

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DAĞITIM ŞEBEKESİNDE KISA DÖNEM YÜK TAHMİNİ

YÜKSEK LİSANS TEZİ

Mert SAÇLI

Elektrik Mühendisliği Anabilim Dalı

Elektrik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Belgin EMRE TÜRKAY

MAYIS 2020

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DAĞITIM ŞEBEKESİNDE KISA DÖNEM YÜK TAHMİNİ

YÜKSEK LİSANS TEZİ

**Mert SAÇLI
(504161067)**

Elektrik Mühendisliği Anabilim Dalı

Elektrik Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Belgin EMRE TÜRKAY

MAYIS 2020

İTÜ, Fen Bilimleri Enstitüsü'nün 504161067 numaralı Yüksek Lisans Öğrencisi Mert SAÇLI, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “DAĞITIM ŞEBEKESİNDE KISA DÖNEM YÜK TAHMİNİ” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı: **Prof. Dr. Belgin Emre TÜRKAY**
İstanbul Teknik Üniversitesi

Jüri Üyeleri: **Dr.Öğr. Üyesi Canan KARATEKİN**
İstanbul Teknik Üniversitesi

Doç. Dr. Ozan ERDİNÇ
Yıldız Teknik Üniversitesi

Teslim Tarihi : **Mayıs 2020**
Savunma Tarihi : **Mayıs 2020**





Değerli Ailem ve Beste'ye,



ÖNSÖZ

Lisans ve yüksek lisans dönemimde yönlendirmeleri, destekleri, tecrübeleri ve bu tez çalışmasının planlanması, oluşumu gibi bütün aşamalarında bilgi, birikim, tecrübe ve yorumları ile sürekli katkıda bulunan ve destekleriyle bana yol gösteren değerli hocam Prof.Dr. Belgin EMRE TÜRKAY'a,

Hayattaki en büyük şansım olan maddi, manevi hayatım boyunca desteklerini esirgemeyen iyi günde ve kötü günde yanımda olan, tezin zorlu oluşum sürecinde motivasyonumun her daim yüksek olmasının mimarı babam Orhan SAÇLI, annem Naciye SAÇLI ve kardeşim İnt. Dr. Deniz SAÇLI'ya,

Yük tahmini için gerekli yük verilerini sağlayan Dr. Murat AYHANOĞLU ve meteorolojik verilerin temini konusunda yardımlarını esirgemeyen İTÜ FBE ve MEVBİS çalışanlarına,

ARIMA modellemelerindeki yardımı, desteği, bilgi birikimi ve paylaşımı için Aydın AYDEMİR'e,

Hayatımı anlamlandıran ve hayatıma değer katan hayat arkadaşım Asistan Dr. Beste YÜKSEL SAÇLI'ya sonsuz teşekkür ve şükranlarımı sunarım.

Mayıs 2020

Mert SAÇLI
(Elektrik Mühendisi)



İÇİNDEKİLER

Sayfa

ÖNSÖZ	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
SEMBOLLER	xiii
ÇİZELGE LİSTESİ	xv
ŞEKİL LİSTESİ	xvii
ÖZET	xix
SUMMARY	xxi
1. GİRİŞ	1
1.1 Literatür Araştırması	3
1.2 Tezin Amacı	7
2. YÜK TAHMİNİ	9
2.1 Zamana Göre Yük Tahmini	9
2.1.1 Çok Kısa Dönem Yük Tahmini (ÇKDYT)	10
2.1.2 Kısa Dönem Yük Tahmini (KDYT)	10
2.1.3 Orta Dönem Yük Tahmini (ODYT)	11
2.1.4 Uzun Dönem Yük Tahmini (UDYT)	11
2.2 Yük Tahminine Etki Eden Faktörler	12
2.2.1 Meteorolojik Faktörler	12
2.2.2 Zaman Faktörü	12
2.2.3 Demografik ve Ekonomik Faktörler	13
2.2.4 Önceki Yük Parametresi	13
2.2.5 Rastgele Değişkenler	13
2.3 Yük Tahmininde Kullanılan Yöntemler	13
2.3.1 Analitik Yöntemler	14
2.3.1.1 Benzer Gün Yaklaşımı	14
2.3.1.2 Regresyon Analizi	14
2.3.1.2.1 Basit Regresyon Modeli	15
2.3.1.2.2 Çoklu Regresyon Modeli	16
2.3.1.3 Zaman Serisi Analizi	16
2.3.1.3.1 AR(p) Modeli (Otoregresif Süreci)	17
2.3.1.3.2 MA(q) Modeli (Hareketli Ortalama Süreci)	18
2.3.1.3.3 ARMA (p, q) Modeli (Otoregresif Hareketli Ortalama Süreci)	19
2.3.1.3.4 ARIMA(p,d,q) Modeli (Bütünleşik Otoregresif Hareketli Ortalama Süreci)	19
2.3.2 Yapay Zekâ Yöntemleri	20
2.3.2.1 Yapay Sinir Ağları	21
2.3.2.2 Bulanık Mantık	22
2.3.2.3 Destek Vektör Makineleri	24
2.3.2.3.1 Doğrusal Ayrılabilen Destek Vektör Makineleri	26

2.3.2.3.2 Doğrusal Ayrılamayan Destek Vektör Makineleri.....	28
2.3.2.3.3 Çekirdek (Kernel) Fonksiyonları	29
2.3.2.4 Destek Vektör Regresyonu.....	31
2.3.2.5 KNN (K-Nearest Neighbor) Algoritması.....	33
2.3.2.6 Karar Ağaçları	34
3. YÜK TAHMİNİ UYGULAMASI.....	39
3.1 Veri Seti.....	39
3.2 Tahmin Programları.....	44
3.3 Tahmin Algoritmaları ve Metodolojileri	44
3.3.1 Tahmin Metodolojileri	45
3.3.1.1 Model-1	47
3.3.1.2 Model-2	47
3.3.1.3 Model-3	47
3.3.1.4 Model-4	48
3.3.1.5 Model-5	48
3.3.1.6 Model-6	48
3.3.1.7 Model-7	48
3.3.1.8 Model-8	49
3.3.1.9 Model-9	49
3.3.1.10 Model-10	49
3.3.1.11 Model-11	50
3.3.1.12 Model-12	50
3.3.1.13 Model-13	50
3.3.1.14 Model-14	50
3.3.1.15 Model-15	50
3.3.2 Tahmin Algoritmaları.....	51
3.3.2.1 ARIMA.....	51
3.3.2.2 Destek Vektör Makineleri	51
3.3.2.3 Destek Vektör Regresyonu.....	51
3.3.2.4 Karar Ağaçları	52
3.3.2.5 KNN- En yakın K Komşu	52
4. TAHMİN ÇIKTILARI VE KARŞILAŞTIRMALAR.....	53
5. SONUÇ VE ÖNERİLER.....	55
KAYNAKLAR.....	59
ÖZGEÇMİŞ.....	83

KISALTMALAR

ANFIS	: Adaptif Yapay Sinir Ağılı Bulanık Mantık Çıkarım Sistemi
AR	: Otoregresif Süreç
ARIMA	: Otoregresif Bütünleşik Hareketli Ortalama
ARMA	: Otoregresif Hareketli Ortalama Süreci
ÇEAŞ	: Çukurova Elektrik Anonim Şirketi
ÇKDYT	: Çok Kısa Dönem Yük Tahmini
DD	: Dalgacık Dönüşümü
DVM	: Destek Vektör Makineleri
DVR	: Destek Vektör Regresyonu
EKK	: En Küçük Kareler
EYS	: Enerji Yönetim Sistemi
FIS	: Fuzzy Inference System
GKA	: Görgül Kip Ayrışımı
GSMH	: Gayrisafi Milli Hasıla
KDYT	: Kısa Dönem Yük Tahmini
KKT	: Karush-Khun-Tucker
KNN	: En yakın K Komşu
MA	: Hareketli Ortalamalar
MAPE	: Mean Absolute Percentage Error
MEVBİS	: Meteorolojik Veri Bilgi Sunum ve Satış Sistemi
MGM	: Meteoroloji Genel Müdürlüğü
ODYT	: Orta Dönem Yük Tahmini
RMSE	: Root Mean Square Error
RTFSA	: Radyal Tabanlı Fonksiyon Sinir Ağları
UDYT	: Uzun Dönem Yük Tahmini
YSA	: Yapay Sinir Ağları



SEMBOLLER

Y	: Bağımlı değişken veya çıkış parametresi
α	: Sabit değer
X	: Bağımsız değişken
ε	: Hatayı ifade eden terim
β	: Regresyon katsayısı
Y_t	: Tahmin edilecek yük
$Y_{p-1}, \dots, Y_{p-i}$	: Y_t ifadesinin t-n zamanda aldığı değerler
$\phi_i, \phi_i, \dots, \phi_p$	: Otoregresif parametreler
c	: Kesme terimi
ε_t	: Hata veya kalıntı değeri
θ_i	: i. mertebeden hatanın ağırlığı
ε_{p-i}	: i gecikmeli hata parametresi
p	: Otoregresif bölüm
q	: Hareketli ortalamalar
d	: Durağanlaştırma
w	: Ağırlık vektörü
b	: Eşik değeri
$a_1, a_2, a_3 \dots a_n$	: Lagrange çarpanları
ε	: Aylak değişken
ξ	: Duyarlı alan dışında kalan hata toleransı
$\emptyset(x)$	: Haritalama fonksiyonu
V	: Veri seti
E	: Entropi



ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 1.1: Literatür özet tablosu.....	7
Çizelge 3.1: Tahmin modelleri yapısı.....	47
Çizelge 4.1: Tahmin sonuçları karşılaştırılması.	53





ŞEKİL LİSTESİ

Sayfa

Şekil 1.1: Elektrik enerji sistem planlaması organizasyon şeması	1
Şekil 2.1: İnsan sinir hücre yapısı	22
Şekil 2.2: Üç girişli çok katmanlı ileri beslemeli ağ yapısı	22
Şekil 2.3: Klasik ve bulanık mantık küme sınıflandırılma farkları	23
Şekil 2.4: Destek vektör makineleri sınıflandırma örneği	25
Şekil 2.5: Destek vektörleri ile sınıflandırma örneği.....	25
Şekil 2.6: Doğrusal ayrılabilen DVM	26
Şekil 2.7: Çekirdek fonksiyonu kullanılan bir destek vektör makinesi	30
Şekil 2.8: Sıklıkla kullanılan çekirdek fonksiyonları	30
Şekil 2.9: Destek vektör regresyonu örneği	32
Şekil 2.10: Karar ağacı mekanizması genel şeması.....	35
Şekil 2.11: Karar ağacı sınıflandırma örneği	36
Şekil 3.1 : 2016 yılı yük tüketim verisi (kWh)	39
Şekil 3.2 : 2017 yılı yük tüketim verisi (kWh)	40
Şekil 3.3 : 2016-2017 yılları yük tüketim verisi (kWh).....	40
Şekil 3.4 : 2016 yılı nispi nem verisi (%)	41
Şekil 3.5 : 2017 yılı nispi nem verisi (%)	41
Şekil 3.6 : 2016-2017 yılları nispi nem verisi (%).....	42
Şekil 3.7 : 2017 yılı sıcaklık verisi (°C).....	42
Şekil 3.8 : 2017 yılı sıcaklık verisi (°C).....	43
Şekil 3.9 : 2017 yılı sıcaklık verisi (°C).....	43
Şekil A.1 : Model-1 tahmin sonuçlarının karşılaştırılması.....	67
Şekil A.2 : Model-2 tahmin sonuçlarının karşılaştırılması.....	68
Şekil A.3 : Model-3 tahmin sonuçlarının karşılaştırılması.....	69
Şekil A.4 : Model-4 tahmin sonuçlarının karşılaştırılması.....	70
Şekil A.5 : Model-5 tahmin sonuçlarının karşılaştırılması.....	71
Şekil A.6 : Model-6 tahmin sonuçlarının karşılaştırılması.....	72
Şekil A.7 : Model-7 tahmin sonuçlarının karşılaştırılması.....	73
Şekil A.8 : Model-8 tahmin sonuçlarının karşılaştırılması.....	74
Şekil A.9 : Model-9 tahmin sonuçlarının karşılaştırılması.....	75
Şekil A.10 : Model-10 tahmin sonuçlarının karşılaştırılması.....	76
Şekil A.11 : Model-11 tahmin sonuçlarının karşılaştırılması.....	77
Şekil A.12 : Model-12 tahmin sonuçlarının karşılaştırılması.....	78
Şekil A.13 : Model-13 tahmin sonuçlarının karşılaştırılması.....	79
Şekil A.14: Model-14 tahmin sonuçlarının karşılaştırılması.....	80
Şekil A.15 : Model-15 tahmin sonuçlarının karşılaştırılması.....	81



DAĞITIM ŞEBEKESİNDE KISA DÖNEM YÜK TAHMİNİ

ÖZET

Günümüzde enerji ihtiyacı teknolojik gelişmeler ile birlikte sürekli olarak artmaktadır. Artan enerji ihtiyacının güvenilir, ekonomik, kaliteli ve sürdürülebilir bir şekilde sağlanabilmesi konusu gün geçtikçe önem kazanmaktadır. Enerjinin ekonomik, kaliteli ve sürdürülebilir bir şekilde sağlanabilmesi enerji sistem planlamasının optimizasyonu ve başarısı ile doğrudan ilişkilidir. Yük tahmini enerji sistem planlamasının ilk ve en önemli aşamalarından biridir. Elektrik yük tahmini geçmiş ve bugünkü elektrik yüklerini analiz ederek, yük tahminine etki eden faktörlerin belirlenmesi ve bunların analizi sonucunda farklı yöntemler, algoritmalar kullanılarak gelecekte olabilecek yükü belirlemek olarak tanımlanabilir.

Yük tahmini farklı amaçlar doğrultusunda zaman kıstasına göre çok kısa dönem, kısa dönem, orta dönem ve uzun dönem olarak sınıflandırılmaktadır. Bir yıldan uzun yük tahminleri uzun dönem, bir haftadan bir yıla kadar olan yük tahminleri orta dönem, bir saatten bir haftaya kadar olan zamanları kapsayan tahminler kısa dönem ve bir saatten daha kısa yük tahminleri çok kısa dönem yük tahmini olarak sınıflandırılmaktadır. Elektrik yük tahmini yapmak için kullanılan giriş parametreleri değişkenlik göstermektedir. Uzun dönem yük tahmininde gayrisafi milli hasıla, nüfus gibi parametreler ağırlıklı olarak kullanılırken kısa dönem yük tahminlerinde meteorolojik veriler ağırlıklı olarak kullanılmaktadır.

Elektrik tüketimi gün içerisinde saatlik bazda değişkenlik göstermektedir. Gün içerisinde değişen tüketim miktarını, devreye girecek ve devreden çıkacak enerji santrallerini belirlemek kısa dönem yük tahmini ile başlar. Kısa dönem yük tahmininin bir başka fonksiyonu ise enerjinin ekonomik olmasını sağlamaktır. Gün geçtikçe artan enerji ihtiyacını karşılamak ve enerjinin sürdürülebilirliğini sağlamak yük tahmini ile başlar. Bu sebeple yük tahmini uygulamaları gün geçtikçe önem kazanmaktadır.

Literatürde, yük tahmini için şu ana kadar kullanılan birçok farklı algoritma ve metodoloji bulunmaktadır. Bu çalışmada literatürde kullanılan analitik yöntemler ve yapay zekâ yöntemleriyle Marmara bölgesinde bulunan bir ile ait 2016-2017 yılları arası veriler kullanılarak saatlik kısa dönem yük tahmini yapılmıştır. Çalışmada kullanılan modeller de temel olarak hafta günleri 3 farklı yöntem kullanılarak ayrılmıştır. Tahmin çalışmalarında kullanılan bağımsız değişken ve bağımlı değişkenler arasında pozitif ve/veya negatif ilişki ne kadar yüksekse model o kadar yüksek tahmin sonuçları vermiştir. Bu sebeple literatürde bazı çalışmalar hafta günleri arasında herhangi bir ayırım yapmadan, bazı çalışmalar hafta içi ve hafta sonu olarak ayırım yapıp iki farklı eğitim ve test verisi kullanılarak, bazı modeller ise hafta içi ve hafta sonu modellerini ayrı eğitim ve test verisi olarak kullanmamak için giriş parametrelerine hafta içi ve hafta sonu ayırımı yapmak için kukla değişken olarak isimlendirilen bağımsız değişken kullanılarak tahmin işlemi gerçekleştirilmiştir. Bu çalışmada gün ayırımlarının tahmin sonucuna etkisini gözlemlemek amacıyla 3 farklı gün ayırımı kullanılmıştır, kullanılan gün ayırımları hafta içi ve hafta sonu için gün

ayrımı yapılmadan, gün ayrımı yapılarak ve hafta içi ve hafta sonu için kukla değişken kullanılarak yapılan tahmin olarak en genel haliyle ifade edilebilir. Bu çalışmada gün modelleri 5 farklı bağımsız değişken seti kullanılarak tahmin işlemi gerçekleştirilmiştir. Kullanılan bütün metodolojilerde kullanılan yük, sıcaklık veri seti gibi veriler aynı veri seti kullanılarak yapılmıştır. Kullanılan farklı metodolojilerin amacı gün ayrımı yapılmasının ve yapılan ayırmda kullanılan giriş parametrelerinin kullanılan yöntemlerle birlikte tahmin doğruluğunun üzerindeki etkisini araştırmaktır. Veri setinde görülebileceği üzere yük ve meteorolojik veriler mevsimsellik gösterdiği için mevsim parametresi kış, ilkbahar, yaz ve sonbahar sırasıyla (0, 1, 2, 3) dönüştürülerek tüm modeller için bağımsız değişken olarak kullanılmıştır. Bu çalışmada kısa dönem yük tahmini saatlik veya bir diğer ifadeyle bir saat aralığı yapıldığı için kullanılan algoritmaların veri seti saatlik olarak tüm modellere tanıtılmıştır. Modellerde kullanılan meteorolojik verilerden nem ve sıcaklık verileri 1 saat önceki nem ve sıcaklık verileridir. Tahmin yöntemleri olarak ARIMA (Oto-regresif bütünleşik hareketli ortalama), destek vektör makineleri, destek vektör regresyonu, KNN (En yakın K Komşu) ve karar ağacı regresyonu algoritmaları kullanılmıştır. Tüm metodolojiler için giriş parametresi olarak bir önceki yük, sıcaklık, nem, mevsim ve saat verisi kullanılmıştır. Çıkışta ise yük tahmin edilmiştir. Verilerin %75'i eğitim, %25'i test için kullanılmıştır. Kullanılan algoritmaların sonucu birbirine ve literatüre yakın çıkmıştır. En iyi tahmin sonucu olarak MAPE (mean absolute percentage error) değeri DVR algoritmasında %3,57 bulunmuştur.

Tezin ilk kısmında literatür çalışmaları incelenmiş, karşılaştırılmış ve tezin yazılma amacı ortaya konmuştur. İkinci kısmında yük tahmini hakkında sınıflandırılmalar, yük tahminine etki eden faktörler, yük tahmininde kullanılan analitik ve yapay zekâ yöntemleri incelenmiştir. Tezin üçüncü kısmında bu çalışmada yapılan yük tahmini uygulaması, modeller ve kullanılan algoritmalar ortaya konmuştur. Tezin dördüncü aşamasında kurulan modeller ve algoritmalar ile sonuçlar karşılaştırılmıştır. Tezin beşinci kısmında sonuçlar detaylı olarak anlatılmış ve çıkarımlar paylaşılmıştır. Tezin Ek-A bölümünde ise model çıktıları grafiksel olarak kıyaslanmıştır.

SHORT-TERM LOAD FORECASTING IN DISTRIBUTION NETWORK

SUMMARY

Load estimation or with other words load forecasting is one of the first and most important stages of energy system planning. Electricity load forecasting can be defined as analyzing the electrical consumptions of the past and present, determining the factors affecting the load estimation, and analyzing them, using different methods and algorithms to determine the future load. The ultimate goal of load forecasting is to supply economic, reliable, and quality energy to the consumer. For this reason, electricity load forecasting at production, transmission, and distribution levels was made by dividing them into different periods.

Load forecasting is classified as very short-term, short-term, mid-term, and long-term according to time criteria for different purposes. Load forecasts longer than one year are classified as long-term, load forecasts from one week to one year are medium-term, forecasts covering one hour to one week are short-term, and forecasts of less than an hour are classified as very short-term load estimates. The input parameters used to make electrical load estimates vary. In the long-term load forecasting, parameters such as gross national product and population are predominantly used, while meteorological data are used predominantly in the short-term load forecasting.

Electricity consumption varies on an hourly basis during the day. Determining the amount of consumption that changes during the day, the power plants that will come in and out starts with the short-term load forecast. Another function of short-term load forecasting is to make energy economical. To provide the increasing energy needs day by day and to ensure the sustainability of the energy starts with the load forecasting. For this reason, load forecasting applications are gaining importance day by day.

The main reasons for the load estimation to be divided into different time zones are that the factors affecting different time electricity consumption vary and the methodologies used vary. For this reason, it is necessary to examine the electrical load estimation in terms of time and factors affecting the load estimation. Load estimation is classified as very short-term, short-term, mid-term, and long-term according to time criteria for different purposes. Electricity consumption varies on an hourly basis during the day. To determine the amount of consumption changing during the day, the power plants to be activated and to be deactivated starts with the short-term load estimation. Another function of short-term load estimation is to make energy economical. To supply the increasing energy consumption and ensure the sustainability of energy starts with load estimation. For this reason, load estimation practices are gaining importance day by day.

In the first part of the thesis, literature studies are examined and compared and the purpose of writing the thesis is explained. In the second part, the classification of load estimation, factors affecting load estimation, analytical, and artificial intelligence methods used in load estimation are examined. In the third part of the thesis, the load

estimation application, models, and algorithms used in this study are presented. The models and algorithms results, which were established in the fourth stage of the thesis, were compared. In the fifth part of the thesis, the results are explained in detail and the inferences are shared. In Appendix-A of the thesis, the model outputs are compared graphically.

In the literature, there are many different algorithms and methodologies used so far for load forecasting. In this study, hourly short-term load forecasting was made by using data from 2016-2017 of a district in the Marmara region with the analytical methods and artificial intelligence methods used in the literature. The models used in the study were separated using 3 different methods on weekdays. As mentioned in the previous sections, the higher the positive and / or negative relationship between the independent variable and dependent variables, the higher the predictive results. For this reason, some studies in the literature do not distinguish between the weekdays, some studies distinguish between weekdays and weekends and use two different training and test data, while some models do not use the weekday and weekend models as separate training and test data. To classify between weekdays and weekends, forecasting was performed using an independent variable called a dummy variable. In this study, 3 different day separations have been used, the day separations used can be expressed in the most general form as the prediction made by making the day separation for weekdays and weekends, by making the day separations and using the dummy variable for weekdays and weekends. In this study, forecasting is performed by using day models with 5 different sets of independent variables. Data such as load, temperature data set used in all methodologies used were made using the same data set. The purpose of the different methodologies used is to investigate the effect of day discrimination and the input parameters used in the discrimination on the prediction accuracy together with the methods used. As can be seen in the data set, since the load and meteorological data show seasonality, the season parameter is converted to winter, spring, summer, and autumn respectively (0, 1, 2, 3) and used as an independent variable for all models. In this study, the data set of the algorithms used were introduced to all models on an hourly basis, since the short-term load forecasting was made hourly or in other words, one hour apart. Humidity and temperature data from the meteorological data used in the models are humidity and temperature data 1 hour before. ARIMA (autoregressive integrated moving average), Support Vector Machines, Support Vector Regression, KNN (K Nearest Neighbour), and decision tree were used as forecasting methods. Previous load, temperature, humidity, season, and hour data were used as input parameters for all methodologies. The load was estimated at the exit. 75% of the data was used for education and 25% for testing. The results of the algorithms used are close to each other and the literature. As a result of the best forecasting, MAPE (mean absolute percentage error) value was found 3.57% in the SVR algorithm.

There are some results about output of comparing forecasting datas. The model, which was created by using the load one hour earlier, gave high accuracy results compared to the model that was created by using the previous day. Consumption load one hour earlier affects the next load more. The models in which the previous load and the previous load were used together gave higher accuracy than the models used separately. In this study, it is not sufficient to use only one hour before a load or 24 hours before loading in the region to be predicted for a high accuracy rate. Using moisture value from meteorological data negatively affected the accuracy of the forecast. For this reason, the characteristics of the region should be well known and

determined in this way regarding the selection of meteorological data. The prediction accuracy is the highest Model-5. Model-5 is the algorithm that is predicted using the load only one hour before and one day before, without meteorological data, without any discrimination.

Models created using meteorological data have higher MAPE value, in other words, the accuracy rate is lower. Meteorological data is a parameter that negatively affects the accuracy of the forecast for the region used in the study. The model created using only the previous load parameters is the model with the highest accuracy rate. In this study, which was carried out using a 2-year data set, using two different test and training data on weekdays and weekends increased the accuracy rate for some models and decreased the accuracy of the prediction for some models. Day separation using dummy variable When compared to the day separation using 2 different training and test data, it increased the accuracy rate for some models and decreased the accuracy rate for some models. The exact effect of the day separation was not observed in all models. For this reason, according to the characteristics of the region to be estimated and the algorithm to be used in the prediction, weekdays should be distinguished. Model-15 is the model with the highest accuracy rate after Model-5. Making the day separation using the dummy variable gave more accurate results for the prediction model without using meteorological data.

The ARIMA model, which was established by making weekday and weekend split, gave more slightly accurate results than the model established without the day distinction. ARIMA model has given higher accuracy results than artificial intelligence algorithms in many models. According to the data set to be used from here, it can be concluded that time series analysis methods are superior in some models and algorithms.

