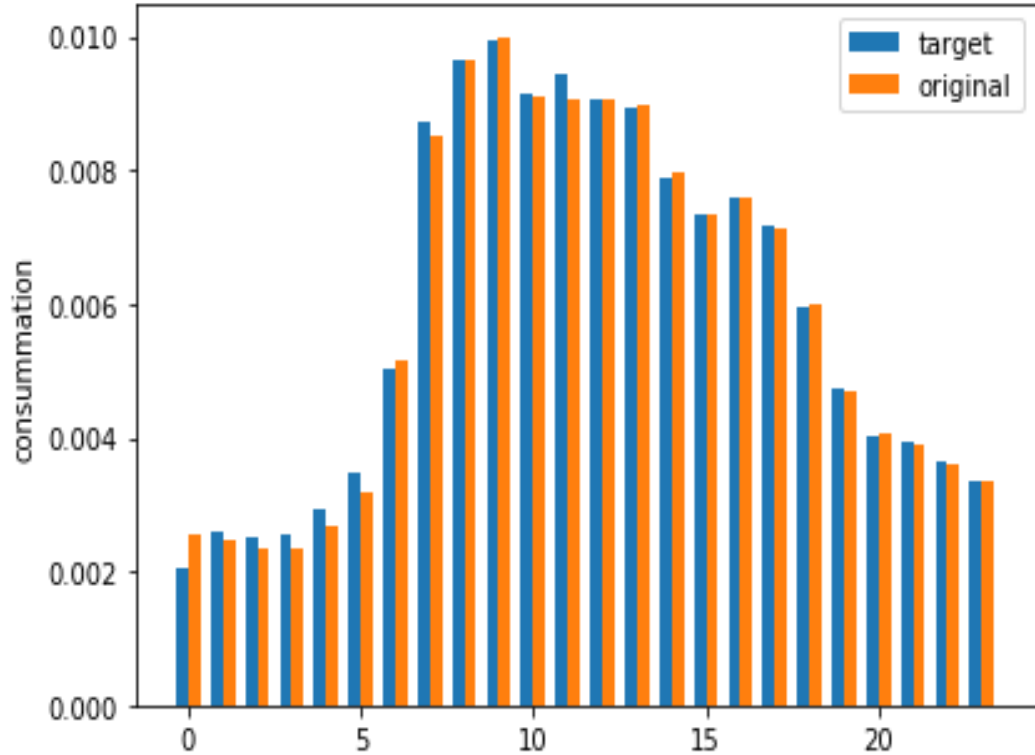
Figure 4.5: Forecasting resaults by using neural networks.

4.4 Forecasting with Other Methods

The primary goal of this thesis is to build a neural network model for forecasting the one-day load. This model could be used for other data sets and other case studies. To make sure that the model is accurate, four other methods were used for the same data set to compare the results. These methods and their results will be described in the next part.

4.4.1 Decision Tree

The decision tree is another supervised machine learning method which is utilized for both regression and classification problems. This method is used to fit for data set. Using these methods has three steps. The first step is fitting the model into the data set. The second step is predicting future values, and the last step is to find accuracy. The important hyper-parameter in this method is the maximum depth. The maximum depth determines how far the tree goes on in splitting. A for loop is the best choice for tuning this parameter.

Table 4.3 shows the results by using the decision tree method. These numbers depicted in curves in Fig. 4.9, the axis-x indicates 24 hours based time, and the axis-y indicates Consumption.
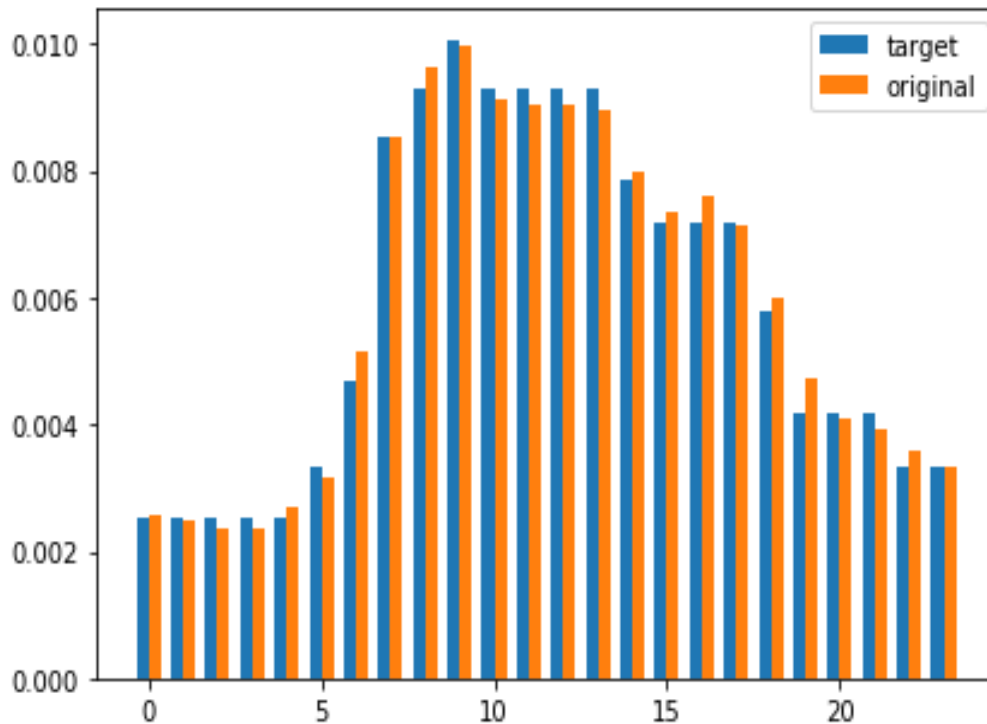
Figure 4.6: Forecasting resaults by decision tree.

