DOKUZ EYLÜL UNIVERSITY

GRADUTE SCHOOL OF NATURAL AND APPLIED SCIENCES

YEKDEM UNIT PRICE PREDICTION WITH ARTIFICIAL NEURAL NETWORKS

by

Akın ERTAYLAN

April, 2021

İZMİR

YEKDEM UNIT PRICE PREDICTION WITH ARTIFICIAL NEURAL NETWORKS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partitial Fulfillment of the Requirements for the Degree of Master of Science
of Department of Computer Engineering

by

Akın ERTAYLAN

April, 2021

İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"YEKDEM UNIT PRICE PREDICTION WITH ARTIFICIAL NEURAL NETWORKS"** completed by **AKIN ERTAYLAN** under supervision of **ASSIST. PROF DR. ÖZLEM AKTAŞ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Özlem AKTAŞ

Supervisor

Assist.Prof.Dr.Kökten Ulaş BİRANT (Jury Member)

Prof.Dr.Ayşegül ALAYBEYOĞLU (Jury Member)

Prof. Dr. Özgür ÖZÇELİK

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENT

Throughout my thesis, I would like to extend my gratitude and respect to my valuable advisors, Assist. Prof. Dr. Özlem Aktaş and Assist. Prof. Dr. Yunus Doğan, who guided me with their valuable knowledge and experience. I would like to thank my friend Gökhan Akgün who helped me in all matters throughout my work. In conclusion, I would like to thank my esteemed family who took all the difficulties with me during my work and supported me in every stage of my life.

Akın ERTAYLAN

YEKDEM UNIT PRICE PREDICTION WITH ARTIFICIAL NEURAL NETWORKS

ABSTRACT

With the developing technology, energy production methods are changing in the world. Fossil energy resources have begun to be abandoned and replaced by renewable energy resources with unlimited capacity. In this way, both green energy is produced and less cost is spent. Improvements are also available for the production of green energy in Turkey. With the YEK Law enacted in 2005, YEKDEM (Renewable energy resources support mechanism) started. YEKDEM is a support mechanism applied to wind, biomass, canal, solar, gas obtained from biomass, wave current energy, geothermal, tidal energy, river and hydroelectric production facilities. Energy produced by the Yekdem method also has a monthly updated unit price. This unit price is named as yekdem unit price. In this present studying, an unit price prediction has been made with ann, which are one of the sub-branches of artificial intelligence. In the project, four dependent variables were estimated and the unit price of the same was obtained with these variables. In the estimation process, 30 independent parameters were used. Different machine learning models were created to compare the results of neural networks and the best result was investigated. MPE and MSE were used as accuracy criteria in the study. It was 0.35 in the MSE criterion in the best model and 53.67 in the MSE criterion in the worst model.

Keywords: Artificial neural networks, regression, energy consumption, price forecast, RERSM (Renewable energy resources support mechanism)

YAPAY SİNİR AĞLARI İLE YEKDEM BİRİM FİYAT TAHMİNLEMESİ

ÖZ

Gelişen teknoloji ile dünya üzerinde enerji üretim yöntemleri de değişmektedir. Fosil enerji kaynakları terk edilmeye başlanmış ve yerini kapasitesi sınırsız olan yenilenebilir enerji kaynakları yer almaya başlamıştır. Bu şekilde hem yeşil enerji üretilmiş olup hem de daha az maliyet harcanmış olmaktadır. Türkiye'de de yeşil enerji üretimi için gelişmeler mevcuttur. 2005 yılında çıkarılan YEK Kanunu ile YEKDEM (Yenilenebilir enerji kaynaklarını destekleme mekanizması) mekanizması başlamıştır. YEKDEM rüzgâr, güneş, jeotermal, biokütle, biokütleden elde edilen gaz, dalga akıntısı enerjisi, gel-git enerjisi, kanal, nehir ve hidroelektrik üretim tesislerine uygulanan destekleme mekanizmasıdır. Yekdem yöntemiyle üretilen enerjinin de her ay güncellenen birim fiyatı bulunmaktadır. Bu birim fiyata yekdem birim fiyatı adı verilmiştir. Bu çalışmada yapay zekânın alt dallarından biri olan yapay sinir ağları ile yekdem birim fiyatı tahmini yapılmıştır. Projede dört adet bağımlı değişken tahminlenmiş ve bu değişkenler ile yekdem birim fiyatı elde edilmiştir. Tahminleme işleminde ise 30 adet bağımsız parametre kullanılmıştır. Sinir ağlarının sonucu karşılaştırmak için farklı makine öğrenmesi modelleri oluşturulmuş ve en iyi sonuç araştırılmıştır. Çalışmada doğruluk kıstası olarak MPE ve MSE kullanılmıştır. En iyi modelde MSE kıstasında 0,35 en kötü modelde ise MSE kıstasında 53,67 değeri elde edilmiştir.

Anahtar Kelimeler: Yapay sinir ağları, regresyon, enerji tüketimi, fiyat tahmini YEKDEM (Yenilenebilir enerji kaynakları destekleme mekanizması)

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

Depending on today's developing technology; Due to rapid industrialization, urbanization, rising living standards and continuing population growth, the need for energy is increasing day by day. The negative economic and environmental impacts that occur as a result of the high dependency on fossil fuels in meeting energy needs direct countries to use sustainable energy resources. These negative effects can be given as an example. Coal, oil and natural gas power plants cause destruction locally in the regions where they are established, as well as threatening the whole world globally. Fossil fuels are harmful gases when burned, the resulting dust and soot, close to the polluted environment while causing death, carbon dioxide and other greenhouse gases cause global climate change is threatening the life of the world and all countries. It is claimed that if global warming continues to increase, the sea level will rise up to one meter by 2040, in which case the world's largest cities will be inundated. These reasons are due to have started a trend toward green energy in the world and in Turkey. As a result of the combination of the studies that increased with this trend, the rapid development of technology and the applied energy policies, there is a decrease in the costs of renewable energy systems that can compete with fossil fuels.

Renewable energy sources have many advantages. Some of these can be listed as follows. Helps protect the environment by reducing carbon dioxide emissions. Since they are domestic resources, it helps to reduce dependence on foreign energy. Solar, Geothermal, Hydraulics, Biomass, Wind are known as renewable energy sources. Turkey is known as a country rich with diversity and potential of that in terms of renewable energy sources. Utilization of renewable energy sources, Turkey has many advantages compared to many other countries. It is very advantageous to have geothermal energy resources that are not available in other countries. Renewable energy sources are explained below.

- Solar Power:

The sun is the most important source of energy for all planets in the solar system. It is an indispensable resource especially for all living things living in our world. Solar panels collect solar rays and generate heat or electricity. Solar energy is evaluated in the form of light, heat and electricity. It is among the most important renewable energy sources today.

- Wind Power:

Wind energy is a type of energy that originates from solar energy. The sun cannot warm the lands and seas at the same rate. As a result of this situation, the pressure difference occurs and creates the wind. Wind turbines installed in regions with high wind speed convert the kinetic energy of the wind into mechanical energy first and then into electrical energy. The energy obtained from the wind varies according to the current speed and blowing time of the wind. Wind energy meets 2% of the world's electricity needs today. The wind turbine generating electricity from wind has very little damage to the environment.

- Bioenergy / Biomass Energy:

Biogas is an unlimited resource. It is an important energy source that leads to economic developments in rural areas. Specially grown algae such as corn and wheat, plants, algae in the sea, herbs, animal excrement, fertilizer and industrial waste, and all organic waste disposed from homes are sources of biomass (Bayrak,2020). The use of biomass as an alternative to the depletion of fossil fuels and the environmental pollution it creates is gaining importance every day to solve the energy problem.

- Geothermal Energy:

Geothermal means surface temperature. Waters formed as a result of natural events and especially rainfall reach the magma level over the crust. The waters heated in the magma layer emerge as hot water and steam. Thanks to the water and steam that reaches the earth, it can be transformed into many types of energy. Thermal energy, which is generally present in the earth's crust, generates geothermal energy. This energy that is extracted to the earth is converted into electrical energy with the established power plants. They can also be used in central heating and cooling systems used in homes and workplaces, in many physical therapy centers and touristic centers preferred by patients.

- Hydroelectric Energy:

Hydroelectric energy is based on using the energy of running water and converting this energy into electrical energy. Hydroelectric energy is in the green energy class. In places with high altitude, the flow rate of water is high. Therefore, hydroelectric power plants will be more useful in these regions. As the energy of flowing water in hydroelectric power plants is used as the main element, it is used in irrigation, facilitating transportation, developing fisheries and mostly in energy production.

- Hydrogen Energy:

Due to the technology and production difficulties used today, its use is not very common yet. However, technology is advancing day by day. However, as a clean energy source, it is one of the most important candidates in meeting the energy needs of the world. In the future, we can mention that hydrogen energy will be used in the production of electricity, heat and fuel cells.

- Wave / Ocean Energy:

One can actually think of the oceans as two different energy sources. The first is thermal energy, the second is mechanical energy fed from waves and tides. Oceans, which cover 70% of the surface of the world, also constitute the world's largest solar collectors. The temperature difference between hot water on the surface of the oceans and cold water deep in the ocean creates a natural thermal energy. If used sufficiently, even a small part of this energy is enough to meet all the energy needs.

The use of renewable energy sources used in power in Turkey, at very low levels. However, Turkey has very significant potential in many renewable energy sources (Bayrak,2020). The installed capacity of power plants that use renewable energy in Turkey are as follows:

- Installed power of wind energy in our country is 7000 MW,
- Hydraulic installed power 28000 MW,
- Installed power of geothermal energy is 1300 MW,
- Installed power of Solar Energy is 5000 MW,

Energy sources such as wave, tide, biomass energy are used more and more every day. Sourced renewable energy installed capacity in Turkey is anticipated in the coming years will increase too much.

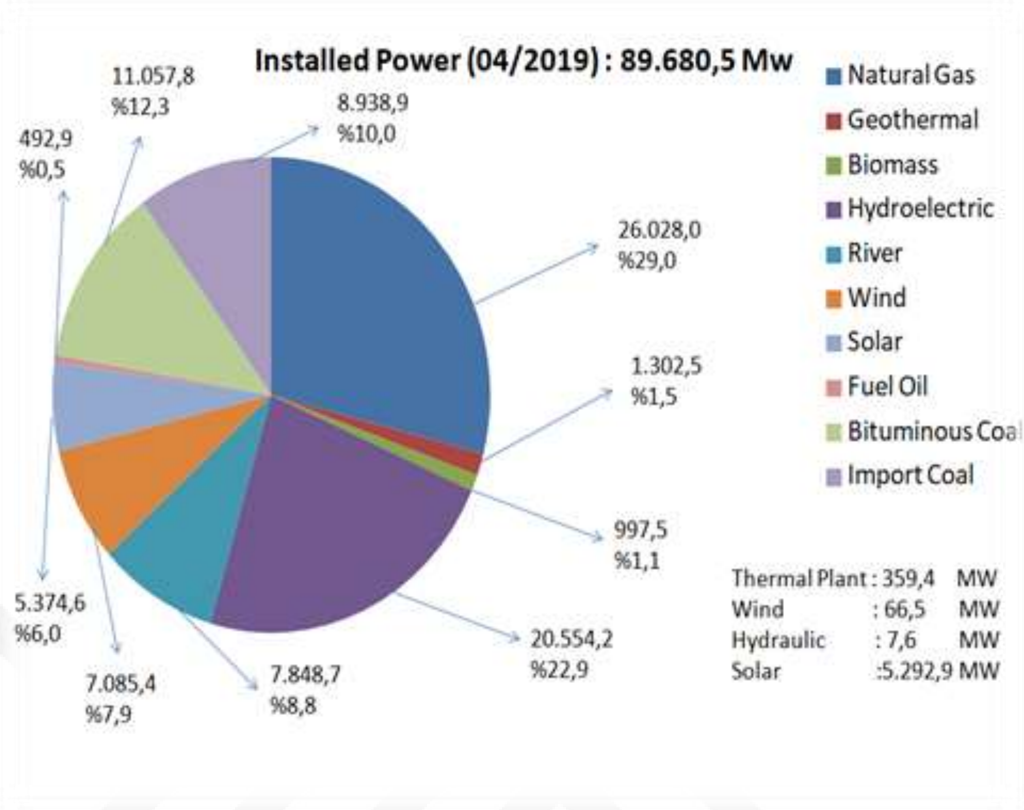

Figure 1.1 Installed power (04/2019)

Renewable energy production in Turkey is increasing day by day. Conscious consumption and renewable energy production studies are also planned and supported by the state. This YEKDEM laws in Turkey as a result of the work was created, and this law was put into effect. Yekdem means Renewable Energy Resources Support Mechanism. YEKDEM is an incentive system for production facilities to switch to renewable energy sources. Its rules were determined by the state and made into law. Within the scope of this law, renewable energy plants are updated every year. The number of renewable power plants has been increasing since 2016.

1.1 Purpose of Thesis

Electrical energy covers many areas of our lives. It is the most used type of energy and therefore has a dynamic nature. Its production and consumption varies according to the time and climate. The pricing structure also adapts to these conditions. Different charges are made in the processes of generating and distributing electrical energy. One of these charges is the Yekdem unit price. Electricity generated and consumed within the scope of renewable energy is priced. This price is calculated every month. It is advantageous for the stakeholders in the electricity market to predict these prices and plan. Therefore, in the thesis it aimed to estimate the Yekdem the unit price. In this study, four dependent variables are estimated, and the unit price of the same is obtained with these variables. In the estimation process, 30 independent parameters were used. These are ptf values, dollar, climate variables, electricity production and consumption values, holidays and time parameters. The data set consists of values between 01.01.2018 - 31.10.2020. Artificial neural networks and machine learning models were used as models.

1.2 Literature Contribution

During the literature research process, many articles concerning the electricity market were read, theses were examined, and blog articles were researched. These studies investigate the factors determining electricity demand, price and income elasticity of electricity demand. However, the literature on the unit price of Yekdem in the electricity market has not developed much. In this study, YEKDEM price formation is analyzed on the basis of YEKDEM market structure and solvency parameters. Different models for unit price estimation have been compared. Determination of factors that impact on this comparison, the YEKDEM market in Turkey and to reveal the results will give the different models are being considered in terms of price forecasts will contribute to the literature.

1.3 Thesis Organization

Renewable energy sources are explained in the introduction part of the study, in which the unit price of Yekdem is discussed, and the situation in the world and in our country is stated. At the same time, the aim of the thesis and its contribution to the literature are explained. In the second chapter, the literature research is given. YEKDEM mechanism is provided in the third chapter. In the fourth chapter of the thesis, the models used in the estimation of the unit price are explained. In the fifth chapter, information is given about the studies done. In the sixth chapter of the thesis, the findings obtained as a result of the study are explained. In the last section, the findings are discussed and future studies are stated.

CHAPTER TWO
LITERATURE REVIEW

Before starting the thesis study, literature was searched, previously written articles on similar subjects were read and their models were examined. Yekdem a specific mechanism has a special place in the electricity sector in Turkey. Therefore, no previous article about yekdem has been found. In general, information about the content was collected, but no studies were found to estimate the unit price of yekdem. The biggest reason for this is that a working mechanism unique to Turkey. For this reason, studies modeled with machine learning and artificial neural networks aimed at price and consumption estimation are taken as an example. Information about the studies reviewed is given below.

Dalgın (2017) made PTF estimation using artificial neural networks and achieved a success of 5.21 in the MAPE criterion. However, no information was given about any pre-processing performed on the data set.

Pençe et al. (2019) 2017 - 2023 Turkey's electricity consumption for the industry were forecasting using artificial neural networks. In the study, only year information is included as input parameter.

Karacan et al. (2016) estimated electricity demand in the program they developed. This prediction was realized by using only simple regression and multiple regression methods. In this study, compared to other articles, the data was preprocessed and then divided into two as testing and training. Then, the machine learning model was created and the estimation process was made.

Gümüş et al. (2015) used Fuzzy Logic and Artificial Neural Network methods to predict monthly mean evaporation. When the outputs were examined, it was observed that the methods used gave successful results. Especially, the combination of the ANFIS method with six inputs has been observed to have a more successful result than other models.

Es et al. (2014) predict the amount of electric power of Turkey on neural network methods. Population, gross domestic product, export, building area, import and number of vehicles were used as independent variables between 1970 and 2010 as the data set in the estimation study.

Toker et al. (2012) proposed a method for estimating the short-term amount of electricity demand and checked the results of the model with the actual values to test the accuracy of this method. The data set included in the prediction model is temperature, actual electricity consumption, date information, radiation days and other necessary meteorological factors.

Considering the seasons and market conditions from Turkey, expert systems and neural network methods are used together, they developed a short-term requirement quantity of electricity high percentage Predictive hybrid method (Başoğlu & Bulut, 2016).

Çayır et al. (2018) used approximately 3-year electricity consumption of 30 different houses in a specific region in London, England as a data set and it was aimed to create a model that predicts short-term consumption with machine learning algorithms. The data used in the study belongs to the smart meter energy consumption dataset in London, England. The model with the lowest absolute error loss score was chosen and the experiments were carried out with these models.

Kocadayı et al. (2017) estimated the annual electrical energy consumption of the TR81 (Bartın, Karabük, Zonguldak) area using the neural network method (ANN). In the study, while estimating energy consumption, export, population, import and surface area information of buildings were used as data set.

Cantürk (2018) in her study, a wind field in Turkey utilizing artificial neural network method with data obtained from the wind farm made production of short-term prediction. He analyzed the data sets taken from the wind farm, made data analysis and created appropriate data sets for the ANN model. The estimation results made with the created models are evaluated according to the estimation results obtained with the naive method, which is accepted as a criterion in short term wind generation estimates. It has been observed that the error rates of the estimations made with two different Artificial Neural Network systems in static and dynamic structure are within an acceptable range compared to the naive method. The lowest average error rates were obtained with the static ANN.

Boltürk (2013) aimed to forecast electricity demand in her study. The data set consists of 12-year electricity consumption values. She divided the consumption values in the data set into three parts. These; Short Term (between 2011), Medium Term (between 2009-2011) and Long Term (between 2000-2011). The methods used in the study have used Moving Average, Exponential Correction, Holt Model, Regression, ARIMA, Fuzzy Logic, Artificial Neural Networks. First of all, a time series was created from 2000-2011 electricity consumption and electricity for the year 2012 was estimated from this time series. These predictions are compared with the actual 2012 electricity consumption data. He used 4 different accuracy criteria to compare the forecast results. These are Mean Square Error, Mean Absolute Error, Mean Absolute Percent Error and Tracking Signal. Small differences were obtained between the result values and the actual values.

Bicil (2015) thesis that the liberalization process in Turkey in the work realized price forecasts hourly electricity market price estimates using different methods in Turkey electricity market. It used seasonal autoregressive moving average model and multi-layer sensors designed with different network architectures in price estimation. Working in considering the transition to a market day-ahead electricity market is an important change for Turkey 2012 January - April 2014 period covering the hourly market price and temperature variables are used. The performance of the prediction models was evaluated according to the criteria of mean absolute error, mean absolute percent error, root mean square error and Theil's coefficient of inequality. The estimation results show that the multi-layer sensor trained with Sigmoid Activation Function and Levenberg Marquardt algorithm provides the highest prediction performance.

Öksüz (2019) made a consumption estimation in Denizli province by using the data covering the years 2008-2019 in her thesis study. In the study, variables known to have an effect on electricity consumption such as residential subscribers' electricity consumption amount, population, electricity unit prices, unemployment and temperature were used. Denizli has made consumption estimation with the help of artificial neural network method and MATLAB program. The results were compared with the current values, and their accuracy was measured.

Biçer (2018) made a short-term electricity demand prediction in his thesis. The data set consists of consumption data, temperature and humidity data in the Urla region of Izmir. He used artificial neural networks as a model. By doing normalization operations in the data set, it enabled the model to learn faster. He divided the data set into two and used 70 percent of it for training. As a result, it fell below ten percent in the MAPE criterion.

Şenol et al. (2017) estimated electricity generation from wind energy by creating an artificial neural network (ANN) model with various wind turbines. In the model created, wind speed data were used in the testing phase, and the output powers of different types of wind turbines were used in the training phase. It is understood that the estimates made by the model formed in the regression curves that emerged after the application are reliable and consistent.

Şenocak & Kahveci (2016) made Market Clearing Price (PTF) estimates by using the production amounts of primary energy sources with Adaptive Network Based Fuzzy Inference System and Artificial Neural Networks, which are among the estimation methods. It is explained which method is more successful and its relationship with the amount of primary energy sources production. The Matlab platform was used for ANFIS and the interface program created in the C# software language for ANN. For the two methods used, the weighted average values of the 2015 Market Clearing Price were calculated by applying the production amount of primary energy resources for the years 2012-2014 as input data during the training phase. MAPE error rates between 1-5% were obtained.

Özden & Öztürk (2018) their studies, time series and artificial neural networks for an industrial zone in Turkey, including using two different approaches have been studied and the results were tested estimated energy needs. In this study, unlike previous studies, a simple model for short-term estimation with limited data has been developed. The model contains consumption data and temperature of the past days as input parameters. Only historical energy consumption data are used in the time series approach. Both approaches have been used in the energy demand estimation, the results are discussed and compared. According to the results obtained; time series method R value 0.939, artificial neural network method R value 0.985 was obtained. Time series method has made a worse estimation than artificial neural networks due to data constraints.