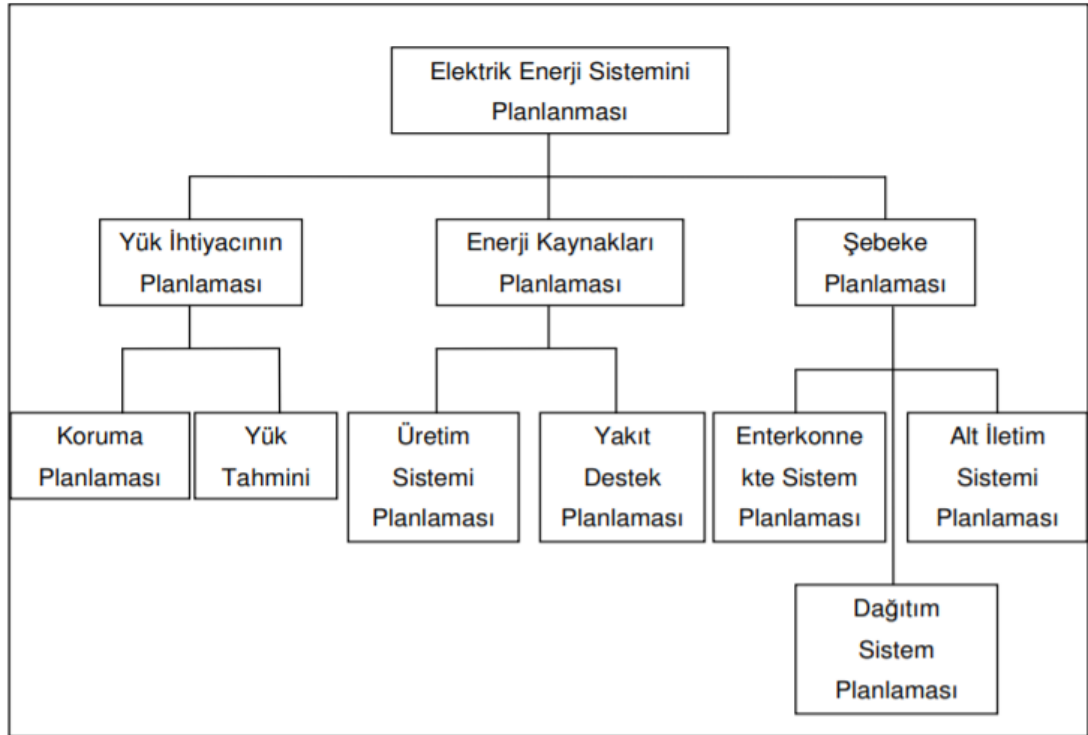
Among the artificial intelligence models, it gave the highest accuracy support vector regression model. The second highest accuracy artificial intelligence algorithm was the decision tree regression algorithm. The accuracy of the nearest neighbor and support vector machines algorithm varies according to the established models. The accuracy rate of the models using regression is higher. For this reason, whether regression should be used in the classification made according to the region to be estimated should be determined according to the characteristics of the region. The parameters affecting the load to be estimated should be decided according to the characteristic of the load. The forecasting method is important in terms of prediction accuracy. For this reason, the forecasting method can be decided as a result of the trend analysis of the load to be estimated.



1. GİRİŞ

Teknolojinin insan hayatına girmesi ve son yıllarda teknolojiye meydana gelen radikal gelişmelerle birlikte teknoloji, insanoğlunun hayatında vazgeçilmez ve önemli bir unsur haline gelmiştir. Teknolojik gelişmelerin reel ve somut karşılığı olan teknolojik ürünlerin tasarımı, üretimi ve kullanımı gibi aşamaların şüphesiz kaynaklarından biri enerjidir. Bu enerjinin kaynağının önemli bir bölümünü elektrik enerjisi oluşturmaktadır.

Elektrik enerjisinin büyük güç ve hacimlerde depolanamaması neticesi elektrik enerjisinin talep karşısında arz edilmesi -üretilmesi- sonucunu doğurmuştur. Günümüz ekonomi koşullarında kısıtlı enerji üretim kaynaklarıyla, elektrik enerjisinin güvenilir, ekonomik, kaliteli ve kesintisiz karşılanabilmesi için elektrik enerji sistem planlaması gün geçtikçe önem kazanmaktadır [1]. Şekil 1.1’de görülebileceği üzere elektrik enerji sistem planlaması genel olarak üç alt kırılımda incelenebilir.



Şekil 1.1: Elektrik enerji sistem planlaması organizasyon şeması [1]

Organizasyon şemasından da açıkça görülebileceği üzere yük tahmini plan aşamalarının başında gelen önemli bir aşamadır. Yük tahmini planlaması diğer bütün planlamaları etkilediği ve burada yapılabilecek bir hata diğer planlamalarda da hataya sebebiyet verebileceği için yük tahmin planlamasının doğruluğu hayati önem taşımaktadır. Yük tahmininin gerçekleşen reel yükten düşük tahmin edilmesi durumunda sistem çöküşünü engellemek amacıyla elektrik kesintilerine ve kalitesiz enerji sağlanmasına, gerçekleşen reel yükten fazla tahmin edilmesi ise optimize kapasitede çalışmayan enerji üretim santrallerine ve bunun getirdiği ekonomik olmayan elektrik teminine neden olacaktır [2]. Bu sebeple elektrik yük tahmininin gerçekleşen reel yüke en yakın oranda tahmin edilmesi gerekmektedir.

Yük tahmini elektrik enerji sistem planlamasında yük ihtiyacının belirlenmesi, alt kırılımın vazgeçilmez unsurlarından biridir. Zaman içerisinde yük ihtiyacının belirlenmesi için yük tahmini farklı zaman dilimlerine ve bölümlerine ayrılmıştır. Farklı zaman dilimleri için farklı amaçlar maksadıyla yapılan yük tahminlerinde tahmin edilecek yüke etki eden faktörler, diğer bir ifadeyle bağımlı değişkenler ve bağımsız değişkenler farklılık göstereceğinden farklı metodolojilerle yapılmıştır. Literatürde elektrik yük tahmini teknolojinin de tahminlerde sıklıkla kullanılmaya başlanmasından sonra 4 alt bölümde yapılmaktadır [3]. Bir saatten daha kısa süren zaman dilimlerini kapsayan tahminlere çok kısa dönem yük tahmini (ÇKDYT), bir saatten bir aya kadar olan zaman dilimlerini kapsayan tahminlere kısa dönem yük tahmini (KDYT), bir aydan bir yıla kadar olan zaman dilimlerini kapsayan yük tahminlerine orta dönem yük tahmini (ODYT) ve bir yıldan başlayan dönemleri kapsayan tahminlere ise uzun dönem yük tahmini (UDYT) adı verilir [4]. Zaman aralıklarına göre sınıflandırılan yük tahminlerinin amaçları birbirinden farklılık göstermektedir.

Yük tahmininin doğruluğu kullanılan yöntem ve yükü etkileyen faktörlere bağlıdır. Yüke ilişkin bilgiler, yükün özelliği yöntemin ve değişkenlerin belirlenmesinde tahmini yapan kişinin yargısını belirler. Yükü etkileyen faktörler; nüfus artışı, gayri safi milli hasıla (GSMH), ekonomik veriler, coğrafik faktörler (sıcaklık, nem, yağış, ...), endüstriyel planlamalar ve şehir planları, sosyokültürel faktörler, insanların hayat tarzlarındaki ve kullanım alışkanlıklarındaki değişiklikler, teknolojik gelişmeler olarak sıralanabilir [5]. Yük tahmininde kullanılacak olan değişkenlerin belirlenmesi,

öncelikle bu verilerin elde edilip edilememesine, daha sonra tahminin özelliğine, yöntem ve tüketici grubuna bağlıdır [6].

Yük tahmininin doğruluğunu etkileyen faktörlerden biri de kullanılan parametrelere ek olarak yük tahmin etmek için kullanılan metodolojilerdir. Yük tahmininde kullanılan yöntemler tahmin edilecek yükün zaman aralığı ve yükün özelliğine göre değişkenlik göstermektedir [7].

Yüke etki eden faktörler ve tahmin algoritmasının belirlenmesi için tahmin edilecek bölgenin özellikleri iyi analiz edilmeli ve analiz sonrasında doğruluk oranı en yüksek sonuca göre belirlenmelidir.

1.1 Literatür Araştırması

Geleneksel yöntemlerin dışında istatistiki yöntemler kullanılarak yük tahmini çalışmasının 1962 yılından beri yapıldığı bilinmektedir [8]. New çalışmasında zaman serisi analizini kullanarak uzun dönem yük tahmini yapmıştır.

G. Gross ve F.D. Galiana 1987 yılında yapmış olduğu “Short-term load forecasting” başlıklı makalesinde kısa dönem tahmini enerji yönetim sisteminin (EYS) çevrimiçi programlaması ve güvenlik fonksiyonlarındaki önemli rolünü gözden geçirip, daha sonra yükün doğasını ve davranışını etkileyen farklı faktörleri tartışmıştır [9].

Papalexopoulos ve Hesterberg tarafından Pasifik Gaz ve Elektrik yük verilerini kullanarak yapılan çalışmada klasik modellere kıyasla önerilen hibrit model kıyaslanmıştır. Oluşturulan lineer regresyon tabanlı hibrit modelin daha tutarlı tahmin sonuçları verdiği ortaya konmuştur [10].

Bakırtzıs ve arkadaşları tarafından Yunanistan enerji sisteminin günlük yük tahminini yapmak amacıyla 63 girişli, 24 çıkışlı çok katmanlı yapay sinir ağı (YSA) modeli uygulanmıştır. Giriş parametreleri olarak önceki yük verileri ile sıcaklık verileri bağımsız değişken olarak kullanılırken, bağımlı değişken olarak ertesi gün olarak seçilmiştir [11].

LIU tarafından Amerika Birleşik Devletleri yük verileri kullanılarak yapılan çalışmada bulanık mantık, sinir ağları ve otoregresif model kıyaslanmış çok kısa dönem tahmin yapılmıştır. Bulanık mantık ve sinir ağları modelleri kullanılan modelin doğruluk oranı daha yüksek sonuç verdiği ortaya konmuştur [12].

Gürsoy tarafından 2000 yılında yapılan uzun dönem yük tahmin çalışmasında regresyon, zaman serisi ve YSA kullanılarak Antalya Kepez bölgesi ve Çukurova Elektrik A.Ş (ÇEAŞ) bölgesi için 2020 yılına kadar uzun dönem tahminler yapılmıştır [13].

Hippert ve arkadaşları tarafından yapılan çalışmada 1991 ve 1999 yılları arasında yayınlanmış kısa dönem yük tahmini yapılmış ve sinir ağları kullanılmış çalışmaları inceleyenmiş ve kullanılan giriş parametrelerin ve sınıf sayıları farklarının etkisi incelenmiştir [14].

Yalçınöz ve arkadaşları tarafından yapılan çalışmada Niğde bölgesinin elektrik yük tahmini aylık olarak tahmin edilmiştir. Çalışmada YSA ve hareketli ortalamalar (MA) yöntemleri kullanılmıştır. Bazı aylar için hareketli ortalamalar metodu daha anlamlı sonuçlar verirken bazı aylar için YSA daha tutarlı sonuç vermiştir. Genel hata ortalamasına bakıldığı zaman ise YSA algoritmasının hareketli ortalamalar metoduna göre daha üstün olduğu sonucuna varılmıştır [15].

Huang tarafından 2005 yılında yapılan çalışmada ARMAX modeli ile saatlik kısa dönem yük tahmini yapılmıştır. Yapılmış olan çalışma parametre seçimi ve belirlenmesi için parçacık sürü optimizasyonu yöntemi kullanılmıştır [16].

Ceylan tarafından Ankara Gölbaşı bölgesi 2002 ve 2003 yılları verileri kullanılarak ertesi günün yükünü tahmin etmek için yapılan kısa dönem yük tahmini çalışmasında metodoloji olarak ileri beslemeli çok katmanlı perseptron ağı YSA ve regresyon teknikleri kullanılmıştır. Çalışma sonucunda YSA'nın regresyon modeline göre daha iyi ve tutarlı sonuçlar verdiği bulunmuştur [17].

Otoregresif bütünleşik hareketli ortalama (ARIMA) ve genelleştirilmiş regresyon yapay sinir ağı ve destek vektör makineleri ile Tayvan elektrik yükünü tahmin için Pai ve Hong tarafından yapılmış olan çalışmada benzetil tavlama ile destek vektör makineleri metodoloji diğer iki yönteme göre doğruluk oranı daha yüksek sonuçlar vermiştir [18].

Demirel tarafından yapılan çalışmada Türkiye'nin yıllık uzun dönem yük verileri tahmin edilmiştir. Giriş parametreleri olarak Gayri Safi Milli Hasıla, tüketilen enerji, üretilen enerji, nüfus sayısı ve kurulu güç verileri kullanılmıştır. Tahmin yöntemleri olarak ANFIS, ARMA, YSA ve regresyon modelleri kullanılmış ve sonuçlar

karşılaştırılmıştır. En iyi tahmin sonucunu ANFIS yönteminin verdiği sonucuna ulaşılmıştır [19].

Demren tarafından İstanbul Avrupa Yakası için orta dönem günlük yük tahmini yapılmıştır. Eğitim verileri olarak 2006-2009 yılları seçilirken, 2010 yılı Nisan ayının puant yük değerleri tahmin edilmiştir. Bağımlı değişkenler olarak geçmiş yük verileri, ortalama hava sıcaklığı, takvim günleri ve elektrik fiyatı kullanılmış ve bağımsız değişkenin tahmin sonucuna etkisini görmek amacıyla hava sıcaklığı ve geçmiş yük verisi ile farklı modeller kurulmuştur. Tahmin metodolojisi olarak DVM ve YSA kullanılmıştır. Model olarak DVM'nin YSA modeline göre daha üstün olduğu sonucu çıkmıştır. Sıcaklık ve geçmiş yük bağımsız değişken olarak seçilen modelin doğruluk oranının daha yüksek olduğu ortaya çıkmıştır [20].

Esener, Yüksel ve Kurban tarafından 72 saatlik giriş verisi kullanılarak 24 saatlik yük verisi tahmin edilmiştir. YSA, Dalgacık Dönüşümü (DD) ve YSA, Radyal Tabanlı Fonksiyon Sinir Ağları (RTFSA), Görgül Kip Ayrışımı (GKA) olmak üzere dört farklı metodoloji ile tahmin yapılmıştır. Oluşturulan hibrit modellerle Türkiye geneli tahmin yapılmış ve sonuç karşılaştırılmıştır. GKA ve RTFSA modelinin tahmin doğruluğu regüleli modelde diğer modellerle kıyaslandığında daha üstün sonuçlar vermiştir [21].

Çevik tarafından yapılmış çalışmada Türkiye'nin saatlik tüketim verileri kullanılmıştır. 2009-2011 yılları eğitim, 2012 için test verisi kullanılarak saatlik elektrik yük tahmini yapılmıştır. Çalışmada tahmin metotları olarak bulanık mantık ve ANFIS kullanılmıştır. Bağımsız değişken olarak mevsim, elektrik yükü ve sıcaklık farkı seçilmiştir. Çalışma sonucunda ANFIS metodolojisinin MAPE değeri bulanık mantık yöntemine göre daha düşük çıkmıştır [22].

İşyapar tarafından yapılan çalışmada elektrik yük tüketim aboneleri yapay zekâ yöntemlerinden kümeleme algoritması yardımıyla sınıflandırılmıştır. İşyapar'a göre bu çalışmada karar ağacı, hiyerarşik metodlar vb. metotlarla elde edilen abone kümelendirme çalışması yük tahmini çalışmasının ön ayağını oluşturmaktadır [23].

Ahmad ve arkadaşları tarafından yapılmış olan çalışmada yük tahmininde kullanılan ARIMA, YSA ve DVM ve bunların hibrit yapıları incelenmiştir. Kullanılan metotların tüketici grubu farklılıkları ve MAPE ve RMSE (root mean square error) parametreleri üzerinden karşılaştırılmıştır. Sonuç olarak ARIMA ve DVM hibrit modelinin diğer modellere nazaran doğruluk oranı yüksek sonuçlar çıkardığı yorumu yapılmıştır [24].

Tutu tarafından yapılan çalışmada günlük kısa dönem yük tahmini yapılmıştır. Giriş parametresi olarak 2012-2015 arası Türkiye günlük elektrik verileri kullanılırken test için 2016 ilk çeyrek parametresi seçilmiştir. Tahmin metodolojisi olarak En Küçük Kareler (EKK) yöntemi kullanılmıştır. Seçilen modelde giriş parametresi olarak sıcaklık, nem ve bir önceki güne ait elektrik verisinin kullanıldığı modelin başarı oranının en yüksek olduğu görülmüştür [25].

Akman tarafından yapılan çalışmada Orta Anadolu Yüz Tevzi İşletme Müdürlüğüne ait 2009-2015 tarihine kadar gerçekleşen saatlik yük verileri, maksimum ve minimum sıcaklık verileri giriş parametreleri için kullanılmıştır. Çıkış olarak günlük tahmini yapılmış ve 10 farklı gün tipi seçilerek test edilmiştir. Tahmin yöntemi olarak YSA ile GA ve YSA içeren hibrit metodolojileri kullanılmıştır. Oluşturulan hibrit modelin gerçeğe daha yakın sonuçlar verdiği çıkarımında bulunulmuştur [26].

Khan tarafından yapılmış olan çalışmada Faisalabad Elektrik Tedarik Şirketi için kısa dönem yük tahmini gerçekleştirilmiştir. Giriş parametreleri Relief ve Korelasyon tekniği ile belirlenmiştir. Tahmin metodolojisi olarak ileri doğru ilerleyen YSA yöntemi kullanılmıştır. Çalışmanın amacı bağımsız değişkenlerin seçiminin YSA'nın tahmin performansına etkisini incelemektir. Çalışma sonucunda literatürde yer alan iletim seviyesi YSA tahmin sonuçlarına benzer sonuçlar elde edilmiştir [27]

Butekin tarafından Yaşar Üniversitesi kampüs binalarından Y-blok için YSA kullanılarak yük tahmini yapılmıştır. Tahmin sonuçlarının daha anlamlı olması açısından gün tipleri ayrılmış ve yük verisi bir takım istatistiksel yöntemlerle işlenmiştir. Sonuç olarak gün tiplerine ayrılan ve işlenmiş verinin doğruluk oranının daha yüksek olduğu gözlemlenmiştir [28].

Doruk tarafından Sakarya bölgesi ait hane halkı verileri kullanılmıştır. Tahmin yöntemi için Durum-Uzay modellerinin Bayesci bir yaklaşımı olan Dinamik Lineer Model ve ARIMA modeli kıyaslanarak işlemin performansı ortaya konmuştur. Sezonluk faktör, lineer büyüme ve ARMA kullanılarak oluşturulan modelin doğruluk oranının en yüksek olduğu ortaya konulmuştur [29].

Literatür özet tablosu Çizelge 1.1'de görülmektedir.

1.2 Tezin Amacı

Yapılan bu çalışmada Marmara bölgesinde bulunan bir il merkezi dağıtım trafosundan alınan 2016-2017 saatlik hane halkı ortalama elektrik yük tüketim verileri kullanılmıştır. Bu çalışmada verilerin %75'i eğitim için kullanılırken 2017 ikinci yarısına tekabül eden %25 veri test için ayrılmıştır.

Bu yüksek lisans tezinde, tahmin yönteminin ve giriş parametrelerinin tahmin doğruluğu üzerinde etkisini incelemek amacıyla istatistiki yöntem ARIMA, destek vektör makineleri (DVM), destek vektör regresyonu (DVR), en yakın k-komşu (KNN) ve karar ağacı regresyonu algoritmaları sıcaklık, nem, mevsim ve saat verileri modellere dahil edilerek tahmin yapılmıştır.

Çizelge 1.1: Literatür Özet Tablosu

Çalışma	Çalışma Yılı	Tahmin Bölgesi	Tahmin Dönemi	Bağımsız değ. Karşılaştırması	Tahmin Algoritmaları	
					Analitik	Yapay Zeka
[10]	1990	A.B.D	Kısa	X	✓	X
[11]	1996	Yunanistan	Kısa	✓	X	✓
[12]	1996	A.B.D	Çok Kısa	X	✓	✓
[13]	2000	Türkiye	Kısa	X	✓	✓
[14]	2001	Farklı bölgeler	Kısa	✓	X	✓
[15]	2002	Niğde	Orta	X	✓	✓
[16]	2003	Tayvan	Kısa	X	✓	X
[17]	2004	Ankara	Kısa	X	✓	✓
[18]	2005	Tayvan	Kısa	X	X	✓
[19]	2009	Türkiye	Uzun	X	✓	✓
[20]	2011	İstanbul	Kısa	✓	X	✓
[21]	2012	Türkiye	Kısa	X	✓	✓
[22]	2013	Türkiye	Kısa	X	X	✓

Çizelge 1.1 (devam): Literatür Özet Tablosu

Çalışma	Çalışma Yılı	Tahmin Bölgesi	Tahmin Dönemi	Bağımsız değ. Karşılaştırması	Tahmin Algoritmaları	
					Analitik	Yapay Zeka
[24]	2014	Malezya	Uzun	X	X	✓
[25]	2017	Türkiye	Kısa	✓	✓	X
[26]	2018	Gölbaşı	Kısa	X	X	✓
[27]	2018	Faisalabad	Kısa	✓	X	✓
[28]	2019	İzmir	Kısa	✓	X	✓
[29]	2019	Sakarya	Kısa	X	✓	X
Bu Çalışma	2020	Dağıtım Bölgesi	Kısa	✓	✓	✓

2. YÜK TAHMİNİ

Günümüz teknolojileri kullanılarak büyük miktarda elektrik enerjisinin depolanamaz olmasından dolayı, elektrik enerjisinin talep doğrultusunda üretilmesi veya diğer bir ifadeyle arz edilmesi gerekir. Bu sebeple elektrik enerji tüketim değerinin önceden tahmin edilmesi önemli ve gereklidir. Yük tahmini, arz-talep dengesini karşılamak amacıyla üretim, iletim ve dağıtım seviyeleri için farklı zaman dilimleri baz alınarak farklı amaçlar doğrultusunda birtakım analitik ve yapay zekâ teknikleriyle yapılan işlemler olarak tanımlanabilir [29]. Yapılan elektrik yük tahminin doğruluğu, bir elektrik yük hizmeti sağlayan kamu hizmeti kuruluşlarının operasyonel ve yönetsel yüklemesi ve bu sürecin başarısı için büyük öneme sahiptir [11].

Yük tahmininin nihai amacı ekonomik, güvenilir ve kaliteli enerjiyi tüketiciye ulaştırmaktır. Enerjinin tüketiciye ulaştırılmasının sürdürülebilirliği yük tahmininin doğru bir şekilde yapılmasıyla sağlanabilmektedir. Üretim, iletim ve dağıtım seviyelerinin dinamikleri ve süreçleri farklılık göstermektedir. Bu nedenle üretim, iletim ve dağıtım seviyelerinde elektrik yük tahmini farklı zaman dilimlerine bölümlendirilerek yapılmıştır. Yük tahmininin farklı zaman dilimlerine ayrılmasının temel sebepleri farklı zaman elektrik tüketimine etki eden faktörlerin ve kullanılan metodolojilerin değişkenlik göstermesidir.

2.1 Zamana Göre Yük Tahmini

Güç sistemlerinin ekonomik, verimli ve güvenilir planlaması; planlanan güç sistemlerinin iletim ve dağıtım seviyelerinde kaliteli, ekonomik ve kesintisiz şekilde işletilmesi arz ve talep dengesinin sağlanması ile mümkündür. Arz ve talep dengesi ise doğru yük tahmini ile sağlanabilir. Yük tahmininin üretim, iletim ve dağıtım seviyelerinde yapılması ve bu yük tahminine eden etmenler farklılık gösterdiği için yük tahmini zamana göre sınıflandırılmıştır. Literatürde enerji sistem planlamasına göre yük tahmini çok kısa, kısa, orta ve uzun dönem olmak üzere 4 dönemde incelenmektedir:

- Çok kısa dönem yük tahmini

- Kısa dönem yük tahmini
- Orta dönem yük tahmini
- Uzun dönem yük tahmini

2.1.1 Çok Kısa Dönem Yük Tahmini (ÇKDYT)

Literatürde bir saatten daha kısa süreli zaman dilimleri için yapılan tahminler çok kısa süreli yük tahmini olarak isimlendirilmiştir. Gelişen teknoloji ve son dönemlerde yapay zekâ algoritmalarının tahmin teknikleri içerisinde kullanılması ile birlikte çok kısa dönem yük tahmini popülerliğinde artış meydana gelmiştir. Çok kısa dönem yük tahmininde trend analizinin yorumlanması tahmin doğruluğu için önem arz etmektedir.

Elektrik tüketiminde radikal oranda meydana gelebilecek artışlar veya azalışlar enterkonnekte sistemde problemlere neden olabilmektedir. Tüketim trendinde bu değişiklikler genellikle puant zamanlarda meydana gelmektedir [30]. Üretim tüketim dengesinde problemlere yol açılabilen bu trend dalgalanmalarının habercisi olarak çok kısa süreli yük tahmini kullanılmaktadır.

Enterkonnekte sisteme dahil olan yenilenebilir enerji sistemleri teknolojik gelişmeler ile gün geçtikçe artmaktadır. Yenilenebilir enerjiler meteorolojik koşullardan direkt etkilenebildiği için enerji sisteminde ciddi dalgalanmalara sebep olabilmektedir [31]. Çok kısa dönem yük tahmini bu dalgalanmaların etkisini incelemek ve dalgalanmanın enterkonnekte sisteme henüz etki etmeden önüne geçilebilmesi amacıyla 15 dakikalık periyotlarla uygulanabilmektedir [31].

2.1.2 Kısa Dönem Yük Tahmini (KDYT)

Bir saatten bir haftaya olan süreleri kapsayan zaman dilimleri için yapılan tahminler kısa dönem yük tahmini (KDYT) olarak sınıflandırılmaktadır.

Kısa dönem yük tahmini enerji sistemi için ekonomik, güvenilir ve güvenli işletme stratejilerinin oluşturulmasında kilit bir rol oynar. Enerji arzı talep doğrultusunda yapılmaktadır. Elektrik tüketimi gün içerisinde gün tipine göre farklılık göstermektedir. Diğer bir ifadeyle saatlik elektrik tüketimi değişmektedir. Kısa dönem yük tahmininin temel amacı, saatlik olarak değişen enerji tüketim piyasasında temel üretim çizelgeleme işlevleri için yük tahminlerini sağlamaktır.

Elektrik güvenilirlik gereklilikleri kapsamında fiziksel, çevresel etkilerle operasyon kısıtlamalarını en ekonomik şekilde belirlemek kısa dönem yük tahmininin birincil fonksiyonu olarak belirtilebilir. Diğer bir deyişle saatlik değişen enerji tüketiminde güvenilirlik gerekliliklerini esas alarak enerji santrallerinin üretim plan ve programlarının ekonomik çerçevede belirlenmesi kısa dönem yük tahmininin ana fonksiyonudur.

2.1.3 Orta Dönem Yük Tahmini (ODYT)

Bir haftalık bir dönemden bir yıllık bir periyoda kadar uzanan tahminler orta dönem yük tahmini olarak adlandırılır. Orta dönem yük tahmini, üretim santralleri yakıt planlaması (ör. hidroelektrik santralleri güç rezervlerinin düzenlenmesi vb.), bakım planlamaları organizasyonu ve düzenlenmesi, dağıtım sistemlerine yapılacak ekleme çalışmaları için önemli bir olgudur.