Table 4.6: Forecasting resaults by decision tree.

Hour	Avr	Predict	Error (%)
0	0.002557	0.002534	0.905634
1	0.002469	0.002534	2.635847
2	0.002345	0.002534	8.049107
3	0.002349	0.002534	7.874246
4	0.00269	0.002534	5.814196
5	0.003174	0.003345	5.402767
6	0.005141	0.004695	8.665988
7	0.008532	0.008543	0.119949
8	0.00965	0.009306	3.557344
9	0.009974	0.010044	0.706799
10	0.00912	0.009306	2.041534
11	0.00905	0.009306	2.837821
12	0.009052	0.009306	2.812183
13	0.008967	0.009306	3.783215
14	0.007967	0.007857	1.374068
15	0.007347	0.007171	2.395922
16	0.007605	0.007171	5.710001
17	0.007118	0.007171	0.740487
18	0.005988	0.005806	3.051066
19	0.00471	0.004182	11.21091
20	0.004077	0.004182	2.562983
21	0.003909	0.004182	6.968876
22	0.003603	0.003345	7.140413
23	0.003335	0.003345	0.310854

4.4.2 Linear regression

The third method used is linear regression. The results and curves are shown in Table 4.4 and Fig. 4.10.
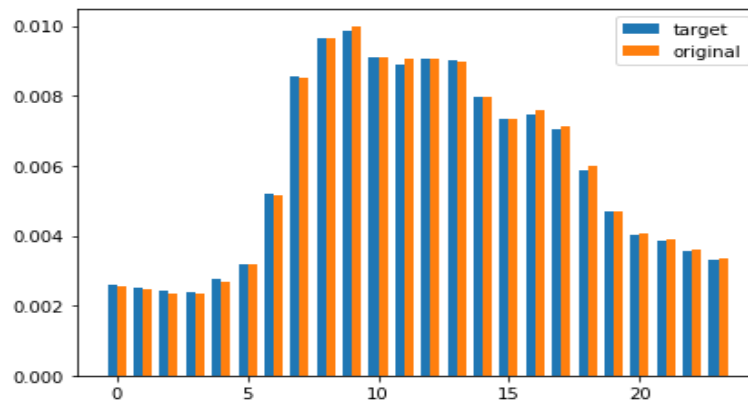
Figure 4.7: Forecasting resaults by linear regression.

Table 4.7: Forecasting resaults by linear regression.

Hour	Avr	Predict	Error (%)
0	0.002557	0.002609	2.049871
1	0.002469	0.002529	2.426774
2	0.002345	0.002419	3.169837
3	0.002349	0.002386	1.579416
4	0.00269	0.002761	2.626574
5	0.003174	0.003178	0.14311
6	0.005141	0.005179	0.740775
7	0.008532	0.00857	0.434115
8	0.00965	0.009629	0.218702
9	0.009974	0.009849	1.251099
10	0.00912	0.009086	0.375227
11	0.00905	0.008907	1.577063
12	0.009052	0.009043	0.095246
13	0.008967	0.00902	0.593915
14	0.007967	0.007965	0.02028
15	0.007347	0.007359	0.154794
16	0.007605	0.007481	1.630937
17	0.007118	0.007042	1.071401
18	0.005988	0.005853	2.256812
19	0.00471	0.004706	0.078143
20	0.004077	0.004037	0.995269
21	0.003909	0.003864	1.16385
22	0.003603	0.00357	0.907743
23	0.003335	0.003304	0.918988