CHAPTER THREE
YEKDEM

With the Law No. 5346 on the Use of Renewable Energy Resources for the Purpose of Generating Electricity, a mechanism (YEKDEM) has been established to support the generation license holders of electricity generation facilities based on renewable energy sources and electricity generation facilities that produce unlicensed electricity. Production facilities that can benefit from the YEKDEM system are wind, solar, wave, biomass, geothermal, biomass gas (including landfill gas), current energy and tidal and canal or river type or hydroelectric with a reservoir area of less than fifteen square kilometers, as defined in this law. They are sources of electricity generation suitable for establishing a production facility. According to the current legal regulation, a considerable amount of incentives are given with the support mechanism for electricity generation from renewable sources. The price has been determined for the enterprises that have entered into operation after the effective date of Law No. 5346 and will be in operation by the end of 2020. The purchase price has been set for 7.3 cents / kWh for hydraulic and wind resources, 10.5 cents / kWh for geothermal resources, 13.3 cents / kWh for solar and biomass resources. The period of benefiting from the incentives defined by the law is limited to 10 years.

3.1 Yekdem Progress

With the start of the first implementation in 2011, the number and capacity of power plants that join this mechanism increased every year. While participation was relatively low until 2015, there was a rapid increase in participation after 2015.

3.1.1 Number of Power Plant

According to the sources, the annual numbers of the power plants within the scope of YEKDEM are shown in Table 3.1. Although there is an increase in the number of power plants generating electricity from all renewable sources, it is clear that the most increase is in hydraulic power plants. In 2015, there was a huge increase in the number of power plants that joined the system. The enactment of the Regulation on Certification and Support of Renewable Energy Resources on 01.10.2013 is considered to be the most important factor in the increase of these applications.

Table 3.1 Number of plant power

	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Hydroelectric	4	44	19	40	126	388	418	447	463	461
Geothermal	4	4	6	9	14	20	29	37	45	49
Wind	9	22	12	21	60	106	141	151	160	165
Solar	0	0	12	0	0	0	2	3	9	17
Biomass	3	8	16	20	34	42	57	70	100	126
Sum	20	78	65	90	234	556	647	708	777	818

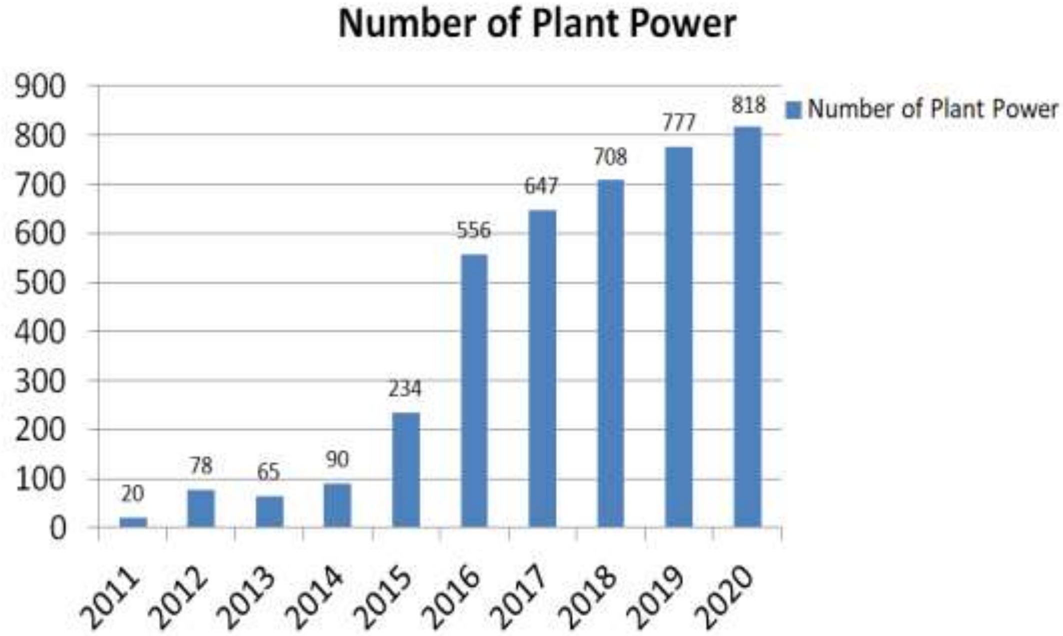

Figure 3.1 Number of plant power

3.1.2 Installed Power

Installed power sizes are more important than the number of power plants registered in the YEKDEM system. Because the amounts to be paid to these power plants within the scope of the mechanism are made according to the amount of production they produce during the year, depending on their installed power. The total capacity, which was 713.1 MW in the first implementation year in 2011, reached 21,872.2 MW in 2020.

Table 3.2 Installed power

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Hydroelectric	21.0	929.7	246.2	598.2	2116.3	9960.0	11096.3	11706.4	12588.5	12445.3
Geothermal	72.4	72.4	140.4	227.8	389.9	599.2	752.1	996.8	1252.7	1503.0
Wind	563.1	685.0	106.5	824.8	2732.1	4319.8	5238.7	6200.0	6495.6	6974.3
Solar	0.0	0.0	51.8	0.0	0.0	0.0	12.9	13.9	81.7	174.9
Biomass	56.6	73.4	101.6	139.7	185.2	203.7	300.0	349.2	203.1	778.7
Sum	713.1	1760.4	646.6	1790.4	5423.6	15082.7	17339.9	19266.3	20921.5	21877.2

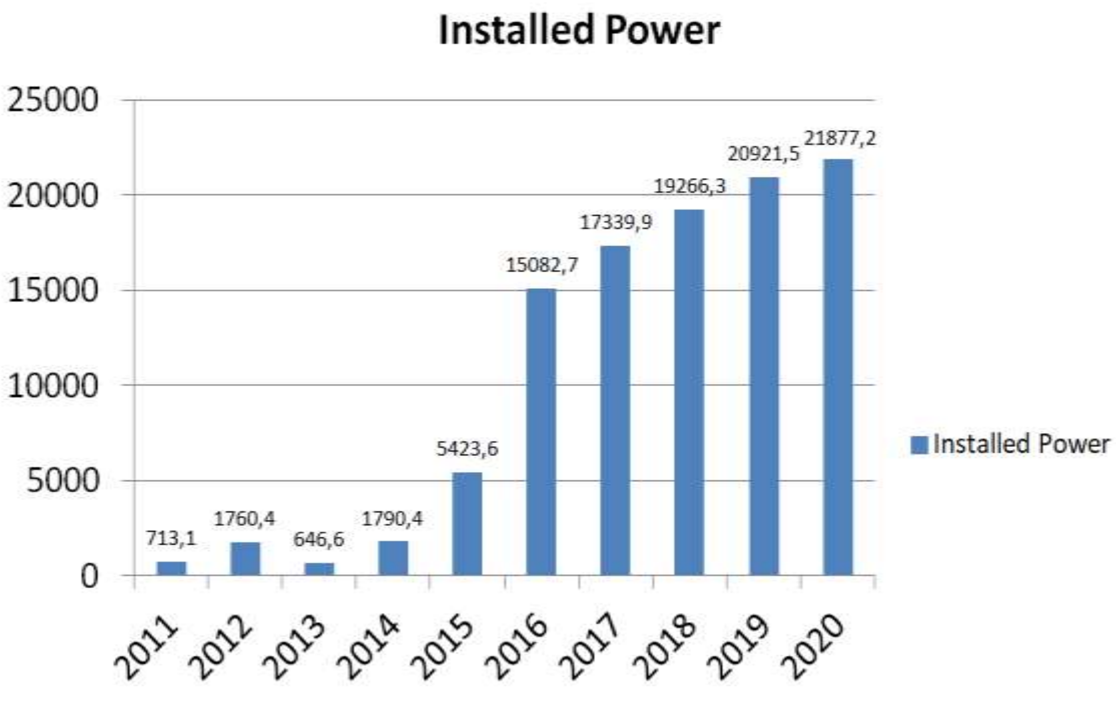

Figure 3.2 Installed power

The total installed power capacity in Turkey and depending on how much of the total renewable resources is important to the YEKDEM system where it is registered. As is known, a large amount of direct tariff support is provided for electricity generation within the scope of this mechanism. Since this support will be paid by electricity consumers, the capacity included in this scope and the amount of electricity generation obtained from this capacity are very important. Any support other than its real purpose will be directly covered by the electricity consumer. For this reason, the YEKDEM system should be established very well and support within this scope should be given to the power plants that are selected very accurately. Otherwise, the electricity produced by the power plant that receives unfair support will be a burden to the consumer.

YEKDEM started to be implemented in 2005 and the utilization period of the power plants benefiting from this year started to expire in 2015. Considering that some of the power plants connected to hydraulic and wind resources, which are not currently included in the YEKDEM utilization list, have benefited from this support for 10 years before, almost all of the capacity based on renewable resources made by private companies benefited from great incentives. Moreover, these incentives were applied directly to the tariff support, and their production was also guaranteed to be purchased.

Giving such large and guaranteed incentives to electricity generation of such capacity will naturally increase the electricity sales prices directly and these high electricity prices will be paid by the consumers. Instead of providing such large supports, the state's direct power generation facilities with the financing resources allocated to this support will undoubtedly reduce the cost of electricity generation to society.

3.1.3 Production Capacity

Although the plants benefit from the YEKDEM system as much as the amount of production they perform, it will be useful to examine the electricity generation capacities of these plants. Thus, it will be understood more clearly how much of the total electricity generation capacity can benefit from the high support in this system.

As can be seen in Table 3.3 and Figure 3.3, the annual production capacity of the facilities based on renewable resources within the scope of YEKDEM has increased very rapidly especially after 2015 and reached the level of 82.5 billion kWh in 2019.

Table 3.3 Production capacity

	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Hydroelectric	97.8	3376.3	785.4	2131.8	7678.7	35320.2	39023.6	39756.9	42514.9	41225.2
Geothermal	515.9	501.5	1051.9	1906.7	4157.4	5004.9	6619.5	8804.8	10975.1	11831.0
Wind	2017.2	2411.6	279.3	3655.2	11163.1	17023.2	21152.3	22926.0	23779.4	23716.7
Solar	0.0	0.0	69.0	0.0	0.0	0.0	23.8	43.8	161.3	345.7
Biomass	455.7	2580.0	809.3	1023.4	1434.0	1630.7	2195.5	2693.6	4001.1	5409.3
Sum	3086.5	8869.5	2994.9	8717.2	24433.2	58978.9	69023.7	74225.1	81431.7	82527.9

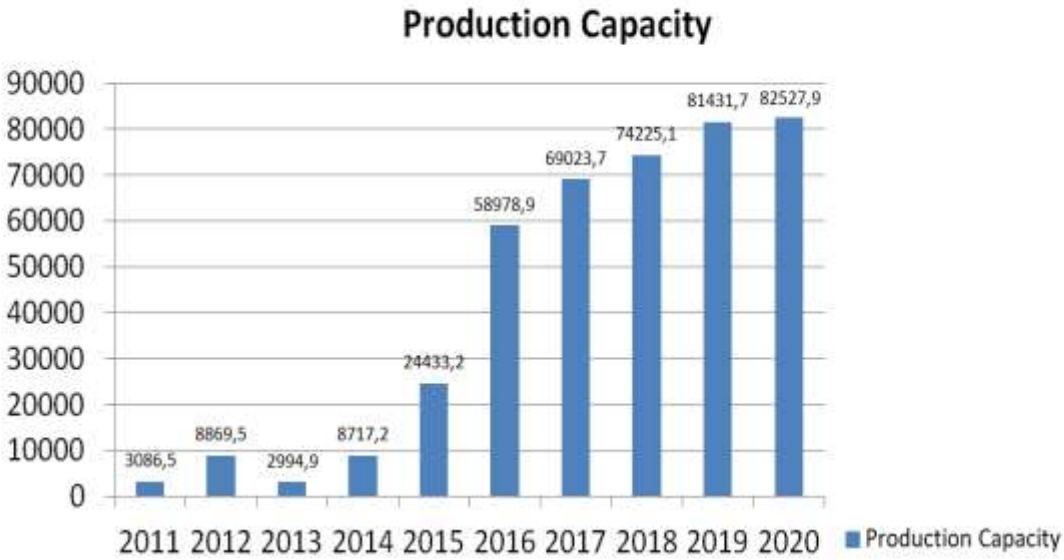

Figure 3.3 Production capacity

3.1.4 YEKDEM Payment

Additional incentive payments are made to the power plants registered in the RES system, based on the unit prices determined according to resource types and the rates of using domestic parts in return for the amount of production realized. As the amount of electricity generation in this system has increased over the years, naturally, the amount of support paid for these productions has also increased. The annual total of the supports paid for the system is shown in Figure 3.4 The total amount of support paid to this system in 2019 was 38 billion TL.

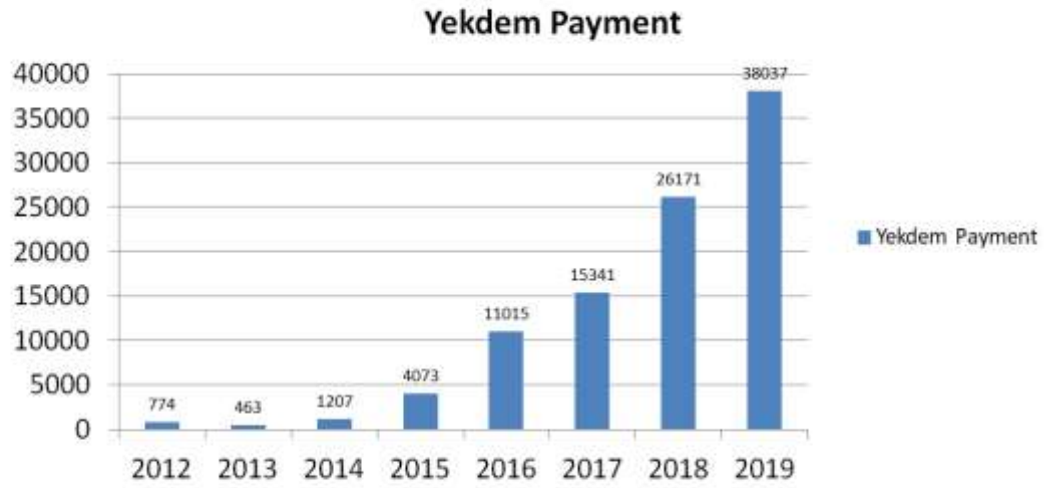

Figure 3.4 Yekdem payment

3.2 Yekdem Terms

In this section, the terms used in the yekdem mechanism are explained.

- YEKDEM Price

RES cost, as defined in the legislation; It is the unit price which is the ratio of the RES cost undertaken by the suppliers to the consumption.

- YEKDEM Cost

It is the total cost incurred by the suppliers as per the legislation within the scope of RER and which can be expressed as a whole. Yekdem cost is the sum of licensed and unlicensed producers. UEVM refers to licensed production. PTF is the market clearing price. YEKF refers to the premium price determined according to the plant type. The exchange rate indicates the purchase price. LUYTOB is the total price of unlicensed producers. p is the tolerance coefficient for unlicensed producers. LUNET refers to the unlicensed production amount. The formula below shows the calculation of the YEKDEM cost.

$$\mathrm{YEKBEDi} = \sum_{b=1}^{n}\sum_{t=1}^{l}\sum_{u=1}^{k}(\mathrm{UEVM\ x\ y\ x\ [\ (YEKF\ x\ KUR)} - (\ \mathrm{PTF\ x\ j}\ )]) \quad (3.1)$$

$$YEKBEDm = LUYTOBm - \sum_{t=1}^{l}\sum_{u=1}^{k}(LUNET\ x\ PTF\ x\ p) \quad (3.2)$$

- YEKDEM Price Difference

It is the amount reflected on the invoice by the supplier as the YEKDEM difference amount of the correct amount that should be reflected in the invoice by multiplying the difference between the YEKDEM unit costs projected in the probability and the realized YEKDEM unit costs by the actual consumption of the customer. Potentially projected unit costs may be the estimated YEKDEM prices announced by EPDK for the next year at the end of each year.

- YEKDEM Price

As explained above, when the financial settlements of all electricity market participants are finalized in a period, on the Suppliers; It is the unit price of YEK, which can be expressed as the ratio of the RES cost of the relevant supplier to its consumption. Uecm is licensed consumption values. ST_uecm is unlicensed consumption values.

$$Price\ = {}^{YG + YEKBEDi + YEKBEDm}/_{UECM + ST\_UECM} \quad (3.3)$$

Table 3.4 Yekdem unit price

	2018	**2019**	**2020**
January	35.11	81.57	59.03
February	39.48	69.97	77.26
March	66.24	90.89	139.70
April	55.72	142.15	233.28
May	69.10	156.42	230.79
June	73.17	111.58	112.10
July	56.46	62.04	103.98
August	68.37	66.75	100.98
September	55.83	60.81	94.01
October	49.55	58.90	92.66
November	47.06	52.06	111.00
December	50.83	54.34	

- YEKDEM Income

YEKDEM income can be expressed as the income of the producers within the scope of YEKDEM resulting from multiplying the incentive prices with the production quantities. At the same time, there is an invoice item that is invoiced between the EPIAS, which is called as YEKDEM income, from the financial settlement structure of the supply companies that are market participants. From the supplier perspective, the ratio of the difference between EPIAS expenditure and YEKDEM income to total consumption expresses the RES price / unit price. It is calculated by the formula below. Refers to UEVM licensed generation plants. PTF is the market clearing price. LUNET refers to the unlicensed production amount. KB is the participation fee. FB is the difference fee for the relevant period. j is the tolerance coefficient for licensed producers. p is the tolerance coefficient for unlicensed producers.

$$YG = \sum_{i=1}^{a}\sum_{b=1}^{n}\sum_{t=1}^{l}\sum_{U=1}^{k}(\mathrm{UEVM\ x\ PTF\ x\ i}) + \sum_{m=1}^{d}\sum_{t=1}^{l}\sum_{u=1}^{k}(\mathrm{LUNET\ x\ PTF\ x\ p}) + \sum_{p=1}^{N}\mathrm{KB} + \sum_{m=1}^{g}\mathrm{FB} \quad (3.4)$$

- YEKDEM Total Expenditures

It refers to the total YEK fee to be paid to participants who own both licensed and unlicensed generation plants within the scope of RER for the relevant invoicing period. It is expressed by the formula below.

$$\mathrm{YEKTOB} = \mathrm{YG} + \mathrm{YEKBEDi} + \mathrm{YEKBEDm} \quad (3.5)$$

3.3 Yekdem Affecting Factors

- Number of Power Plants

The high number of applications will increase the cost of YEKDEM. Participation is directly related to the previous year's market prices and dollar rate. If the PTF prices are very low in the previous year and the dollar rate is expected to rise, participation in YEKDEM will be high. If the dollar rate increases, the unit price will also increase.

- Increase in Dollar Rate

Renewable Sources engaged in Manufacturing and energy is produced by all plants included in the yekdem, plant type (wind, hydroelectric, geothermal, solar, etc.) and nativism contribution rate determined according to currency is calculated by multiplying over the price of the total amount, the other in Turkey suppliers are paid in proportion to the size of the customer portfolio. The purchase price figures guaranteed by the RES mechanism to investors on the basis of resources are as follows:

Biomass	: 13.3 dollar / kWh
Geothermal	: 10.5 dollar / kWh
Wind	: 7.3 dollar / kWh
Hydroelectric	: 7.3 dollar / kWh
Solar	: 13.3 dollar / kWh

- Market Clearing Price (PTF)

If PTF prices are high in the electricity market and the exchange rate is low, even if they are within the scope of RES, it will be more profitable to sell their electricity to the market, so they will not be subject to YEKDEM and will sell them to the stock market. If PTF prices are too low, power plants within the scope of RES will increase the number of plants to be included in YEKDEM due to purchase guarantee. Therefore, PTF prices may cause YEKDEM prices to decrease or increase.

- Climate Condition

In cases where precipitation is less, electricity usage and market will increase. Assuming that the dollar exchange rate is also very low in the same year, market conditions will not be suitable for a power plant generating electricity from renewable energy to enter YEKDEM. On the other hand, high precipitation will increase the use of renewable energy sources. At the same time, the high exchange rate of the dollar will make YEKDEM suitable.

CHAPTER FOUR
METHODS

The methods used in unit price estimation are explained in this section. Artificial neural networks, linear regression, decision tree and random forest models were used in Yekdem unit price estimation. Theoretical information about these models is given, respectively.

4.1 Artificial Neural Networks

Artificial Neural Networks have been created by adopting the working system of the human brain. Therefore, they are algorithms designed to fulfill the characteristics of the brain such as remembering, making decisions, learning, and discovering new information by generalizing. In short, ANN are computer programs that imitate biological nervous systems (Çayıroğlu, 2019).

The purpose of the development and use of artificial neural networks is data association, optimization, recognition, classification and predictions for the future. Since these operations are not easy to do with basic algebra methods, they can be done mathematically. Therefore, artificial neural networks are actively used in many areas.

The structure of artificial neural networks is based on the human nervous system. The structure of the nerve cell, which is the basis of the biological nervous system, consists of four main parts; nucleus, axon, dendrite, and connection. Dendrites have a structure that is located at the border of neurons and has a plant root appearance. The purpose of dendrites is to send signals from neighboring nerve cells to the nucleus. The nucleus collects the signals from the dendrite and transmits it to the axon. The axon processes the collected signals and transmits them to the connections on the other part of the neuron. The connections transmit the newly formed signals to the cells. The working logic of artificial neural networks is like biological nerve cells.

Çayıroğlu explains that an artificial nerve cell consists of five parts. These are Inputs, Weights, Addition Function, Loss and Activation Function and Outputs.