Günümüz dünyasında birçok ülkede enerji sistem fiyatlandırması bağımsız olarak ve serbest piyasada yapılmaktadır. Orta dönem yük tahmininin son zamanlarda önemli amaçlarından biri serbest piyasada oyuncuların fiyat belirleme aşamalarından olan orta vadede yükü tahmin etmektir.

Orta dönem yük tahminine etki eden faktörlerin başında mevsimsel etkiler, tüketici fiyat endeksi gibi tahmin edilen ülkenin yıllık ekonomik parametrelerini gösteren veriler, tahmin edilen yük tipi ve bölgeye göre yeni yerleşim yerleri kurulması vb. gibi faktörler bulunmaktadır.

2.1.4 Uzun Dönem Yük Tahmini (UDYT)

Bir yıldan uzun süreleri kapsayan zaman aralıklarının tahminleri uzun dönem yük tahmini (UDYT) olarak isimlendirilir. Uzun dönem yük tahmini ülkelerin enerji politikasının belirlenmesinde önemli yer tutmaktadır. Uzun dönem yük tahmini ile üretime girmesi ve üretimden çıkması gereken üretim santralleri ve bu üretim santralleri tipi, iletim, dağıtım sistemlerinin planlanması ve ülkelerin bütün uzun vadeli kalkınma planlarında kullandıkları enerji sistem politikasının yapı taşları belirlenir. Uzun dönem yük tahmini serbest piyasada büyük oyuncuların enerji fiyat anlaşmalarını etkileyen önemli faktörlerden biridir.

Uzun dönem yük tahmininde yaşanabilecek olası sapmalar anlaşma fiyatının karlılık ve zararlılık oranında değişikliklere neden olacaktır. Bu sebeple uzun vadeli fiyat

anlařmaları için büyük oyuncuların uzun vadeli tüketim miktarlarının belirlenmesi önem arz etmektedir.

Uzun dönem yük tahminine gayri safi milli hasıla, iç borçlanma, dış borçlanma, döviz endeksi gibi ekonomik göstergeler, nüfus artış hızı, sisteme dahil olacak büyük tesisler, ülkenin büyüme oranını gösteren parametreler ve ülkenin dış politikası ve bunun ekonomiye etkisi gibi birçok parametre etki etmektedir.

2.2 Yük Tahminine Etki Eden Faktörler

Yük tahmini yapılırken dikkat edilmesi gereken önemli faktörlerden biri yüke etki eden faktörlerin belirlenmesidir. Yük tahminine etki eden faktörler yükün tipine ve yük tahmininin dönemine göre değişmektedir.

2.2.1 Meteorolojik Faktörler

Meteorolojik koşullar kısa ve orta dönem yük tahminine etki eden önemli faktörlerden birisidir. Yük tipine göre tüketilen yük miktarı mevsimlere göre değişkenlik göstermektedir. Dünya genelinde genellikle mesken yük tüketici grubu için en yüksek tüketim genellikle kış aylarında gerçekleşmektedir. Sıcaklık, nem, bulutluluk oranı, rüzgâr hızı gibi faktörler kısa dönem yük tahminine etki eden önemli faktörlerin başında gelmektedir. Tahmin edilen bölgenin koşullarına göre kullanılacak meteorolojik veri değişkenlik göstermektedir. Nem oranı yüksek bölgelerde nem faktörü önemli bir faktör iken karasal bölgelerde ise nem oranının etkisi çok düşüktür. Bulutluluk oranı, rüzgâr hızı gibi meteorolojik parametreler de nem oranı gibi tahmin edilecek bölgenin koşullarına göre değişkenlik göstermektedir. Bu sebeple tahmin edilecek bölgenin özelliğı iyi bilinmelidir.

2.2.2 Zaman Faktörü

Zaman faktörü yük tahminine etki eden faktörlerden biridir. Özellikle mesken yükü için yük karakteristikleri zamana göre değişkenlik göstermektedir. Elektrik insanoğlunun yaşamının önemli bir parçası olduğundan dolayı insanoğlunun günlük yaşamından birebir etkilenmektedir. Mesai saatleri, uyku süresi/ uyanma saatleri gibi insanoğlunun günlük yaşam faktörleri tüketim trendini birebir etkilenmektedir. Hafta içi ve hafta sonu, resmi ve dini bayramlar, tatil günleri gibi gün tipleri yük tüketim trendine direkt etki etmektedir. Saat dilimi etkileyen faktörlerden biridir. Gün

içerisinde elektrik tüketimi farklılık göstermektedir. Bu sebeple elektrik tüketimi gece, gündüz ve puant talep olarak üç döneme ayrılmıştır.

2.2.3 Demografik ve Ekonomik Faktörler

Ekonomik ve demografik faktörler özellikle orta ve uzun dönem yük tahminine etki eden faktörlerin başında gelmektedir. Ülkelerin enerji tüketimleri ile ülkelerin büyüme endeksi parametreleri arasında nedensellik ilişkileri vardır. Demografik faktörler de uzun ve orta dönem yük tahminine etki eden faktörlerdir. Yeni bir yerleşim yeri kurulması, nüfus artışı, bulunan bölgeye ticari yatırım yapılması gibi faktörler doğrudan yük miktarına etki eder. Bu sebeple uzun ve orta dönem yük tahmini uygulamalarında yük tahmini yapılan ülke ve bölgenin özellikleri iyi belirlenmelidir ve bu özellikler tahmin algoritmasında kullanılmalıdır.

2.2.4 Önceki Yük Parametresi

Yük tahminine etki eden başlıca parametrelerin başında önceki yük parametresi gelmektedir. Önceki yük parametresini belirlemek amacıyla tahmin edilen yük eğrisi üzerinde trend analizi ve oto korelasyon çalışması yapılması gerekmektedir. Bazı yükler bir, iki, üç saat önceki yükten etkilenirken bazı yükler 24 veya 48 saat önceki yükten etkilenebilmektedir. Bu sebeple tahmin edilecek bölgenin yük karakteristiği ve trend analizi iyi bilinmelidir.

2.2.5 Rastgele Değişkenler

Yük tahmininde tahmin edilmesi en güç parametreler rastlantısal olaylardır. Rastlantısal olaylar öngörülemez olduğundan dolayı bu faktörlerin olduğu günlerde yük karakteristiği trend eğrisi dışına çıkıp farklı özellik sergiler. Mesken yükü için bireysel ve toplumsal olaylar ve bunların sonuçları rastgele değişkenlere örnek verilebilir. Sanayi yükü içinse grev veya lokavt olayları yük tüketimine direkt etkisi olan rastlantısal değişkenlerdendir.

2.3 Yük Tahmininde Kullanılan Yöntemler

Elektriksel yük tahmini konusunda farklı yaklaşımlar/metodolojiler kullanılarak birçok farklı çalışma yapılmıştır. Yük tahmini çalışmaları kronolojik olarak öncelikle geleneksel yaklaşımlarla sonra istatistiki veya başka bir deyişle analitik yöntemlerle günümüzde ise teknolojinin hayatımıza girmesi ve efektif olarak kullanılmaya

başlanması ile birlikte yapay zekâ yaklaşımları ile yapılmaya başlanmıştır. Günümüzde yük tahmini çalışmalarında yöntemler en genel kapsamda analitik yöntemler ve yapay zekâ içeren yöntemler olarak 2 sınıfta incelenebilir.

2.3.1 Analitik Yöntemler

Analitik yöntemler istatistiki veya matematiksel yöntemler olarak ifade edilebilir. Literatürde sıklıkla kullanılan bazı analitik yöntemler şu şekildedir:

- Benzer gün yaklaşımı
- Regresyon
- Zaman serisi analizi
- En küçük kareleri yöntemi

2.3.1.1 Benzer Gün Yaklaşımı

Elektrik yük tahminleri arasında kompleksitesi en düşük yöntem olan benzer gün yaklaşımı, tahmin edilen günün karakteristik özelliklerine benzer başka bir günü referans alıp tahmin etmeye yarayan yöntemdir. Benzer gün yaklaşımı değişken koşullara cevap vermediği için bu yaklaşımın doğruluk oranı düşüktür.

2.3.1.2 Regresyon Analizi

Regresyon metodu, bağımlı ve bağımsız değişken olarak tanımlanan iki farklı değişken arasındaki ilişkiyi incelemek ve tahmin etmek amacıyla kullanılan istatistiksel tahmin yaklaşımlarından biridir.

Regresyon yaklaşımında bağımlı değişkenin bağımsız değişkenlerin değişimi sonucunda nasıl etkilendiği incelenir ve formülize edilir. Bir diğer ifadeyle bağımsız değişkenler ile bağımlı değişkenler arasında çeşitli formülasyon bağıntıları kurma tekniği regresyon olarak tanımlanabilir.

Elektrik yük tahmini yaklaşımlarından biri olan regresyon yaklaşımında tahmin edilecek yük ile bu yüke etki eden faktörler arasında matematiksel bir ilişki kurularak tahmin yapılır.

Tahmin edilecek yük Y , tahmin edilecek olan yüke etki eden giriş parametreleri ise $X_1, X_2, X_3 \dots$ olarak tanımlanırsa,

$$Y = f(X_1, X_2, X_3) \quad (3.1)$$

Tahmin edilecek yük Y giriş parametrelerinin bir fonksiyonudur ve bu fonksiyon en genel ifadesiyle regresyon yaklaşımını ifade etmektedir. Regresyon bağlantısında bağımlı ve bağımsız değişkenler arasında korelasyon arttıkça tahmin doğruluğu artacaktır. Değişken yüklerde regresyon tekniğinin doğruluk oranı düşüktür.

Bağımsız değişken parametresinin iki veya daha fazla olmasına göre regresyon teknikleri literatürde sınıflandırılmıştır. Bağımsız değişken parametresinin iki değişken olduğu denklemler basit regresyon; bağımsız değişken sayısının ikiden fazla olduğu denklemler ise çoklu regresyon modeli olarak literatürde isimlendirilmiştir.

2.3.1.2.1 Basit Regresyon Modeli

İki bağımsız değişkenin ve bir bağımlı değişkenin olduğu regresyon modelleridir. Basit regresyon modellerinde bağımlı ve bağımsız değişken arasındaki ilişki:

$$Y = \alpha + \beta X + \varepsilon \quad (3.2)$$

şeklinde formülize edilebilir. Formülde yer alan ifadeler:

- Y : Bağımlı değişken veya çıkış parametresi
- α : Sabit bir değer
- X : Bağımsız değişken
- ε : Hatayı ifade eden terim
- β : Regresyon katsayısı

olarak tanımlanabilir. 3.2 denkleminde de görülebileceği üzere basit regresyon modelinde kurulan bağlantı lineer bir denklemdir. Bağımsız değişken X bir birim değiştiğinde, bağımlı değişken çıkış parametresi olan Y , β birim kadar değişir. Bu katsayıların belirlenmesinde sıklıkla en küçük kareler yöntemi kullanılır. En küçük kareler yöntemi tahmin edilen değer ile gerçek değerlerin farklarının toplamını minimize etmeye dayanan teknik olarak tanımlanabilir.

Basit regresyon modelinde bağımlı değişken sayısının 2 olması ve bağımlı ve bağımsız değişkenler arasında lineer bir ilişki kurulması sebebiyle elektrik yük tahminlerinde doğruluk oranı düşüktür. Bu sebeple basit regresyon modeli elektrik tahmin çalışmalarında günümüzde çok tercih edilen bir yöntem değildir. Genellikle basit problemlerin çözümünde tercih edilmektedir.

2.3.1.2.2 Çoklu Regresyon Modeli

Elektrik yük tahminlerinde bağımlı değişken ve bağımsız değişken arasında direkt lineer bir ilişki olmaması ve elektrik yüküne etki eden faktörlerin birden fazla olması sebebiyle çoklu regresyon modeli basit regresyon modeline göre doğruluk oranı daha yüksek sonuçlar verir.

Çoklu regresyon modelinde veya literatürdeki yaygın kullanım adı olan doğrusal çoklu regresyon modelinde bağımlı değişken ile birden çok bağımsız değişken arasında lineer doğrusal ilişki kurulur. Çoklu regresyon modeli en genel olarak:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n + \varepsilon \quad (3.3)$$

şeklinde formülize edilebilir. Çoklu regresyon modelinde denklemden görülen hata ve sabit değişken basit regresyon modeli ile aynıdır. Basit regresyon modeline kıyasla çoklu regresyon modelinde birden fazla bağımsız değişken yer almaktadır. Denklemden yer alan katsayılar bağımsız değişkenlerin değişiminde bağımlı değişkenin ne kadar değiştiğini ifade eder. Bağımsız değişkenler sıcaklık, nem, saat vb. gibi elektrik yük tüketimine etki eden faktörler olarak ifade edilebilirken, bağımlı değişken ise bu faktörlerden etkilenen elektrik yükü olarak ifade edilebilir. Bağımsız değişkenlerin katsayıları korelasyona göre değişebilmektedir. Katsayılar en küçük kareler yöntemiyle hesaplanabilir.

2.3.1.3 Zaman Serisi Analizi

Zaman serisi belirli bir zaman aralığında ölçülmüş olan gözlem değerlerinin ardışık şekilde sıralanmış kümesi olarak tanımlanabilir. Bir diğer ifadeyle bir verinin farklı zaman dilimlerinde gözlemlenen farklı değerlerinin sıralanması sonucu oluşan veri setleridir [32]. Newbold ve diğ. yapmış oldukları çalışmada bir ekonometrik zaman serisinin trend, mevsim bileşeni, döngüsel bileşen ve düzensiz bileşen olmak üzere dört bileşenden oluştuğunu çalışmasında belirtmiştir [33]. İlgili çalışma her ne kadar ekonometrik veriler özelinde olsa da zaman serileri için genellenebilir. Bu bileşenler:

- **Trend Bileşeni:** Bir zaman serisinin uzun zaman diliminde gösterdiği durum trend bileşeni olarak isimlendirilir. Diğer bir ifadeyle ise genel eğilim bileşeni de denilmektedir. Bir zaman serisi kararlı hale geldikten sonra trend bileşeni artma, azalma veya sabit olarak eğim göstermektedir.

- Düzensiz Bileşen: Zaman serisinde herhangi bir kural ve düzene uymayan durumları temsil eder. Bir diğer ifadeyle hata terimleri olarak da isimlendirilirler.
- Mevsimsel Bileşen: Zaman serisinin verilerinin farklı mevsimlerde farklı trendler gösterdiğini ifade eder.
- Çevrimsel Bileşen: Zaman serisinde mevsimsel olmayan fakat bir sebepten dolayı verilerin o dönem içerisinde farklı bir döneme kıyasla farklı bir trend eğrisi izlediği bileşendir. Örnek verecek olursak tatil günlerinde elektrik yük tüketimi kendi içerisinde farklı bir trend göstermektedir.

Zaman serisinin tanımı ve özellikleri üst kısımlarda açıklanmıştır. Zaman serisi analizi ise bir zaman serisinin bu an ve önceki verilerini çeşitli yöntem ve yaklaşımları analiz edip, zaman serisinin gelecekteki davranışını tahminde bulunmak olarak tanımlanabilir. Zaman serisi analizleri için literatürde kullanılmış birçok farklı yöntem vardır. Literatürde kullanılan en yaygın modellerden biri Box-Jenkins yöntemleridir. Box- Jenkins yöntemlerinin yaygın olarak kullanılmasının temel sebebi bir zaman serisinin istatistiksel özelliklerini analiz ederek ve kullanarak çözümleme yapısında verilerin olasılık özelliklerinin ana planda olmasıdır [34].

Zaman serisi modelleri literatürde genellikle 3 sınıfa ayrılarak incelenmiştir. AR, MA, ARMA modelleri literatürde sıklıkla kullanılan ve doğruluk oranları yüksek olan zaman serisi analizi yöntemleridir.

2.3.1.3.1 AR(p) Modeli (Oto regresif Süreç)

Zaman serisi analizi yaklaşımlarından olan, oto regresif süreç olarak tanımlanan İngilizce ifadesiyle Autoregressive-AR modeli bağımlı değişkenin geçmişteki değerinin bir fonksiyonu olarak tanımlanan süreçtir. AR modeli 1926 yılında Yule tarafından geliştirilmiştir [35]. AR modelinde zaman serisi kendi gecikmeli değerinin farklı ağırlıklarda toplamı ve denkleme eklenecek rassal hata teriminin toplamı olarak ifade edilir. P dönem geçmiş veriler kullanılarak veya bir diğer ifadeyle p. mertebeden ifade edilen bir AR modeli:

$$Y_t = c + \sum_{i=1}^p (\phi_i * Y_{t-i}) + \epsilon_t \quad (3.4)$$

olarak ifade edilebilir.

Denkleimde yer alan Y_t ifadesi tahmin edilecek parametreyi diğer bir deyişle bağımlı değişkeni temsil etmektedir. Elektrik yük tahmini açısından değerlendirilirse Y_t ifadesi tahmin edilecek yüke denk gelmektedir. Denkleimde yer alan toplam sembolünde ifade edilen $Y_{p-1}, Y_{p-2}, \dots, Y_{p-i}$ ifadeleri bağımsız değişkenleri diğer bir ifadeyle bağımlı değişken olan Y_t ifadesinin geçmiş zamana ait gecikmiş değerleridir, diğer bir ifadeyle Y_t ifadesinin t-n zamanda aldığı değerlerdir. Elektrik yük tahmini açısından değerlendirilirse saatlik tahmin x saat önceki tüketim yüküne denk gelmektedir. $\phi_i, \phi_i, \dots, \phi_p$ ifadeleri otoregresif parametreleri temsil eder [36].

Otoregresif ifadelerin aldığı değer bağımlı değişken ile bağımsız değişken arasında korelasyon ilişkisine göre değişkenlik göstermektedir. Elektrik yük tahmini açısından değerlendirilirse tahmin edilecek yükün t saat önceki değeri ile tahmin edilecek yük arasındaki korelasyona göre değişkenlik gösterir. Denkleimde görülen c terimi sabit bir değeri diğer bir ifadeyle kesme terimini ifade etmektedir. Literatürde c teriminin denklemini etkilemediği durumlarda denklemini sadeleştirmek amacıyla c terimi çoğu zaman ihmal edilmektedir. Denkleimde yer alan ε_t değeri hata veya kalıntı değerini temsil etmektedir. Hata teriminin varyansının sabit, ortalamasının sabit olması ve herhangi bir sebeple açıklanamayan rassal değerler içermesi gerekmektedir.

2.3.1.3.2 MA(q) Modeli (Hareketli Ortalama Süreci)

Zaman serisi analizi yaklaşımlarından olan MA modeli veya İngilizce ifadesiyle Moving Average hareketli ortalama süreci olarak tanımlanabilir. MA modeli 1937 yılında Slutsky tarafından geliştirilmiştir [35]. MA sürecinde tahmin edilecek olan bağımlı değişken değişkenin hata terimlerinin ortalamalarının bir fonksiyonu olarak ifade edilir. Q dönem geçmiş veriler kullanılarak veya diğer bir ifadeyle q. mertebeden MA modeli:

$$Y_t = c + \sum_{i=1}^q (\theta_i * \varepsilon_{t-i}) + \varepsilon_t \quad (3.5)$$

olarak ifade edilebilir. Denkleimde yer alan Y_t, c, ε_t ifadeleri AR sürecinde olduğu gibi sırasıyla bağımlı değişkeni, sabit terimi ve hata terimini temsil etmektedir. θ_i ifadesi hareketli ortalama parametresidir, diğer bir ifadeyle i. mertebeden hatanın ağırlığıdır. ε_{p-i} parametresi i gecikmeli hata parametresini temsil eder. Denklem 3.5'ten anlaşılacağı üzere bir MA modelinde t anındaki bağımlı değişken değeri q adet

i gecikmeli hataların farklı ağırlıkta toplanması olarak ifade edilir. Ağırlık oranı kurulan modele ve veri setine göre değişkenlik göstermektedir.

MA modelinde hata terimlerinin ortalamasının sıfır ve varyansının sabit olması gerekmektedir. MA modeli bu özelliğinden dolayı tersine çevrilebilir bir modeldir. Gecikme uzunluğu arttıkça denklemde yer alan katsayıların ağırlığı azaltılabilir, bir diğer ifadeyle denklemde q ifade edilen denklemin mertebe sayısı arttıkça θ_i ile ifade edilen katsayıların ağırlığı azaltılabilir.

2.3.1.3.3 ARMA (p, q) Modeli (Oto regresif Hareketli Ortalama Süreci)

İngilizce ifadesi olan Autoregressive Moving Avarage ifadelerinin kısaltılmışı olan ARMA modeli zaman serisi yaklaşımlarından AR ve MA modellerinde olduğu gibi bir lineer denklemle ifade edilmektedir. Zaman serisi analizi yaklaşımlarından olan ARMA modeli 1954 yılında Wold tarafından geliştirilmiştir [35].

ARMA modeli yukarıdaki bölümlerde bahsedilen AR ve MA modellerinin bir birleşimi, kombinasyonu şeklinde ifade edilebilir. P dönem geçmiş veri kullanılan q. mertebeden hata oranı barındıran bir ARMA modeli:

$$Y_t = c + \sum_{i=1}^p (\phi_i * Y_{t-i}) + \sum_{i=1}^q (\theta_i * \varepsilon_{t-i}) + \varepsilon_t \quad (3.6)$$

şeklinde ifade edilir.

Denklem 3.6'danda görülebileceği üzere denklem AR modeli ve MA modellerini içermektedir. 3.6 denkleminde yer alan parametreler 3.4 ve 3.5'te yer alan parametreler ile aynıdır.

ARMA modeli durağan ve tersine çevrilebilir bir modeldir. ARMA modelinin durağan olması denklem 3.6'da görülen AR modelinden gelen toplam parametresine, tersine çevrilebilir olması ise MA modelinden gelen toplam parametresine bağlı olmaktadır. Bu durumun diğer bir ifadesi ARMA modeli durağanlığı birebir olarak modeli oluşturan oto regresif bölümle(p), tersine çevrilebilirliği ise modeli oluşturan hareketli ortalamalar(q) bölümüyle aynıdır.

2.3.1.3.4 ARIMA(p,d,q) Modeli (Bütünleşik Oto regresif Hareketli Ortalama Süreci)

ARIMA modeli veya İngilizcesi Autoregressive Integrated Moving Average olarak ifade edilen Bütünleşik Oto regresif Hareketli Ortalama süreci literatürde sıklıkla tercih

edilen zaman serisi analizi yaklaşımlarındandır. 1926 yılında Yule tarafından geliştirilen model ARIMA modeline baz teşkil etmiştir. AR ve MA süreçlerini birlikte barındıran ayrıca I(integrated) terimini barındıran süreçtir. I bağımlı değişkenin serisinin barındırdığı trend olarak tanımlanabilir. ARIMA modeli box-jenkins sürecinin temelini oluşturmaktadır.

Zaman serisinde zaman serisinin varyansının, kovaryansının ve ortalamasının serinin farklı dilimlerinde sabit olduğu diğer bir ifadeyle zaman serisinin durağan olduğu durumlarda AR modeli, MA modeli veya ARMA modeli veri setine göre kullanılabilir. Literatürde sıklıkla karşılaşılan durağan olmayan zaman serisi için AR, MA veya ARMA modeli kullanılamaz. Öncelikle veri setinin durağanlaştırılması gerekmektedir. Durağan olmayan zaman serilerinde ARIMA yaklaşımı kullanılmaktadır. Zaman serisinin durağanlaştırması için farklı yöntemler izlenmektedir. Bu yöntemlerden en yaygın olanı fark almadır. ARIMA modeli (p, d, q) terimlerini içermektedir. P terimi AR modelinden, q terimi MA modelinden d terimi durağan olmayan zaman serilerini durağanlaştırmak maksadıyla uygulanan durağanlaştırma veya bir diğer ifadeyle fark teriminden gelmektedir.

ARIMA modelinin kullanılmasının çeşitli avantajları ve dezavantajları vardır. ARIMA modelinin minimal sayıda parametre içeren yapıyı bulması amacına dayanması önemli avantajlarından. Bir diğer avantajı ve tercih edilme sebebi ise zaman serisinin durağan/durağan olmama durumu gibi özelliklerine bakılmaksızın hazır kütüphaneler vasıtası ile birçok zaman serisine uygulanabilirliği ve uygulama kolaylığıdır. ARIMA modeli (p, d, q) parametreleri deneysel metotlarla belirlenir. Bir diğer ifadeyle birçok farklı parametre denenir ve en az sayıda parametre içeren uygun model belirlenir. En optimum model ve değer belirlenirken denemelerin birer birer yapılması ve sonucun bu şekilde elde edilmesi modelin en büyük dezavantajlarından biridir.

2.3.2 Yapay Zekâ Yöntemleri

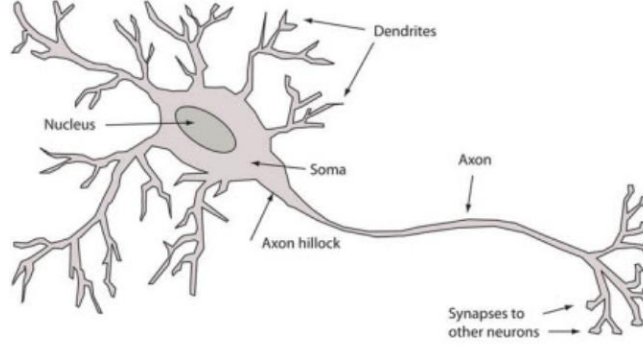
Son dönemlerde bilgisayarların işlemcilerinin gücünün artması, süper işlemcili bilgisayarların kullanılmaya başlanması ile milyon satırlık matrislerin kısa sürede çözülebilmesi ile birlikte yapay zekâ yöntemleri elektrik yük tahminlerinde sıklıkla kullanılmaya başlanmıştır. Literatürde yöntemlerin ayrı ayrı ve birlikte kullanılmaya başlanarak oluşturulan hibrit modeller birçok durumda analitik yöntemlere nazaran bilhassa değişken yük ve koşul durumlarında doğruluk oranı yüksek sonuçlar

vermiştir. Literatürde kullanılan birçok yapay zekâ yöntemi vardır. Temelde makine öğrenmesi ve alt kümesi olarak derin öğrenme olarak sınıflandırabilen bu algoritmalar bu tez kapsamında denetimli makine öğrenmesi algoritması kullanıldığı için denetimli makine öğrenmesi algoritmalarına yer verilmiştir. Makine öğrenmesi denetimli ve denetimsiz öğrenme olarak iki ayrı gruba ayrılan algoritmalar grubudur. Denetimli öğrenme algoritmaları, sınıflandırma ve regresyon yöntemleri olarak gruplandırılabilir. Denetimsiz öğrenme metotları ise kümeleme ve ilişki yöntemleri olarak gruplandırılabilir. Denetimsiz öğrenmede bağımlı ve bağımsız değişkenler belirlidir. Sınıflandırma (classification) ve regresyon(regression) algoritmaları kullanılarak eğitilen algoritma yeni bir x giriş değerinde tahmin olarak çıkış y değerini verecektir. Denetimsiz öğrenme metodolojilerinde ise sonuç belirli değildir. Yalnızca giriş değişkenleri veya diğer bir ifadeyle bağımsız değişkenler bulunmaktadır. Amaç veri kümesini tanımak genel yapıyı ifade etmektir. Bu çalışmada denetimli öğrenme metodolojileri kullanıldığı için bu yöntemlerin üzerinde durulacaktır.