4.4.3 K-nearest neighbors (KNN)

The KNN regression method is used when target values are continuous. These property values are means of k nearest neighbours. The method uses three different distance functions to find the nearest neighbours.

$$Manhattan = \sum_{i=1}^{\infty}|x_i - y_i| \tag{4.2}$$

$$Minkowski = \left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q} \tag{4.3}$$

$$Euclidean = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{4.4}$$

Finding the right value for k is very important, and could be done using one simple "for" loop. Analyzing the result of this loop reveals the best k value, which avoids over fitting and under-fitting problems. The k for this case is 7. Fig. 4.11 and Table 4.5 shows the results for this method.

Table 4.8: Forecasting resaults by nearest neighbors.

Hour	Avr	Predict	Error (%)
0	0.002557	0.002884	12.8037
1	0.002469	0.002787	12.91126
2	0.002345	0.002709	15.51883
3	0.002349	0.002711	15.39779
4	0.00269	0.002839	5.537941
5	0.003174	0.003079	2.973973
6	0.005141	0.004996	2.814776
7	0.008532	0.008509	0.272676
8	0.00965	0.009536	1.182074
9	0.009974	0.010101	1.274145
10	0.00912	0.008958	1.777715
11	0.00905	0.009115	0.720233
12	0.009052	0.009329	3.065411
13	0.008967	0.009278	3.46163
14	0.007967	0.008041	0.934237
15	0.007347	0.007234	1.543367
16	0.007605	0.007608	0.027404
17	0.007118	0.007058	0.852768
18	0.005988	0.005907	1.353775
19	0.00471	0.004733	0.500619
20	0.004077	0.003856	5.435458
21	0.003909	0.004025	2.970912
22	0.003603	0.003698	2.648632
23	0.003335	0.003042	8.777209

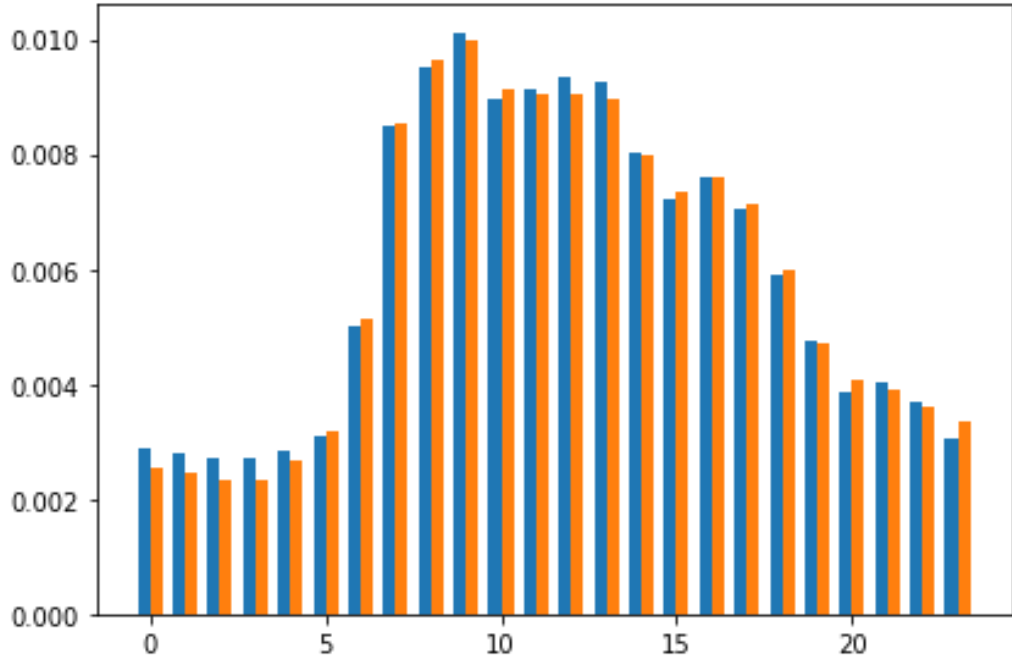**Figure 4.8**: Forecasting resaults by nearest neighbors.

Table 4.6 shows the accuracy of four different methods.

Table 4.9: Forecasting errors for different methods.

Methods	MAE	MSE	RSG
Decision Tree	0.000224	7.64e-08	0.97476
KNN	0.000418	2.55e-07	0.91549
Linear Regression	0.0000655	8.19e-09	0.977295
ANN	0.0003469	1.95e-07	0.985306

5. CONCLUSIONS AND RECOMMENDATIONS

The primary purpose of power systems is to maintain reliable and secure electricity power. The power system attributes must learn to accomplish this aim. Consumer behavior is an essential element in system management. Short term load forecasting is an excellent opportunity to study this subject under the impact of different features—this thesis decades forecasting in the power systems in the first chapter. After introducing different methods for forecasting, the artificial neural networks were proposed and this is because the load curves are not linear.