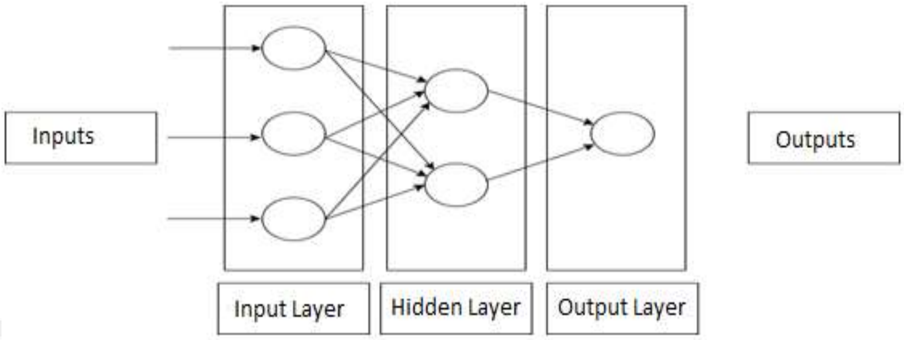

Figure 4.1 Artificial neural networks

A neuron collects the incoming signals by multiplying it with the weights of the neurons it is linked to, and adds the bias term to this total value, if any, and transmits the result to the defined activation function. The general formula of artificial neural networks is given below.

$$y = w * x + b \tag{4.1}$$

4.1.1 Inputs

Inputs are information that comes to nerve cells and comes first to artificial nerve cells from the outside world and then transmitted to cells in other layers.

4.1.1 Weights

The inputs to the cells in artificial neural networks reach after multiplying by the weight of the connections they come from. With this method, the influence of inputs on output data can be adjusted. Weight values multiplied by entries can be positive or negative, or zero. If the weight is zero, inputs have no effect on output.

4.1.3 Addition function

Addition method is the method that calculates the net input of that neuron by taking the total input data multiplied by the weights. There is no system that determines the optimum addition method to be used in a model. In the literature, mostly the addition function is selected by testing and comparison.

4.1.4 Loss and Activation Function

The activation method is the function that determines the output of the neuron by processing the input reaching the nerve nodes. It can also be called a transfer function. The activation method is mostly a nonlinear function chosen. This is due to the fact that artificial neural networks are often non-linear. Another point to consider when choosing the activation method is that it is an easily derivative function. The easy availability of the derivative both increases the processing speed and makes the computers less tiring. Recently, the activation function commonly used in multilayer neural network models is Sigmoid function. Below is information about the basically used activation functions.

4.1.4.1 Step Function

The step function is a function that takes a binary value. Therefore, it is used for classification purposes and is generally preferred in output layers. It is not used in hidden layers because its derivative is zero. The result of the neuron takes the value of 1 or 0, depending on whether the input is greater or less than zero.

$$f(x) = \begin{cases} 1\ if\ x > 0 \\ 0\ if\ x \leq 0 \end{cases} \tag{4.2}$$

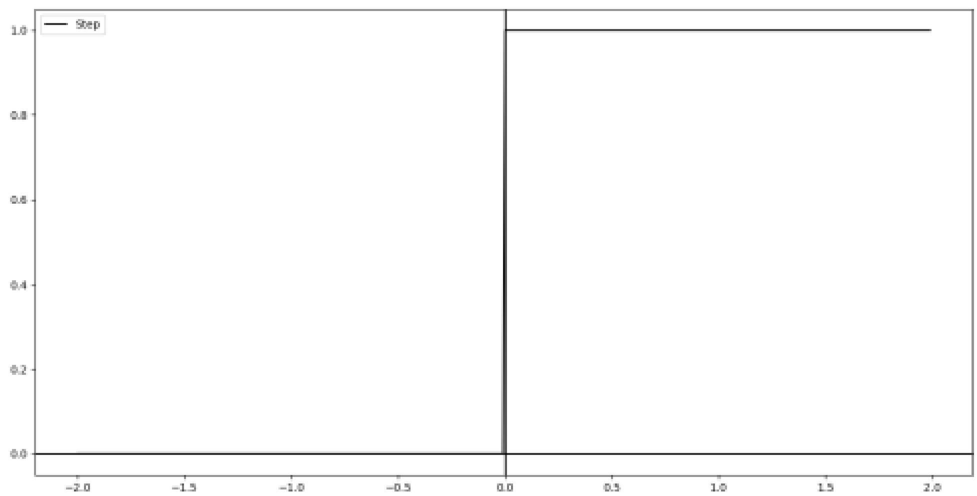

Figure 4.2 Step function

4.1.4.2 Linear Function

An activation function that does not contradict the nature of the model can be chosen to solve linear problems. The result from the linear function is determined as the output of the neuron by multiplying it by a constant coefficient (Pençe,2019). But the linear function has a disadvantage. The derivative of the function is constant. There is no learning in systems with fixed derivatives.

$$f(x) = x * c \tag{4.3}$$

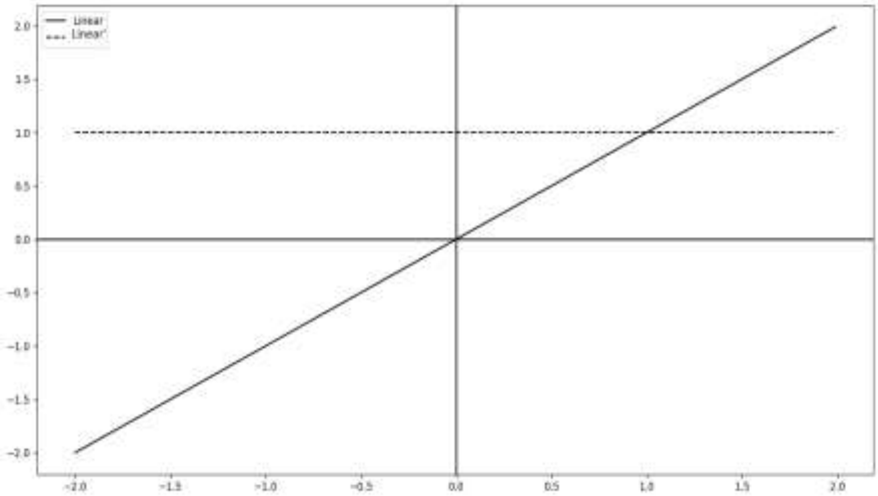

Figure 4.3 Linear function

4.1.4.3 Sigmoid Function

The Sigmoid method is a non-discrete and easily derived function. It is the most commonly used function in nonlinear and classification applications. Sigmoid function produces results between 0 and 1 according to each input data (Çayır, 2018).

Although it is a good classifier function, it also has a small disadvantage. In the function graph, f (x) values react very little to changes in x. In these regions, the derivative values are very low and converge to 0. The convergence of the derivative to zero stops learning.

$$f\,(x) = {}^{1}\!/_{1\,+\,e^{-x}} \tag{4.4}$$

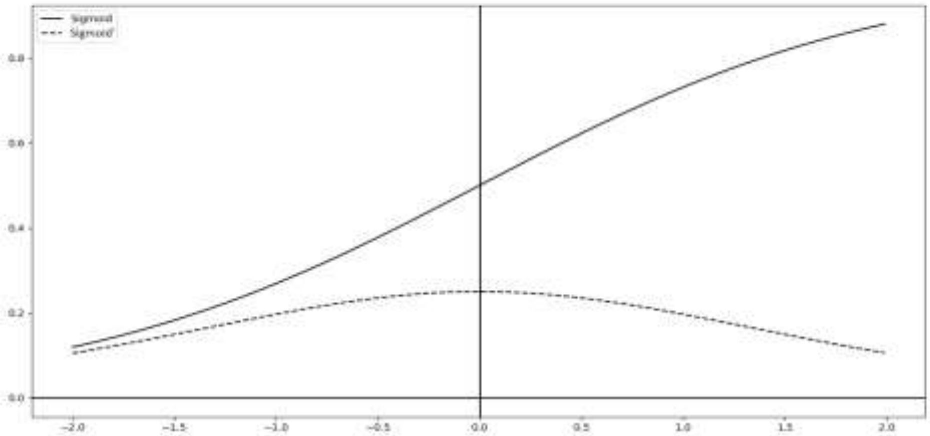

Figure 4.4 Sigmoid function

4.1.4.4 Tangent Hyperbolic Function

The hyperbolic tangent method is a type of function similar to the sigmoid method. In the Sigmoid method, the output results vary between 0 and 1, while in the hyperbolic tangent method the output results are between -1 and 1 (Öksüz, 2019). The advantage of this function over the sig moid function is the derivative result.

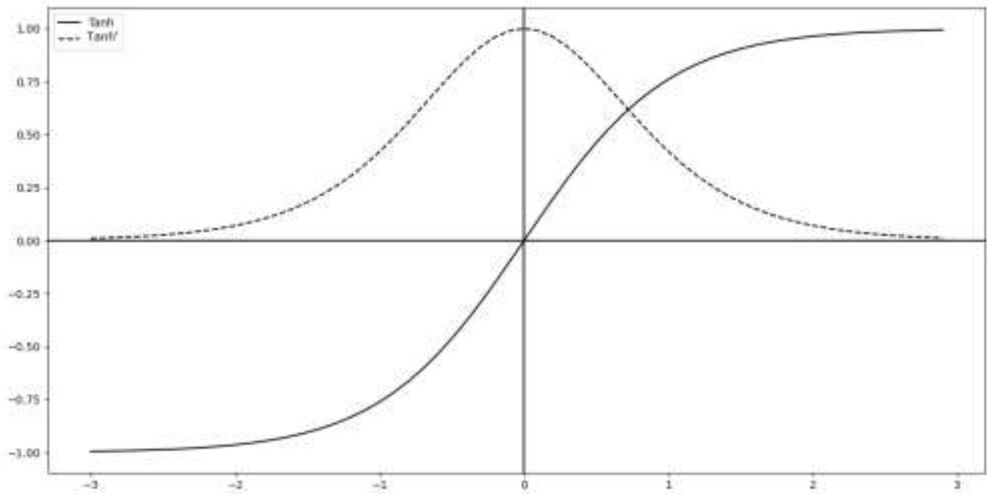

Figure 4.5 Tangent hyperbolic function

4.1.4.5 ReLU (Rectified Linear Unit) Function

If the data coming to the function is less than zero, the result is 0, and if it is between zero and one, it produces the results that give itself. For values less than zero, the result is 0, which makes the network run fast. The fact that the computation load is less than the sigmoid and hyperbolic tangent functions makes it more preferable in multilayer networks.

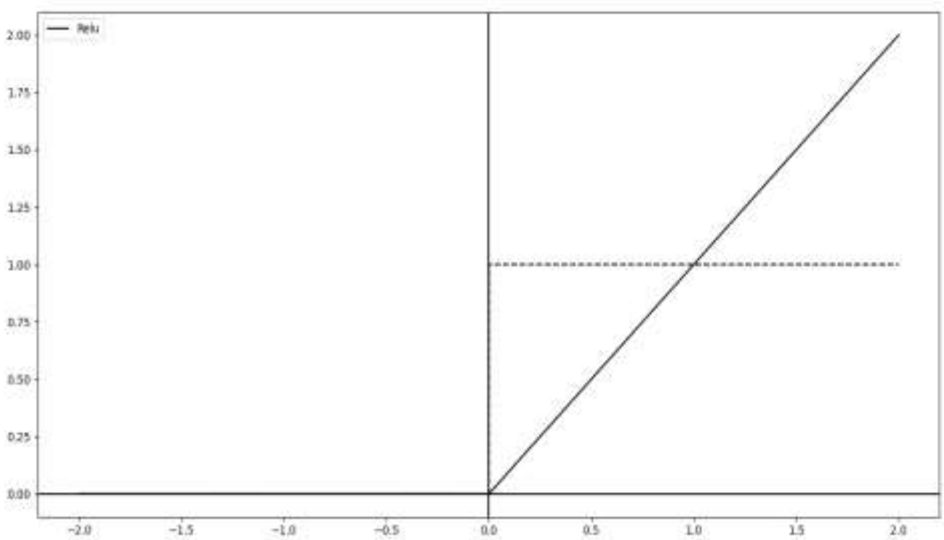

Figure 4.6 ReLU function

$$f(x) = \begin{cases} x \, if \; x > 0 \\ 0 \; if \; x \leq 0 \end{cases} \quad (4.5)$$

4.1.4.6 Softmax Function

It has a similar structure with the Sigmoid function and is used for the same purpose. It performs well when used in classification problems. The most important difference is that it is used in the output layer of artificial neural networks in situations that require more than two classifications. It enables the probability of the input to belong to a certain class to be determined by generating values between 0 and 1. That is, a probabilistic result is returned. The class with the highest probability returns as a result.

The loss function is the method that measures the error coefficient and success of the artificial neural network model. Loss function is generally defined in the output layer. The main job of this function is to calculate how the estimate of the model differs from the actual value. In the case where the actual value and the predicted value are the same, the loss value will be 0. The loss value of a good model should be a value that converges to 0.

4.1.5 Outputs

The result of the activation function of the cell is the output result of the neuron. The output value can be used as the result value of artificial neural networks, if necessary, or used as input to other layers and used again in the network.

Artificial neural networks have a layered structure. It can have many hidden layers or the input layer can be directly connected to the output layer. Depth in artificial neural networks is generally measured by the number of hidden layers. Each layer has its own neuron and activation functions. Neural networks generally consist of 3 separate layers. These layers are; Input Layer is Hidden (Intermediate) Layer and Output Layer.

Input layer; This is the layer from which the input (initial) data of the arguments come from. This layer contains as many neurons as the number of entries (independent variable).

Hidden Layer is located between the input layer and the output layer. The number of layers and the number of neurons may vary from model to model. Number of layers is a criterion that affects computational complexity, and if it is too large, it also increases computation time. In order to achieve successful solutions in complex problems, the number of layers and the number of cells are kept high. Although this situation negatively affects the working time, it increases the success rate in a correctly established network.

Output Layer; It is the layer that calculates the data received from the hidden layer and enables the data produced in the direction of the model to be returned as a result.

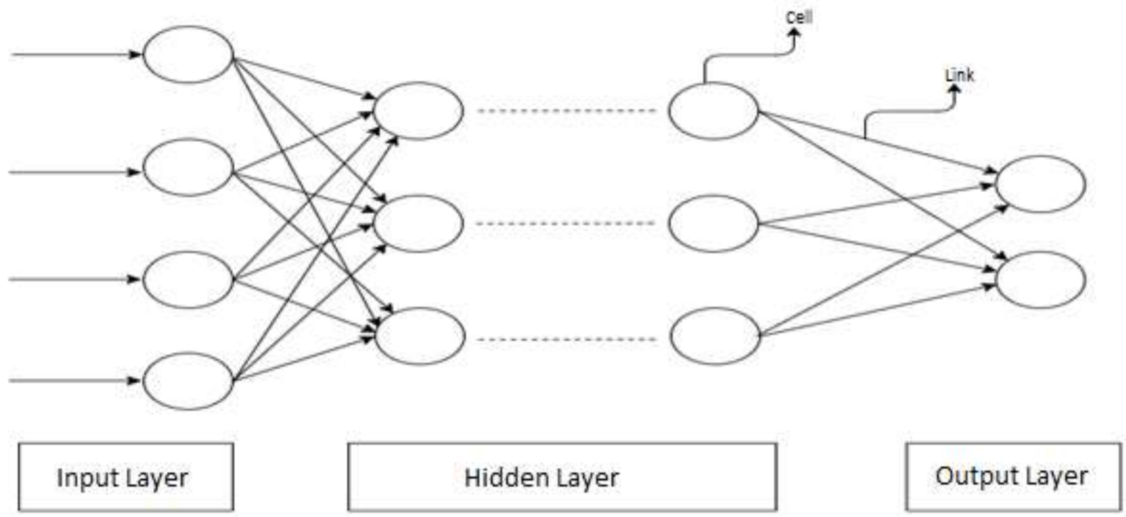

Figure 4.7 Layers of artificial neural networks

4.1.6 ANN Advantages and Disadvantages

Advantages of ANN:

- They perform machine learning.
- The information is stored on the network, not in the database.
- They learn by using examples. The examples are events that have happened. It is important that the samples obtained can show the event completely.

- In order for ANNs to be used safely, they must first be trained and tested.
- It can be used in events related to perception.
- They can make associations and classifications.
- They can supplement missing information.
- Has self-organizing and learning features (adaptation)
- They can work with missing information. According to the state of the network, it can be determined how important the missing information is for the network.

Disadvantages of ANN:

- They can deteriorate gradually.
- It requires hardware that can work in parallel.
- There are no rules governing how a network should be created. The appropriate network is found by trial and error.
- ANNs can produce acceptable solutions. But it cannot be claimed that this is the optimum solution without being tested.

4.2 Linear Regression

It is not possible to explain any dependent variable with a single independent variable in the fields of economy, energy and business administration. Economic models often arise from a combination of causes. Often times, more than one variable can affect a single variable. These variables can also affect each other among themselves. For this reason, it is not possible to perform a single regression analysis in cases where more than one such variable should be used. Regression analysis of dependent variable with more than one independent variable is called multiple regression analysis.

Multiple linear regression examines the linear relationship between two or more independent variables and one dependent variable. There is a correlation between dependent and independent variables in multiple regression. The most general multiple regression equation is as follows to show the X's free variables and Y the dependent variable.

$$Y = a0 + \sum_{i=1}^{r} aiXi + e \qquad (4.6)$$

In the multiple regression model, just as in the single model, the deviations of the free variables from the mean are used when calculating the coefficients. The reliability of the estimated coefficients in the model is tested by looking at the standard error and the smallness of the variance. This partially gives us a ratio for the fit of the predicted values to the real values.

Although the coefficients of the free variables in the regression model give information about the state, importance and power of the model, they do not show the value of the relationship between dependent and free variables. Therefore, we measure the relationship between dependent and free variables or variables with correlation analysis.

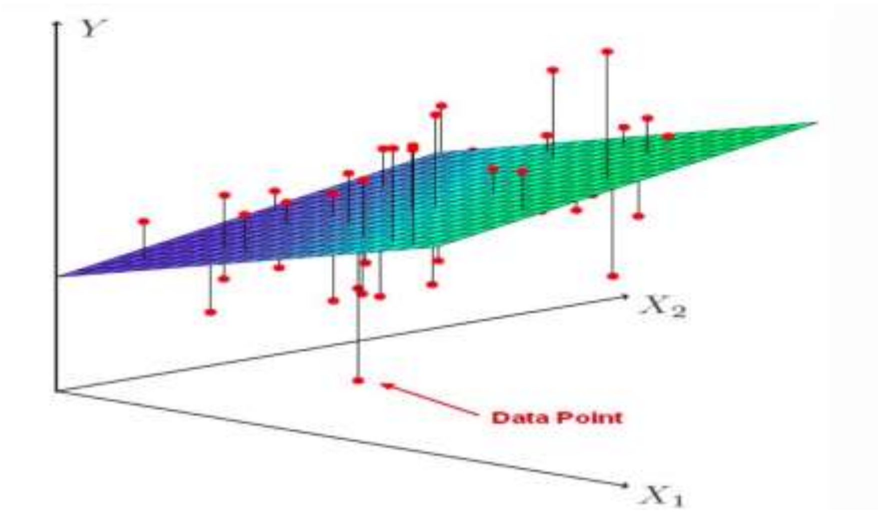

Figure 4.8 Multiple regression

4.3 Decision Tree

Tree-based learning algorithms are among the most widely used supervised learning algorithms. In general, they can be adapted to the classification and regression solution of all the problems discussed (Emel et al., 2019).

Decision trees are among the widely used machine learning algorithms. It is widely used in many data science problems. Therefore, it is very important for data scientists to learn this method.

The decision tree is used in data sets containing very large records. Because it is used to divide into smaller clusters by applying certain rules. The decision tree is built according to the method of gaining by dividing the data recursively from the root to the leaves. Initially, all data is collected at the root of the tree. The choice of variables is determined by the knowledge acquisition value. In order for the iterative algorithm to exit the loop, all elements in that node must be included in the same class. If the remaining values belong to only one class or there is no more classifiable value, the recursive algorithm is terminated and the decision tree is created. Each element in the resulting classes shows similar properties with other elements of the same class. Using decision trees has advantages and disadvantages from case to case.

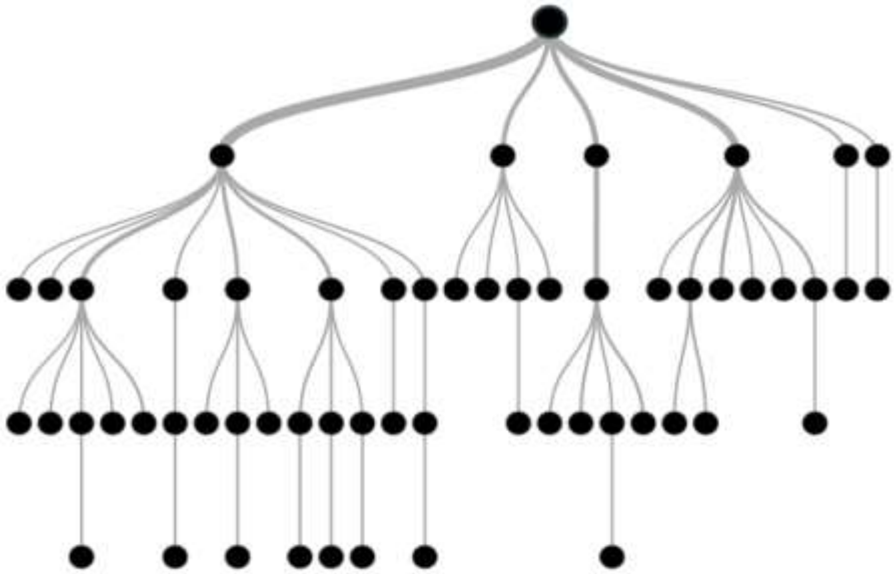

Figure 4.9 Decision tree

The decision tree is built according to the method of gaining by dividing the data recursively from the root to the leaves. Initially, all data is collected at the root of the tree. The choice of variables is determined by the knowledge acquisition value. In order for the iterative algorithm to exit the loop, all elements in that node must be included in the same class. If the remaining values belong to only one class or there is no more classifiable value, the cyclic algorithm is terminated and the decision tree is created. In decision tree algorithms, where it is divided is a feature that affects the accuracy of the tree. The division criteria for classification and estimation problems are different from each other.