2.3.2.1 Yapay Sinir Ağları

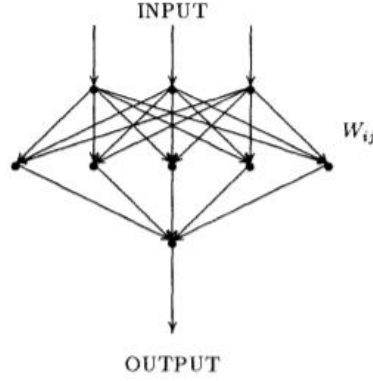
Yapay zekâ algoritması olarak sınıflandırabilen ve kullanım ağı çok geniş olan yapay sinir ağları günümüzde literatürde sıklıkla kullanılmaktadır. İlk yapay sinir ağı modellemesi 1943 yılında nörofizyolog Warren McCulloch ve matematikçi Walter Pitts tarafından yapılmıştır [37]. McCulloch ve Pitts tarafından yapılan çalışmada insan beyninde bulunan merkezi sinir sistemi referans alınarak bir elektronik devre modellenmiştir. Yapılan bu çalışma yapay sinir ağları için mihenk taşı olmuştur ve ileride yapılacak referans olmuştur.

Yapay sinir ağları (YSA), insan beyninin özelliklerinden olan öğrenme ve öğrenilen bu bilgiyi referans alarak yeni bilgiyi oluşturma, biçimlendirme ve yorumlama yeteneklerine sahiptir ve bu işleri tanımlanan algoritma doğrultusunda manuel müdahaleye gerek olmadan otomatik olarak yapabilmektedir. Yapay sinir ağları insan beyni referans alınarak modellendiği için birçok özelliği merkezi sinir sistemi gibi benzerlik gösterir. Yapay sinir ağları paralel olarak birden fazla işlemi yapabilir, Farklı sistemler için uyarlanabilir, bir diğer ifadeyle esnektir. Şekil 2.1’de insan hücre yapısı görülmektedir.



Şekil 2.1: İnsan sinir hücresi yapısı [37].

Yapay sinir ağları ileri beslemeli, geri beslemeli, tek katmanlı ve çok katmanlı yapıda olabilir. Çeşitli kurallar silsilesinde birbirine bağlanması sonucunda oluşturulan veri kümeleridir. Şekil 2.2’de üç girişli çok katmanlı bir ağ yapısı görülmektedir [34].

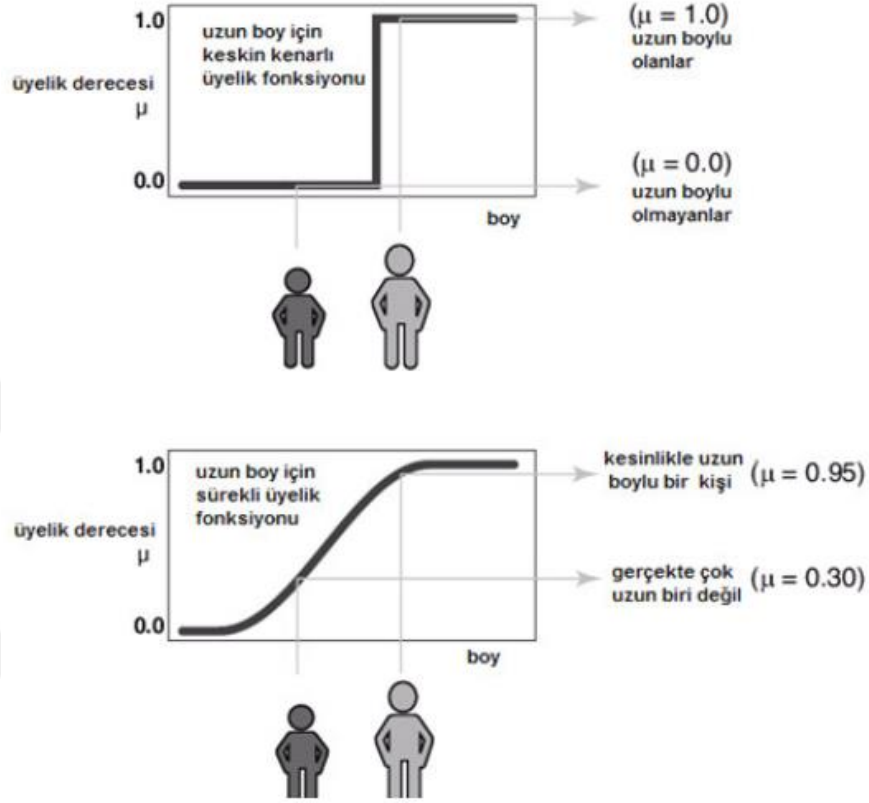


Şekil 2.2: Üç girişli çok katmanlı ileri beslemeli ağ yapısı [38].

2.3.2.2 Bulanık Mantık

Klasik mantık, ikili mantık veya literatürde yaygın olarak bilinen diğer bir ifadesiyle Aristo mantığında bir evrensel kümede bulunan bir eleman bir kümeye aittir veya değildir. Klasik mantıkta bir diğer ifadeyle elemanın kümeye ait olma durumu 0 veya 1 değeri olarak tanımlıdır. 0 ise o kümeye ait değildir; 1 ise o kümeye aittir. Klasik mantıkta değerler kesinlik belirtir ve yaklaşım ifadesi barındırmaz. Gerçek yaşamda verilerin klasik mantık yaklaşımıyla kesin olarak sınıflandırılmayacağını iddia eden Lütü Aliasker Zade tarafından bulanık mantık yaklaşımı ortaya koyulmuştur [39]. Bulanık mantık yaklaşımı, klasik mantıkta bulunan 0 veya 1 değerlerini net değer olarak değil sınır bölgesi olarak kabul eden ve klasik mantığa aykırı olarak değerlerin 0 ve 1 arasında değişebileceği ve verilerin bu şekilde sınıflandırılabileceği bir tekniktir.

Klasik mantık yaklaşımında demografik özellikler olan boy bilgisi uzun ve kısa olarak sınıflandırılır 0 veya 1 değerini almaktadır. Bulanık mantıkta ise boy bilgisi 0 ve 1 olarak kesin net değerler almaz ve değerler 0 ve 1 arasında sınıflandırılırlar. Klasik ve bulanık mantık kümeleme farklılığı Şekil 2.3'te açık bir şekilde görülebilmektedir [40].



Şekil 2.3 Klasik ve bulanık mantık küme sınıflandırılma farkları [40].

Bulanık mantık yaklaşımı yapay zekâ algoritmaları arasında popülerliğini sürdüren bir algoritmadır. Bunun başlıca sebepleri:

- Kesin net ifade yerine yaklaşım düşünme metodunu tercih etmesi
- İnsan düşünce yapısına paralel bir yapısı olması
- Karmaşık matematiksel modellere ihtiyaç duymaması
- Veri aralıklarının sözel değişkenlerle ifade edilebilmesi

olarak söylenebilir. Bunların yanında bulanık mantık algoritmasının birtakım dezavantajları da vardır. Bunlar başlıca:

- Kullanılacak olan FIS (fuzzy inference system) yapısı ve belirlenen kuralların belirlenmesinin uzman görüşüne göre yapılması

- Uygun yapının deneyerek bir başka ifadeyle deneme/yanılma yaklaşımı ile bulunması
- Üyelik fonksiyonu, işlevi ve derecelerinin o veri setine özel olup genele indirgenememeleri olarak tanımlanabilir.

2.3.2.3 Destek Vektör Makineleri

Destek vektör makineleri (DVM) veya İngilizce ifadesiyle support vector machines (SVM) bir veri kümesini hiper düzlem üzerinde farklı türde vektörler vasıtasıyla sınıflandırma amacıyla kullanılan denetimli makine öğrenmesi algoritmalarından biridir.

Destek vektör makineleri yaklaşımının temeli 1936 yılında Fisher tarafından yapılan çalışmada ortaya koyulan doğrusal diskriminant yöntemine dayanır [41]. Vapnik ve Cortes tarafından destek vektör makineleri algoritmasının yapı taşlarından olan soft marjın yaklaşımı 1995 yılında yapılan çalışmayla ortaya konmuştur [42]. 1995 yılında Vapnik soft marjın yaklaşımını “The Nature of Statistical Learning Theory” adlı kitabında genişleterek yayınlamıştır [43]. Destek vektör makineleri matematiksel olarak 1936’lara kadar uzansa da gerçek manada temelleri Vapnik ve arkadaşları tarafından yapılan çalışmaya dayanmaktadır. Vapnik’in ifadesine göre destek vektör makineleri evrensel bir öğrenme algoritmasıdır.

Evrensel olarak destek vektör makineleri algoritmasının birçok farklı fonksiyon ve amaç için kullanılabileceğini ifade eder. Günümüzde destek vektör makineleri sağlık, turizm, enerji, ekonomi, iktisat gibi birçok farklı sektör ve sınıflandırma probleminin çözümünde kullanılmaktadır.

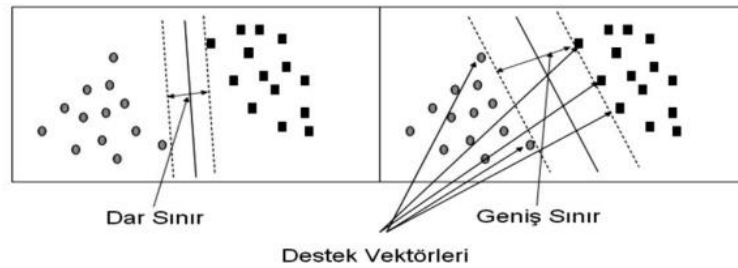
Destek vektör makineleri en yalın ifadesiyle istatistiksel öğrenme metodolojisine dayanan bir hiper düzlem üzerinde bulunan verileri hiper düzlem üzerinde çizilebilecek sonsuz sayıda doğru arasından birbirinden ayıran en geniş aralığı (margin) oluşturan çekirdek fonksiyonunu kullanarak sınıflandıran ve tahmin etmek için kullanılabilen yüksek doğruluk oranlı makine öğrenmesi algoritmasıdır.

Basit bir sınıflandırma örneği şekil 2.4’ten görülebilir. Şekil 2.4’ten de anlaşılacağı üzere iki veri kümesi bir vektör yardımıyla iki ayrı sınıfa ayrılmıştır. Sarı ve kırmızı renklerle renklendirilen ve düzlem üzerinde rastgele veriler atılarak oluşturulan iki ayrı veri kümesi LIBSVM algoritması yardımıyla sınıflandırılmıştır.



Şekil 2.4: Destek vektör makineleri sınıflandırma örneği [44].

Destek vektör makineleri yukarıda da açıklandığı üzere hiper düzlem yaratır ve bu yaratılan hiper düzlemde vektörler vasıtasıyla sınıflandırılır. Bu sınıflandırıcı vektörler 1-boyutta nokta, 2-boyutta doğru ve 3-boyutta boyutta ise düzlem olabilir. Doğruluk oranı en yüksek sınıflandırma için sınıflandırılan iki veri kümesinin veya diğer bir ifadeyle boyut ayırıcısı hiper düzleminin sınırları arasında mesafe maksimize edilmelidir. Şekil 2.5'ten de açıkça görülebileceği üzere doğrular sınıflandırılan hiper düzlemlerdir. Doğruluk oranı yüksek sınıflandırmaya için sınıflandırılan dataların en yakın ikisi seçilir ve bunlara paralel doğrular oluşturulur. Şekilde görülen en yakın veriler üzerinden geçenler destek vektörü olarak isimlendirilir [45].



Şekil 2.5: Destek vektörleri ile sınıflandırma örneği [45].

Destek vektör makineleri doğrusal olarak ayrılabilen, doğrusal olarak ayrılamayan mekanizmalar olarak sınıflandırabilir. Destek vektör makineleri ile sınıflandırma algoritmasında çekirdek fonksiyonları veya İngilizce ifadesiyle kernel fonksiyon giriş parametrelerinin iç çarpımları ile lagrange çarpanlarının kombinasyonları sonucunda çıkış parametresi oluşturulur. Çekirdek fonksiyonları direkt olarak çıkış parametresine etki eder. Bu sebeple doğruluk oranı yüksek tahmin için çekirdek fonksiyonlarının seçimi önem arz etmektedir [46].

2.3.2.3.1 Doğrusal Ayrılabilen Destek Vektör Makineleri

Bir destek vektör makineleri probleminde öğrenme eğitim verisi vasıtasıyla gerçekleşir. Doğrusal olarak ayrılabilen destek vektör makineleri modelleme açısından en basit olan yapıdır.

Ağırlık vektörü(w) ve eşik değeri(b) olarak ifade edilen $\{-1,1\}$ şeklinde sınıf etiketi kullanılarak ayrılabilen N tane veri barındıran bir eğitim verisi eşitsizliklerini sağlaması gerekmektedir [47]:

$$y_i[(wx_i + b) \geq 1, i = 1,2,3 \dots N \quad (3.7)$$

$$b + wx = 0 \quad (3.8)$$

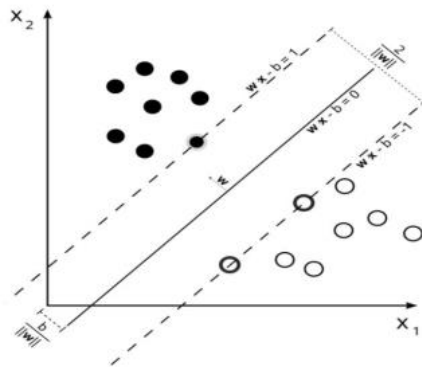
Doğruluk oranı yüksek sınıflandırma için önemli olan hiper düzlemin uygun şekilde belirlenmesidir. İki sınıfa ayrılmış olan ve mesafe olarak birbirine en yakın iki nokta seçilirse iki nokta arası mesafe diğer bir ifadeyle margin 3.8 ve 3.9 denklemleri çözülerek 3.10 denklemindeki gibi ifade edilebilir. Destek vektörleri, sınır doğruları ve denklemleri ve en yakın iki nokta arası mesafe şekil 2.6'da özetle görülebilmektedir.

$$b + wx = 1 \quad (3.9)$$

$$b + wx = -1 \quad (3.10)$$

$$m = 2/|w| \quad (3.11)$$

Şekil 2.6'dan da açıkça görülebileceği üzere en yakın iki nokta sonsuz sayıda doğru ile ayrılabilir. Destek vektör makineleri veya mekanizmaları algoritmasının temel amacı en yüksek margin ile bu iki noktanın ayrılmasıdır.



Şekil 2.6: Doğrusal ayrılabilen DVM [48].

En yakın iki nokta arası mesafe belirlendikten sonra doğruluk oranı yüksek sınıflandırma ve tahmin çıktısı elde edebilmek için bu iki nokta arası mesafeyi maksimum aralıkta tutmak gerekir. Maksimum aralıkta tutmanın amacı iki verinin birbirinden olabildiğince keskin sınırlarla ayrılmasıdır. Maksimum uzaklıkta sınıflandırma yapılabilmesi için $|w|$ ifadesinin minimize edilmesi gereklidir [49].

En yakın iki nokta arasında mesafenin denklem 3.12’de görüldüğü üzere minimize edilmesi gerekmektedir. Optimizasyon problemlerini maksimize ve minimize etmek için Joseph Louis Lagrange tarafından isimlendirilen pratik bir yöntem olan Lagrange çarpanları metodu kullanılabilir [50]. Lagrange çarpanları $a_1, a_2, a_3 \dots a_n$ olarak adlandırılırsa doğrusal olarak ayrılabilen destek vektör makine problemleri için Lagrange fonksiyonu denklem 3.13’teki gibi ifade edilebilir. Denkleminde yer alan Lagrange çarpanları pozitif değerler olup negatif değerler alamaz.

$$\min Q(w) = |w|/2 \quad (3.12)$$

$$Q(a, b, w) = \frac{1}{2} |w|^2 - \sum_{i=0}^l (a_i [Y_i(wX_i + b) - 1]) \quad (3.13)$$

Kurulan Lagrange denklemlerinde Lagrange çarpanlarına göre maksimum w ve b değerlerine göre ise minimum değerlerin bulunması amaç fonksiyonudur. Kurulan Lagrange denkleminde tanımlanan amaç fonksiyonu çerçevesinde $\min Q(w)$ ’yi elde etmek için w ve b değişkenlerine göre türev alınır. Alınan türev sonrasında elde edilen denkleme literatürde sıklıkla kullanılan Karush-Khun-Tucker (KKT) koşullarının uygulanabilirliği denetlenmelidir. Karush-Khun-Tucker (KKT) koşulları optimizasyon problemlerinde kullanılan lineer olmayan programlamalarda tercih edilen, elde edilen sonuçların optimal değer olup olmadığı çıktısını veren testlerdir [51].

$$[Y_i(wX_i + b) - 1] \geq 0 \quad (3.14)$$

$$(a_i [Y_i(wX_i + b) - 1]) = 0 \quad (3.15)$$

Denklemlerinden görülebildiği üzere denklemin sağlanabilmesi için 2 koşul bulunmaktadır. Denklemin sağlanabilmesi için Lagrange çarpanlarının sıfır veya $Y_i(wX_i + b)$ ifadesinin 1’e eşit olması gerekmektedir. Lagrange çarpanlarının sıfıra

eşit olduğu durumda optimizasyon denklemi amaç denklemini sağlayamayacaktır. Bu sebeple $Y_i(wX_i + b)$ ifadesinin 1'e eşit olduğu yorumu yapılabilir. Bu koşulları sağlayan tüm durumlar problem için destek vektörleri olarak ifade edilebilir.

Kısıt denklemleri, Lagrange fonksiyonları ve Karush-Khun-Tucker kriterleri çerçevesinde hiper düzlem ve destek vektörleri eğitim verileri üzerinden uygulandıktan sonra test verilerinin veya bir diğer ifadeyle yeni gelen verilerin sınıflandırılması aşamasıdır. Yeni gelen verilerin sınıflandırılması için sigmoid fonksiyonu Lagrange çarpanları ile birlikte kullanılır. Eğitim verileri sonucu kurulmuş olan algoritması sonucunda test verilerinin veya bir diğer ifadeyle yeni gelen verilerin sınıflandırılması amacıyla kurulmuş olan denklem 3.16'da görülebilmektedir.

$$f(x) = \text{sign} \left(\sum_{i=0}^l (a_i Y_i X_i X + b) \right) \quad (3.16)$$

2.3.2.3.2 Doğrusal Ayrılamayan Destek Vektör Makineleri

Doğrusal olan ayrılabilen destek vektör makinelerinde görüleceği üzere veriler dağılımı hasebiyle doğrusal bir vektör tarafından hatasız ayrılabilir. Endüstride ve gündelik yaşamda kullanılan veri setlerinin birçoğunun doğrusal olarak hatasız bir şekilde ayrılması neredeyse imkansızdır. Gündelik hayatta kullanılan veri setleri doğrusal olarak ayrılırsa hata oranı yüksek, doğruluk oranı düşük bir sınıflandırma ve tahmin çıktısı ile karşı karşıya kalınabilir.

Doğrusal olarak ayrılamayan veri setleri için bu gibi sorunların önüne geçmek ve doğruluk oranını artırmak amacıyla birtakım yöntemler geliştirilmiştir. Doğrusal olarak ayrılmaması veya diğer bir ifadeyle ayrılırsa yüksek oranda hata olması sebebiyle sınıflandırma yapılırken yakınsama yapılarak sınıflandırma yapılmaktadır. Bu sebeple en optimum hiperdüzlemi belirlerken belirli bir hata tolere edilmelidir. Bu sebeple denkleme pozitif değer alabilen bir aylak değişken eklenir [52].

Literatürde bu değişken gevşek, aylak veya atık değişken olarak isimlendirebilir. Aylak değişken ε olarak tanımlanırsa:

$$[Y_i(wX_i + b)] \geq 1 - \varepsilon \quad (3.17)$$

olarak eşitsizlik elde edilir.

Görülebileceği üzere denklem 3.15 ve 3.17 arasında tek fark aylak değişkenin eşitsizliğe dahil edilmesidir. Aylak değişkenin eşitsizliğe dahil edilmesiyle birlikte hata toleransı ile birlikte sınırlar belirlenecek sonucu çıkarımı yapılabilir.

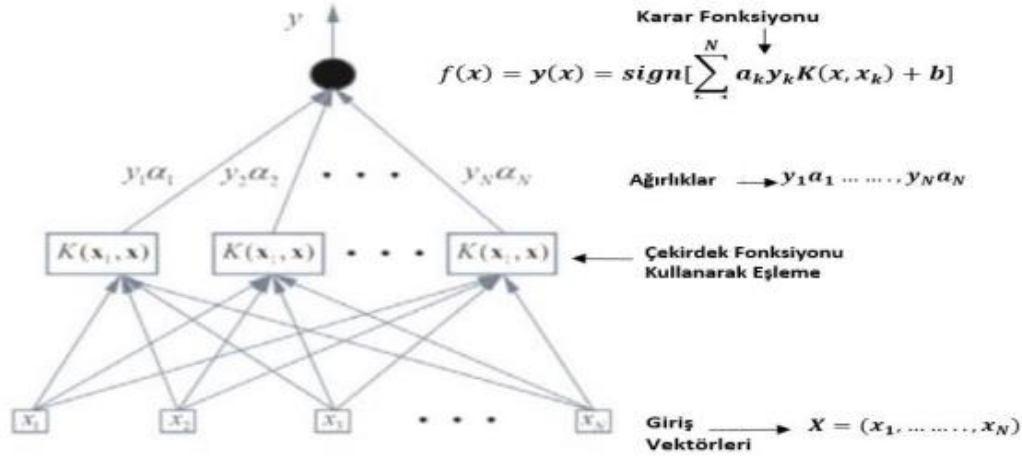
Eşitsizlikte yer alan aylak terimin (ϵ) alacağı değerlere göre sınıflandırma hakkında çeşitli çıkarımlarda bulunabilir:

- Aylak terim sıfıra eşitse sınıflandırma (X, Y) doğru olarak gerçekleşmiştir.
- $0 < \epsilon < 1$ ise sınıflandırma (X, Y) doğru yapılmıştır fakat vektör yüzeyinin içerisinde değil sınırları belirlenen marjin içerisinde yapılmıştır.
- $\epsilon \geq 1$ ise sınıflandırma yanlış olarak gerçekleşmiştir. Bir diğer ifadeyle ayırma yapılan marjin sınırının dışına çıkmıştır.

2.3.2.3.3 Çekirdek (Kernel) Fonksiyonları

Veri setlerinin doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan destek vektör makineleri ile sınıflandırılması doğrusal fonksiyonlarla gerçekleşir. Doğrusal olarak ayrılamayan vektör mekanizmalarında doğrusal fonksiyonlar kullanılır; fakat belirli bir hata sınırı toleransı varsayımı yapılır. Bahsedilen bu iki sınıflandırma metodolojisinde giriş parametresi ile çıkış parametresi arasında lineer bir bağımlılık olduğu varsayımı yapılır ve metodoloji bunun üzerine inşa edilir. Giriş ve çıkış parametresi arasında lineer bir ilişki yoksa kurulan model doğrusal olarak ayrılabilen veya ayrılamayan destek vektör makineleri ile gerçekleştirilirse hata oranı yüksek ve doğruluk oranı düşük olacaktır. Bu sebeple çekirdekli hilber uzayları teorisine dayanan farklı çekirdek fonksiyonları destek vektör makineleri uygulamalarında kullanılmaya başlanmıştır [53].

Bir uzayda tanımlanmış olan bir veri kümesi bir taşıma fonksiyonu kullanarak tanımlanan başka bir uzay alanına taşınabilir. Bu taşıma işlemine literatürde haritalama veya İngilizce ifadesiyle mapping işlemi denilmektedir. Yapılan haritalama işleminde bahsedildiği üzere taşıma fonksiyonları kullanılır. Kullanılan bu taşıma fonksiyonları çekirdek (Kernel) fonksiyonları olarak isimlendirilirler. Çekirdek fonksiyonları giriş vektörlerinden sonraki aşamada kullanılır. Çekirdek fonksiyonu kullanılan bir destek vektör makinesi örneği Şekil 2.7’de görülebilmektedir.



Şekil 2.7: Çekirdek fonksiyonu kullanılan bir destek vektör makinesi [47].

Literatürde sıklıkla kullanılan çekirdek fonksiyonları Çomak tarafından yapılan çalışmada ortaya konmuştur. Kullanılan çekirdek fonksiyonları Şekil 2.8’de görülebilmektedir.

Çekirdek İsimleri	Çekirdek Fonksiyonları
Doğrusal	$K(x, y) = x \cdot y$
d. dereceden yüksek dereceli fonksiyon	$[K(x, y) = (x \cdot y) + 1]^d$
Çok katmanlı algılayıcı	$K(x, y) = \tanh(kxy - \theta)$
Gauss Yarıçap Temelli Fonksiyon	$K(x, y) = K(x - y) = \exp(-\frac{\ x-y\ ^2}{2\sigma^2})$
Üssel Yarıçap Temelli Fonksiyon	$K(x, y) = K(x - y) = \exp(-\frac{\ x - y\ }{2\sigma^2})$
Ters İkinci Dereceden	$K(x, y) = K(x - y) = (\ x - y\ ^2 + c^2)^{1/2}$
İnce Tabaka Spline	$K(x - y) = \ x - y\ ^{2n+1}$ $K(x, y) = K(x - y) = \ x - y\ ^{2n} \ln(\ x - y\)$
Oransal İkinci Dereceden	$K(x, y) = 1 - \frac{\ x - y\ ^2}{\ x - y\ ^2 + \theta}$
Dalga Yapılı	$K(x, y) = \frac{\theta}{\ x - y\ } \sin(-\frac{\ x - y\ }{\theta})$
B-spline	$K(x, y) = B_{2n+1}(x - y)$
Fourier Serisi (trigonometrik çok dereceli)	$K(x, y) = \frac{\sin(d + \frac{1}{2})(x - y)}{\sin \frac{(x - y)}{2}}$
Küresel	$1 - \frac{3}{2} \frac{\ x-y\ }{\theta} + \frac{1}{2} (\frac{\ x-y\ }{\theta})^3$ eğer $\ x - y\ < \theta$

Şekil 2.8: Sıklıkla kullanılan çekirdek fonksiyonları [54].