The neural networks were widely used in classification problems with very high accuracy, but they were also used in regression problems. Chapter three discusses different neural networks and their architecture.

There are many kinds of neural networks, but four of them are widely used. The fully connected networks are the most common models. The convolutional neural networks are the first choice in image processes. The program of neural networks was written in the Python program. The model consists of some simple functions.

The fourth chapter revealed the result of forecasting. The number of needed features was calculated in the first step. Three different error functions were used to compare accuracy.

The primary step in using machine learning methods is data preprocessing. For instance, most of the algorithms do not accept Nan values. This thesis uses two station data for gathering weather data because these data have lots of Nan values. For increasing the accuracy, these Nan values were filled by comparing the data from two different stations. After preprocessing, the data could be used in different machine learning methods.

Chapter four uses four other different methods to make sure that the proposed model is accurate. The thesis forecasted 24 hours. The errors were calculated for each hour. The best method for this data set was chosen from comparing these error functions.

REFERENCES

[1] Ozturk I. (2010). A literature survey on energy-growth nexus, Elsevier, Energy policy, 38 (1), 340-349.

[2] Payne J.E. (2010). A survey of the electricity consumption-growth literature, Elsevier, Appl. Energy, 87 (1), 723-731.

[3] Mahmood, A., Javaid, N., Zafar, A., Riaz, R., A., Ahmed, S., Razzaq, S. (2014). Pakistan's overall energy potential assessment, comparison of TAPI, IPI and LNG gas projects, Elsevier, Renew Sustain Energy Rev (31), 182–93.

[4] RazaKhan, A., Mahmood, A., Safdar, A., A, Khan., Z., AhmedKhan, N. (2015). Load forecasting, dynamic pricing and DSM in smart grid: A review, Elsevier, (54), 1311-1322.

[5] K. Vu, M. M. Begouic, D. Novosel. (1997). Grids get smart protection and control, IEEE Computer Applications in Power, 10(4), 40-44.

[6] NIST(2010). National Iinstitute of Standards and technology. Framework and Roadmap for smart grid interoperability standard. National Institute of Standard and technology special publication 1108

[7] S. M. Elgarhy, M. M. Othman, A. Taha, H. M. Hasanien. (*2017).* Short term load forecasting using ANN technique, *Nineteenth International Middle East Power Systems Conference (MEPCON)*, Cairo, 2017, 1385-1394.

[8] A. D. Papalexopoulos, T. C. Hesterberg. (1990). A regression-based approach to short-term system load forecasting, IEEE Transactions on Power Systems, 5(4), 1535-1547.

[9] G. E. Huck, A. A. Mahmoud, R. B. Comerford, J. Adams, E. Dawson. (1980). Load Forecast Bibliography Phase I, IEEE Transactions on Power Apparatus and Systems, vol. PAS-99, 53-58.

[10] M. S. Sachdev, R. Billinton, C. A. Peterson. (1977). Representative bibliography on load forecasting, IEEE Transactions on Power Apparatus and Systems, (96), 697-700.

[11] A. D. Papalexopoulos, T. C. Hesterberg. (1990). A regression-based approach to short-term system load forecasting, IEEE Transactions on Power Systems, (5), 1535-1547.

[12] H. S. Hippert, C. E. Pedreira, R. C. Souza. (2001). Neural networks for short-term load forecasting: a review and evaluation, IEEE Transactions on Power Systems, (16), 44-55.

[13] M. A. Abu-El-Magd, N. K. Sinha. (1982). Short-Term load demand modeling and forecasting: A review, IEEE Transactions on Systems, Man and Cybernetics, (12), 370-382.

[14] I. Moghram, S. Rahman. (1989). Analysis and evaluation of five short-term load forecasting techniques, IEEE Transactions on Power Systems, (4),1484-1491.

[15] T. Haida, S. Muto. (1994). Regression based peak load forecasting using a transformation technique, IEEE Transactions on Power Systems, (9), 1788-1794.

[16] B. Krogh, E. S. de Llinas, D. Lesser. (1982). Design and implementation of An on-Line Load Forecasting Algorithm, IEEE Transactions on Power Apparatus and Systems, (PAS-101), 3284-3289.