The choice of algorithm varies according to the type of the result variable. The most frequently used algorithms in decision trees are as follows. Entropy, Gini and Classification Error are used for categorical variables. Least Squares method is used for continuous variables.

Entropy is a measure of the uncertainty associated with the data. Therefore, the data must be divided in a way that minimizes entropy. The better the splits, the better the forecast.

$$H = -\sum p(x) \log p(x) \tag{4.7}$$

Here, p (x) is the percentage of the group belonging to a particular class and H is the entropy.

The decision tree is intended to make divisions that minimize the entropy value. Information gain is used to determine the best choice. Information gain is calculated as follows (Emel et al., 2019).

$$Gain(S, D) = H(S) - \sum \frac{|V|}{|S|} H(V) \tag{4.8}$$

S is the original data set and D is a split part of the set. Each V is a subset of S. All of V is discrete and constitutes S. In this case, the information gain is equal to the difference between the entropy of the original data set and the entropy value of each feature set (Emel et al., 2019).

The most used decision tree algorithms are listed below.

- ID3
- C4.5
- CHAID
- CART

4.3.1 Decision Tree Algorithms

4.3.1.1 ID3 Algorithm

The ID3 algorithm was developed by J. Ross Quinlan (1986). The ID3 algorithm can be used if there are many qualifications in the database and if the training set contains too many records and at the same time it is desired to create a reasonable decision tree without much calculation. ID3 is an iterative algorithm. ID3 is based on CLS (Concept Learning System) algorithm (Quinlan 1986).

4.3.1.2 C4.5 Algorithm

It is an algorithm proposed by Quinlan (1993). It creates a classification decision tree by recursively dividing the data into subsets. The C4.5 algorithm can be thought of as an improved version of the ID3 algorithm. The algorithm can be used for properties with continuous values, pruning can be done and decision making can be performed (Berson 2000).

4.3.1.3 CART Algorithm

CART (Classification and Regression Trees) is an algorithm developed by Breiman. In the CART algorithm, the relevant cluster is separated into two more homogeneous subsets at each stage. Discrimination is done according to gini, twoing for categorical variables and least squares deviation for continuous variables according to index calculations. Various flexibility provided in these calculations such as the definition of the priorities between profit, cost values and variable categories cause the CART algorithm to be preferred intensively today (Breiman et al., 1984).

4.3.1.4 CHAID Algorithm

Another decision tree algorithm is the CHAID (Chi-Square Automati Interaction Detector) algorithm. CHAID is similar to the CART algorithm but uses a different way to segment the data. Instead of the entropy or gini metrics used to select the optimum sections, a technique that applies the chi square test is used (Berson 2000).

4.3.1.5 SLIQ Algorithm

SLIQ (Supervised Learning In Quest) was developed by IBM Quest. It uses the pre-sorting technique to create a decision tree by dividing large data sets into sections. This technique greatly avoids the sequencing cost on each node. SLIQ maintains a discrete sorted list for each node called the class list. Each element in this list corresponds to the attributes in the data and has a class label. SLIQ uses the width-first path when building the decision tree. It scans the appropriate sorted list for each attribute and calculates the entropy value for each value. After calculating the entropy for each attribute, one attribute is chosen to divide the data. This process continues recursively until data is divided into classes.

4.3.2 Pruning

The most common problems encountered in decision trees are the situations called over-learning and faulty learning. If the tree is too large and the error rate for training samples is low, but the classification error for test data is large, this is called over-learning. If the model gives erroneous results when the decision tree is not big enough, this situation is called less learning. In order to prevent such situations from occurring and to make the tree a suitable value, processes called pruning are performed.

As a rule, pruning is the removal of branches that are in the tree but will disturb the balance and reliability of the tree. Two methods can be used for pruning. The first of these is pruning, which is called pre-pruning and is provided by stopping the growth of the tree when it reaches a certain size. Another method is to prune the tree after it is fully formed.

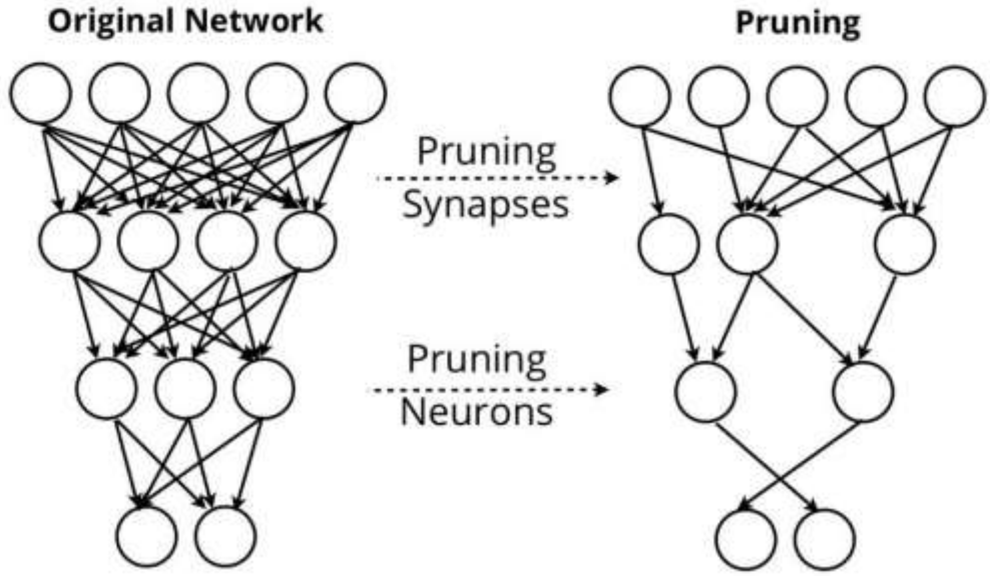

Figure 4.10 Pruning

4.4 Random Forest

The RF method was developed in 2001 by Leo Breiman. Breiman created a new method by combining the Bagging method developed by him in 1996 and The Random Subspace technique proposed by Ho in 1998 and used to randomly select subgroups. While Breiman developed this method; It was also influenced by a study described by Amit and Geman in 1997, where the best distinction for each node was determined through a random selection.

Random Forest method provides a valid prediction validity and interpretability of the model among machine learning methods. Because it includes random sampling and improved features of techniques in batch methods, the RF method makes good generalizations and makes valid predictions. The reasons for the success of the random forest estimates are low bias and poor correlation between trees. The low amount of bias is achieved by creating very large trees. By creating trees that are as different from each other as possible, a collection of low correlation is obtained.

The RF model is based on 2 parameters. These parameters; it is the number of trees to be created (B) and the number of estimators to be randomly selected at each node separation (m). Bootstrap method is used while creating each decision tree. In the Bootstrap method, samples are created using the same size as the number of observations (n) in the original data set. 2/3 of this sample is divided into two as the training data set (inBag) used to create the tree and the remaining 1/3 is the test data set (out of bag or OOB) to test the internal error rate of the model.

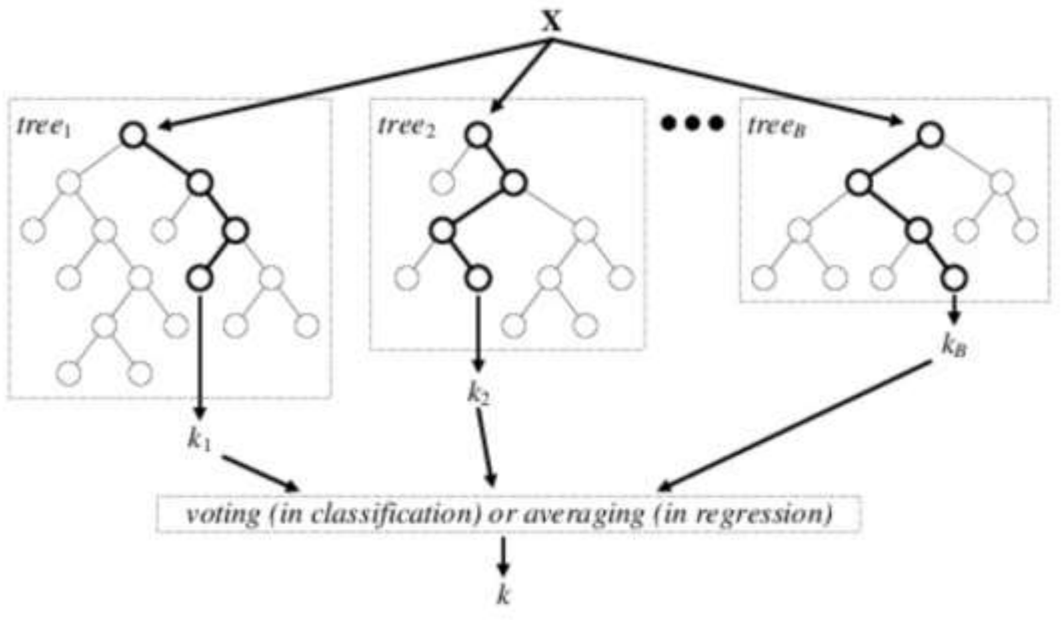

Figure 4.11 Random forest

4.4.1 Advantages and Disadvangates

The features of RF method that make it superior in tree-based community methods can be summarized as follows:

- Generally, the number of estimators in regression analysis should be smaller than the number of observations in the data set. There is no such requirement in the RF method.
- The use of large numbers of trees makes the RF application function more complex than the CART application function. However, it also calculates the error rate by using the OOB data set to evaluate the model performance. Thus, it compensates for overfitting which is a delicate problem for CART.
- Accuracy is quite high compared to many classifiers.
- Makes generalized error estimation in the forest formation process.
- It is an effective method for estimating lost data.
- It is a method that compensates for errors in unbalanced classified data sets.
- Derived forests can be saved for use in other datasets.
- It gives estimates of which variable is important in classification.
- Calculates distances between pairs of observations for clustering, determining deviating values, or scaling.

Apart from its superior features, the RF method has some limitations. These can be listed as follows:

- As with a single decision tree, the resulting result is not visually visible in the tree structure.
- It cannot give a confidence interval for the result.

4.4.2 Algorithm Description

1. With the Bootstrap method, n-volume data set is selected. This data set is divided into two as training data set (inBag) and test data set (OOB).
2. With the training data set (inBag), a decision tree with the largest width (CART) is created and this decision tree is not truncated. While creating this tree, m out of a total of p estimator variables are randomly selected when dividing each node. Here, the condition m <p must be met. Because it is undesirable to observe excessive growth of the tree and excessive adaptation. Branching occurs with the one with the highest information gain from these m estimators selected. The value of this variable is decided by the Gini index. This process is repeated for each node until there are no new branches to be created.
3. A class is assigned to each leaf node. Then the test data set (OOB) is dropped from the top of the tree and the assigned class is recorded for each observation in this data set.
4. All stages from the 1st to the 3rd step are repeated B times.
5. An evaluation is made with observations (OOB) that are not used while creating the tree. The number of times an observation is classified in which categories is counted.
6. Class assignment is made with a majority of votes determined over each observation, tree sets. For example, in a classification model with 2 categories, an observation carries the label of the class that receives at least 51% majority of votes over all trees, and this class becomes its estimated class value.

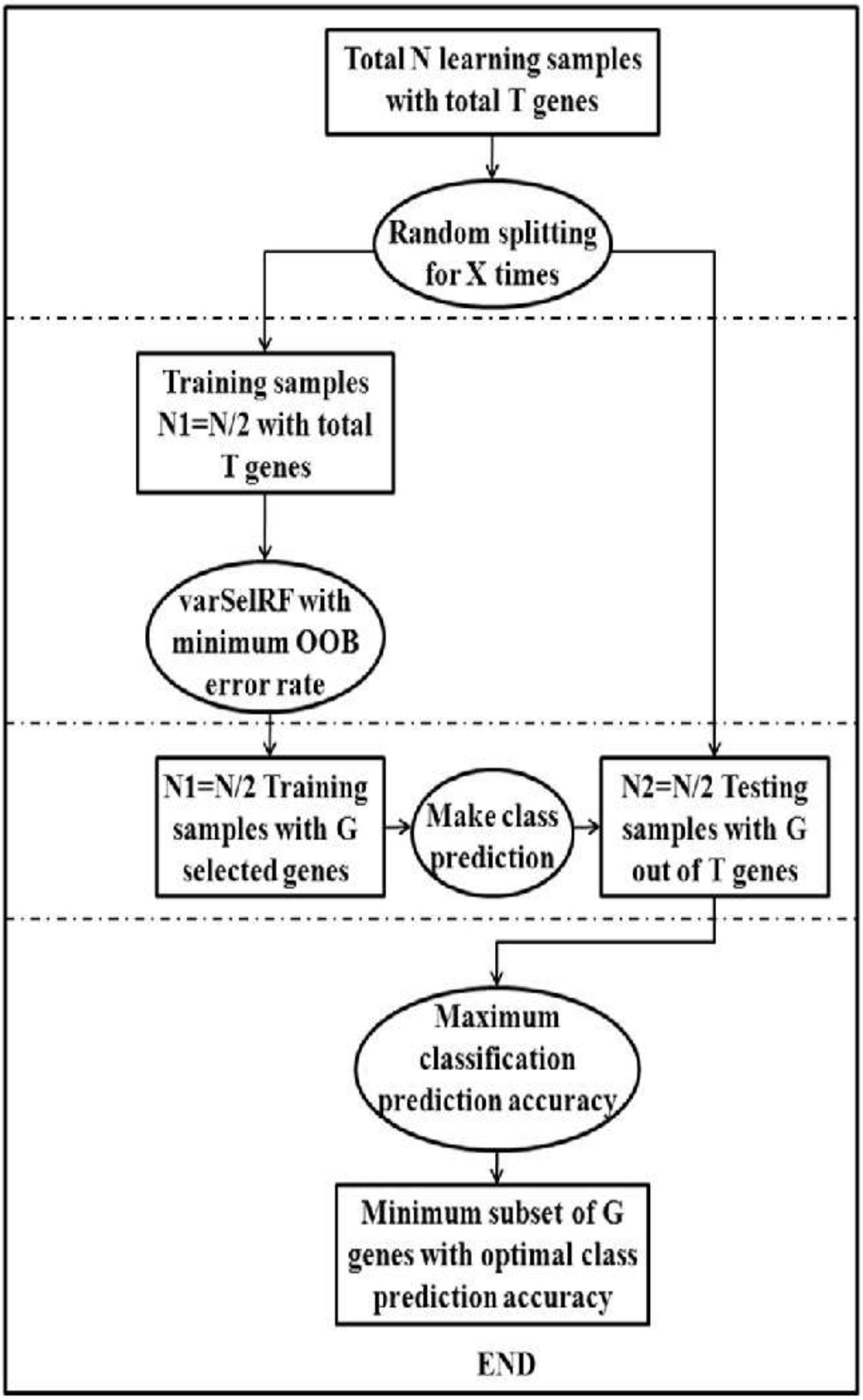

Figure 4.12 Flowcharts of random forest

CHAPTER FIVE
APPLICATION

In this section, the studies carried out in the estimation of the unit price of redemption are mentioned. The parameters and preprocessing steps in the data set are explained. Information was given about the artificial neural network models established.

5.1 Dataset

The data set used in the thesis study includes the values between 01.01.2018 - 31.10.2020. The data set includes date variables, climate data, electricity consumption values, dollar rate, PTF values, holidays and production values. There are numeric values as well as label values in the data set. The parameters in the data set are determined based on previously read articles. Values are generally taken from the transparency platform of EPIAS. Temperature, pressure and humidity variables were taken from a foreign website by parallel programming. Links are shared from the resources section. It contains 24840 data. Some columns in the data set have missing values. These columns are temperature, humidity, pressure, PTF and Yekdem cost and Yekdem income columns. In the data set, UECM, ST_UECM, Yek cost and Yek income variables are dependent and other columns are independent variables. Below are the graphs showing the distributions of independent parameters.

- Yekdem income mininum value: 98.72
- Yekdem income maximum value: 7,022,124.0
- Yekdem income mean value: 1,983,606.642
- Yekdem_i cost mininum value: -1,554,670.37
- Yekdem_i cost maximum value: 7,517,559.49
- Yekdem_i cost mean value: 1,612,050.876

- Yekdem_m cost mininum value: 20.76
- Yekdem_m cost maximum value: 5,511.46
- Yekdem_m cost mean value: 1,121.243
- Uecm mininum value: 14,578.41
- Uecm maximum value: 46,654.12
- Uecm mean value: 32,135.00
- St_uecm mininum value: 23,028.18
- St_uecm maximum value: 2,691.76
- St_uecm mean value: 11,239.674