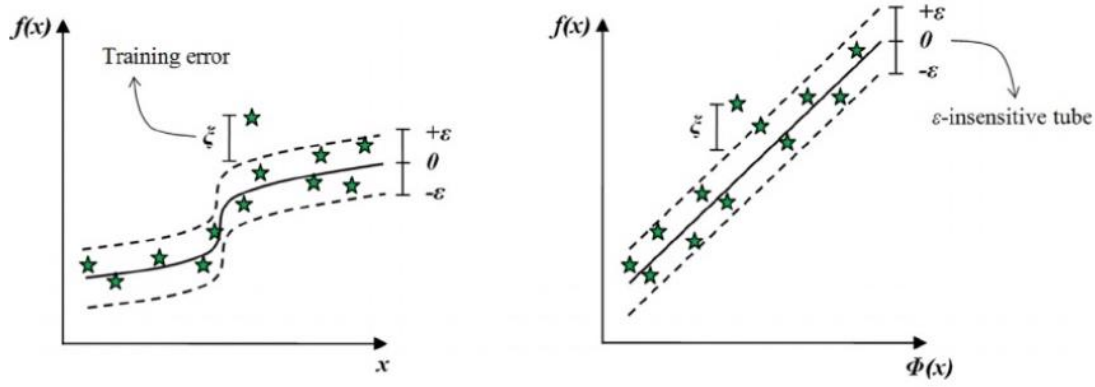
2.3.2.4 Destek Vektör Regresyonu

Destek vektör makineleri bölüm 2.3.2.3'te bahsedildiği üzere sınıflandırma problemlerinin çözülmesi amacıyla kullanılmaktadır. Destek vektör makineleri farklı birçok alanda ve veri setlerinde kullanılıyor ve doğruluk oranı yüksek çıktılar veriyor olsa da sınıflandırma bütün veri setleri için optimum sonucu vermeyebilir. Bu sebeple destek vektörleri kullanılarak Vapnik ve arkadaşları tarafından "Destek Vektör Regresyonu", İngilizce ifadesiyle ise "Support Vector Regression" isimli bir model ortaya konmuştur [55]. Yapılan çalışmada destek vektör regresyonu modelinin destek vektör makineleri modeline göre doğruluk oranı yüksek çıktılar verdiği ortaya konmuştur. Literatür çalışmaları incelendiğinde bazı veri setleri için destek vektör regresyonu modelinin destek vektör makinelerine göre daha iyi sonuçlar verdiği görülebilir.

Destek vektör regresyonu algoritmaları sınıflandırma için destek vektör makineleri ile aynı metodolojiyi kullanır. DVM ile arasındaki temel fark sınıflandırma vektörlerinin regresyonlar için kullanılıyor olmasıdır. Doğrusal olarak ayıramayan destek vektör makineleri için ε olarak tanımlanan, aylak değişken olarak ifade edilen, belirli bir hata oranının tolere edilmesi için kullanılan terim kullanıldığı bölüm 2.3.2.3.2'de bahsedilmişti. Destek vektör regresyonunda ise aynı doğrusal olarak ayıramayan destek vektör makinelerinde olduğu gibi ε olarak bir hata tolerans terimi tanımlanır [55]. Bu tanımlanan terimin amacı duyarsız kayıp olarak isimlendirilen yaklaşık değerlerin bulunması amacını güden bir tür risk terimi olarak tanımlanabilir [56].

Şekil 2.9'da ε ile ifade edilen bir duyarsız bant tanımlanmış sırasıyla doğrusal olan ve doğrusal olmayan iki vektör regresyonu sınıflandırılması görülebilmektedir. ε olarak tanımlanan alanın içinde kalan veriler için hata toleransı sıfırdır. Bu genişlik içinde kalan veriler duyarlı alan olarak tanımlanmaktadır. Duyarlı alan dışında kalan alanda şekilde açıkça görülebileceği üzere bir ξ değişkeni tanımlanmıştır. ξ değişkeni duyarlı alan dışında kalan hata toleransını simgelemektedir. ξ değeri literatürde eğitim hatası olarak da geçmektedir. Duyarlı alan içinde kalan bölgede ξ değeri sıfırdır. Hata toleransı denklem 3.18'deki gibi ifade edilebilir [57]:

$$L_{\varepsilon}(x, y) = \begin{cases} |x - y| - \varepsilon, & |x - y| > \varepsilon \\ 0, & \text{diğer} \end{cases} \quad (3.18)$$



Şekil 2.9: Destek vektör regresyonu örneği [54].

Destek vektör regresyonu kullanılarak sınıflandırılacak olan fonksiyon:

$$f(x, w) = w^T \phi(x) + b \quad (3.19)$$

şeklinde ifade edilebilir. Denklemden yer alan w ağırlık simgesi, $\phi(x)$ ise yukarıda mapping function veya bir diğer ifadeyle haritalama fonksiyonu olarak tanımlanmış fonksiyondur. Denklem 3.20’de tanımlanan hata fonksiyonunun minimize edilmesi ise w ve b değerleri bulunabilir.

$$R = \frac{1}{2} |w|^2 + C \sum_{i=1}^N (|y_i - f(x_i, w)|_{\varepsilon}) \quad (3.20)$$

Denklemden yer alan ağırlık vektörü $\frac{1}{2} |w|^2$ ifadesi modelin genelleştirilebilirliği veya bir diğer ifadeyle güven aralığını temsil etmektedir. ε ifadesi tolere edilebilir hata miktarını simgelemektedir. Yukarıda da açıklandığı üzere tanımlanan aralık içerisinde ise İngilizce ifadesi ile cost ve/veya error olarak tanımlanan hata sıfıra eşit olmaktadır. ε ifadesinin değeri destek vektör regresyonu için kullanılacak olan vektörlerin boyutunu doğrudan etkilemektedir [20]. Eğer hata oranı yeterince büyük seçilirse kullanılacak destek vektörlerinin boyutu küçük ve hesaplama zamanı o kadar az olacaktır. Hata oranı ne kadar küçük olursa hesaplama zamanı o kadar fazla kullanılan destek vektörlerinin boyutu o kadar büyük olacaktır, bu da hesaplama için kullanılacak olan bilgisayarın işlemcisinin güçlü olmasını gerektirir. C terimi İngilizce cost olarak ifade edilen deneysel hata ve ağırlık vektörü arasında dengelemeyi sağlayan bir sabittir. C teriminin belirli bir dereceye kadar artması deneysel hata terimi olarak tanımlanan ξ değerinin düşmesini sağlamaktadır. Yüksek doğruluklu destek vektör regresyonu modellemesi için bu değerlerin belirlenmesi önemlidir.

2.3.2.5 KNN (K-Nearest Neighbor) Algoritması

KNN İngilizce ifadesiyle K-Nearest Neighbor algoritması Türkçe en yakın K komşu olarak literatürde sıklıkla tercih edilen makine öğrenmesi algoritmalarından biridir. K-en yakın komşu algoritmasının temelleri 1950’li yıllara dayanmaktadır. K-en yakın komşu algoritması günümüzde görüntü işleme, biyoistatistik gibi birçok farklı alanda tercih edilmektedir [58].

K-en yakın komşu algoritması denetimli öğrenme algoritmaları arasında yapısı daha basit olup sistematiği kompleks değildir. KNN algoritması makine öğrenmesi algoritmalarından olup sınıflandırma problemlerinin çözümlerinde kullanılmaktadır. KNN algoritması eğitim veri setini işlerken bir verinin en yakın k tane çevresindeki diğer verileri bulur ve bu verileri işaretler. Bütün verilerin işaretleme işlemi bittikten sonra en yakın komşu noktaların numerik açıdan üstünlüğüne göre sınıflandırma yapmaktadır. K-en yakın komşu algoritmasının en yakın verileri etiketleme ve yapılan etiketleme neticesinde sınıflandırması bir genelleme kabulüdür. Bu sebeple literatürde K-en yakın komşu algoritması tembel öğrenci olarak da isimlendirilmiştir [59].

K-en yakın komşu algoritması belirtildiği üzere denetimli makine öğrenmesi algoritmasıdır. Bu sebeple parametrelerinin kullanıcı tarafından belirlenmesi gerekmektedir. Algoritma performansını belirleyen en önemli iki parametre bulunmaktadır. Bunlardan ilki belirlenmesi gereken k adet komşudur. K adet komşunun yakınlığının değerlendirilmesi sonucu sınıflandırma yapılacağı için bu veri sonucu diğer bir ifadeyle doğruluk oranını direkt olarak etkilemektedir. Belirlenecek k sayısı veriden veriye değişkenlik göstereceği için veri setini iyi tanımak bu noktada önemlidir. İkili sınıflandırma problemlerinde k muhakkak tek sayı olarak belirlenmelidir.

K-en yakın komşu algoritmasında başlangıç parametresi olarak tanımlanması gereken ve model doğruluğuna direkt olarak etki eden ikinci parametre ise işaretleme yaparken hesaplanan komşu uzunluklarının hesaplanma metodolojisidir. Literatürde sıklıkla kullanılan 3 mesafe hesaplama yöntemi vardır:

- Öklid mesafesi
- Manhattan mesafesi
- Minkowski mesafesi

N boyutlu bir uzayda X ve Y olarak tanımlanan verileri içeren bir sistem tanımlanırsa Öklid mesafesi şu şekilde hesaplanabilir:

$$d_{\text{öklit}} = \sqrt{\sum_{i=0}^n (X_i - Y_i)^2} \quad (3.21)$$

Manhattan mesafesi iki mesafe arasındaki farkın mutlak değeriyle hesaplanır ve denklem 3.22'deki gibi ifade edilebilir.

$$d_{\text{Manhattan}} = \sum_{i=0}^n |X_i - Y_i| \quad (3.22)$$

Minkowski ise yapmış olduğu çalışmada hesaplamaya p olarak ifade edilen $(0, \infty)$ arasında tanımlanan değişkeni ekleyerek denklem 3.23'teki gibi tanımlanmıştır:

$$d_{\text{Minkowski}} = (\sum_{i=0}^n |X_i - Y_i|^p)^{1/p} \quad (3.23)$$

Mesafe ölçüm metodolojisine göre modelin çıkış parametresi ve doğruluk oranı değişkenlik gösterir. Doğru hesaplama modeli ise kullanılan veri seti ve dağılımına göre değişkenlik göstermektedir.

2.3.2.6 Karar Ağaçları

Karar ağaçları denetimli makine öğrenmesi algoritmalarından olup sınıflandırma problemlerinin çözümlerinde kullanılan algoritmalarından biridir. Karar ağaçları algoritmalarının temeli Morgan ve Sanguist tarafından 1963 yılında yapılan çalışmaya dayanmaktadır. Bu çalışma regresyon ağaçları modelinin temelini oluşturmuştur. İstatikçi olan Breiman ve arkadaşları tarafından 1984 yılında yapılan çalışmada ise veriye göre karar ağaç metodolojisini özetleyen ve belirleyen bir çalışma yapılmıştır [60].

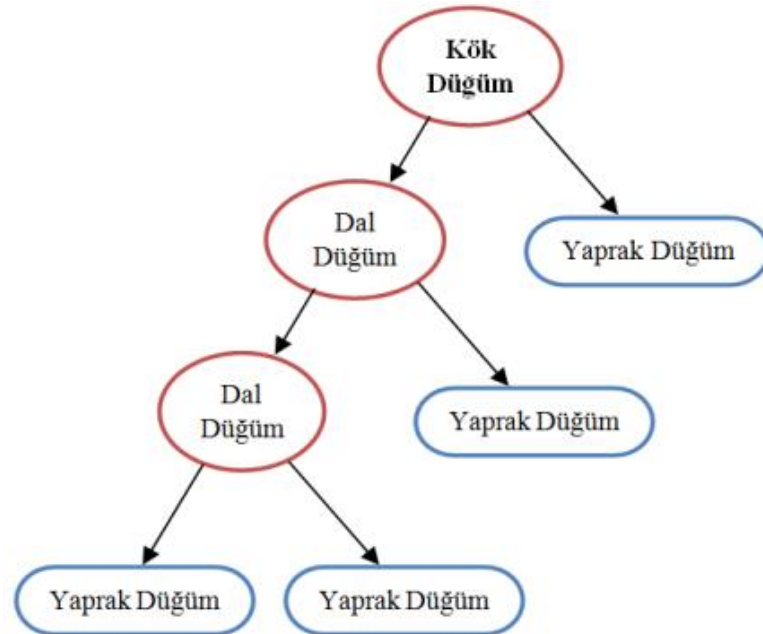
Karar ağaçları algoritmaları genelden özele doğru belirli kurallar çerçevesinde veriyi sınıflandırarak karar veren bir tür makine öğrenmesi algoritmasıdır. Karar ağacı algoritmalarının temeli büyük küme barındıran veriyi belirli kurallar doğrultusunda alt kümelerle ayırmaktır. Alt kümeler de daha sonra yine kurallarla kümelerle ayrılır. Böylece yeni tanımlanan bir x parametresinin karar ağacı kuralları uygulanarak hangi sınıfa ait olduğu bulunabilir.

Karar ağaçları olarak sınıflandırılacak olan veriye göre karar ağacı yapısı değişkenlik göstermektedir. Kullanılan veri setinin özelliği nitelik bir veriyi içeriyor ve çıktı

kategorik olacak ise sınıflandırma ağacı veya İngilizce ifadesiyle classification tree olarak isimlendirilir. Kullanılan veri setinin nicelik, sayısal bir problem özelinde yapılacaksa regresyon ağacı veya İngilizce ifadesiyle regression tree olarak isimlendirilir [61]. Regresyon ve sınıflandırma ağaçları için kullanılan karar ağaçları temel prensipleri aynıdır [62].

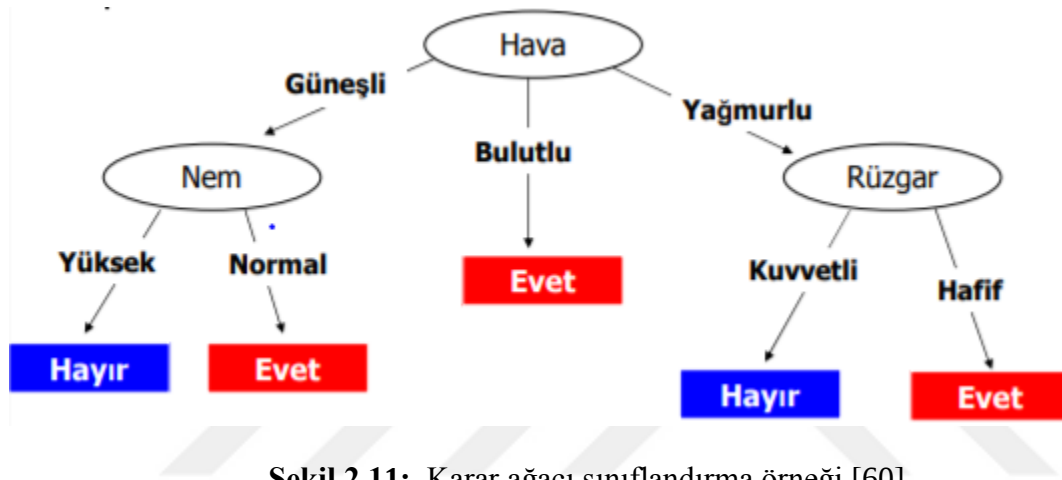
Karar ağaçları algoritmaları düğüm, yaprak, dal ve kök yapılarından oluşur. Karar ağaçlarında bulunan değişkenler düğümlerle temsil edilir. Yapılacak olan sınıflandırmanın alt tabakasından üst tabakasına doğru yaprak düğüm, dal düğüm ve kök düğüm yapısı bulunur. Kök düğüm karar ağaçları algoritmasında kullanılacak olan veriyi temsil etmektedir.

Kök ve yaprak düğümler arasında kalan kısımlar ise dal olarak ifade edilir. Diğer bir ifadeyle ağacın yapısı; verileri temsil eden bir kök düğüm, homojen olmayan dal düğümler ve homojen olan yaprak (terminal) düğümlerden oluşur. Karar ağacı, kök düğümünden başlayarak devam eden ve deney ünitelerine uygulanan evet-hayır gibi cevaplara göre oluşan yollardan oluşur. Genelden özele doğru kök düğümlerden sonra dal düğümler bulunur. Dal düğümleri koşulların tanımlandığı verilerin filtrelendiği kısımlar olarak tanımlanabilir. Dal düğümlerinden sonra ise yaprak düğümleri gelmektedir. Yaprak düğümleri belirlenen kurallar sonucunda sınıflandırılmış veya diğer bir ifadeyle ayrıştırılmış veriyi ifade etmektedir. Bir karar ağacı mekanizması genel şeması Şekil 2.10'da görülmektedir [63].



Şekil 2.10: Karar ağacı mekanizması genel şeması [63].

Bir karar ağacı sınıflandırma örneği Şekil 2.11’de görülebilmektedir [64]. Örnekte hava durumuna göre dışarıya çıkma ve çıkmama kararı evet veya hayır olarak sınıflandırılmıştır. Şekil 2.11’de de açıkça görülebileceği üzere hava kök düğüm olarak tanımlanmıştır. Hava durumunun alacağı olasılıklar öncelikle güneşli, bulutlu ve rüzgârlı olarak sınıflandırılmıştır ve bunlar dal düğüm değerleridir. 2.dal düğüm değerleri ise nem ve rüzgâr değişkenleridir. Hava durumunun güneşli ve nem durumunun yüksek olduğu veya yağmurlu ve rüzgârın kuvvetli olduğu durumlarda dışarı çıkma kararı hayır olarak belirlenirken diğer durumlar için evet olarak sınıflandırılmıştır.



Şekil 2.11: Karar ağacı sınıflandırma örneği [60].

Karar ağaçları diğer makine öğrenmesi algoritmalarına benzer şekilde çalışır. İlk aşama eğitim olarak isimlendirilen İngilizce training olarak literatürde geçen verilerin eğitilmesi aşamasıdır. Verilerin eğitilmesi aşamasında önceki paragraflarda belirtilen dal düğümleri algoritmanın doğruluk oranını direkt etkilendiğinden önem arz etmektedir. Karar ağaçları için literatürde birçok farklı algoritma geliştirilmiştir [63]. Literatürde kullanılan algoritmaların bazıları aşağıdaki şekildedir:

- CART (classification and regression trees) [64]
- CHAID (Chi-squared automatic interaction detector) [65]
- Ayrıntılı (exhaustive) CHAID (Bigss ve ark., 1991) [66]
- QUEST (quick, unbiased, efficient statistical trees) [67]
- Rastgele Orman (Random Forest) [68]
- M5 (Learning with continuous class) karar ağacı [69]

Karar ağaçları algoritmasında dal düğümleri belirli kuralları içerir ve bu kurallar entropiye dayanmaktadır. Entropi en genel manasıyla belirsizlik olarak tanımlanabilir. Bir veri setinde bütün veriler bir sınıf olarak ayrıştırılabiliyorsa o veri seti için entropi sıfır değerini alır. Entropi değeri 0 ile 1 arasında tanımlanırsa, değer 0'a yakınsadıkça belirsizlik azalır. Değer 1'e yaklaştıkça belirsizlik artar. Karar ağaçları algoritmasında temel amaç yaprak düğümün entropi değerini sıfıra yakınlaştırmaktır.

V olarak ifade edilen bir veri setinde olasılık $P(p_1, p_2, \dots, p_i)$ olarak tanımlanırsa N tane veri içeren veri setinin i tane sınıfının E olarak tanımlanan entropi değeri denklem 3.24'teki gibi ifade edilebilir [70]:

$$E(V) = - \sum_{i=0}^N p_i \log_2 (p_i) \quad (3.24)$$

V olarak ifade edilen veri seti $V = \{a, a, a, a, b, b, b, b, b, b, b, b\}$ olarak tanımlanırsa bu veri setinde 4 adet a, 8 adet b ve 2 adet c bulunmaktadır. Bu veri seti için:

- $\{a\}$ elemanının bulunma olasılığı $4/12 = 0.33$
- $\{b\}$ elemanının bulunma olasılığı $8/12 = 0.66$

olarak hesaplanır. Bu elemanları için entropi hesabı ise aşağıda görüldüğü gibi yapılabilir.

$$E(V) = - \frac{4}{12} \log_2 \frac{4}{12} - - \frac{8}{12} \log_2 \frac{8}{12} = 0.92 \quad (3.25)$$

Görüldüğü üzere entropi değeri 1'e yakındır. Olayın gerçekleşme ihtimalinde belirsizlik yüksektir.

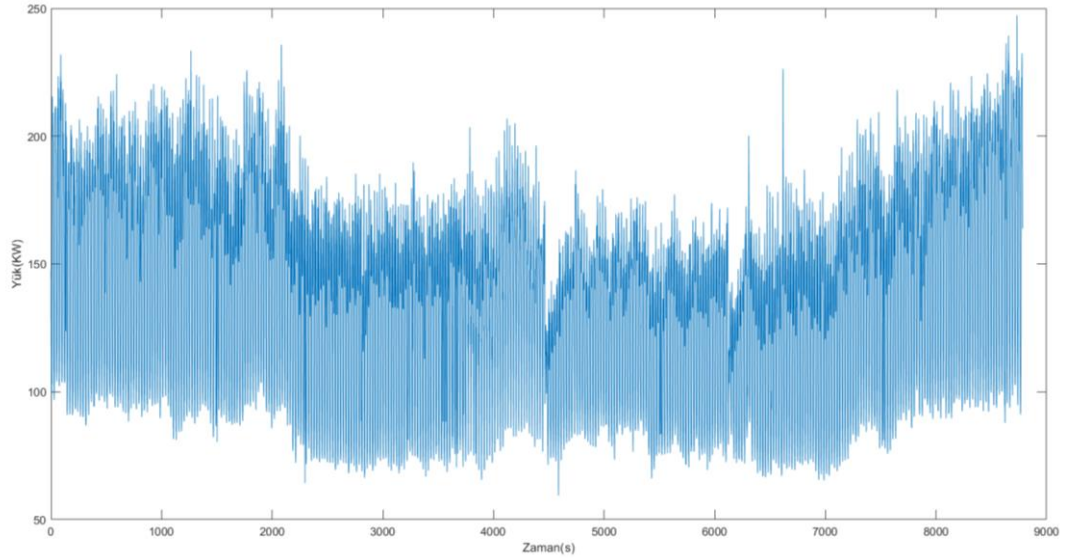


3. YÜK TAHMİNİ UYGULAMASI

3.1 Veri Seti

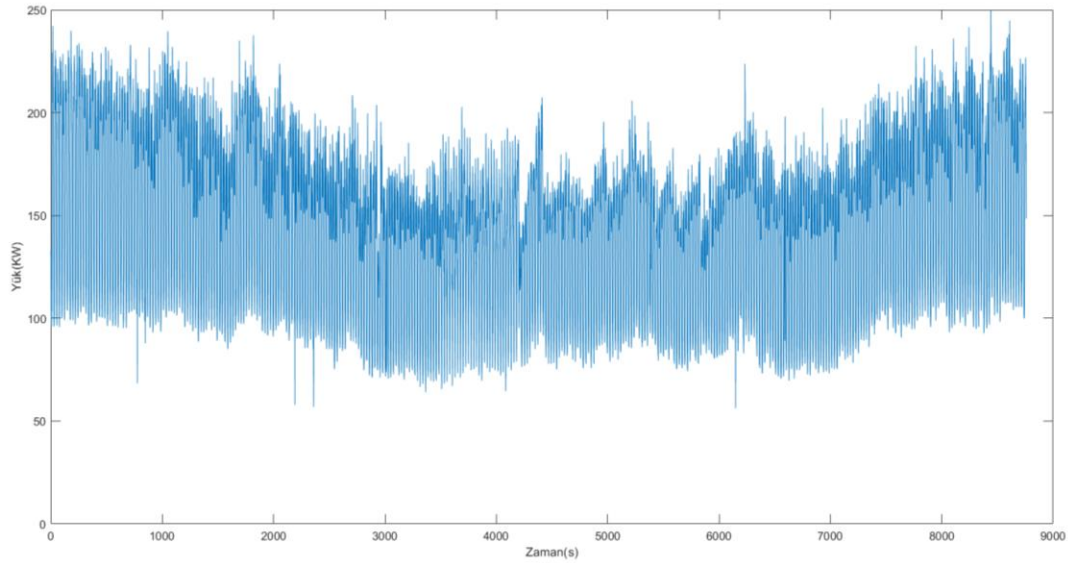
Yük tahmini için kullanılan veri seti Marmara bölgesinde bulunan bir il merkez bölgesinde bulunan dağıtım trafolarından alınan 2016-2017 yıllarına ait saatlik elektrik tüketim verileridir. Veriler dağıtım şirketi gerçek tüketici kullanıcı verileridir. Tüketim yükü ağırlık olarak mesken tüketim yük verilerini içermektedir. Tüketim verileri saatlik olarak kWh cinsinden alınmıştır.

2016 yılı verileri içeren saatlik yük dağılımı Şekil 3.1’de görülebilmektedir. 2016 yılı için saatlik yük tüketim ortalaması 138,81 kWh olup 2016 yılı içerisinde toplam elektrik yük tüketimi 1219,329 MWh olarak gerçekleşmiştir. 2016 yılı yük tüketim verisi 8784 satır-artık yıldan dolayı- saatlik yük tüketimini içermektedir.



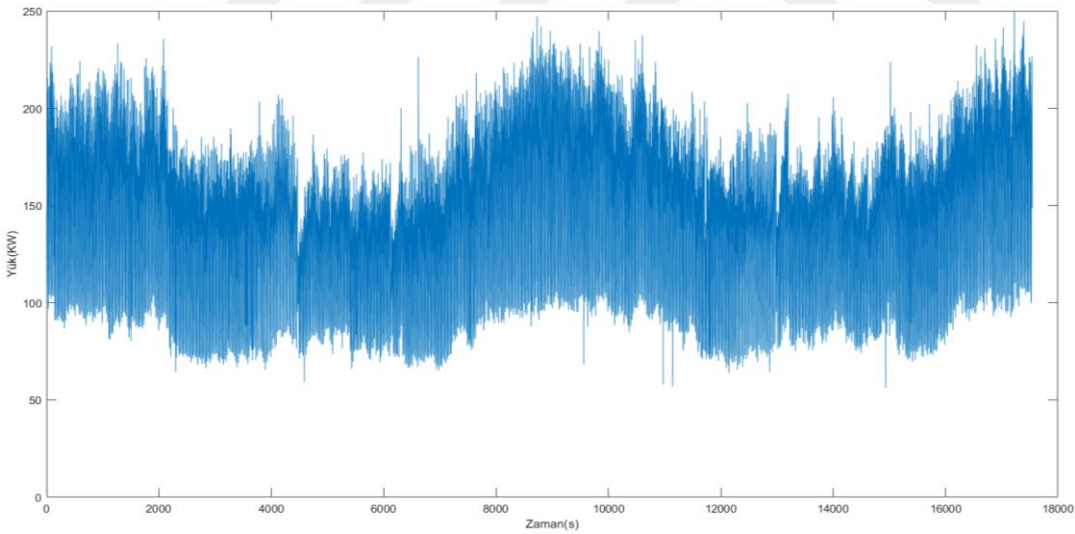
Şekil 3.1 : 2016 yılı yük tüketim verisi (kWh)

2017 yılı verileri içeren saatlik yük dağılımı Şekil 3.2’de görülebilmektedir. 2017 yılı için saatlik yük tüketim ortalaması 146,53 kWh olup 2017 yılı içerisinde toplam elektrik yük tüketimi 1283,639 MWh olarak gerçekleşmiştir. Saatlik ve yıllık elektrik tüketimde 2016 yılına kıyasla 2017 yılında %5,27’lik bir artış olduğu yorumu yapılabilir. 2017 yılı tüketim verisi 8760 satır saatlik yük tüketimini içermektedir.