[17] A. D. Papalexopoulos, H. Shangyou, P. Tie-Mao. (1994). An implementation of a neural network based load forecasting model for the EMS, IEEE Transactions on Power Systems, (9), 1956-1962.

[18] D. K. Ranaweera, N. F. Hubele, A. D. Papalexopoulos. (1995). Application of radial basis function neural network model for short-term load forecasting, IEE Proceedings Generation, Transmission and Distribution, (142), 45-50.

[19] M. H. Choueiki, C. A. Mount-Campbell, S. C. Ahalt, (1997). Implementing a weighted least squares procedure in training a neural network to solve the short-term load forecasting problem, IEEE Transactions on Power Systems, (12), 1689-1694.

[20] K. Methaprayoon, W.-J. Lee, S. Rasmiddatta, J. R. Liao, R. J. Ross. (2007). Multistage Artificial Neural Network Short-Term Load Forecasting Engine with Front-End Weather Forecast, IEEE Transactions on Industry Applications, (43)1410-1416.

[21] A. Demir, N. Hadžijahić. (2018). Power System Planning: Part I, Basic Principles, springer, 178-188.

[22] J. Wu, J. Liu, J. Ma, K. Chen, C. Xu. (2020). Classification of power loads based on an improved de noising de convolutional auto-encoder, Elsevier, (87), 105959.

[23] K. Zor, O. Timur, A. Teke. (2017). A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting, *6th* International Youth Conference on Energy (IYCE), Budapest, 1-7.

[24] T. Hong, M. Gui, M. Baran, H.L. Willis. (2010). Modeling and forecasting hourly electric load by multiple linear regression with interactions, Power and Energy Soc. General Meeting, Minneapolis, MN, 1-8.

[25] N. Amral, C.S. Ozveren, D. King. (2007). Short term load forecasting using multiple linear regression, Universities Power Engineering Conference, Brighton, 1192-1198.

[26] K. Chongqing, CHENG Xu, XIA Qing, et al. (1999). A new unified approach to short-term load forecasting considering correlated factors. Automation of Electric Power Systems, 23 (18): 32-35.

[27] Papaioannou, George P.; Dikaiakos, Christos; Dramountanis, Anargyros; Papaioannou, Panagiotis G. (2016). Analysis and modeling for short-to medium-term load forecasting using a hybrid manifold learning principal component model and comparison with classical statistical models (SARIMAX, Exponential Smoothing) and artificial intelligence models (ANN, SVM): The case of Greek electricity market, *Energies* 9, no. 8: 635.

[28] Tang, An & Tam, Roger & Cadrin-Chênevert, Alexandre & Guest, Will & Chong, Jaron & Barfett, Joseph & Chepelev, Leonid & Cairns, Robyn & Mitchell, Joseph & Cicero, Mark & Gaudreau Poudrette, Manuel & Jaremko, Jacob & Md, Caroline & Gallix, Benoit & Gray, Bruce & Geis, Raym & O'Connell, Timothy & Babyn, Paul & Koff, David & Shabana, Wael. (2018). Canadian association of radiologists white paper on artificial intelligence in radiology, Canadian Association of Radiologists Journal, 69.

[29] Salunkhe, Sachin & Rajamani, D & Esakki, Balasubramanian & Chandrasekhar, U. (2019). Prediction of life of piercing punches using artificial neural network and adaptive neuro fuzzy inference systems, International Journal of Materials Engineering Innovation 10.

[30] Chen, Chien-Sheng & Lin, Jium-Ming. (2011). Applying Rprop Neural Network for the Prediction of the Mobile Station Location. Sensors 11, no. 4: 4207-4230.

[31] Abugabah, Ahed & Ali, Ahmad & Al-Obeidat, Feras & Alarifi, Abdulaziz & Alwadain, Ayed. (2020). Data mining techniques for analyzing healthcare conditions of urban space-person lung using meta-heuristic optimized neural networks. Cluster Computing.

[32] Han J., Moraga C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, vol 930.

[33] Ye, Chen & Toyoda, Kentaroh & Ohtsuki, Tomoaki. (2018). Robust heartbeat detection with doppler radar based on stochastic gradient approach, IEEE International Conference on Communications (ICC), Kansas City, 1-6

[34] V S, Mahendran. (2015). Gripping force measurement and EMG classification for hand functions. International journal of Applied Engineering Research, 10. 17315-17326.

CURRICULUM VITAE

Name Surname   : Ali Ghadiriasl Nobari

Place and Date of Birth : Iran, 1993

E-Mail   : nobari18@itu.edu.tr

EDUCATION:

- **B.Sc.:** 2017, Azerbaijan Shahid Madani University, Electronic Engineer Faculty, Electrical Engineering Department.