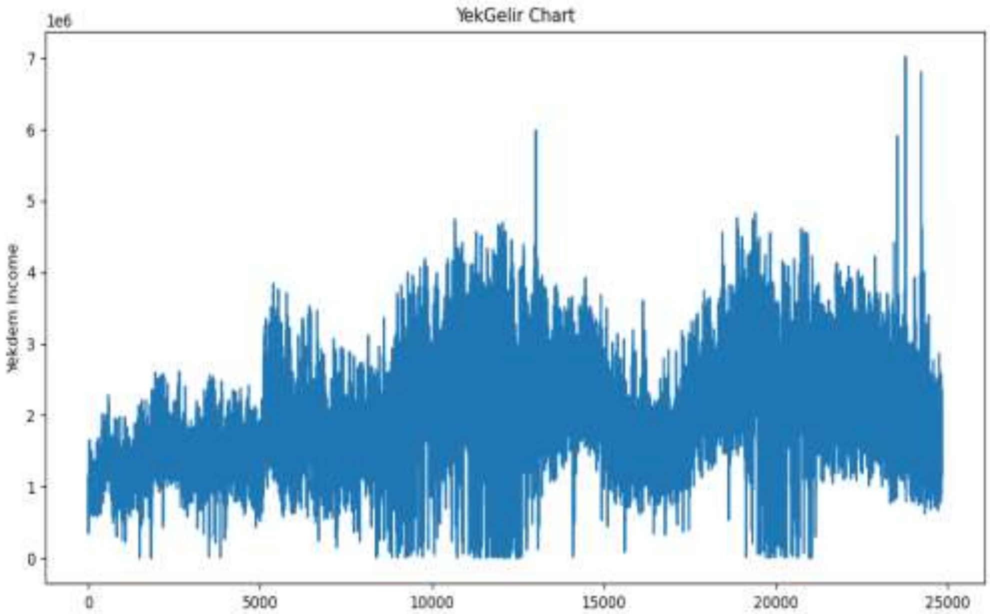

Figure 5.1 Yek invoice values

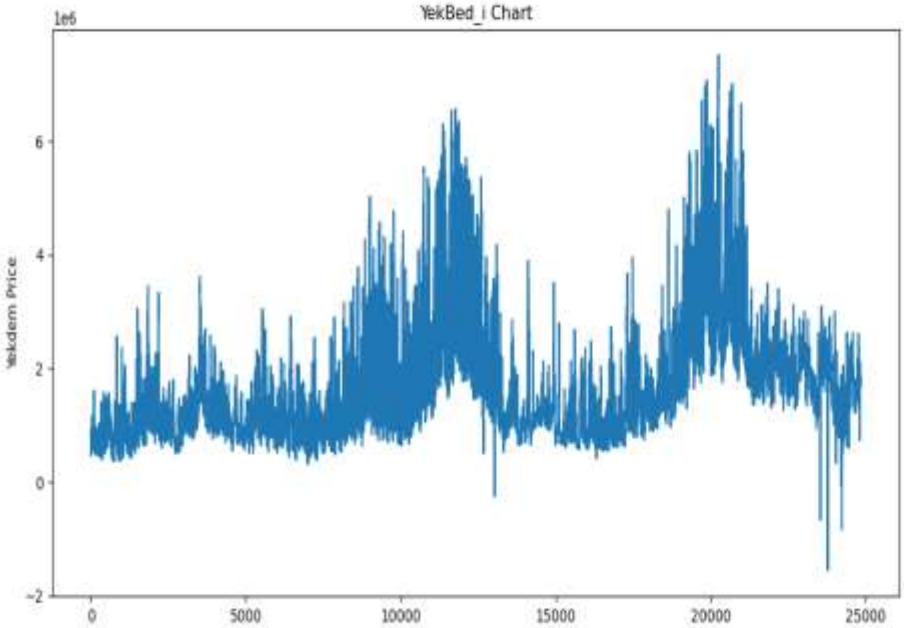

Figure 5.2 YekCost_i values

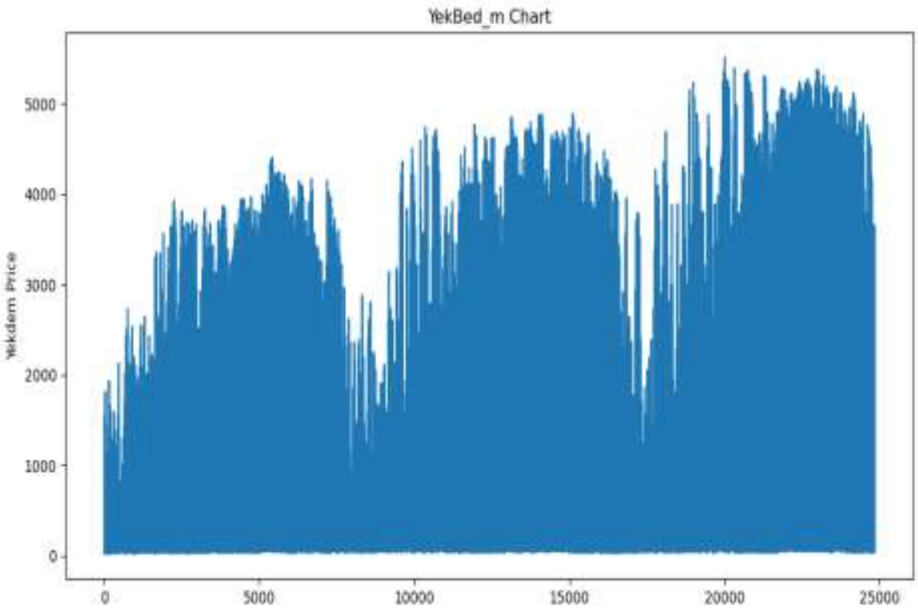

Figure 5.3 YekCost_m values

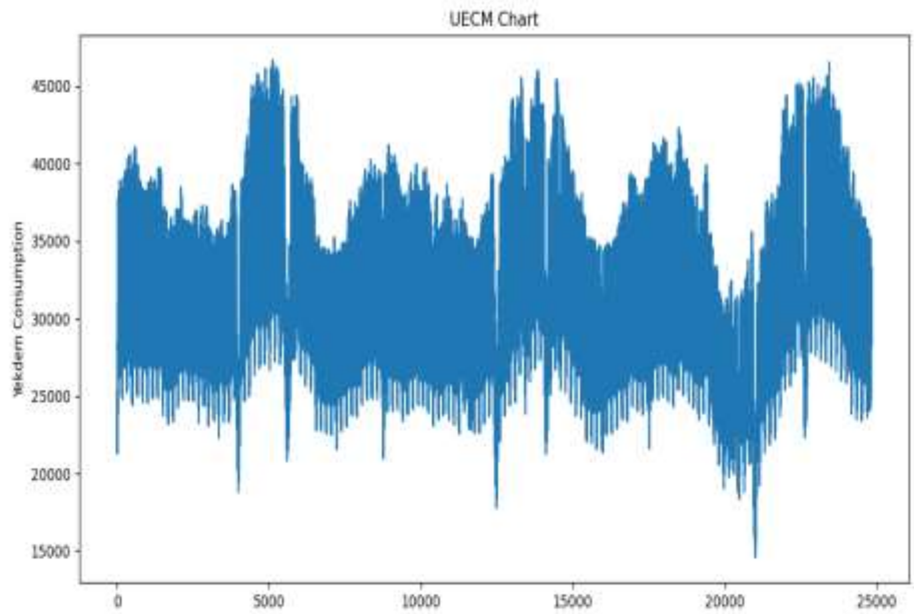

Figure 5.4 Uecmvalues

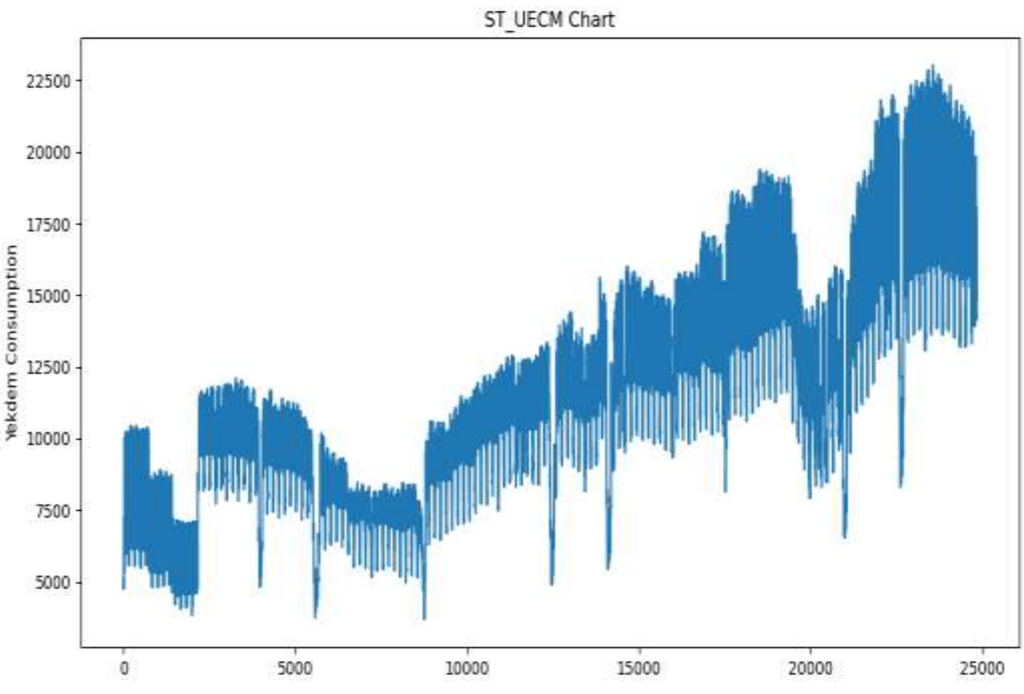

Figure 5.5 St_uecmvalues

5.2 Preprocessing and Feature Extraction

First of all, missing value control was made for each field containing numerical value. Samples with zero were determined and these samples were filled with the average value of their variables in their column. Missing values are filled in with the average of the relevant hour. Since the energy sector has a dynamic structure, there may be big differences between the two hours. Therefore, it is a more reliable method to average the same hours.

It has been observed that artificial neural networks are more successful in learning values between 0 and 1. Therefore, each column containing a numerical value is scaled. Each column's own maximum and minimum values were used in scaling. The values are converted to zero and one range with the formula below.

$$f(x) = {}^{x - min}/_{\max - min} \tag{5.1}$$

Encoding operations were done after these processes. Month, Hour, Day, Vacation class is used for class value rather than numerical value. Therefore, ecncoding process has been applied to these fields. Month values increased to 12 columns as a result of the encoding process, 24 columns at the end of encoding the clock hand, and 7 columns for the day class. After these processes, the whole data set increased to 73 columns.

After the encoding process, there were 73 columns in the data set. 4 of them are independent and others are dependent parameters. Separate models were created to estimate 4 independent parameters. The sub-data sets created with each model are divided into many columns. For this, feature elimination was made by examining the correlation values.

Correlation shows the orientation and value of the linear relationship between two or more variables. The amount of relationship between two variables is calculated using correlation techniques called prearson and spearman. The relationship of a variable to two or more variables is called multiple correlation. One of these variables is fixed and its relation with the other variables is calculated by partial correlation techniques (Şen et al., 2019).

The correlation coefficient is denoted by r and takes values between -1 and +1. If R = -1, there is a strong negative linear relationship. When R = +1, there is a strong positive linear relationship. Where R = 0, there is no relationship between variables (Şen et al., 2019).

- Very weak correlation or no correlation if r <0.2
- Weak correlation between 0.2-0.4
- Moderate correlation between 0.4-0.6
- High correlation between 0.6-0.8
- If 0.8> r, it is interpreted that there is a very high correlation.

Spearman and Pearson correlation matrices were examined, and features below 0.2 in each correlation type were eliminated. Correlation values are given in Table 5.1, Table 5.2, Table 5.3 and Table 5.4. Parameters below 0.2 are not included.

Data preprocessing stages are over with feature extraction. Later, the data set was divided into two parts as training and testing. In the literature, 20% of the data set is generally used for testing and 80% for training. But in this study, the time parameter is very important. Yekdem price forecast is a price that is updated every month. Therefore, only the data of the month to be estimated are included for the test process and the remaining part is used for training. For example, if a forecast for August 2020 is to be made, the data of the relevant month is included in the test part and the other part of the data set is reserved for training.

In order to calculate the Pearson correlation, the variables must be continuous, that is, at least equally spaced at the scale level. In order to talk about the significance of the Pearson correlation coefficient, the assumption that the sample distribution is normal must be fulfilled. In cases where the variables do not have a normal distribution, Spearman Rank correlation coefficient is preferred.

Table 5.1 Yek invoice correlation

Pearson		Sperman	
Yek Invoice	1.0	Yek Invoice	1.0
Day1_YekInvoice	0.21	Day1_YekInvoice	0.75
Week1_YekInvoice	0.68	Week1_YekInvoice	0.71
EUVM_Sum	0.53	EUVM_Sum	0.58
EUVM_Unlicensed	0.45	EUVM_Unlicensed	0.49
EUVM_Solar	0.42	PTF	0.46
EUVM_LandfillGas	0.33	EUVM_Solar	0.44
Dollar	0.28	EUVM_LandfillGas	0.38
EUVM_Wind	0.26	Dollar	0.30
EUVM_Biogas	0.59	EUVM_Wind	0.28
EUVM_Reservoir	0.23	EUVM_Biogas	0.27
EUVM_Biomass	0.22	EUVM_Biomass	0.27
PTF	- 0.21	EUVM_Reservoir	0.26
		EUVM_River	0.20

Table 5.2 Yek cost correlation

Pearson		Sperman	
YekPrice	1.0	YekPrice	1.0
Day1_YekPrice	0.79	Day1_YekPrice	0.80
Week1_YekPrice	0.71	Week1_YekPrice	0.72
EUVM_Reservoir	0.65	EUVM_Sum	0.64
EUVM_Sum	0.65	EUVM_Reservoir	0.57
EUVM_Canal	0.57	EUVM_LandfillGas	0.55
EUVM_River	0.50	Dollar	0.49
EUVM_Geothermal	0.44	EUVM_Biomass	0.48
EUVM_LandfillGas	0.43	EUVM_Geothermal	0.48
EUVM_Biomass	0.41	EUVM_Canal	0.47
Dollar	0.38	EUVM_River	0.45
PTF	0.32	EUVM_Wind	0.34
Month_08	0.32	EUVM_Other	0.32
Month_07	0.30	PTF	-0.30
EUVM_Other	0.27	Month_08	0.28
EUVM_Biogas	0.26	Month_03	-0.24
EUVM_Wind	0.24	EUVM_Biogas	0.22
EUVM_Solar	0.23		

Table 5.3 Eucm correlation

Pearson		Sperman	
EUCM	1.0	EUCM	1.0
Day1_EUCM	0.86	Day1_EUCM	0.85
Week1_UECM	0.86	Week1_UECM	0.85
Humidty	-0.50	Humidty	-0.48
Temperature	0.38	Temperature	0.34
Thursday	-0.28	Thursday	-0.29
Time_5	-0.25	Time_5	-0.25
Time_4	-0.24	Time_4	-0.25
Month_10	0.23	Time_3	-0.24
Time_3	-0.23	Time_2	-0.21
Time_2	-0.20	Month_10	0.21
Time_6	-0.20	Time_6	-0.20

Table 5.4 St_eucmcorrelation

Pearson		Sperman	
ST_EUCM	1.0	ST_EUCM	1.0
Week1_ST_EUCM	0.91	Day1_ST_EUCM	0.91
Day1_ST_EUCM	0.91	Week1_ST_EUCM	0.90
Humidty	-0.35	Humidty	-0.32
Temperature	0.29	Temperature	0.26
Thursday	-0.22	Thursday	-0.22

5.3 Models

A separate model is established for each independent variable in the data set. This section provides information about models.

5.3.1 Yekdem Income Models

In this model, time-valued, production values, dollar rate and historical YEKDEM income values are used as dependent parameters while estimating the benefit income. Parameter numbers have changed after feature extraction. Therefore, different artificial neural networks are used before and after feature extraction. The most successful model was chosen by changing the hidden and neuron numbers of the models.

There are 7 layers in the model before the elimination by correlation. 60 neurons were used in the input layer. In other layers, the number of neurons was divided by half and advanced. Relu was used as activation energy in these layers. There is 1 neuron in the output layer and the activation function is linear. There are 6 layers in the model after the elimination by correlation. 40 neurons were used in the input layer. In other layers, the number of neurons was divided by half and advanced. Relu was used as activation energy in these layers. There is 1 neuron in the output layer and the activation function is linear. In the Random Forest model, 10 was assigned to the estimators variable. Then the data set was divided into two as training and testing. After the models were trained, the estimation process was made. The models have been advanced 100 iterations.

Table 5.5 Yekdem income models

Model	Layer	Number of Cell	Activation Function
Model-1	7	60-30-15-7-5-3-1	Relu / Linear
Model-2	6	40-20-10-5-3-1	Relu / Linear

5.3.2 Licensed Yekdem Cost Models

In this model, time value, production values, dollar rate and historical Yekdem cost value are used as dependent parameters while estimating the Yekdem cost. Unlike the Yekdem income model, unlicensed production values are not taken in this model. Parameter numbers have changed after feature extraction. Therefore, different artificial neural networks were used before and after feature extraction. The most successful model was selected by changing the hidden and neuron numbers of the models.

There are 7 layers in the model before the elimination by correlation. 60 neurons were used in the input layer. In other layers, the number of neurons was divided by half and advanced. Relu was used as activation energy in these layers. There is 1 neuron in the output layer and the activation function is linear. There are 5 layers in the model after the elimination by correlation. 16 neurons were used in the input layer. In other layers, the number of neurons was divided by half and advanced. Relu was used as activation energy in these layers. There is 1 neuron in the output layer and the activation function is linear. In the Random Forest model, 10 was assigned to the estimators variable. Then the data set was divided into two as training and testing. After the models were trained, the estimation process was made. The models have been advanced 100 iterations.

Table 5.6 Yekdem licensed cost models

Model	Layer	Number of Cell	Activation Function
Model-1	7	60-30-15-7-5-3-1	Relu / Linear
Model-2	5	16-8-4-2-1	Relu / Linear

5.3.3 Unlicensed Yekdem Cost Models

Machine learning or artificial neural network models were not used in this model. The ptf and unlicensed production values in the data set were multiplied. Before multiplying, the missing values in both parameters are filled.

5.3.4 Eucm Models

In this model, while estimating eucm consumption data, time-valued, past eucm consumption values, temperature, humidity, pressure and holiday values are used as dependent parameters. Parameter numbers have changed after feature extraction. Therefore, different artificial neural networks are used before and after feature extraction. The most successful model was chosen by changing the hidden and neuron numbers of the models.

There are 6 layers in the model before the elimination by correlation. 58 neurons were used in the input layer. In other layers, the number of neurons was divided by half and advanced. Relu was used as activation energy in these layers. There is 1 neuron in the output layer and the activation function is linear. There are 4 layers in the model after sieving by correlation. 13 neurons were used in the input layer. In other layers, the number of neurons was divided by half and advanced. Relu was used as activation energy in these layers. There is 1 neuron in the output layer and the activation function is linear. In the Random Forest model, 10 was assigned to the estimators variable. Then the data set was divided into two as training and testing. After the models were trained, the estimation process was made. The models have been advanced 100 iterations.

Table 5.7 Uecm models

Model	Layer	Number of Cell	Activation Function
Model-1	6	56-28-14-7-3-1	Relu / Linear
Model-2	5	13-7-3-1	Relu / Linear

5.3.5 St_eucm Models

In this model, while estimating eucm consumption data of eligible consumers, time values, past eucm consumption values, temperature, humidity, pressure and values of holidays are used as dependent parameters. Parameter numbers have changed after feature extraction. Therefore, different artificial neural networks are used before and after feature extraction. The most successful model was chosen by changing the interlayer and neuron numbers of the models.

There are 5 layers in the model before the elimination by correlation. 35 neurons were used in the input layer. In other layers, the number of neurons was reduced by 10. Relu was used as activation energy in these layers. There is 1 neuron in the output layer and the activation function is linear. There are 3 layers in the model after sieving by correlation. 6 neurons were used in the input layer. In other layers, the number of neurons was divided by half and advanced. Relu was used as activation energy in these layers. There is 1 neuron in the output layer and the activation function is linear. In the Random Forest model, 10 was assigned to the estimators variable. Then the data set was divided into two as training and testing. After the models were trained, the estimation process was made. The models have been advanced 100 iterations.

Table 5.8 St_uecm models

Model	Layer	Number of Cell	Activation Function
Model-1	5	35-25-15-5-1	Relu / Linear
Model-2	5	6-3-1	Relu / Linear

CHAPTER SIX
RESULTS

In this section, the results of 5 different models mentioned earlier are given. Results are visualized with graphics and mse / mpe values are explained. The blue dots in the graph represent the estimates and the red dots represent the actual values.

Two methods were used as a performance criterion in the project. The first of these is root mean square error (MSE) and mean absolute error (MAE).

Average absolute error is the measure of the difference between two different variables. MAE is the average vertical distance between the true value and the predicted value. Models with an MAE value converging to zero perform better. N is the number of samples and e is the difference between the actual value and the estimated value.

MSE is used to measure the performance of a machine learning model, the predictive model, and is always positive. It can be said that models with an MSE value close to zero show a better performance.

$$f(x) = \frac{1}{n}\sum_{J=1}^{n}|e_J| \tag{6.2}$$

$$f(x) = \frac{1}{n}\sum_{j=1}^{n} e_j^2 \tag{6.3}$$

6.1 Yekdem Income Model Results

In this model, it is estimated with 4 models for the last 3 months in the data set. The results were shared before and after the feature was extracted with the correlation. The charts contain values for the period 09.2020. Graphics are shared within all 4 models.

MSE was found 0.047 in the ANN model. The model has been advanced 100 steps. There are 25 values whose difference between the estimate and the actual value is greater than 0.10. The ANN Model is successful in estimating the income values of YEKDEM at the level of 0.0 - 0.7, and is insufficient at the extreme values. MSE was found to be 0.02 after elimination with correlation in the ANN model. The model has been advanced 100 steps. There are 3 values whose difference between the estimate and the actual value is greater than 0.10. The ANN Model is successful in predicting the income values of YEKDEM at the level of 0.0 - 0.6, and it is insufficient at the extreme values.

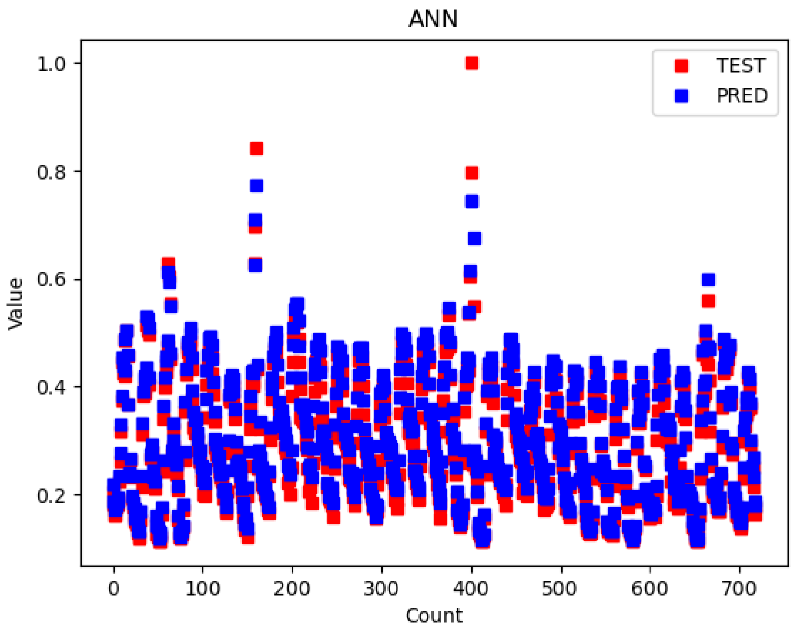

Figure 6.1 ANN result without elimintion

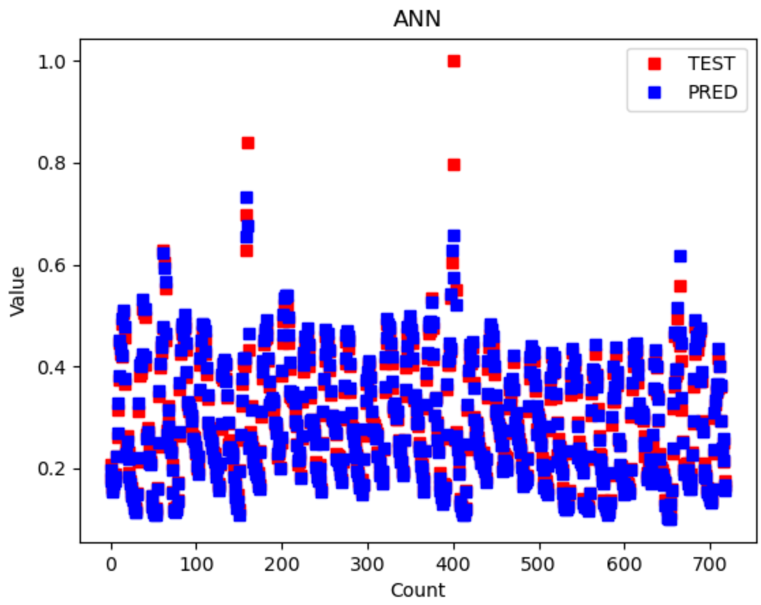

Figure 6.2 ANN result with elimination

In the linear regression model, the MSE was found to be 0.054. The model is successful in predicting the income values of YEKDEM at the level of 0.0 - 0.5, and it is insufficient at the extreme values. After screening with correlation in the linear regression model, the MSE was found to be 0.05. The model is successful in predicting the income values of YEKDEM at the level of 0.0 - 0.6 and it is insufficient at the extreme values.