Şekil 3.2 : 2017 yılı yük tüketim verisi (kWh)

Şekil 3.3'te 2016 ve 2017 yılları birlikte yük tüketim verileri görülmektedir. Şekil 3.1'den de görülebileceği üzere 2016 ve 2017 yılları yük tüketim verileri trendinin benzerlik gösterdiği yorumu yapılabilir. Yük tüketim verisi incelendiğinde tüketim verisinin mevsimsellik trendi içerdiği yorumu yapılabilir.

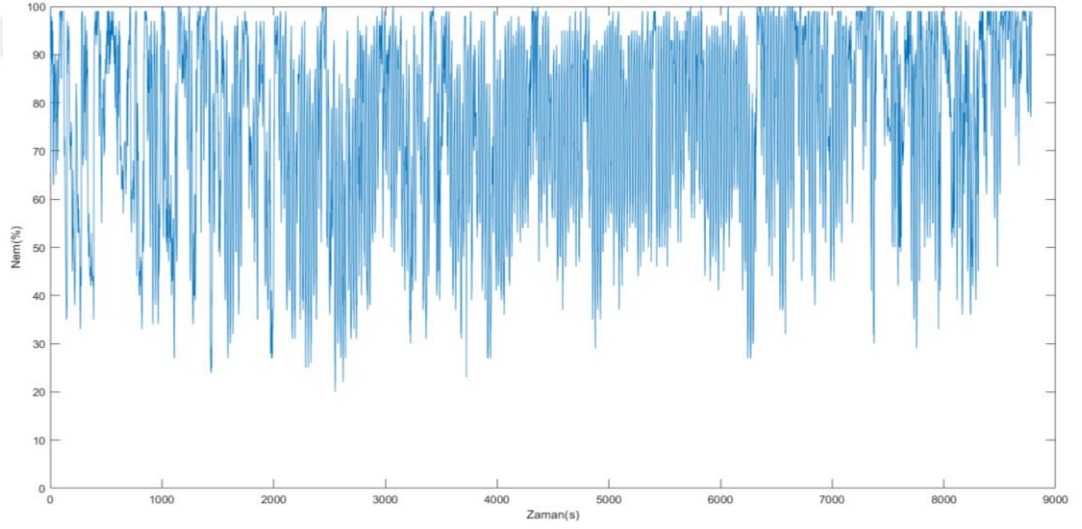


Şekil 3.3 : 2016-2017 yılları yük tüketim verisi (kWh)

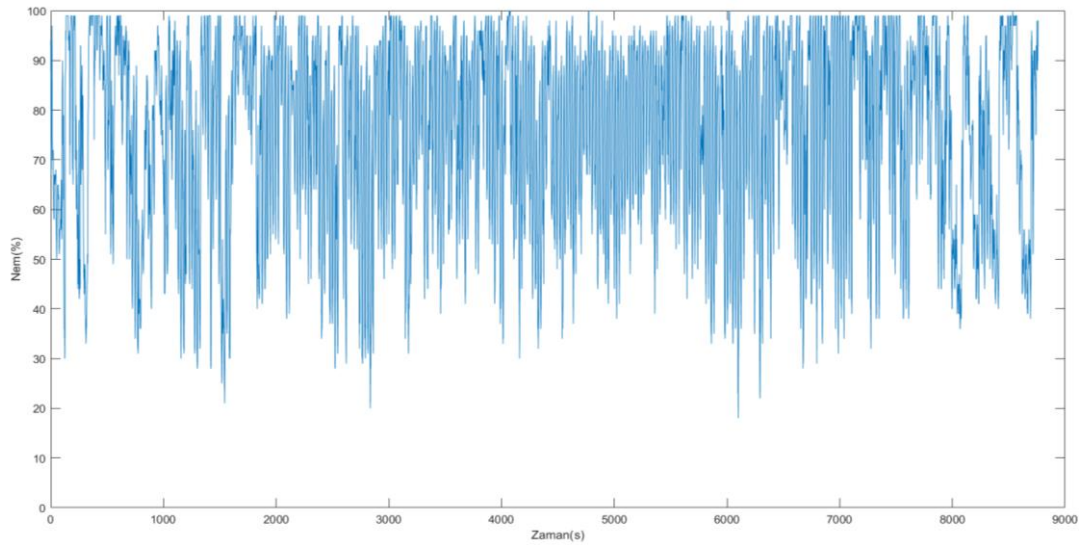
Şekil 3.3'ten de açıkça görülebileceği üzere veri seti kendi içerisinde bir trend barındırmaktadır fakat veri setinin içinde beyaz gürültü olarak tanımlanan veya literatürde rastlantısal faktörler olarak isimlendirilen trend dışı hareketleri de gözlemlemek mümkündür. 2016-2017 yük verisi 17544 satır veri içermektedir. 2016-2017 yılları için saatlik ortalama 142,51 kWh olup toplam tüketim verisi 2504,475 MWh'tır.

Önceki bölümlerde bahsedildiği üzere kısa dönem yük tahmininde meteorolojik veriler kullanılmaktadır. Bu çalışma kapsamında tahmin edilecek il için meteorolojik veriler Meteoroloji Genel Müdürlüğüne ait Meteorolojik Veri Bilgi Sunum ve Satış Sistemi (MEVBİS) vasıtasıyla İl merkezi bölgesi için 2016-2017 saatlik sıcaklık ve nem bilgileri alınmıştır.

Merkez istasyonundan alınan 2016 yılı nispi nem değeri % cinsinden şekil 3.4'te görülmektedir. 2016 yılı için ortalama nispi nem miktarı %75,96 olarak gerçekleşmiştir. 2017 yılı tahmin ili merkezi MEVBİS Merkez İstasyonu'ndan alınan nem değeri grafiği Şekil 3.5'te görülebilmektedir. 2017 yılı ortalama nem miktarı %76 olarak gerçekleşmiştir

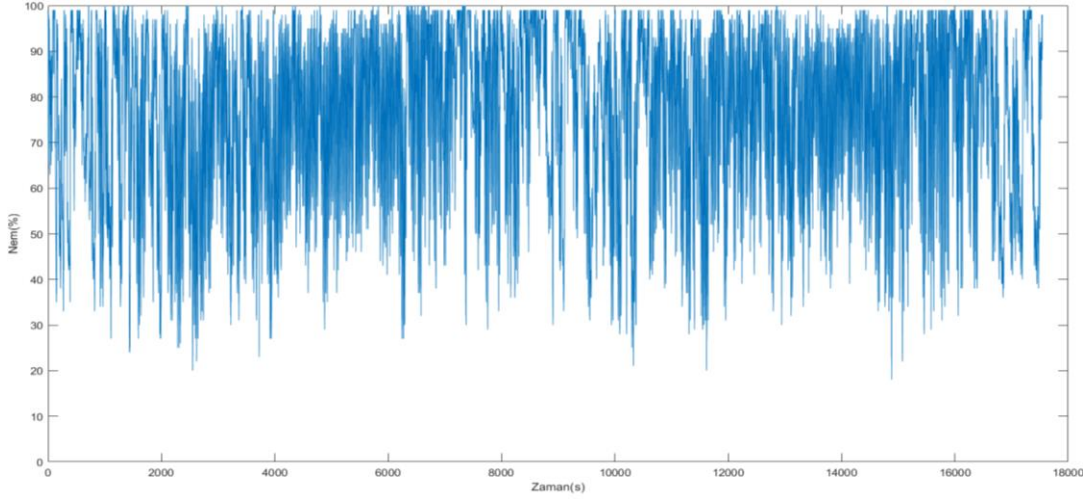


Şekil 3.4 : 2016 yılı nispi nem verisi (%)



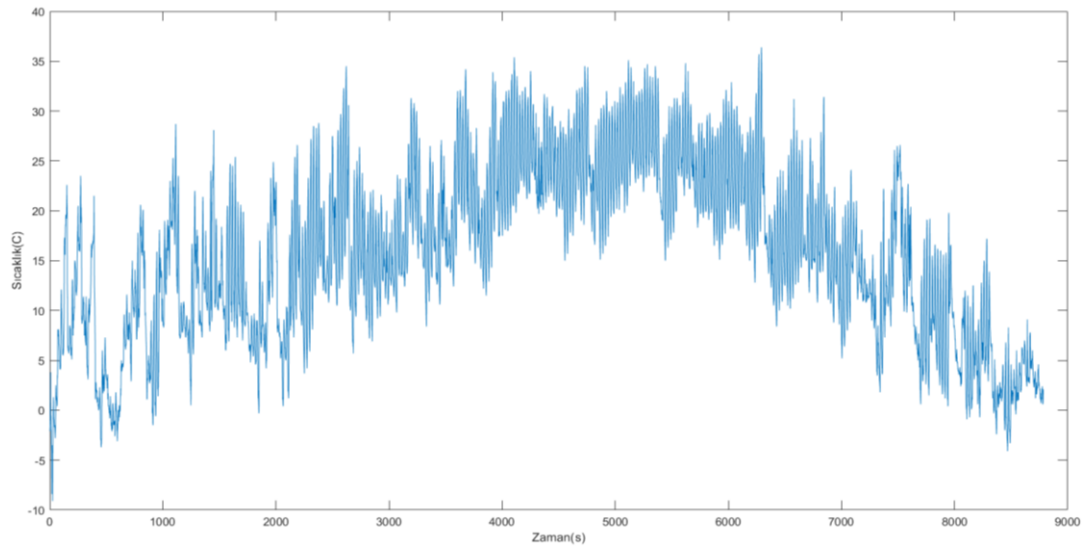
Şekil 3.5 : 2017 yılı nispi nem verisi (%)

Şekil 3.6’te 2016-2017 yılları için nispi nem miktarı (%) görülmektedir. Şekilden de anlaşılabileceği üzere nispi nem miktarı verisi trend içermemektedir, bir diğer ifadeyle nispi nem verisinin belirli bir karakteristiği yoktur. 2016 yılı için minimum nispi nem miktarı %20 olarak gerçekleşmişken, 2017 yılı için minimum nispi nem miktarı %18 olarak gerçekleşmiştir. 2016 ve 2017 yılları için maksimum nispi nem miktarı %100 olarak gerçekleşmiştir. 2016 ve 2017 yılı için totalde 17544 satır nispi nem miktarı (%) veri içermektedir.



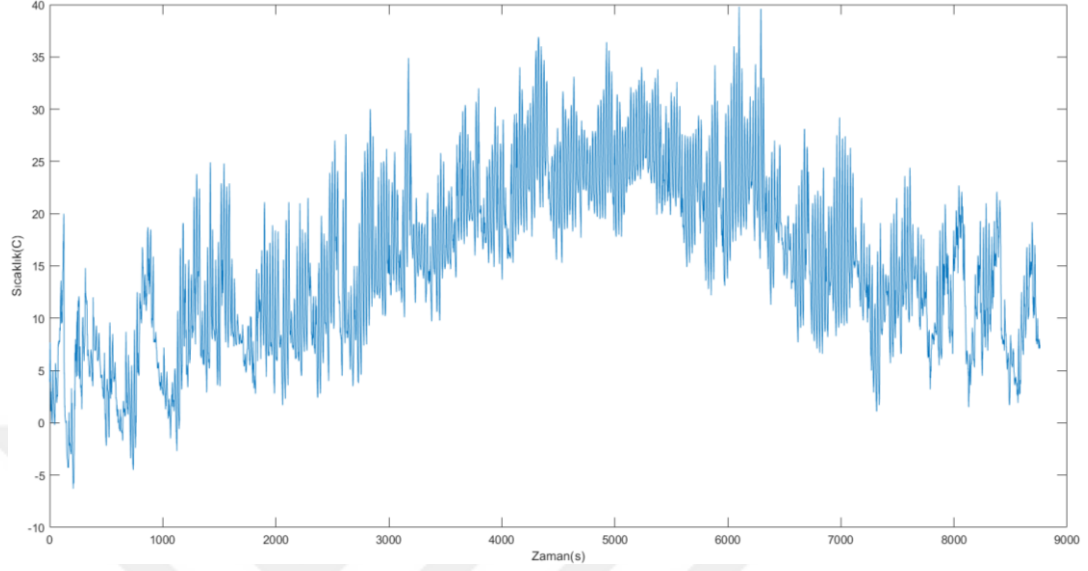
Şekil 3.6 : 2016-2017 yılı nispi nem verisi (%)

Yük tahmini algoritmaları için MEVBIS sisteminden sıcaklık verisi de alınmıştır. Tahmin ili MEVBIS Merkez istasyonundan alınan 2016 yılı sıcaklık verileri Şekil 3.7’de görülebilmektedir. 2016 yılı için saatlik sıcaklık minimum -9.1 °C, maksimum 36.4 °C ve ortalama sıcaklığı 15.8 Celcius olarak gerçekleşmiştir.



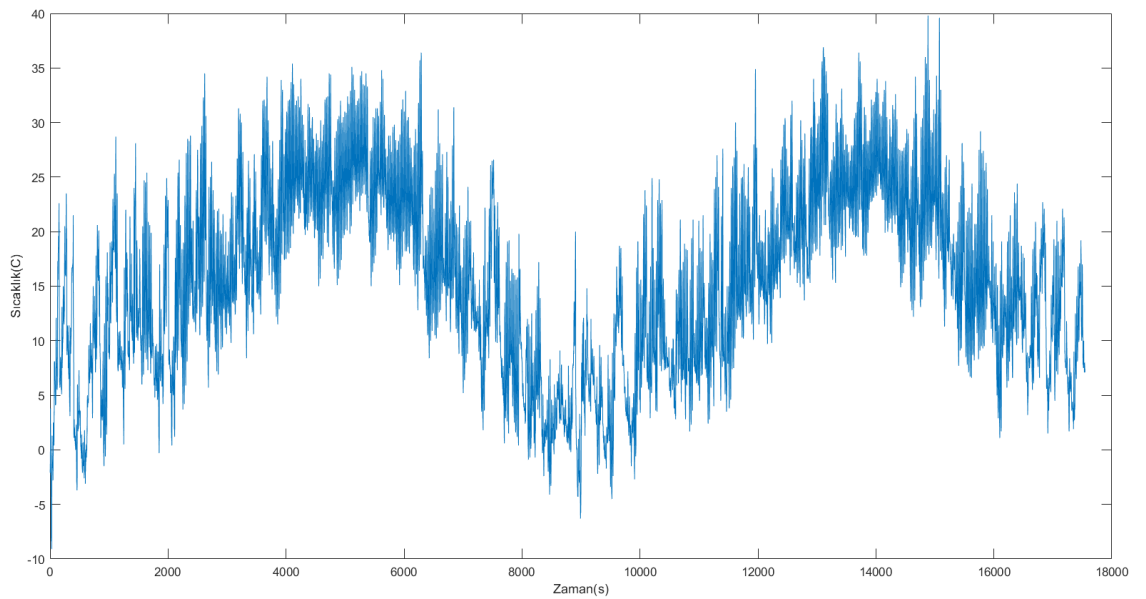
Şekil 3.7 : 2016 yılı sıcaklık verisi (°C)

2017 yılı sıcaklık verileri Şekil 3.8’de görülebilmektedir. 2017 yılı için saatlik sıcaklık minimum -6.3 °C, maksimum 39.8 °C ve ortalama sıcaklığı 15.31 Celcius olarak gerçekleşmiştir



Şekil 3.8 : 2017 yılı sıcaklık verisi (°C)

Şekil 3.9’da 2016 ve 2017 yılları için sıcaklık verisi görülmektedir. Şekilden de açıkça görülebileceği üzere 2016 ve 2017 yılı için sıcaklık verileri benzerlik göstermektedir. 2016 ve 2017 yılları birlikte incelendiği zaman sıcaklık verisinin mevsimsel trend içerdiği ve sıcaklık grafiğinin kendi içerisinde bir karakteristiği olduğu yorumu yapılabilir. Grafikte görülen uç trend dışı verilere bakılarak 2016 ve 2017 yılı için sıcaklık verisinin gürültü içerdiği yorumu yapılabilir.



Şekil 3.9 : 2016-2017 yılı sıcaklık verisi (°C)

3.2 Tahmin Programları

Tahmin programları için EViews 10, WEKA 3.8 ve MATLAB R20107B kullanılmıştır. Kurulan algoritmalar i7 işlemcili 6500 serili 12 GB RAM'e sahip bir bilgisayarla çözülmüştür.

Bu çalışma kapsamında kullanılan analiz programlarından biri EViews isimli analiz programıdır. EViews program giriş ekranı Şekil 3.10'da görülmektedir. EViews programının ilk sürümü 1994 yılında piyasaya sürülmüştür. EViews programı ekonometrik analiz yapmak için genellikle iktisat çalışmaları için kullanılan bir programdır. Bu çalışma kapsamında EViews 10 sürümüne sahip paket program tercih edilmiştir [71].

Tahmin çalışması için kullanılan analiz programlarından biri WEKA isimli programdır. WEKA Waikato Üniversite tarafından geliştirilmiş Waikato Environment for Knowledge Analysis kelimelerinin baş harflerinin kısaltmasından oluşan bir paket programdır. WEKA program java dili altyapısı kullanılarak yazılan bir programdır. WEKA programı veri madenciliği için günümüzde sıklıkla tercih edilen bir paket programdır. WEKA kullanılarak sınıflandırma, bölütleme ve ilişkilendirme işlemleri yapılabilmektedir. WEKA açık kaynak kodlu bir yazılımdır. Bu çalışmada WEKA 3.8.3 sürümü kullanılmıştır [72].

Bu çalışmada matematiksel hesaplar ve program çıktıları sonucu grafikler MATLAB adlı program kullanılarak çizdirilmiştir. MATLAB Matrix Laboratory kelimelerinin kısaltılmışı olan bir programdır. MATLAB birçok mühendislik uygulamaları için kullanılan kullanım yelpazesi çok geniş olan bir programdır. Bu çalışmada MATLAB R2017B sürümü kullanılmıştır [73].

3.3 Tahmin Algoritmaları ve Metodolojileri

Bu çalışmada 2016-2017 tahmin ili merkez dağıtım trafosu saatlik verileri ve MGM verileri kullanılarak saatlik yük tahmini yapılmıştır. Verilerin %75'i eğitim verisi olarak kullanılırken %25'i test verisi olarak ayrılmıştır. Diğer bir ifadeyle 2016 tamamı ve 2017 yılı ilk 6 ayı eğitim verisi olarak kullanılmıştır. 2017 2. Altı ayı ise tahmin edilmiş ve doğruluğu kontrol edilmiştir. Tüm metodolojilerde ortak olarak mevsimsellik ve saat parametresi kullanılmıştır. Sadece ARIMA zaman serisi analizi

olduğu için bağımlı değişken kullanılmış olup diğer veriler kullanılmamıştır. Tahmin metodolojisi olarak

- ARIMA
- SVM
- SVR
- ANFIS
- Karar Ağaçları

algoritmaları kullanılmıştır.

Bu çalışmada ARIMA, SVM, SVR, ANFIS ve Karar Ağacı algoritmaları kullanılmıştır. Model karşılaştırmaları için MAPE (Mean Absoulte Percentage Error) değeri seçilmiştir. $\bar{Y}$ tahmin edilen değer veya diğer bir ifadeyle çıktı değeri, Y gerçek değer olarak ifade edilirse, MAPE değeri denklem 3.1'deki gibi hesaplanabilir:

$$MAPE = \frac{\%100}{N} \sum_{i=1}^N \left(\frac{|\bar{Y} - Y|}{Y} \right) \quad (3.1)$$

3.3.1 Tahmin Metodolojileri

Bu çalışmada 15 farklı model kullanılarak tahmin çalışması yapılmıştır. Hafta günleri 3 farklı yöntemle ayrılmış, giriş parametreleri ise 5 farklı veri kümesi şeklinde kullanılarak 15 farklı metodoloji ile 5 farklı tahmin algoritması kullanılarak tahmin gerçekleştirilmiştir. 15 farklı model ile 5 farklı tahmin algoritması ile 62 farklı sonuç karşılaştırılmıştır.

Çalışmada kullanılan modeller de temel olarak hafta günleri 3 farklı yöntem kullanılarak ayrılmıştır. Önceki bölümlerde de bahsedildiği üzere bağımsız değişken ve bağımlı değişkenler arasında pozitif ve/veya negatif ilişki ne kadar yüksekse model o kadar yüksek tahmin sonuçları vermiştir. Bu sebeple literatürde bazı çalışmalar hafta günleri arasında herhangi bir ayırım yapmadan, bazı çalışmalar hafta içi ve hafta sonu olarak ayırım yapıp iki farklı eğitim ve test verisi kullanılarak, bazı modeller ise hafta içi ve hafta sonu modellerini ayrı eğitim ve test verisi olarak kullanmamak için giriş parametrelerine hafta içi ve hafta sonu ayırımı yapmak için kukla değişken olarak isimlendirilen bağımsız değişken kullanılarak tahmin işlemi gerçekleştirilmiştir. Bu çalışmada 3 farklı gün ayırımı kullanılmıştır, kullanılan gün ayırımları hafta içi ve hafta

sonu için gün ayrımı yapılmadan, gün ayrımı yapılarak ve hafta içi ve hafta sonu için kukla değişken kullanılarak yapılan tahmin olarak en genel haliyle ifade edilebilir. Bu çalışmada gün modelleri 5 farklı bağımsız değişken seti ile kullanılarak tahmin işlemi gerçekleştirilmiştir. Kullanılan bütün metodolojilerde kullanılan yük, sıcaklık veri seti gibi veriler aynı veri seti kullanılarak yapılmıştır. Kullanılan farklı metodolojilerin amacı gün ayrımı yapılmasının ve yapılan ayırmda kullanılan giriş parametrelerinin kullanılan yöntemlerle birlikte tahmin doğruluğunun üzerindeki etkisini araştırmaktır. Veri setinde görülebileceği üzere yük ve meteorolojik veriler mevsimsellik gösterdiği için mevsim parametresi kış, ilkbahar, yaz ve sonbahar sırasıyla (0, 1, 2, 3) dönüştürülerek tüm modeller için bağımsız değişken olarak kullanılmıştır. Bu çalışmada kısa dönem yük tahmini saatlik veya bir diğer ifadeyle bir saat arayla yapıldığı için kullanılan algoritmaların veri seti saatlik olarak tüm modellere tanıtılmıştır. Modellerde kullanılan meteorolojik verilerden nem ve sıcaklık verileri 1 saat önceki nem ve sıcaklık verileridir. Bu çalışmada kullanılan modellerde hafta günleri sınıflandırması ve kullanılan bağımsız değişkenler Çizelge 3.1’de özetle ifade edilmiştir. Çizelgede yer alan hafta günleri ayrımı G olarak ifade edilirse modeller en genel haliyle şu şekilde özetlenebilir:

- G1: Hafta günleri arasında herhangi bir ayırım yapılmamıştır.
- G2: Hafta günleri pazartesi, salı, çarşamba, perşembe, cuma hafta içi ve hafta sonu ayrımı yapılarak 2 ayrı test ve eğitim verisi kullanılarak tahmin yapılmıştır.
- G3: Hafta günleri pazartesi, salı, çarşamba, perşembe, cuma hafta içi ve hafta sonu ayrımı yapılarak bağımsız değişken eklenerek kukla değişken kullanılmıştır.

Tahmin modellerinde kullanılan ve çizelgede yer alan bağımsız değişkenler BD olarak ifade edilirse kullanılan bağımsız değişkenler şu şekildedir:

- BD1: Nem, sıcaklık ve bir saat önceki tüketim yükü
- BD2: Nem, sıcaklık, yirmi dört saat veya diğer ifadeyle bir gün önceki tüketim yükü
- BD3: Nem, sıcaklık, bir saat önceki ve bir gün önceki tüketim yükü
- BD4: Sıcaklık, bir saat önceki tüketim yükü ve bir gün önceki tüketim yükü

- BD5: Bir saat önceki tüketim yükü ve bir önceki tüketim yükü

Çizelge 3.1: Tahmin modelleri yapısı

	G1	G2	G3
BD1	Model-1	Model-6	Model-11
BD2	Model-2	Model-7	Model-12
BD3	Model-3	Model-8	Model-13
BD4	Model-4	Model-9	Model-14
BD5	Model-5	Model-10	Model-15

Çizelge 3.1’de görülebileceği üzere kurulan modellerde yapılan hafta günleri ayrımları sütunlarda, kullanılan bağımsız değişkenler ise satırlar ile ayrılmıştır. Çizelge 3.1’den de anlaşılacağı üzere Model-1’de bağımsız değişken olarak BD-1 yani nem, sıcaklık, bir saat önceki tüketim yükü hafta günleri ayrımı yapılmadan kullanılarak tahmin işlemi gerçekleştirilmiştir. Çizelge 3.1’de den de anlaşılacağı üzere 15 farklı Model bu çalışmada kullanılmıştır. Tüm modellerde mevsimsellik ve saatlik veri olarak tanıtılmıştır. Bu sebeple aşağıda yer alan model ayrıntılarında yer verilmemiştir.

3.3.1.1 Model-1

Model-1’de elektrik yük tahmini için kullanılan veri seti hafta içi ve hafta sonu olarak ayrımı yapılmadan veri seti bütün olarak değerlendirmeye alınmıştır. Kullanılan yapay zekâ yöntemleri için bağımsız değişken olarak mevsim, nem sıcaklık ve bir saat önceki yük parametreleri seçilmişken tahmin edilecek saat yükü bağımlı değişken olarak seçilmiştir.