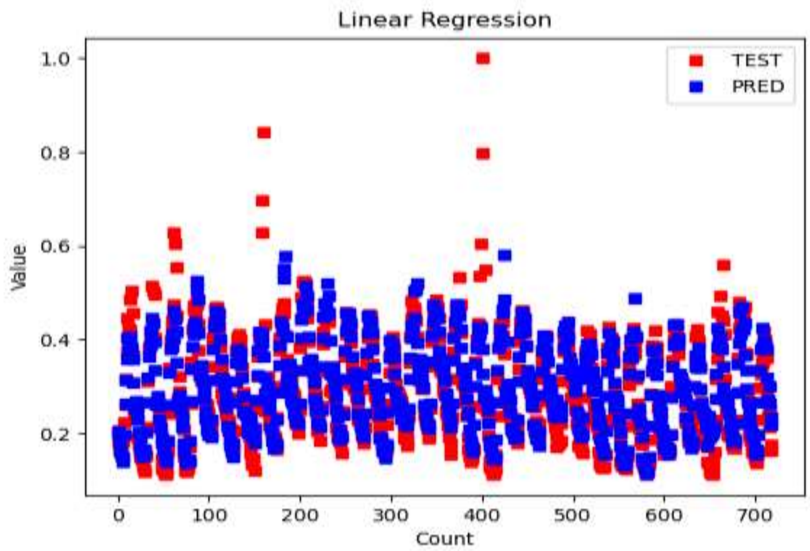

Figure 6.3 LR result without elimination

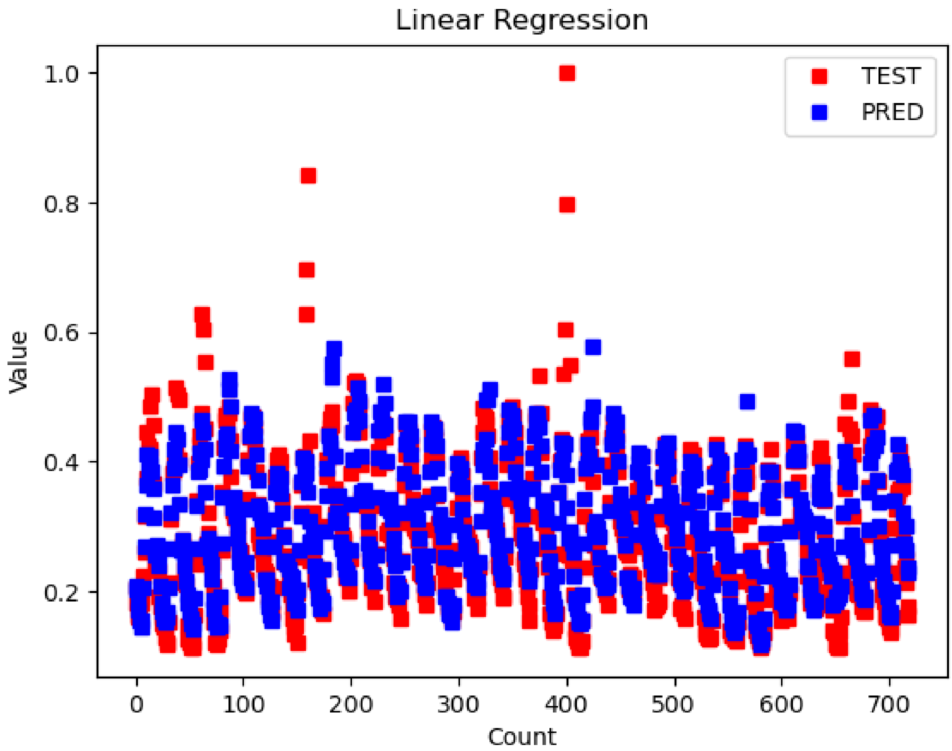

Figure 6.4 LR result with elimination

In the Decision Tree regression model, the MSE was found to be 0.029. The model is successful in predicting the income values of YEKDEM at the level of 0.0 - 0.6, and it is better at extreme values than the other models. After screening with correlation in Decision Tree model, MSE was found to be 0.031. The model was successful in estimating YEKDEM income values at the level of 0.0 - 0.6, and it gave good results in extreme values.

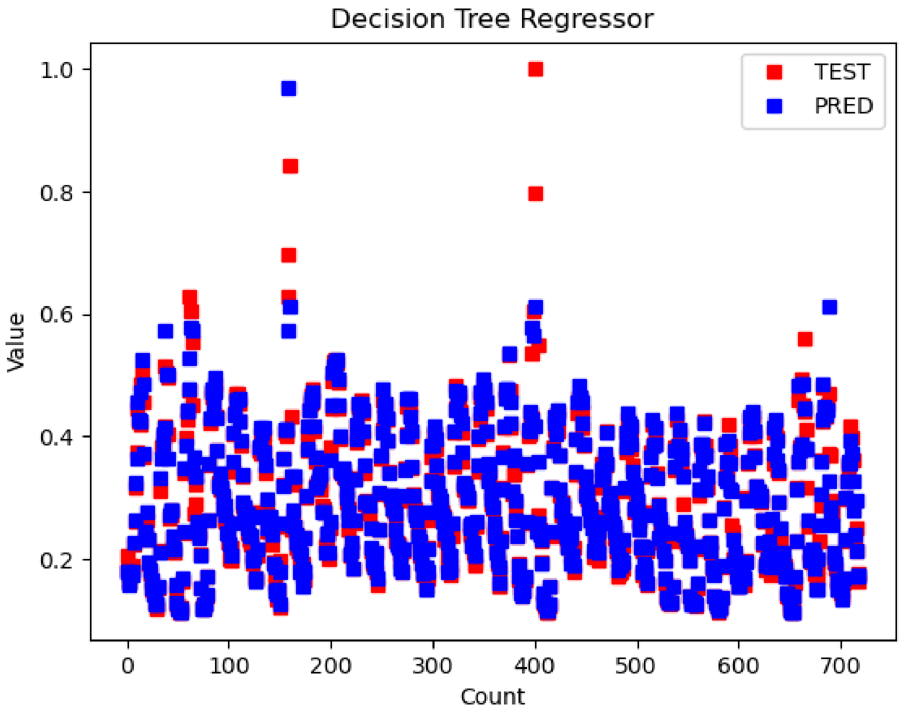

Figure 6.5 DT result without elimination

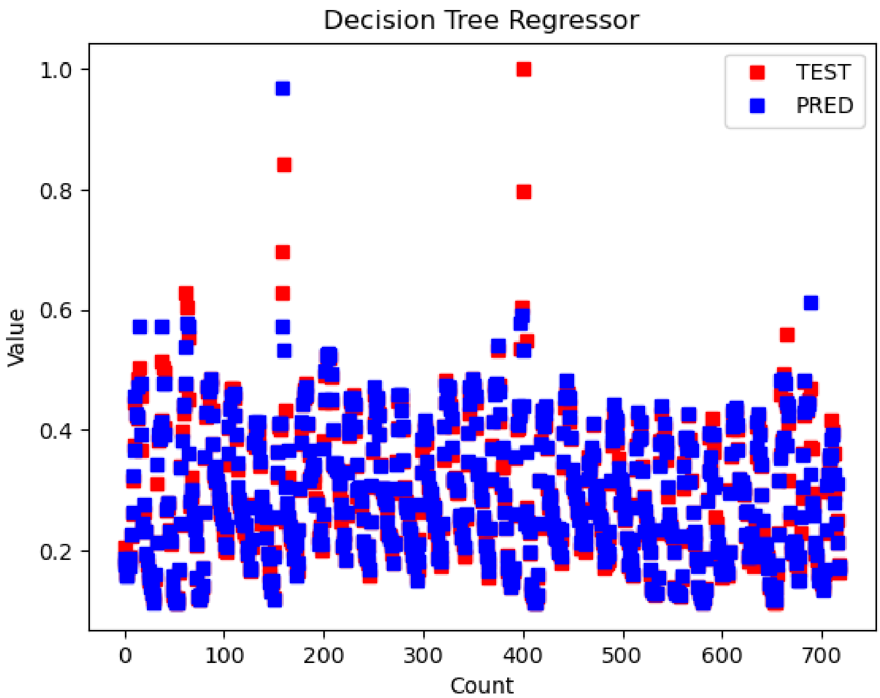

Figure 6.6 DT result with elimination

MSE was found to be 0.030 in the Random Forest regression model. The model is successful in predicting the income values of YEKDEM at the level of 0.0 - 0.7, and it is better than other models in extreme values. The MSE was found to be 0.029 after screening with correlation in the Random Forest model. The model was successful in predicting YEKDEM income values at the level of 0.0 - 0.7, and it gave good results in extreme values.

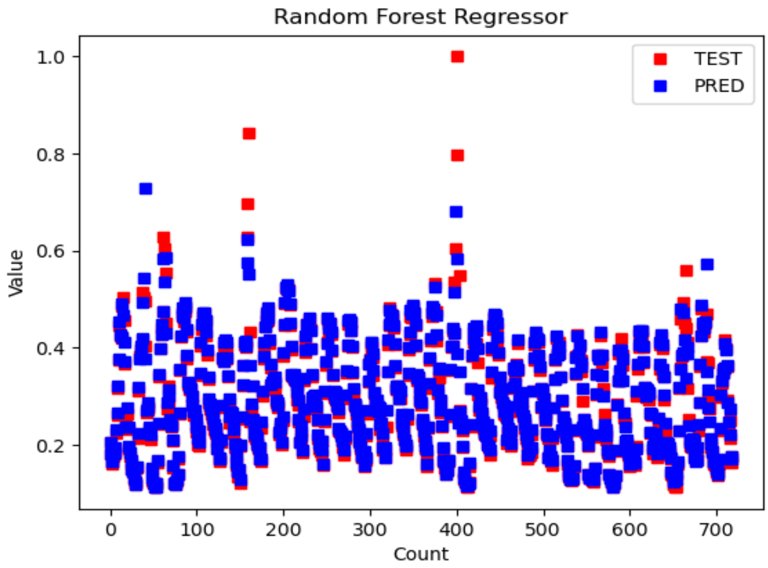

Figure 6.7 RF result without elimination

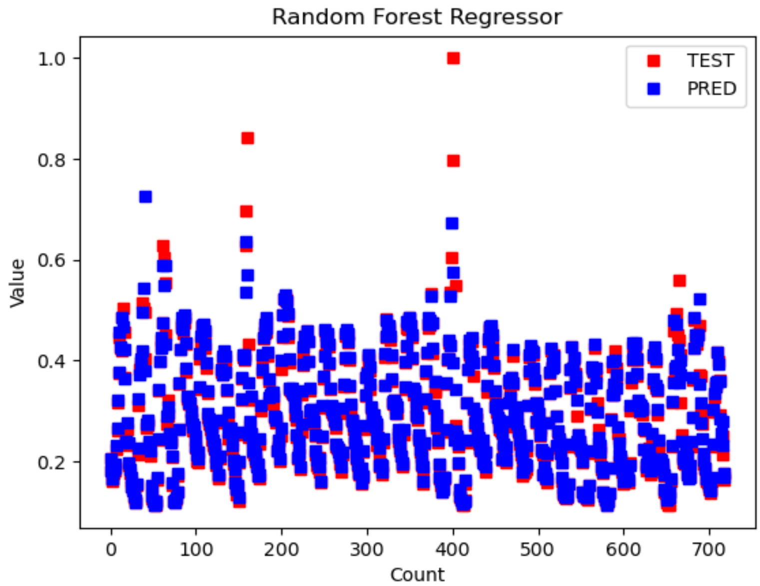

Figure 6.8 RF result with elimination

6.2 Yekdem Cost Model Results

In this model, it is estimated with 4 models for the last 3 months in the data set. The results were shared before and after the feature was extracted with the correlation. The charts contain values for the period 09.2020. Graphics are shared within all 4 models.

MSE was found 0.030 in the ANN model. The model has been advanced 100 steps. There are 9 values whose difference between the estimate and the actual value is greater than 0.10. The ANN Model is successful in estimating the values of YEKDEM Cost at the level of 0.3 - 0.4 and it is insufficient for values less than 0.4. The MSE was found to be 0.03 after elimination with correlation in the ANN model. The model has been advanced 100 steps. There are 8 values whose difference between the estimate and the actual value is greater than 0.10. The ANN Model is successful in estimating the values of YEKDEM Cost at the level of 0.3 - 0.4, and remains insufficient at the extreme values.

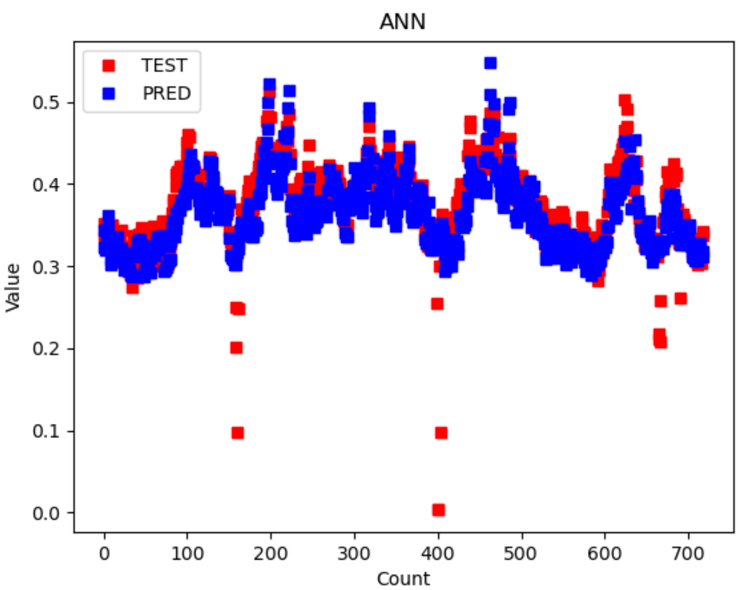

Figure 6.9 ANN result without elimination

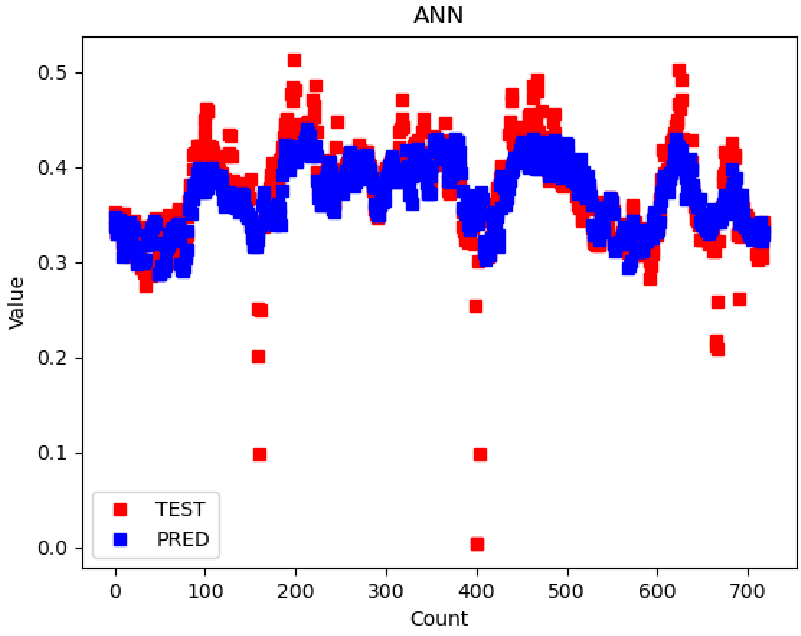

Figure 6.10 ANN result with elimination

In the linear regression model, the MSE was found to be 0.035. The model is successful in estimating the values of YEKDEM Cost at the level of 0.3 - 0.5 and is insufficient at the extreme values. The MSE was found to be 0.038 after elimination with correlation in the linear regression model. The model is successful in estimating the values of YEKDEM Cost at the level of 0.3 - 0.4 and falls short at the extreme values. ANN predicts similar results.

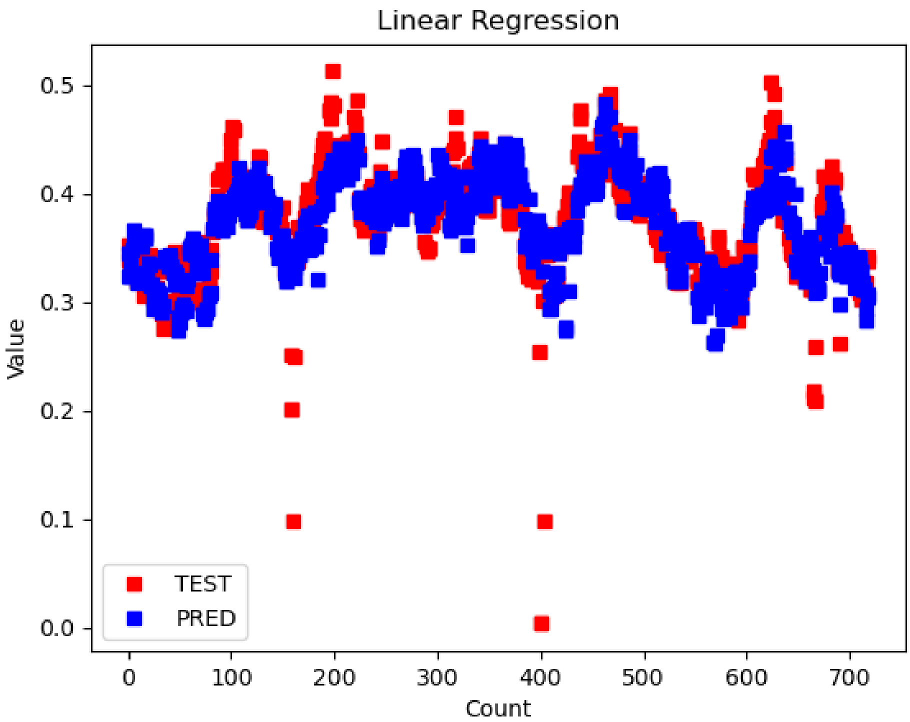

Figure 6.11 LR result without elimination

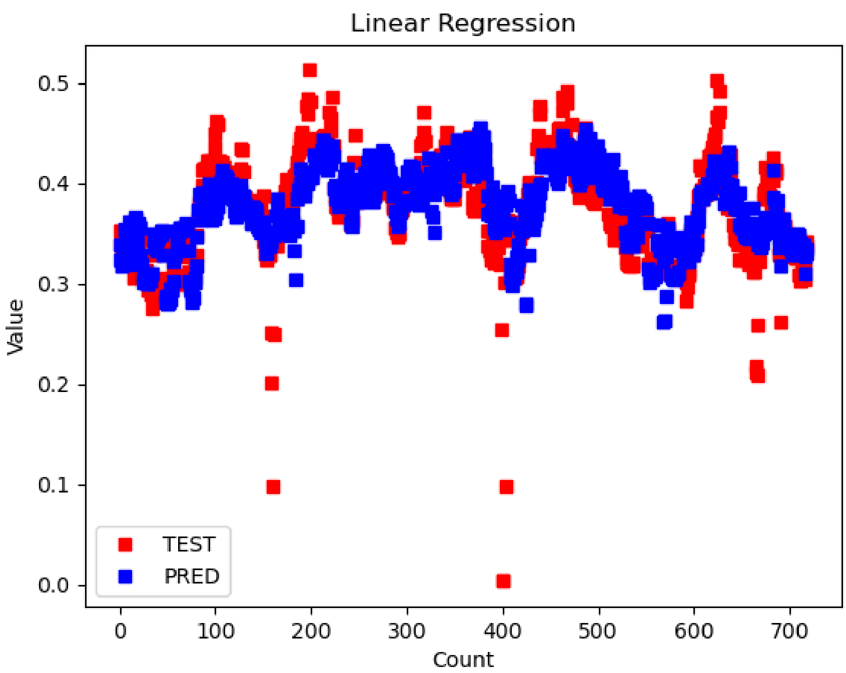

Figure 6.12 LR result with elimination

In the Decision Tree regression model, the MSE was found to be 0.031. The model is successful in estimating the value of YEKDEM Cost at the level of 0.2 - 0.5 and is better than other models at extreme values. After screening with correlation in Decision Tree model, MSE was found to be 0.028. The model was successful in estimating the values of YEKDEM Cost at the level of 0.2 - 0.5 and gave good results in extreme values.

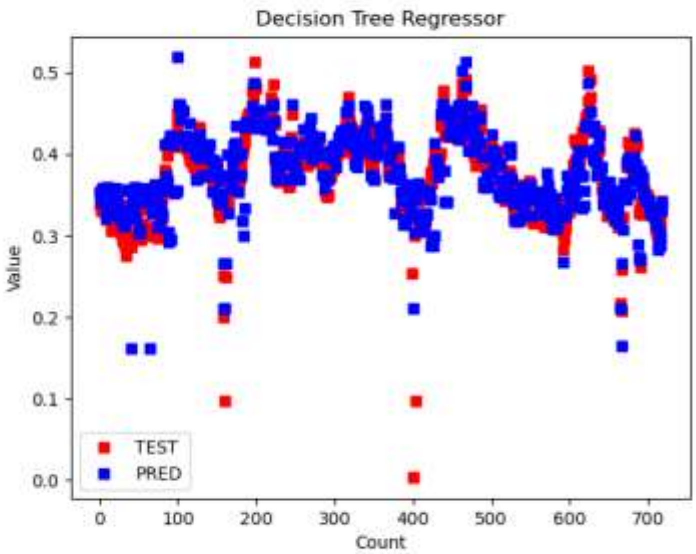

Figure 6.13 DT result without elimination

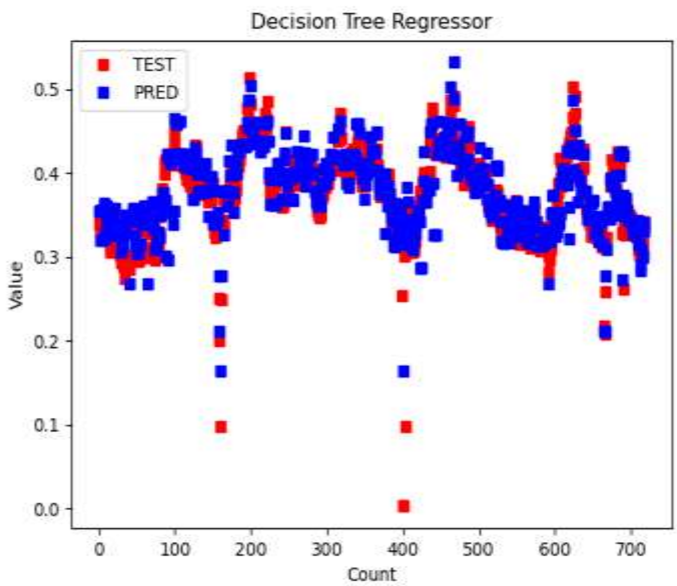

Figure 6.14 DT result with elimination

MSE was found 0.033 in the Random Forest regression model. The model is successful in estimating the values of YEKDEM Cost at the level of 0.3 - 0.5 and is better than other models at extreme values. MSE was found to be 0.033 after elimination with correlation in the Random Forest model. The model was successful in estimating the values of YEKDEM Cost at the level of 0.3 - 0.5 and gave good results in extreme values.