3.3.1.2 Model-2

Model-2’de veri setinde Model-1’de olduğu gibi hafta içi hafta sonu gibi herhangi bir ayırım yapılmamıştır. Model-2’de bağımsız değişken olarak mevsim, nem, sıcaklık ve 24 saat önceki yük parametresi yani bir gün önceki yük parametresi kullanılarak tahmin yapılmıştır. Model-2’nin amacı kullanılan önceki yük parametrelerinden bir saat önceki yük ve bir gün yani 24 saat önceki yük parametrelerinin tahmin doğruluğuna etkisinin karşılaştırılmasıdır.

3.3.1.3 Model-3

Model-3’te veri setinde Model-1’de olduğu gibi hafta içi hafta sonu gibi herhangi bir ayırım yapılmamıştır. Model-3’te bağımsız değişken olarak mevsim, nem, sıcaklık ve

bir saat önceki yük parametresi ve 24 saat önceki yük parametresi yani bir gün önceki yük parametresi kullanılarak tahmin yapılmıştır. Model-3'te önceki yük parametrelerinin her ikisi de kullanılarak tahmin doğruluğu araştırılmıştır.

3.3.1.4 Model-4

Model-4'te veri setinde Model-1'de olduğu gibi hafta içi hafta sonu gibi herhangi bir ayırım yapılmamıştır. Model-3'ten farklı olarak giriş parametrelerinden nem değeri çıkarılmıştır. Model 4'te bağımsız değişken olarak mevsim, sıcaklık ve bir saat önceki yük parametresi ve 24 saat önceki yük parametresi yani bir gün önceki yük parametresi kullanılarak tahmin yapılmıştır. Model-4'te meteorolojik verilerden nem çıkartılarak nem verisinin tahmin sonuçlarına etkisi araştırılmıştır.

3.3.1.5 Model-5

Model-5'te veri setinde Model-1'de olduğu gibi hafta içi hafta sonu gibi herhangi bir ayırım yapılmamıştır. Model-5'te Model-(1...4)'ten farklı olarak giriş parametrelerinden meteorolojik veriler çıkarılmıştır. Bu modelde amaç meteorolojik veriler kullanılmadan yapılan tahminin doğruluğunun araştırılmasıdır. Model 5'te giriş parametresi olarak sadece önceki yük değerleri kullanılmıştır. ARIMA modelinde giriş parametresi olmadığından ve veri seti bütün olarak kullanıldığından dolayı Model-5 kapsamında değerlendirilmiştir. Model-5'te kullanılan yapay zekâ algoritmalarında ise bağımsız değişken olarak mevsim, sıcaklık ve bir saat önceki yük parametresi ve 24 saat önceki yük parametresi yani bir gün önceki yük parametresi kullanılarak tahmin yapılmıştır.

3.3.1.6 Model-6

Model-6'da hafta günleri ayırımı yapılmıştır. Hafta günleri pazartesi, salı, çarşamba, perşembe ve cuma günleri hafta içi olarak; cumartesi ve pazar günleri ise hafta sonu olarak ayrılmıştır. Hafta içi ve hafta sonu olarak 2 ayrı eğitim ve test verisi kullanılarak tahmin işlemi gerçekleştirilmiştir. Bağımsız değişken olarak Model-6'da nem, sıcaklık ve bir saat önceki tüketim yükü kullanılırken t anındaki tüketim yükü tahmin edilmiştir

3.3.1.7 Model-7

Model-7'de Model-6'da olduğu gibi aynı şekilde hafta günleri ayırımı yapılmıştır. Hafta günleri pazartesi, salı, çarşamba, perşembe ve cuma günleri hafta içi

olarak; cumartesi ve pazar günleri ise hafta sonu olarak ayrılmıştır. Hafta içi ve hafta sonu olarak 2 ayrı eğitim ve test verisi kullanılmıştır. Giriş parametresi olarak nem, sıcaklık, bir gün önceki tüketim yükü kullanılırken t anındaki tüketim yükü tahmin edilmiştir.

3.3.1.8 Model-8

Model-8’de Model-6 ve Model-7’de olduğu gibi veri seti günleri pazartesi, salı, çarşamba, perşembe ve cuma günleri hafta içi olarak ayrılmıştır. Cumartesi ve pazar günleri ise hafta sonu olarak ayrılmıştır. Hafta içi ve hafta sonu olarak ayrılan veri setleri iki ayrı küme halinde eğitim ve test verisi olarak kullanılmıştır.

Kullanılan yapay zekâ yöntemleri için giriş parametresi veya diğer ifadeyle bağımsız değişken olarak nem, sıcaklık, bir saat önceki yük ve bir gün önceki yük parametresi kullanılarak hafta içi ve hafta sonu ayrımı yapılarak iki ayrı eğitim ve iki ayrı test kümesi oluşturularak tahmin yapılmıştır.

3.3.1.9 Model-9

Model-9’de Model-6’da olduğu gibi aynı şekilde hafta günleri ayrımı yapılmıştır. Hafta günleri günleri pazartesi, salı, çarşamba, perşembe ve cuma günleri hafta içi olarak; cumartesi ve pazar günleri ise hafta sonu olarak ayrılmıştır. Hafta içi ve hafta sonu olarak 2 ayrı eğitim ve test verisi kullanılmıştır. Giriş parametresi olarak sıcaklık, bir saat önceki ve bir gün önceki tüketim yükü kullanılırken t anındaki tüketim yükü tahmin edilmiştir.

3.3.1.10 Model-10

Model-10’da Model-6 da olduğu gibi hafta günleri aynı şekilde hafta içi ve hafta sonu olarak 2 ayrı eğitim ve test verisi kullanılmıştır.

ARIMA yöntemi için giriş parametresi olarak yük harici bir parametre kullanılmadığından ve zaman serisi analizi yöntemlerinden olduğu için tahmin edilecek yük hafta içi ve hafta sonu olarak ayrı olarak kullanılmış ve tahmin gerçekleştirilmiştir. ARIMA modelinde hafta içi ve hafta sonu verisi ayrı, bağımsız veri seti gibi kullanılmıştır. Bu sebeple Model-10 kapsamında incelenmiştir. Kullanılan yapay zekâ yöntemlerinde ise giriş parametresi olarak nem, sıcaklık, bir gün önceki tüketim yükü kullanılırken t anındaki tüketim yükü tahmin edilmiştir.

3.3.1.11 Model-11

Model-11’de hafta günleri ayrımı yapılmıştır fakat şu ana kadar kadar bahsedilen modellerden farklı olarak bu ayrım giriş parametrelerine kukla değişken olarak kullanılarak tek eğitim ve test veri seti kullanılmıştır. Hafta içi modelleri 0 olarak tanımlanırken hafta sonu modellerine 1 değişkeni kukla değişken olarak atanmıştır. Model 11’de bağımsız değişken seti 1 kullanılarak nem, sıcaklık ve bir saat önceki tüketim yükü kullanılırken t anındaki tüketim yükü tahmin edilmiştir.

3.3.1.12 Model-12

Model-10 ve Model-11’de olduğu gibi hafta günü ayrımı kukla değişken kullanılarak tahmin işlemi gerçekleştirilmiştir. Model-12’de bağımsız değişken seti olarak nem, sıcaklık ve bir gün önceki yük kullanılarak t anındaki yük tahmin edilmiştir. Bu modelde kukla değişken kullanılarak yapılan gün ayrımında bir saat önceki yük ve 24 saat önceki yük kullanımının tahmin sonucu üzerindeki etkisi karşılaştırılmıştır.

3.3.1.13 Model-13

Model-13’te Model-11 gibi hafta günleri ayrımı yapılmıştır. Bağımsız değişkenler veya diğer ifadeyle giriş parametreleri olarak nem, sıcaklık, bir önceki saat yükü ve 24 saat önceki elektrik yükü seçilirken o saatteki elektrik yükü tahmin edilmiştir.

3.3.1.14 Model-14

Model-14’te kullanılan sınıflandırma yapısı Model-10 ile birebir aynıdır. Hafta içi hafta sonu günleri ayrımı kukla değişken kullanılarak yapılmıştır ve tek bir eğitim ve test verisi kullanılarak tahmin işlemi gerçekleştirilmiştir. Model-14’te Model-10’dan farklı olarak bağımsız değişkenlerden nem parametresi çıkarılmıştır. Bağımsız değişken olarak sıcaklık, bir önceki elektrik tüketim değeri ile bir gün yani 24 saat önceki elektrik tüketim değeri kullanılmıştır.

3.3.1.15 Model-15

Model-15’te Model-10...14’te olduğu gibi hafta içi hafta sonu ayrımı kukla değişkenler ile sağlanmıştır. Model-15’te Model-10...14’ten farklı olarak kullanılan tahmin algoritmalarında tahmin edilecek yük için meteorolojik veriler kullanılmamıştır. Bağımsız değişken olarak mevsim, bir saat önceki yük parametresi ve 24 saat önceki yük parametresi kullanılmış ve tahmin yapılmıştır. Bu modelde amaç

meteorolojik veriler bağımsız değişkenler arasından çıkarıldığı zaman sadece önceki yük parametreleri ile yapılan tahmin işleminin tahmin doğruluğu üzerindeki etkisini araştırmaktır.

3.3.2 Tahmin Algoritmaları

Bölüm 3.3.1’de dağıtım seviyesi için yapılan yük tahminlerinde kullanılan bağımsız değişkenler ve sınıflandırılması açıklanmıştır. Belirlenmiş olan 6 farklı metot baz alınarak 5 farklı yöntem ile tahmin yapılmıştır. Bu çalışma kapsamında ARIMA, destek vektör makineleri, destek vektör regresyonu, en yakın k komşu ve karar ağaçları regresyonu algoritmaları kullanılarak tahmin gerçekleştirilmiştir

3.3.2.1 ARIMA

ARIMA modeli önceki bölümlerde anlatıldığı üzere zaman serisi modellerinden birisidir. Bu çalışmada ARIMA modeli için EViews 10.0 programı kullanılmıştır. Durağanlık testi için Dickey-Fuller testi seçilmiştir. Veri setinin durağan olduğu sonucu modelin uygulanabilir olduğu sonucuna varılmıştır. Zaman serisi modellerinde AR ve MA parametreleri deneme yanılma yöntemleriyle seçilmektedir. Deneme sonucunda AR (1) AR (23) ve MA (23) köklerinin MAPE değerinin en düşük olduğunu sonucuna ulaşılmıştır.

3.3.2.2 Destek Vektör Makineleri

Destek vektör makineleri literatürde sıklıkla tercih edilen makine öğrenmesi algoritmalarından biridir. Bu çalışmada DVM modeli için WEKA kullanılmıştır. DVM için LIBSVM algoritması seçilmiştir. LIBSVM hazır kütüphanesi olan kullanımı pratik bir DVM algoritmasıdır. Çekirdek fonksiyonu RBF-Kernel seçilmiştir. Cost 10^6 , epsilon 0.5 ve gamma değeri 10^{-8} olarak belirlenmiştir. Cost, epsilon ve gamma değeri grid.py algoritması kullanılarak hesaplanmıştır.

3.3.2.3 Destek Vektör Regresyonu

Destek vektör regresyonu (DVR) veya İngilizce ifadesiyle Support Vector Machines (SVR) modeli destek vektörler makinelerini regresyon için kullanan hibrit makine öğrenmesi algoritmalarındandır. Bu çalışma da destek vektör regresyon yaklaşımı için için WEKA da bulunan SMOreg algoritması kullanılmıştır. SMOreg algoritması sıralı

minimum optimizasyon regresyonu olarak isimlendirebilir. C değeri 100 olarak gridsearch algoritması kullanılarak optimize edilmiştir.

3.3.2.4 Karar Ağaçları

Karar ağacı algoritması bahsedildiği üzere bağımsız değişkenleri kazanç oranına göre ayıran bu ayırma sonucu sınıflandıran veya regresyon uygulayan bir algoritmadır. Bu çalışmada karar ağacı yaklaşımı için WEKA kullanılmıştır. WEKA’da M5P olarak karar ağacı regresyonları olarak bilinen algoritma kullanılmıştır. Karar ağacı regresyonları düğümler oluşturan ve bu düğümlerin her birinde farklı regresyon denklemleri kuran bir yaklaşımdır.

3.3.2.5 KNN- En yakın K Komşu

En yakın K komşu algoritması tembel veya İngilizce ifadesiyle lazy algoritmalar olarak tanımlanan makine öğrenmesi algoritmalarındandır. En yakın K komşu algoritması için WEKA kullanılmıştır. WEKA’da IBK algoritması tercih edilmiştir. IBK algoritması ile 3, 5, 7, 9, 11 parametreler değerlendirilmiş ve en yüksek doğruluk sonucunun 3 yakın komşu olduğu sonucuna varılmıştır. Bu sebeple IBK3 kullanılarak en yakın 3 komşu ve öklit uzaklığı seçilerek tahmin yapılmıştır.

4. TAHMIN ÇIKTILARI VE KARŞILAŞTIRMALAR

Tezin bu bölümünde 2016-2017 yılı Marmara bölgesinde bulunan tahmin ili merkez bölgesi yük verileri ve meteorolojik veriler kullanılarak farklı tahmin yöntemleriyle yapılan tahminlerin sonuçları ve karşılaştırılmaları tartışılmıştır.

Çizelge 4.1 : Tahmin sonuçları karşılaştırılması.

Modeller	ARIMA	DVR	Karar ağacı	KNN	DVM
Model-1	-	%4,43	%4,86	%6,31	%7,01
Model-2	-	%5,86	%6,00	%7,47	%7,90
Model-3	-	%3,95	%4,21	%5,88	%6,41
Model-4	-	%3,96	%4,13	%5,33	%5,88
Model-5	%4,50	%3,57	%3,60	%4,69	%4,86
Model-6	-	%4,45	%4,78	%7,27	%8,85
Model-7	-	%5,79	%5,92	%8,22	%8,29
Model-8	-	%3,90	%4,11	%6,59	%6,79
Model-9	-	%3,97	%4,14	%5,62	%6,07
Model-10	%4,45	%3,88	%4,05	%4,88	%5,41
Model-11	-	%4,43	%4,95	%7,23	%7,61
Model-12	-	%5,80	%6,05	%8,35	%7,85
Model-13	-	%3,84	%4,12	%6,13	%6,37
Model-14	-	%3,82	%4,10	%5,33	%5,74
Model-15	-	%3,80	%4,05	%4,82	%5,21



5. SONUÇ VE ÖNERİLER

Elektrik enerjisinin güvenilirliği, kalitesi, ekonomikliği ve sürdürülebilirliği için elektrik yük tahmini önem arz etmektedir. Sürekli artan enerji ihtiyacının sürdürülebilir bir şekilde sağlanması açısından yük tahminin doğruluğu gün geçtikçe önem kazanmaktadır. Bu çalışmada Marmara bölgesinde bulunan bir ilde yer alan veriler kullanılarak saatlik elektrik yük tahmini çeşitli yöntemler kullanılarak yapılmıştır. Tahmin verisi olarak 2016-2017 yılları dağıtım trafosunun saatlik elektrik verisi kullanılmıştır. Eğitim verisi olarak bütün modellerde 1,5 yıllık veri kullanırken 2017 ikinci yarıyıl 6 aylık süre test verisi olarak kullanılmıştır.

Bu çalışmada gün modelleri 3 ayrı sınıfa ayrılarak, bu dağıtım bölgesi için gün modelleri ayrımının tahmin sonucu üzerine etkisi incelenmiştir. Bağımsız değişken olarak 5 farklı veri seti oluşturularak 3 ayrı sınıfa ayrılmış gün modeli ile 15 farklı tahmin modeli oluşturulmuştur. Oluşturulan 15 farklı tahmin modeli ARIMA, SVM, SVR, ANFIS ve Karar Ağaçları tahmin algoritmaları kullanılarak tahmin işlemi gerçekleştirilmiştir. Gerçekleştirilen bu tahminlerden çizelge 4.1’de görülen 62 farklı tahmin sonucu karşılaştırılmıştır. Kullanılan yöntemlerden MAPE oranı en düşük diğer bir ifadeyle gerçek değere en yakın sonucu %3,57 MAPE değeriyle SVR vermiştir.

Bu çalışmadan şu sonuçlar çıkarılabilir:

- Bir saat önceki yük kullanılarak oluşturulan model bir gün önceki yük kullanılarak oluşturulan modele göre doğruluk oranı daha yüksek sonuçlar vermiştir. Bir saat önceki tüketim yükü tahmin yük değerini daha fazla oranda etkilemektedir.
- Bir saat önceki ve bir gün önceki tüketim yükünün birlikte kullanıldığı modeller ayrı ayrı kullanıldığı modellere göre doğruluk oranı daha yüksek sonuçlar vermiştir. Bu çalışmada tahmin yapılacak bölgede sadece bir saat önceki yük veya 24 saat öncesi yükün kullanılması yüksek doğruluk oranı için yeterli değildir.

- Meteorolojik verilerden nem değerin kullanılması tahmin doğruluğunu negatif olarak etkilemiştir. Bu sebeple meteorolojik verilerin seçimi konusunda bölgenin özellikleri iyi tanınmalı bu şekilde belirlenmelidir.
- Tahmin doğruluğu en yüksek Model-5'tir. Model-5 gün ayrımı yapılmadan meteorolojik veriler olmadan sadece bir saat önceki ve bir gün önceki yük kullanılarak tahmin yapılan algoritmadır.
- Meteorolojik veriler kullanılarak oluşturulan modellerin MAPE değeri daha yüksektir diğer bir ifadeyle doğruluk oranı daha düşüktür. Meteorolojik veriler bu çalışmada kullanılan bölge için tahmin doğruluğunu negatif yönde etkileyen bir parametredir. Sadece önceki yük parametreleri kullanılan oluşturulan model doğruluk oranı en yüksek modeldir.
- 2 yıllık veri kümesi kullanılarak yapılan bu çalışmada hafta içi ve hafta sonu olarak iki ayrı test ve eğitim verisi kullanmak bazı modeller için doğruluk oranını artırmış bazı modeller için tahmin doğruluğunu azaltmıştır.
- Kukla değişken kullanılarak yapılan gün ayrımı 2 ayrı eğitim ve test verisi kullanılarak yapılan gün ayrımı kıyaslandığında bazı modeller için doğruluk oranını arttırmış bazı modeller için doğruluk oranını azaltmıştır.
- Gün ayrımının etkisinin bütün modellerde kesin bir sonucu gözlenmemiştir. Bu sebeple tahmin yapılacak bölgenin karakteristiğine göre ve tahminde kullanılacak algoritmaya göre hafta günleri ayrımı yapılmalıdır.
- Model-15 Model-5'ten sonra doğruluk oranı en yüksek olan modeldir. Gün ayrımının kukla değişken kullanılarak yapılması meteorolojik veriler kullanılmadan yapılan tahmin modeli için daha doğru sonuçlar vermiştir.
- Gün ayrımı yapılarak kurulan ARIMA modeli gün ayrımı yapılmadan kurulan modele çok küçük farklı olsa da daha doğru sonuçlar vermiştir.
- ARIMA modeli birçok modeldeki yapay zekâ algoritmalarından daha yüksek doğruluk oranlı sonuçlar vermiştir. Buradan kullanılacak olan veri setine göre zaman serisi analizi yöntemlerinin bazı modellerde ve algoritmalarda daha üstün olduğu çıkarımı yapılabilir.

- Yapay zekâ modelleri arasında en yüksek doğruluk destek vektör regresyonu modeli vermiştir. Doğruluk oranı en yüksek 2. Yapay zekâ algoritması karar ağacı regresyonu algoritması olmuştur.
- En yakın k komşu ve destek vektör makineleri algoritması doğruluk oranları üstünlüğü kurulan modellere göre değişkenlik göstermektedir.
- Regresyon kullanılan modellerin doğruluk oranı daha yüksektir. Bu sebeple tahmin yapılacak bölgeye göre yapılan sınıflandırmada regresyon kullanıp kullanılmayacağı bölgenin özelliğine göre iyi belirlenmelidir.
- Tahmin edilecek yüke etki eden parametreler yükün karakteristiğine göre karar verilmelidir.
- Tahmin yöntemi tahmin doğruluğu açısından önemlidir. Bu sebeple tahmin yöntemi tahmin edilecek yükün trend analizi sonucunda karar verilmelidir.

Çalışma kısıtlılıkları olarak tahmin yapılan bölgenin tek bölge olması ve farklı bölge yük karakteristiği ve iklim koşullarının etkisinin incelenememesi ve kullanılan algoritmaların tek kullanmaları ve hibrit yapılar kullanılmaması sayılabilir.

Bu çalışmaya ek olarak meteorolojik faktörlerden rüzgâr hızı, bulutluluk oranı gibi faktörlerin yük tahminine etkisi incelenip karşılaştırma yapılabilir. Ayrıca kullanılan bağımsız değişkenlerde farklılaştırılarak kullanılan giriş parametrelerin tahmin doğruluğu üzerinde etkisi araştırılabilir. Başka bir bölgenin yük verileri kullanılarak meteorolojik verilerin yük tahminine etkisi araştırılabilir. Ayrıca tek bir algoritma yapısı kurulup farklı bölgeleri tahmin etmek için tak-çalıştır tarzında bir uygulama geliştirilebilir.

Bu çalışma kapsamında Marmara bölgesinde yer alan bir dağıtım şebekesinde 2 yıllık veriler kullanılarak tahmin işlemi gerçekleştirilmiştir. Farklı bölgedeki dağıtım şebekesine modeller ve algoritmalar uygulanıp karşılaştırmalar yapılabilir.



KAYNAKLAR

- [1] **Yoldaş, U. C.** (2016) “Elektrik enerjisinde yük tahmini yöntemleri ve Türkiye’nin 2005-2020 yılları arasındaki elektrik enerjisi talep gelişimi ve arz planlaması”, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara,
- [2] **Aksel, F.** (2000). Regresyon Analizi ve Yapay Sinir Ağı Yöntemleri ile Uzun Dönem Yük Tahmini, Yüksek Lisans Tezi, İ.T.Ü. Fen Bilimleri Enstitüsü, İstanbul
- [3] **Feinberg, E.A., Genethliou, D.** (2005). *Load forecasting. In Applied Mathematics for Restructured Electric Power Systems: Optimization, 54 Control and Computational Intelligence*, Power Electronics and Power Systems. Springer, USA, pp. 269–285
- [4] **Soliman, S. A. H., & Al-Kandari, A. M.** (2010). Electrical load forecasting: modeling and model construction. Elsevier.
- [5] **İşık, Esener İ., Yüksel T., Kurban M.** (2012). Sıcaklık Verisi Olmadan Kısa Dönem Yük Tahmini İçin Yapay Zekâ Tabanlı Melez Yapılar, *Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu*, 29 Kasım- 01 Aralık 2012, Bursa.
- [6] **Akman T., Yılmaz C., Sönmez Y.** (2018). Elektrik Yüğü Tahmin Yöntemlerinin Analizi. *Gazi Mühendislik Bilimleri Dergisi*, 4(3),168-175.
- [7] **Singh A. ve diğ.** (2012). Load forecasting techniques and methodologies: A review, *2nd International Conference on Power, Control and Embedded Systems*, (pp.1-10). December 17-19, Allahabad, INDIA.
- [8] **New W. R.** (1962). Load Forecasting on the TVA System Par I-Substation Loads, *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, 81(3), 101-105.
- [9] **Gross G., Galiana F. D.** (1987). Short-term load forecasting, *Proceedings of the IEEE*, 75(12), 1558-1573.
- [10] **Papalexopoulos, A. D., & Hesterberg, T. C.** (1990). A regression-based approach to short-term system load forecasting. *IEEE Transactions on Power Systems*, 5(4), 1535-1547.
- [11] **Bakirtzis, A. G., Petridis, V., Klartzis, S. J., Alexiadis, M.C.** (1996). A Neural Network Short Term Load Forecasting Model For Greek Power System, *IEEE Transactions on Power Systems*, 11(2), 858-863.
- [12] **Liu, K., Subbarayan, S., Shoults, R. R., Manry, M. T., Kwan, C., Lewis, F. I., & Naccarino, J.** (1996). Comparison of very short-term load forecasting techniques. *IEEE Transactions on power systems*, 11(2), 877-882.

- [13] **Gürsoy, E.** (2000). *Yük Tahmini Yöntemleri ve Çukurova Elektrik AŞ, Kepez Elektrik T.A.Ş. Bölgelerine Uygulanması*. (Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [14] **Hippert, H. S., Pedreira, C. E., & Souza, R. C.** (2001). Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on power systems*, 16(1), 44-55.
- [15] **Yalçınöz, T., Herdem, S. ve Eminoğlu, U.** (2002). Yapay Sinir Ağları ile Niğde Bölgesinin Elektrik Yük Tahmini, ELECO'2002 Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu ve Fuarı, sayfa 25-29, Bursa, 18-22 Aralık 2002.
- [16] **Huang, C. M., Huang, C. J., & Wang, M. L.** (2005). A particle swarm optimization to identifying the ARMAX model for short-term load forecasting. *IEEE Transactions on Power Systems*, 20(2), 1126-1133.
- [17] **Ceylan, G.** (2004). *Yapay Sinir Ağları ile Kısa Dönem Yük Tahmini*. (Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [18] **Pai, P. F., & Hong, W. C.** (2005). Support vector machines with simulated annealing algorithms in electricity load forecasting, *Energy Conversion and Management*, 46(17), 2669-2688.
- [19] **Demirel, Ö.** (2009). *Anfis ve Arma Modelleri ile Elektrik Enerjisi Yük Tahmini*. (Yüksek Lisans Tezi). Marmara Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [20] **Demren, D.** (2011). *Electrical Load Demand Forecasting Application Using Support Vector Machines*. (Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [21] **Esener, I. I., Yüksel, T., & Kurban, M.** (2012). Artificial intelligence based hybrid structures for short-term load forecasting without temperature data. *Proceedings 11th International Conference on Machine Learning and Applications*, Turkey: Bursa 29 Nov-01 Dec.
- [22] **Çevik, H.** (2013) *Türkiye'nin Kısa Dönem Yük Tahmini*. (Yüksek Lisans Tezi). Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya.
- [23] **İşyapar, M. T.** (2013) *In Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Engineering*. (Yüksek Lisans Tezi). Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.
- [24] **Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R.** (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, 33, 102-109.
- [25] **Tutu, E. B.** (2017). *Türkiye için Kısa Vadeli Elektrik Enerjisi Talep Tahmini*. (Yüksek Lisans Tezi). TOBB Ekonomi ve Teknoloji Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- [26] **Akman, T.** (2018). *Yapay Zekâ Modelleri Kullanarak Ankara Bölgesinin Kısa Dönem Elektrik Enerjisi Yük Tahmini*. (Yüksek Lisans Tezi). Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

- [27] **Khan, N. U.** (2018). *Short-Term Load Forecasting by Using Artificial Neural Networks*. (Yüksek Lisans Tezi). Bahçeşehir Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [28] **Butekin Ç, T.** (2019) *Ann Based Electricity Consumption Forecasting in Yaşar University*. (Yüksek Lisans Tezi). Yaşar Üniversitesi, Fen Bilimleri Enstitüsü, İzmir.
- [29] **Doruk, E.** (2019). *Sakarya Bölgesi Hanehalkı Elektrik Tüketiminin Dinamik Lineer Model ile Tahmini*. (Yüksek Lisans Tezi). Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya.
- [30] **Anwar, T., Sharma, B., Chakraborty, K., & Sirohia, H.** (2018). Introduction to Load Forecasting. *International Journal of Pure and Applied Mathematics*, Volume 119, No 15, 1527-1538.
- [31] **Walther, J., Spanier, D., Panten, N., & Abele, E.** (2019). Very short-term load forecasting on factory level–A machine learning approach. *Procedia CIRP*, 80, 705-710.
- [32] **Hamilton, J. D.** (1994). *Time series analysis*. New Jersey: Princeton.
- [33] **Newbold, P., Carlson, W. L., & Thorne, B.** (2013). *Statistics for business and economics*. Boston, MA: Pearson.
- [34] **Gujarati, D. N.** (2009). *Basic econometrics*. Tata McGraw-Hill Education.
- [35] **Makridakis, S., Whellwright, S.C** (1989). *Forecasting Methods of Management* (5th ed.) John Wiley&Sons, Inc.
- [36] **Zuhairi B., Nidal, K.** (2009). Autoregressive method in short term load forecast. PE Con 2008- 2008 IEEE 2nd International Power and Energy Conference. 1603- 1608.
- [37] **McCulloch WS, Pitts W.** (1943). A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys.* 1943;5(4):115–133.
- [38] **Park, D. C. ve diğ.** (1991). Electric load forecasting using an artificial neural network. *IEEE transactions on Power Systems*, 6(2), 442-449.
- [39] **Zadeh, L. A.** (1965). Fuzzy Sets, *Information and Control*, Vol. 8., pp. 338-353.
- [40] **Klir, G. J., & Yuan, B.** (1995). *Fuzzy Sets and Fuzzy Logic. Theory and Applications*, Prentice Hall.
- [41] **Fisher, R. A.** (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2), 179-188
- [42] **Boser, B. E., Guyon, I. M., & Vapnik, V. N.** (1992). A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory* (144-152).
- [43] **Vapnik, V.** (2013). *The Nature of Statistical Learning Theory*. New York, Springer Science & Business Media.
- [44] **Chang, C.-C., Lin, C.-J., LIBSVM: a Library for Support Vector Machines** from <http://www.csie.ntu.edu.tw/~cjlin/Libsvm/>
- [45] **Erdal, H. İ.** (2011). *Destek Vektör Makineleri İle Tahmine Dayalı Modelleme Ve Bir Uygulama*. (Doktora Tezi), İstanbul Üniversitesi, İstanbul.