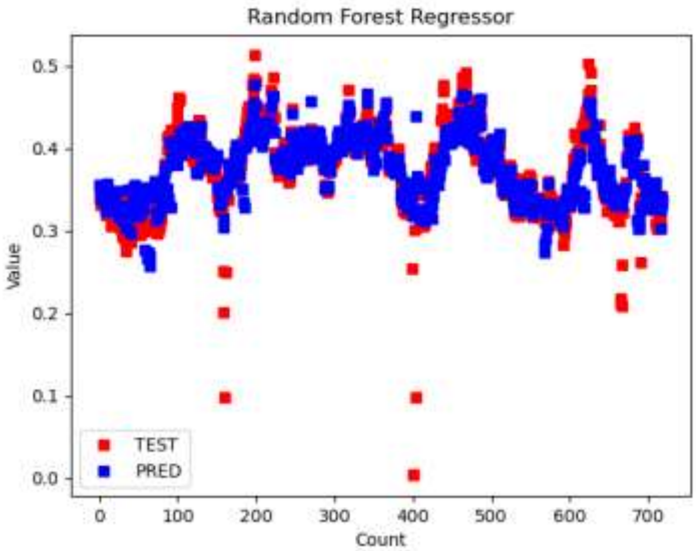

Figure 6.15 RF result without elimination

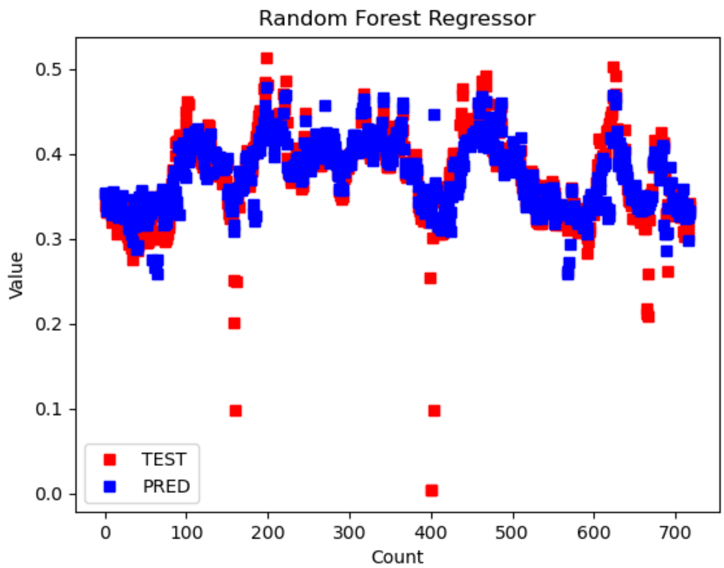

Figure 6.16 RF result with elimination

6.3 UECM Model Results

In this model, it is estimated with 4 models for the last 3 months in the data set. The results were shared before and after the feature was extracted with the correlation. The charts contain values for the period 09.2020. Graphics are shared within all 4 models.

MSE was found 0.020 in the ANN model. The model has been advanced 100 steps. There is no value with a difference greater than 0.10 between the estimate and the actual value. The ANN Model is successful in predicting uecm values at 0.4 - 1.0 level and is generally a good model. The MSE was found to be 0.03 after elimination with correlation in the ANN model. The model has been advanced 100 steps. There are 8 values whose difference between the estimate and the actual value is greater than 0.10. ANN Model has been successful in predicting uecm values of 0.3 - 1.0 level.

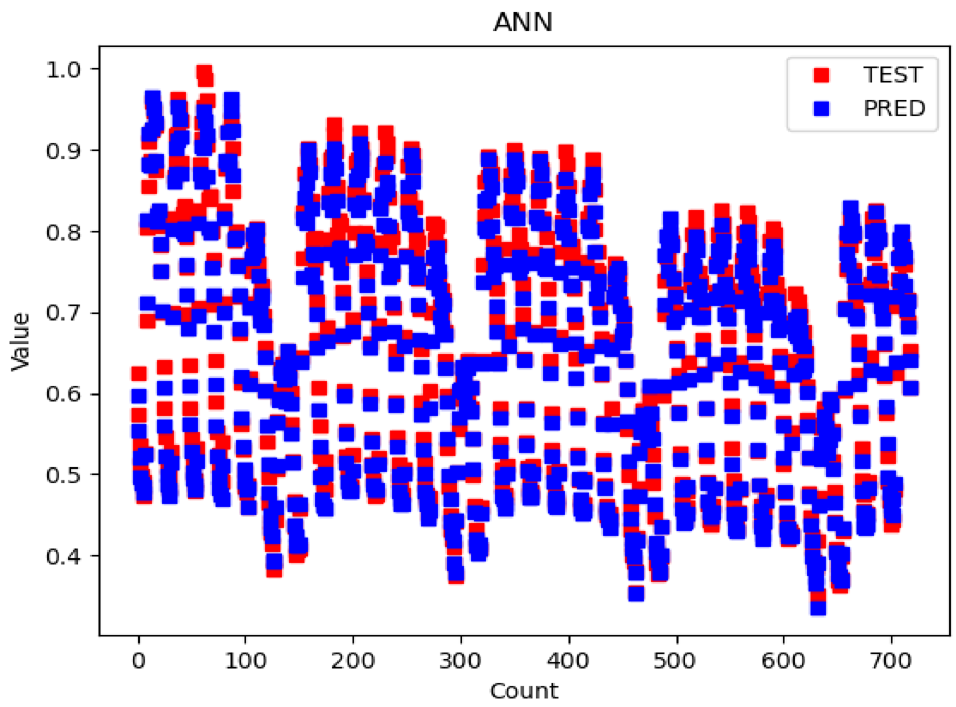

Figure 6.17 ANN result without elimination

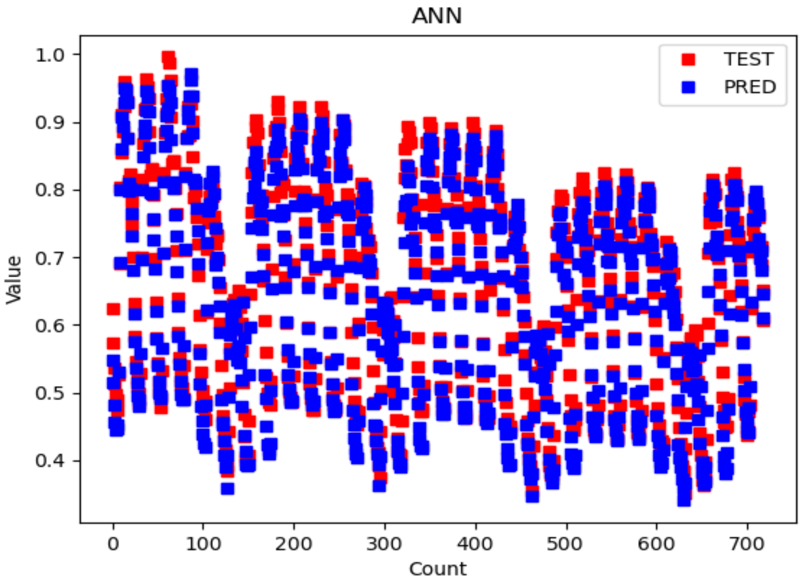

Figure 6.18 ANN result with elimination

In the linear regression model, the MSE was found to be 0.048. The model is successful in predicting uecm values at the level of 0.3 - 0.9, but falls short at extreme values. After screening with correlation in the linear regression model, the MSE was found to be 0.48. The model was successful in predicting uecm values between 0.4 and 0.9.

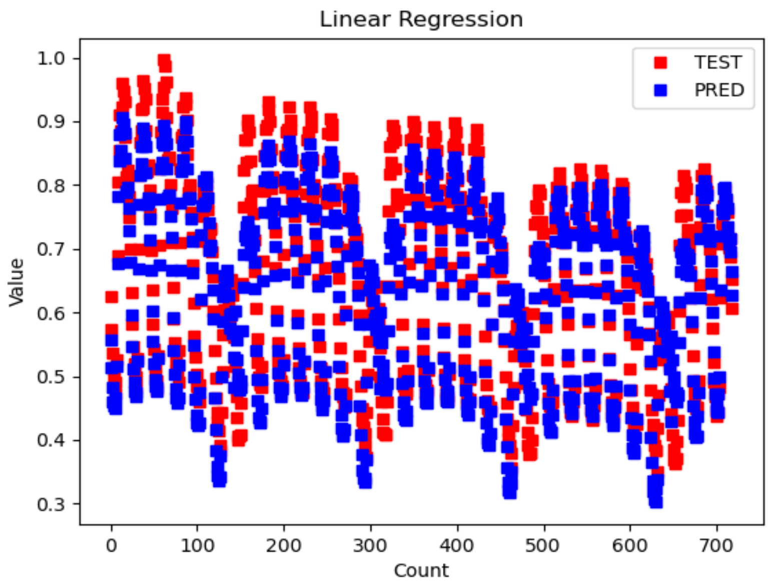

Figure 6.19 LR result without elimination

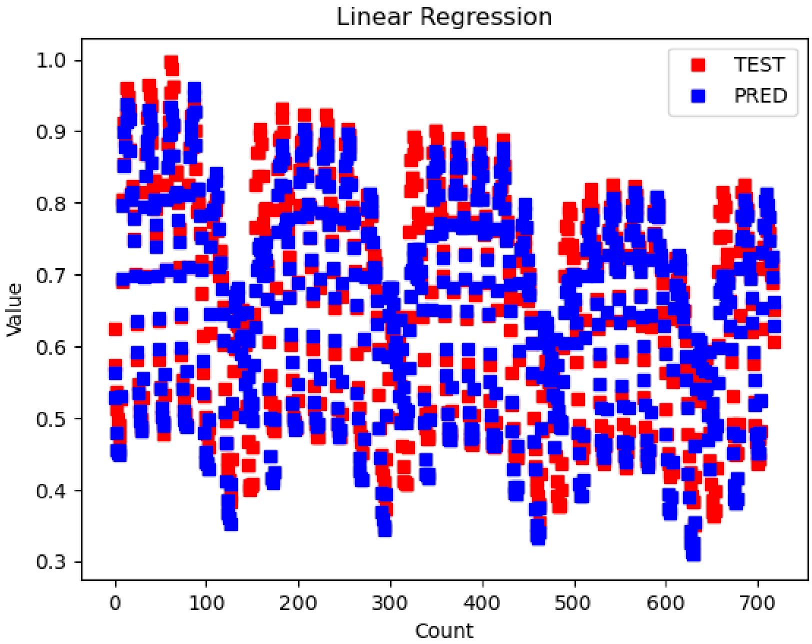

Figure 6.20 LR result with elimination

In the Decision Tree regression model, the MSE was found to be 0.031. The model is successful in predicting uecm values at the level of 0.3 - 1.0, and is better at extreme values than other models. After screening with correlation in Decision Tree model, MSE was found 0.056. The model was successful in predicting uecm values at the level of 0.3 - 1.0, and gave good results at extreme values. Feature extraction had a negative effect on this model.

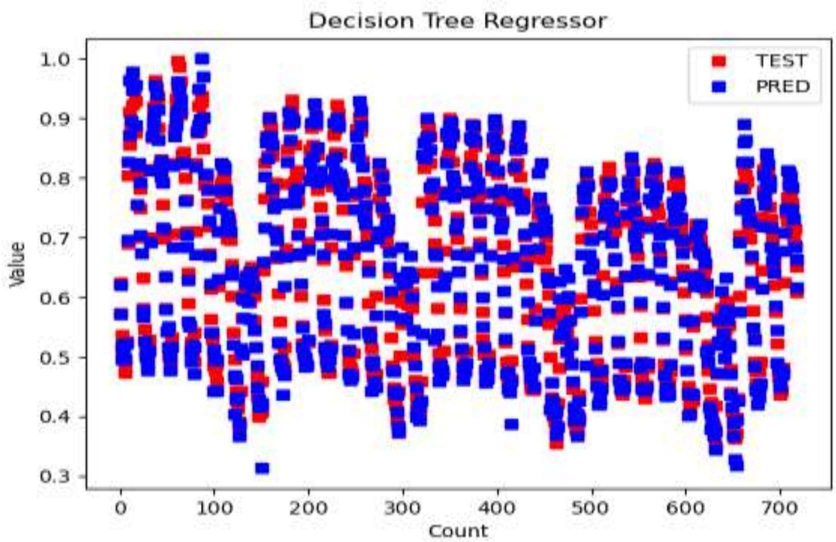

Figure 6.21 DT result without elimination

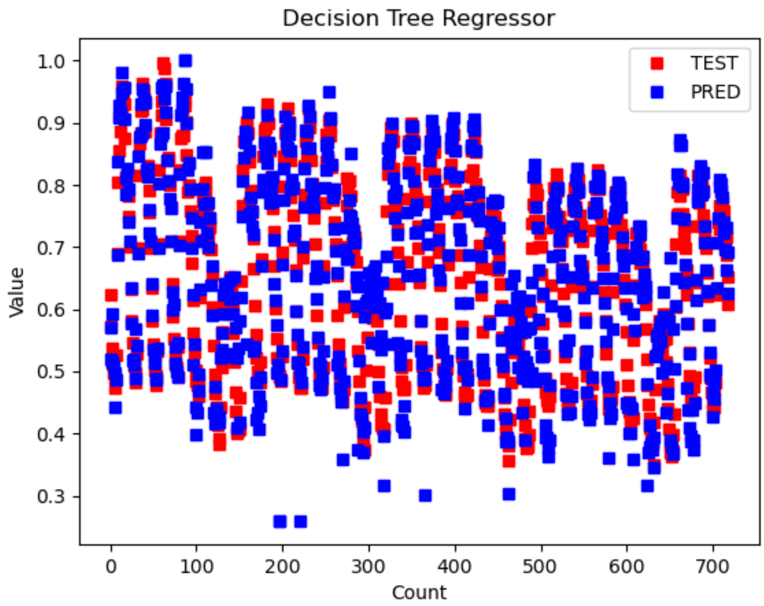

Figure 6.22 DT result with elimination

MSE was found to be 0.022 in the Random Forest regression model. The model is successful in predicting uecm values at the level of 0.3 - 1.0, and is better at extreme values than other models. MSE was found to be 0.034 after screening with correlation in the Random Forest model. The model was successful in predicting uecm values at the level of 0.3 - 1.0, and gave good results at extreme values.

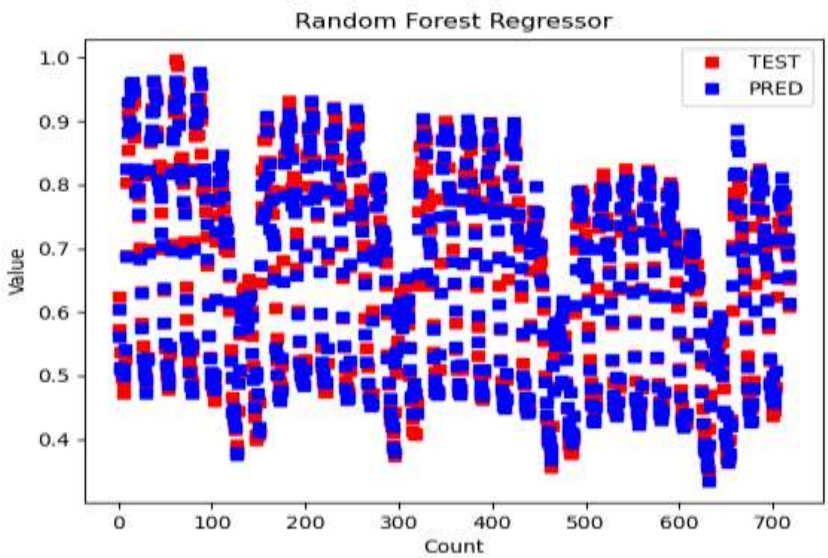

Figure 6.23 RF result without elimination

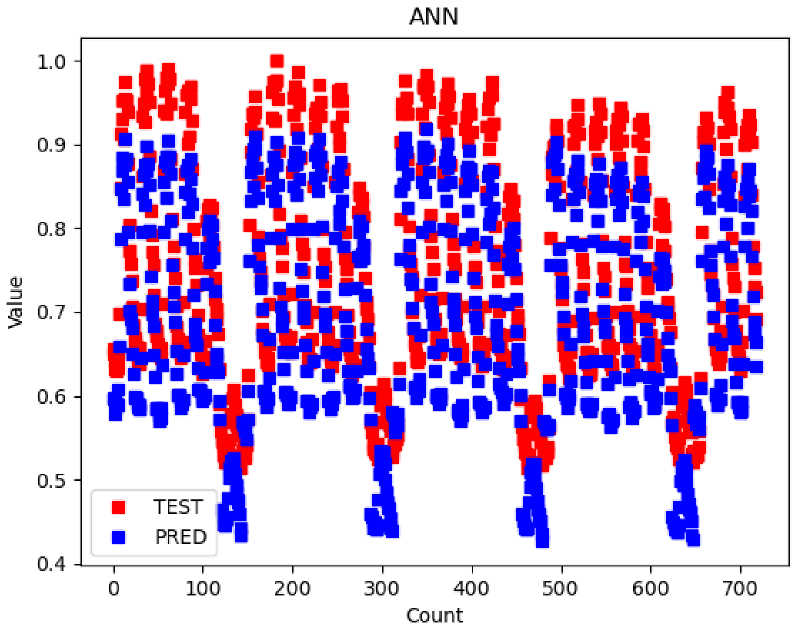

Figure 6.24 RF result with elimination

6.4 ST_UECM Model Results

In this model, it is estimated with 4 models for the last 3 months in the data set. The results were shared before and after the feature was extracted with the correlation. The charts contain values for the period 09.2020. Graphics are shared within all 4 models.

MSE was found to be 0.070 in the ANN model. The model has been advanced 100 steps. There are 24 values whose difference between the estimate and the actual value is greater than 0.10. The ANN Model is successful in predicting st_uecm values at the level of 0.6 - 0.9, and is insufficient for values less than 0.6. The MSE was found to be 0.03 after elimination with correlation in the ANN model. The model has been advanced 100 steps. There are 10 values whose difference between the estimate and the actual value is greater than 0.10. The ANN Model is successful in predicting the st_uecm values at the level of 0.6 - 1.0, and is insufficient at the extreme values.

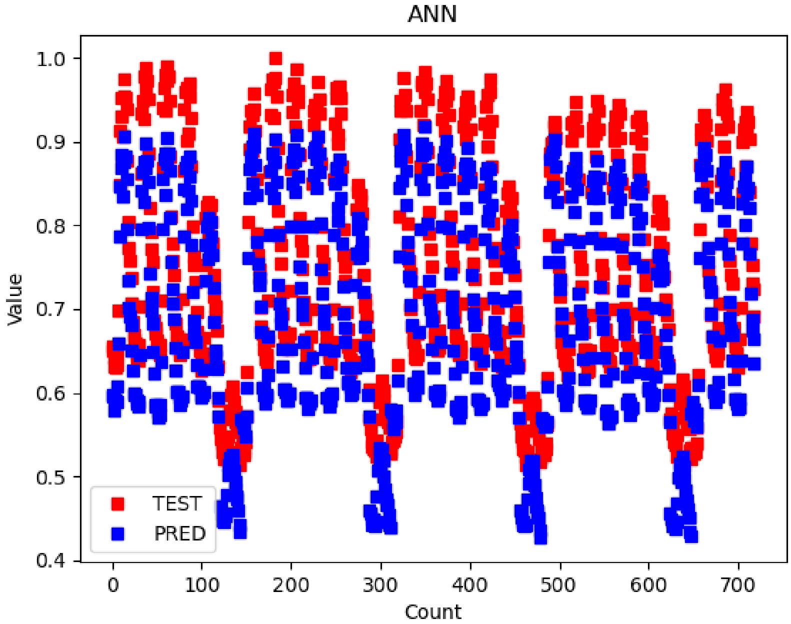

Figure 6.25 ANN result without elimination

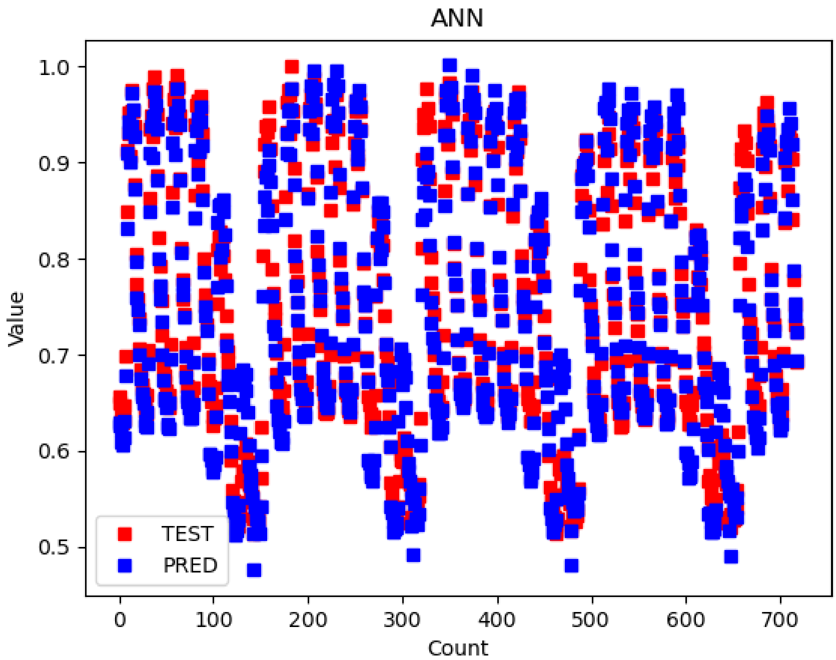

Figure 6.26 ANN result with elimination

The MSE was found to be 0.045 in the linear regression model. The model is successful in predicting 0.6 - 1.0 st_uecm values, but falls short at extreme values. After screening with correlation in the linear regression model, the MSE was found to be 0.053. The model is successful in predicting the st_uecm values at the level of 0.5 - 0.9 and falls short at the extreme values.

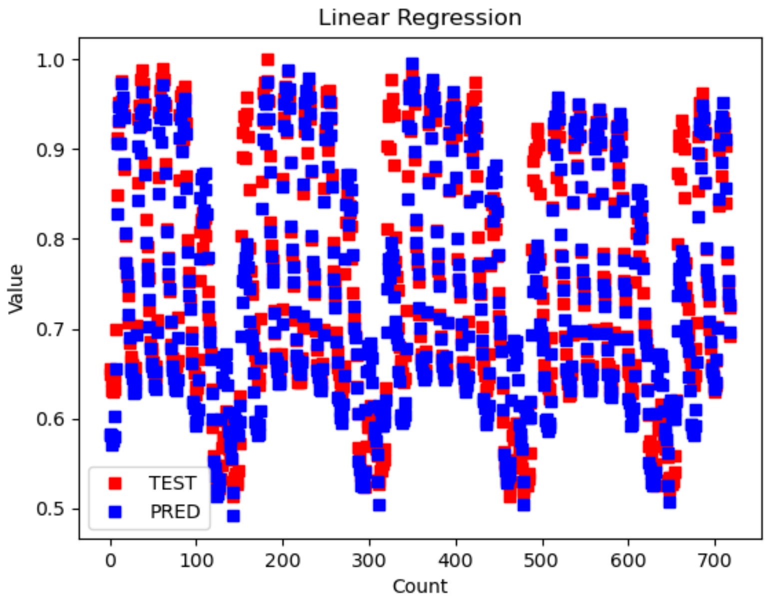

Figure 6.27 ANN result without elimination

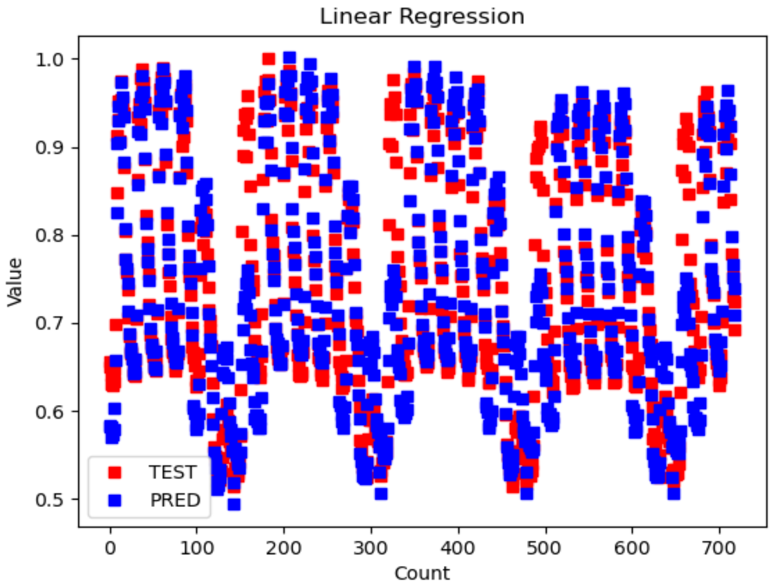

Figure 6.28 ANN result with elimination

In the Decision Tree regression model, MSE was found to be 0.024. The model is successful in predicting 0.6 - 1.0 st_uecm values and is better at extreme values than other models. After screening with correlation in Decision Tree model, MSE was found to be 0.028. The model was successful in predicting st_uecm values between 0.6 and 0.9, and it did well at extreme values.

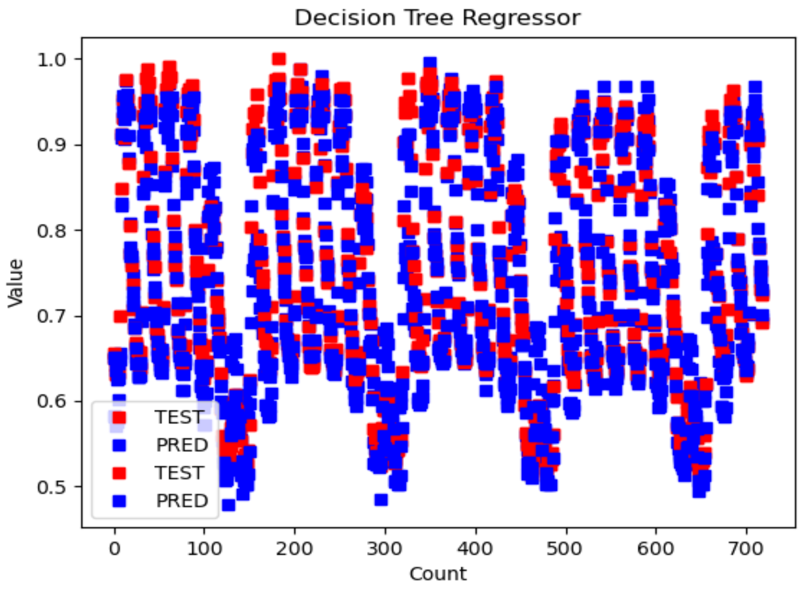

Figure 6.29 DT result without elimination

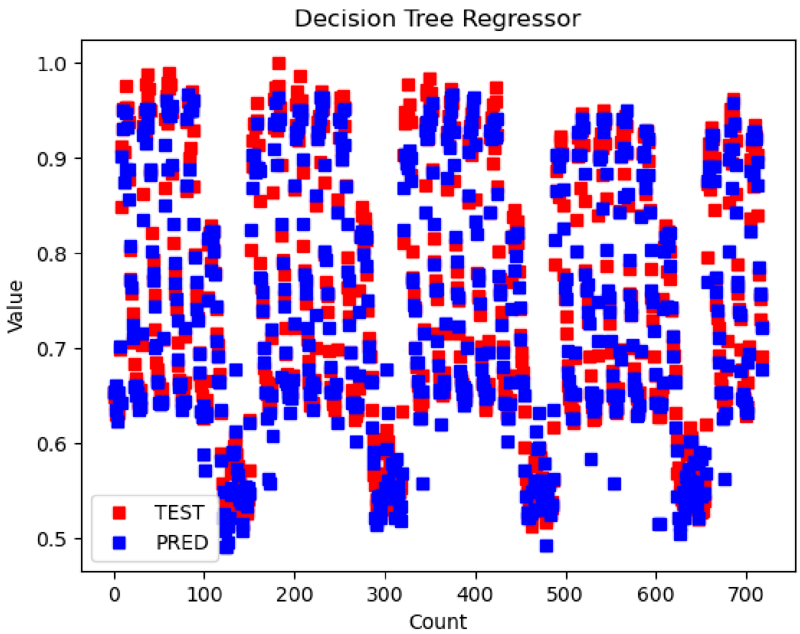

Figure 6.30 DT result with elimination

MSE was found to be 0.021 in the Random Forest regression model. Its model is successful in predicting st_uecm values at the level of 0.6 - 0.9, and is better at extreme values than other models. The MSE was found to be 0.021 after screening with correlation in the Random Forest model. The model was successful in predicting st_uecm values of 0.6 - 1.0 level, and it gave good results at extreme values.

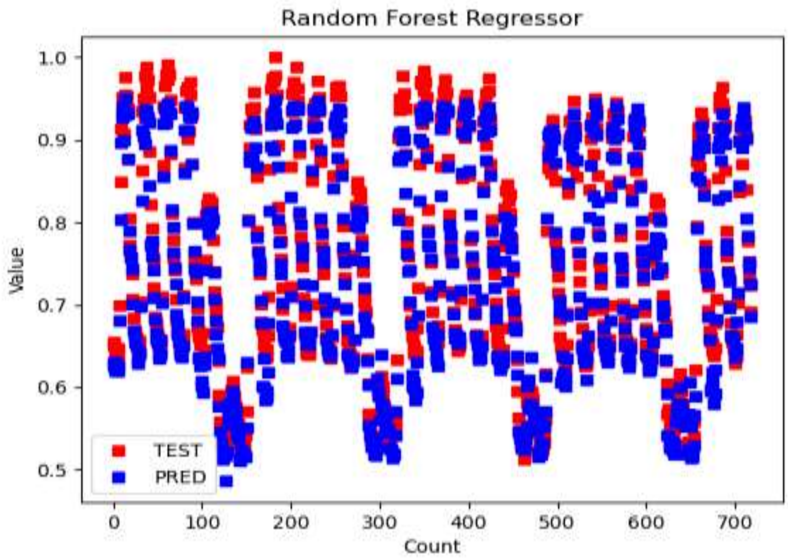

Figure 6.31 RF result without elimination

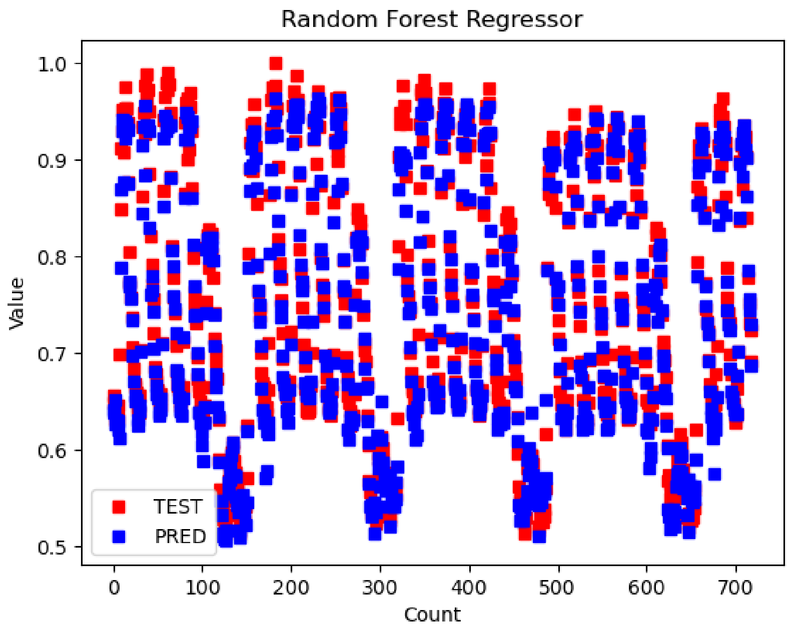

Figure 6.32 RF result with elimination

After collecting the return results from the models, the YEKDEM unit price formula was applied and the final values were found. Average values for the last 3 months are given in the table below. ANN has produced very successful results except for November 2020. In addition, machine learning models have also produced successful results.

Table 6.1 Model result

Type	***Model Result without Feature Extraction***			***Model Result with Feature Extraction***		
Date	01.10.2020	01.09.2020	01.08.2020	01.10.2020	01.09.2020	01.08.2020
True Valeu	92.66	94.01	100.97	92.66	94.01	100.97
Pred_ANN	88.73	92	101.98	88.44	93.30	100.38
MPE_ANN	4.23	1.22	-1.00	4.54	0.74	0.58
MSE_ANN	15.42	1.32	1.02	17.74	0.49	0.35
Pred_LR	88.94	93.50	101.22	84.03	93.77	103.65
MPE_LR	4.01	0.54	-0.24	9.30	0.25	-2.64
MSE_LR	13.84	0.25	0.05	74.40	0.05	7.15
Pred_DT	84.12	92.58	102.31	84.99	92.63	102.18
MPE_DT	9.21	1.51	-1.32	8.27	1.46	-1.19
MSE_DT	72.93	2.03	1.78	58.76	1.89	1.46
Pred_RF	83.76	92.29	102.26	89.84	92.31	101.91
MPE_RF	9.60	1.82	-1.27	9.51	1.80	-0.92
MSE_RF	79.15	2.93	1.65	77.72	2.88	0.87

CHAPTER SEVEN
CONCLUSIONS AND FUTURE WORKS

Electric charge estimation is a subject that has been studied for a long time. While statistical methods were mostly used in this field in the early days, with the development of artificial intelligence, the use of artificial intelligence systems has gained importance. Before making an electrical load estimation, the analysis of historical load data should be done well. Excessive historical load data will help to create a good prediction model.

Today, hybrid methods are also used in electrical load estimation. Hybrid methods consisting of a combination of various artificial intelligence applications or both artificial intelligence and statistical methods are used. Successful results are obtained with the opinions of experts.

Estimates of Yekdem Mechanism can be made by statistical methods. It is very successful in price estimation with machine learning and artificial neural networks. For this purpose, YEKDEM unit price tried to estimating using artificial neural networks and machine learning method.

In this study, YEKDEM unit price estimates were made using artificial neural Networks, Linear Regression, Decision Tree and random forest methods by using electricity consumption and production values, dollar rate, ptf, temperature, pressure, humidity data between 01.01.2018 - 31.10.2020. Consumption and production values, dollar exchange rate, ptf, temperature, pressure, humidity, holidays are used as input parameters. Using these data, YEKDEM income, YEKDEM price, consumption and consumption data of eligible consumers were estimated. Temperature values were found with the average values of 80 provinces. The data were taken from the website by parallel programming. Test data of the month to be estimated, all remaining values are reserved for training.

MPE and mse are used as accuracy criteria. Artificial neural networks and machine learning methods have produced values close to each other. Model estimates were analyzed by qualifying correlation and without any screening. After the feature extraction, more meaningful high performance models with less data were established. Checks have been made over the last 3 months in the data set and successful results have been obtained.

Despite the preporcess and feature extraction in the study, the unit price of YEKDEM for November 2020 is weak compared to other results. Accuracy can be done in various processes to increase performance. The parameters that will affect the electricity consumption and production can be increased. As an example, the installed power variable can be added to the data set. However, it should not be forgotten that while performing such applications, while error rates may decrease, the complexity of the system increases and the modeling of the system becomes difficult.

REFERENCES

Aslan, Y., Yaşar, C.& Nalbant, A. (2006). Kütahya ilinin yapay sinir ağları kullanılarak elektrik puant yük tahmini . *Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, (11), 63-74

Bayrak, Y . (2020). Yenilenebilir enerji kaynaklarından elektrik üretimini destekleme mekanizması. *Türkiye'nin Enerji Görünümü-2020*, (717), 347-364 .

Başoğlu, B. & Bulut, M. (2016). Kısa dönem elektrik talep tahminleri için yapay sinir ağları ve uzman sistemler tabanlı hibrit sistem geliştirilmesi. *Journal of the Faculty of Engineering and Architecture of Gazi University*, *32*(2), 575-583.

Bicil, İ. (2015). *Elektrik piyasasında fiyatlandırma ve türkiye elektrik piyasası fiyat tahmini.* Ph.D Thesis, Balıkesir Üniversitesi, Balıkesir.

Biçer, A. (2018). *Enerji talep tahminine yönelik program geliştirme ve bir bölge için uygulaması.* M.Sc. Thesis, Afyon Kocatepe Üniversitesi, Afyon.

Boltürk, E. (2013). *Elektrik talebi tahmininde kullanılan yöntemlerin karşılaştırılması.* M.Sc. Thesis, İstanbul Teknik Üniversitesi, İstanbul.

Cantürk, S. (2018). *Bir rüzgâr çiftliğinden yapay sinir ağlarıyla kısa süreli elektrik üretim tahmini.* M.Sc. Thesis, Hacettepe Üniversitesi, Ankara.

Çam, S. & Kılı, S. (2018). Altın fiyatı günlük getirilerinin yapay sinir ağları algoritması ve markov zincirleri modelleri ile tahmini. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (18), 681-694.

Çarkacı, N. (2017). *Derin öğrenme uygulamalarında temel kavramlar*. Retrieved November 1, 2017, from https://www.linkedin.com/pulse/derin-ogrenme uygulamalarinda-temel-kavramlar-skor-ve carkaci.

Çayır, A. , Yenidoğan, I. & Dağ, H. (2018). Konutların günlük elektrik güç tüketimi tahmini için uygun model seçimi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *30*(3), 15-21.

Çayıroğlu , İ. (2019). *İleri algoritma analizi*. Retrieved January 1, 2017, from http://www.ibrahimcayiroglu.com/dokumanlar/ilerialgoritmaanalizi/ilerialgoritma analizi-5.hafta yapaysiniraglari.pdf

Çevik, H. (2013). *Türkiye'nin kısa dönem elektrik yük tahmini*. M.Sc. Thesis, Selçuk Üniversitesi, Konya.

Dalgın, A. (2017). *Türkiye gün öncesi piyasası elektrik fiyat tahmini*. M.Sc. Thesis, İstanbul Teknik Üniversitesi, İstanbul.

Es, A. , Kalender, F. & Hamzaçebi C. (2014). Yapay sinir ağları ile türkiye net enerji talep tahmini. *Gazi Üniversitesi Mimarlık Mühendislik Fakültesi Dergisi, 29*(3), 495-504.

Ekinci, F. (2019). Ysa ve anfis tekniklerine dayalı enerji tüketim tahmin yöntemlerinin karşılaştırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi,* (7), 1029-1044.

Emel, G. & Taşkın, Ç. (2005). Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi,* (6), 222-239.

Gök, A., Yıldız, C. & Şekkeli, M. (2019). Yapay sinir ağları kullanarak kısa dönem güneş enerjisi santrali üretim tahmini: Kahramanmaraş örnek çalışması. *Uluslararası Doğu Anadolu Fen Mühendislik ve Tasarım Dergisi*, *1*(2), 186-195.

Gümüş, V., Şimşek, O., Soyda, N., Aköz, M. & Yenigün, K. (2016). Adana istasyonunda buharlaşmanın farklı yapay zekâ yöntemleri ile tahmini. *VIII. Ulusal Hidroloji Kongresi – 2015, 7*(2), 310-318.

Hamzaçebi, C. & Kutay, F. (2004). Yapay sinir ağları ile türkiye elektrik enerjisi tüketimini 2010 yılına kadar tahmini. *Gazi Üniversitesi Mimarlık Mühendislik Fakültesi Dergisi, 19*(3), 227-233.

Karacan, H. & Karacan, C. (2016). Çoklu regresyon metoduyla elektrik tüketim talebini etkileyen faktörlerin incelenmesi. *Selçuk Üniversitesi Mühendislik Bilim ve Teknoloji Dergisi, 4*(3), 182-195.

Kılıç, E., Özbalcı, Ü. & Özçalık, H. (2012). Lineer olmayan dinamik sistemlerin yapay sinir ağları ile modellenmesinde MLP ve RBF yapılarının karşılaştırılması. *ELECO '2012 Elektrik - Elektronik ve Bilgisayar Mühendisliği Sempozyumu,* (1), 694-698.

Kocadayı, Y., Erkaymaz, O. & Uzun, R. (2017). Yapay sinir ağları ile Tr81 bölgesi yıllık elektrik enerjisi tüketiminin tahmini. *Bilge International Journal of Science and Technology Research 2017,* (1), 59-64.

Makas, Y. (2016). Yapay sinir ağlarıyla hidroelektrik enerji üretiminin çok dönemli tahmini. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 21*(3), 757-772.

Öksüz, S. (2019). *Yapay sinir ağları ile denizli ili mesken aboneleri elektrik tüketim tahmini.* M.Sc. Thesis, Pamukkale Üniversitesi, Denizli.

Özden, S. & Öztürk, A. (2018). Yapay sinir ağları ve zaman serileri yöntemi ile bir endüstri alanının (ivedik OSB) elektrik enerjisi ihtiyaç tahmini. *Bilişim Teknolojileri Dergisi, 11*(3), 255-261.

Pençe, İ., Kalkan, A. & Çeşmeli, M. (2019). Türkiye sanayi elektrik enerjisi tüketiminin 2017-2023 dönemi için yapay sinir ağları ile tahmini. *Mehmet Akif Ersoy Uygulamalı Bilimler Dergisi, 3*(2), 206-228.

Pınarbaşı, M. (2009). *Elektrik enerji sistemlerinde talep tahmin yöntemleri ve yapay sinir ağları uygulaması*. M.Sc. Thesis, İstanbul Üniversitesi, İstanbul.

Sinecan, M., Kaya, B. & Yıldız, Ö. (2017). Aydın ilinde insan sağlığını birincil dereceden etkileyen hava değişkenlerine yönelik yapay sinir ağı tabanlı erken uyarı modeli. *Gazi Üniversitesi Fen Bilimleri Dergisi, 5*(4), 121-131.

Şeker, A., Diri, B. & Balık, H. (2017). Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme. *Gazi Üniversitesi Mühendislik Bilimleri Dergisi, 3*(3), 47-64.

Şen, S. & Yıldırım, İ. (2019). *Eğitimde araştırma yöntemleri* (1.Baskı). Ankara: Nobel Akademik Yayıncılık Eğitim ve Danışmanlık Tic. Ltd. Şti.

Şenocak, F. & Kahveci, H. (2018) . *Gün öncesi piyasasında PTF dönemsel ağırlıklı fiyat ortalama tahmini*. Retrieved January 1,2018, from https://www.emo.org.tr/ekler/f7b2fa56e559dc6_ek.pdf

Şenol, Ü. & Musayev, Z. (2017). Rüzgâr enerjisinden elektrik üretiminin yapay sinir ağları ile tahmini. *Bilge Uluslararası Fen ve Teknoloji Araştırmaları Dergisi*, *1*(1), 23-31.

Toker, C. & Korkmaz, O. (2012). *Türkiye kısa süreli elektrik talebinin saatlik olarak tahmin edilmes*. Retrieved January 1, 2012, from https://www.emo.org.tr/ekler/d6d1c493ddc8662_ek.pdf

Ünal, S., Çınarer, G., Karaman, İ. & Yurttakal, A. (2016) . Yapay sinir ağları yönetimi ile burs analizi. *EEB 2016 Elektrik-Elektronik Ve Bilgisayar Sempozyumu 11-13 May 2016*, (1), 132-136.