- [46] **Ayhan, S., & Erdoğmuş, Ş.** (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 9(1), 175-201.
- [47] **Yürük, M. F.** (2019). *Yapay Zekâ Yöntemleri ile İşletmelerin Finansal Başarısızlığının Tahmin Edilmesi: BİST İmalat Sanayi Uygulaması*, Doktora Tezi, Gaziantep Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme Ana Bilim Dalı, Gaziantep.
- [48] **Kaya, K.** (2016). *Destek Vektör Makineleri Yardımıyla Tüketici Kredilerinin Sınıflandırılması*, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- [49] **Sun, J. & Li, H.** (2012). Financial Distress Prediction Using Support Vector Machines: Ensemble vs. Individual, *Applied Soft Computing* 12, (p.2254-2265)
- [50] **Rockafellar, R. T.** (1993). Lagrange multipliers and optimality. *SIAM review*, 35(2), 183-238.
- [51] **Kecman, V.** (2001). *Learning and Soft Computing: Support Vector Machines, Neural Networks and Fuzzy Logic Models*, MIT Press, Cambridge, Massachusetts, England
- [52] **Duman, A.** (2010). *Multiple Criteria Sorting Methods Based on Support Vector Machines*. (Yüksek Lisans Tezi). Orta Doğu Teknik Üniversitesi, Ankara.
- [53] **Rosipal, R., & Trejo, L. J.** (2001). Kernel partial least squares regression in reproducing kernel hilbert space. *Journal of machine learning research*, 2, 97-123.
- [54] **Çomak, E.** (2008). *Destek Vektör Makinelerinin Etkin Eğitimi İçin Yeni Yaklaşımlar*, (Doktora Tezi), Selçuk Üniversitesi, Konya.
- [55] **Drucker, H. ve diğ.** (1997). Support vector regression machines. *In Advances in neural information processing systems* (pp. 155-161).
- [56] **Destek vektör regresyonu ve destek vektör makineleri**, Erişim: 9 Eylül 2018, <https://yavuz.github.io/destek-vektor-regresyonu-ve-makineleri/>
- [57] **Sharma, A. and Suryawanshi, A.** (2016). A Novel Method for Detecting Spam Email using KNN Classification with Spearman Correlation as Distance Measure, *International Journal of Computer Applications*, 136, 28–35
- [58] **Han, J. and Kamber, M.** (2006), *Data Mining: Concepts And Techniques*, University of Illinois, San Francisco.
- [59] **Bramer, M.** (2016), *Principles of Data Mining*, Springer, London.
- [60] **Breiman L, Friedman J, Stone CJ, Olshen RA.** (1984) *Classification and Regression Trees*. Taylor and Francis, Chapman&Hall/CRC.
- [61] **Gehrke & Nong Ye** (2003), Decision Trees, *The Handbook of Data Mining*, New Jersey, Lawrence Erlbaum Associates Inc, pp. 3-4
- [62] **Omitaomu, OA.** (2006), Decision Trees. In Michael W. Berry and Murray Browne (Eds). *Lecture Notes in Data Mining*. World Scientific Publishing of Hackensack, New Jersey, USA: 2006:39-51.

- [63] **Elasan, S.** (2019), *Veri Madenciliğinde Farklı Karar Ağaçları ve K-En Yakın Komşuluk Yöntemlerinin İncelenmesi: Kadın Hastalıkları ve Doğum Verisinde Bir Uygulama*, (Doktora Tezi), Yüzüncü Yıl Üniversitesi Sağlık Bilimleri Enstitüsü, Van.
- [64] **Orhan, U.** (t.y.). *Makine Öğrenmesi* [PowerPoint slides]. Retrieved from <https://bmb.cu.edu.tr/uorhan/DersNotu/Ders03.pdf> Ders notu.
- [65] **Kass, G.V.** (1980), An exploratory technique for investigating large quantities of categorical data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2), 119-127.
- [66] **Bigss D, Ville B, Suen E.** (1991), A method of choosing multiway partitions for classification and decision trees. *Journal of Applied Statistics*, 18(1):4, 9-62.
- [67] **Loh WY, Shih YS.** (1997), Split selection methods for classification trees, *Statistica Sinica*, 7, 815-840.
- [68] **Breiman L.** (2001), Random Forests, *Machine Learning* 45(1), 5-32.
- [69] **Quinlan, J. R.** (1992), *Learning with continuous classes*, In 5th Australian joint conference on artificial intelligence (Vol. 92, pp. 343-348).
- [70] **Quinlan J. R.** (1987), Simplifying decision trees, *International Journal of Man-Machine Studies*, 27(3), 221-234.
- [71] **Eviews** (Version 10) [Computer software] Irvine, CA USA.
- [72] **Frakn E., Hall M.A., and Witten I.** (2016). *The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Fourth Edition, 2016
- [73] **MATLAB and Statistics Toolbox** (Version R2012b), [Computer software]. The MathWorks, Inc., Natick, Massachusetts, United States.



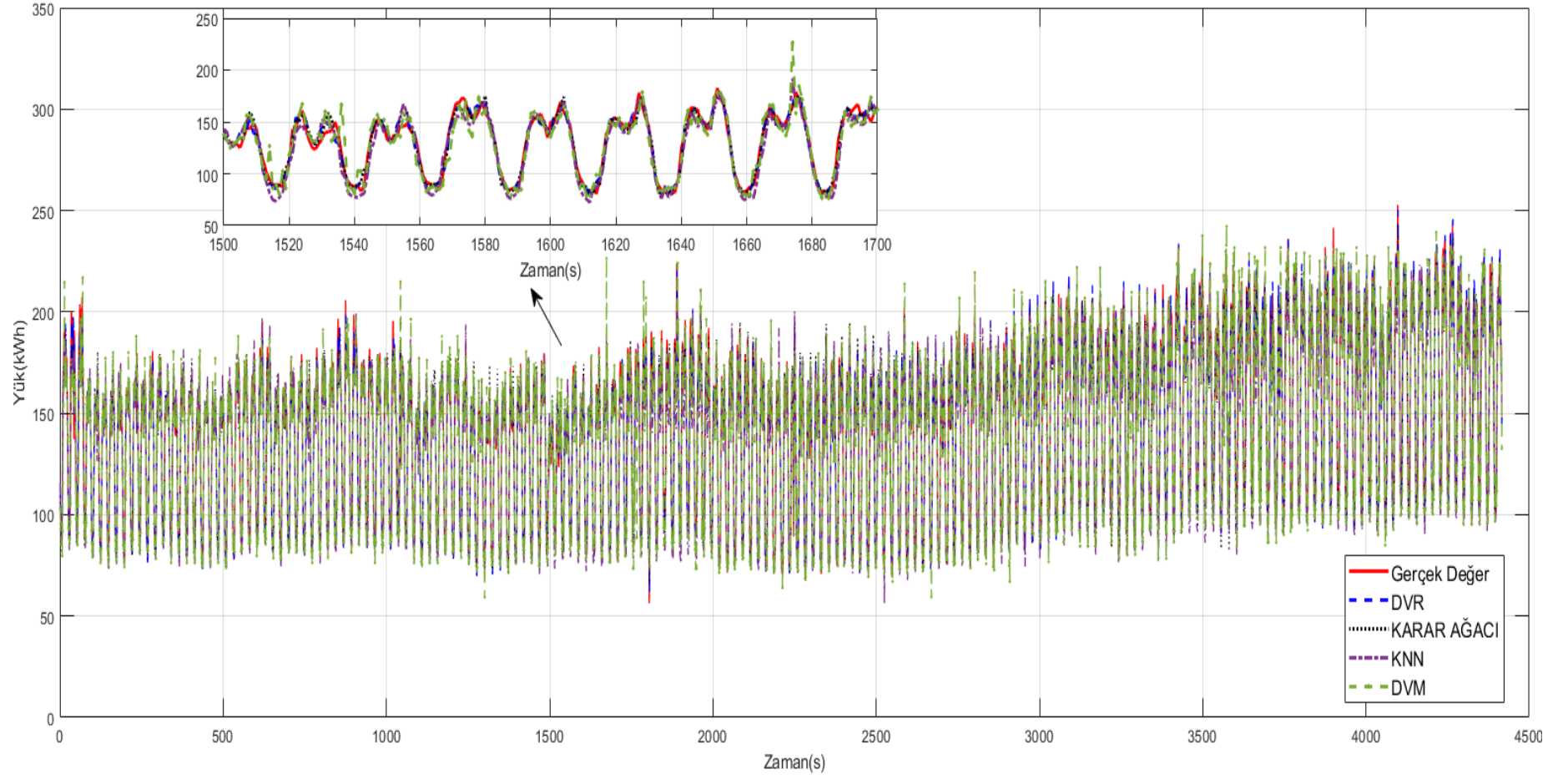
EKLER

EK A: Tahmin Sonuçları

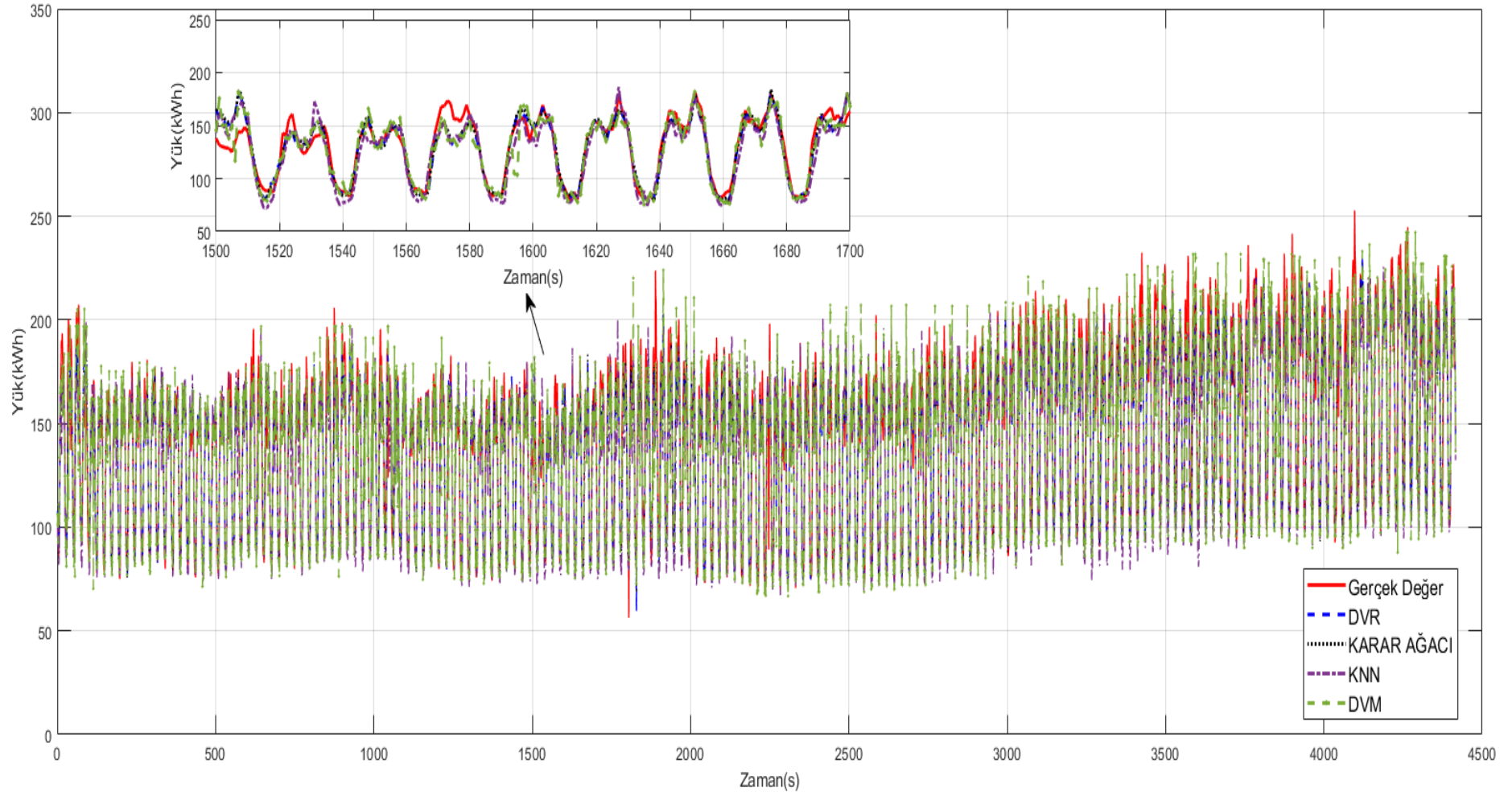




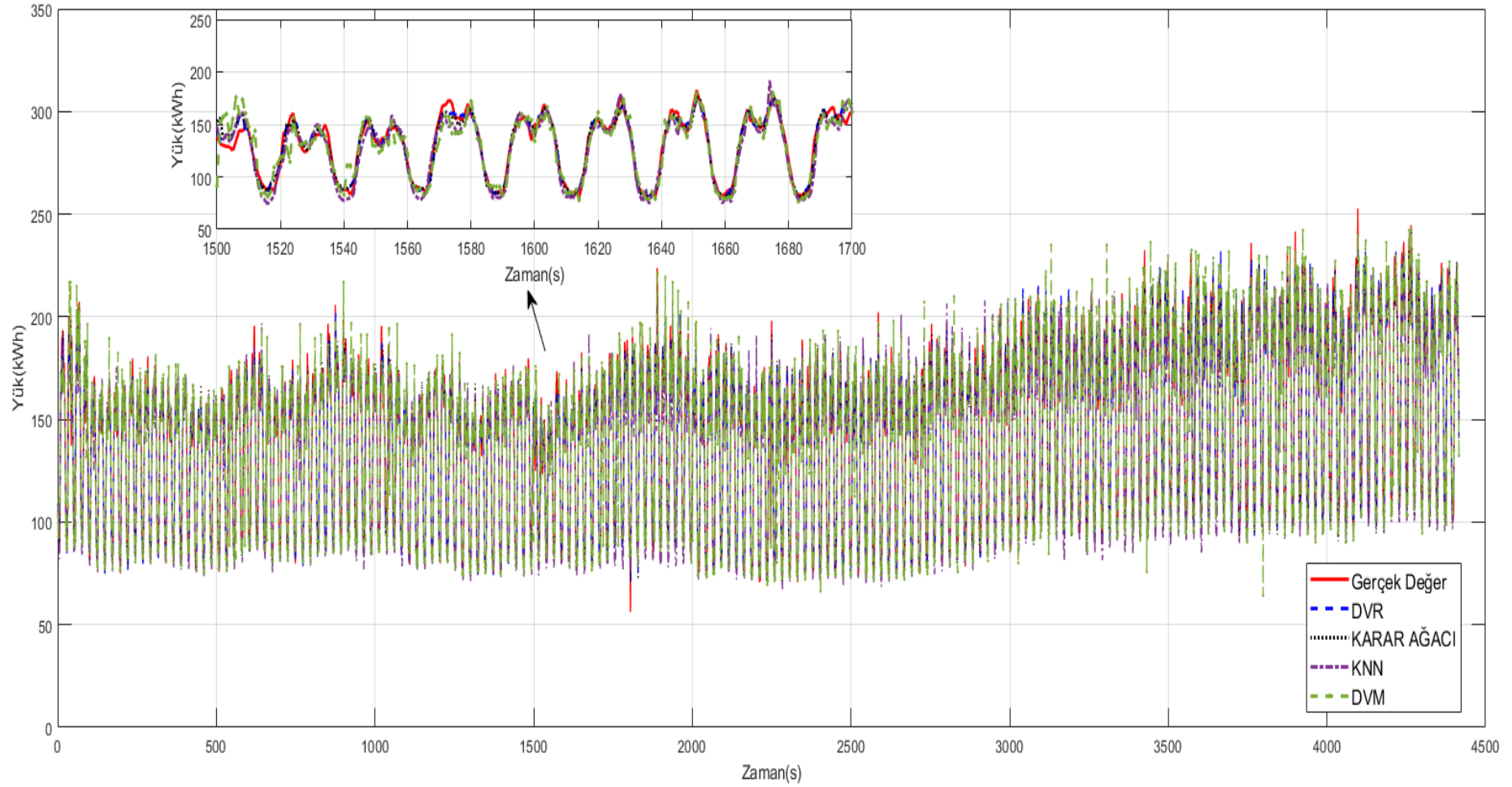
EK A



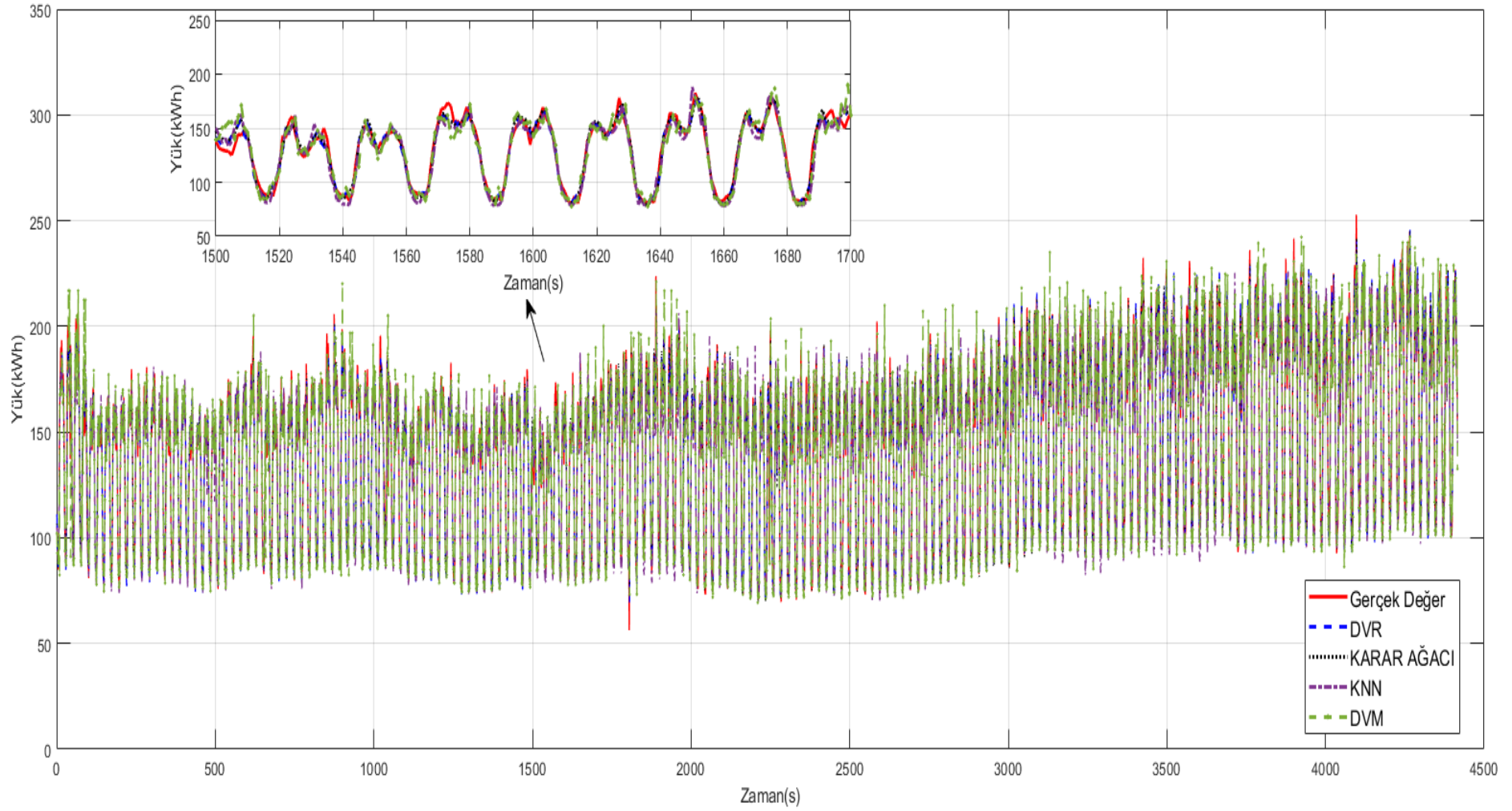
Şekil A.1 : Model-1 tahmin sonuçlarının karşılaştırılması



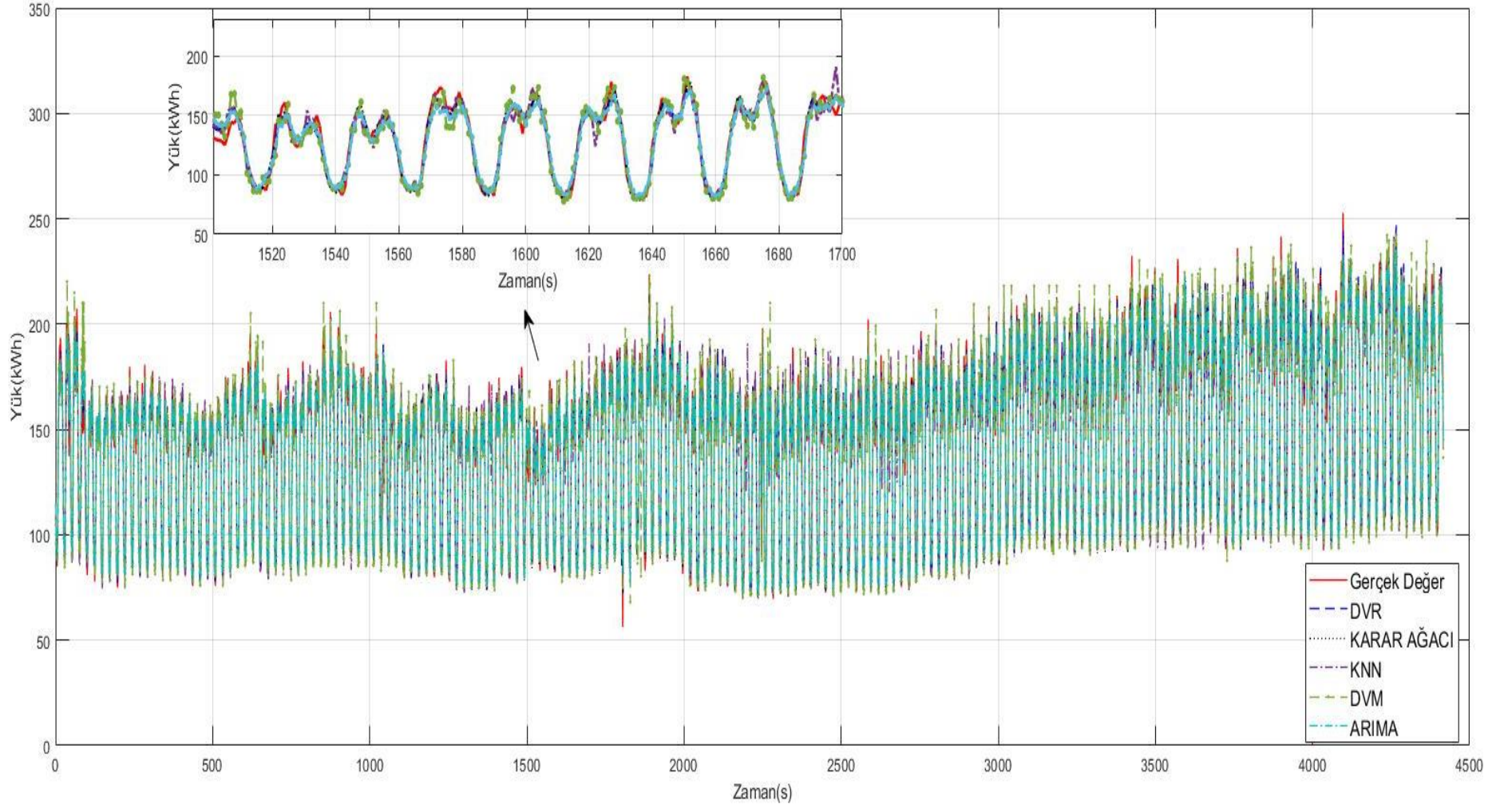
Şekil A.2 : Model-2 tahmin sonuçlarının karşılaştırılması



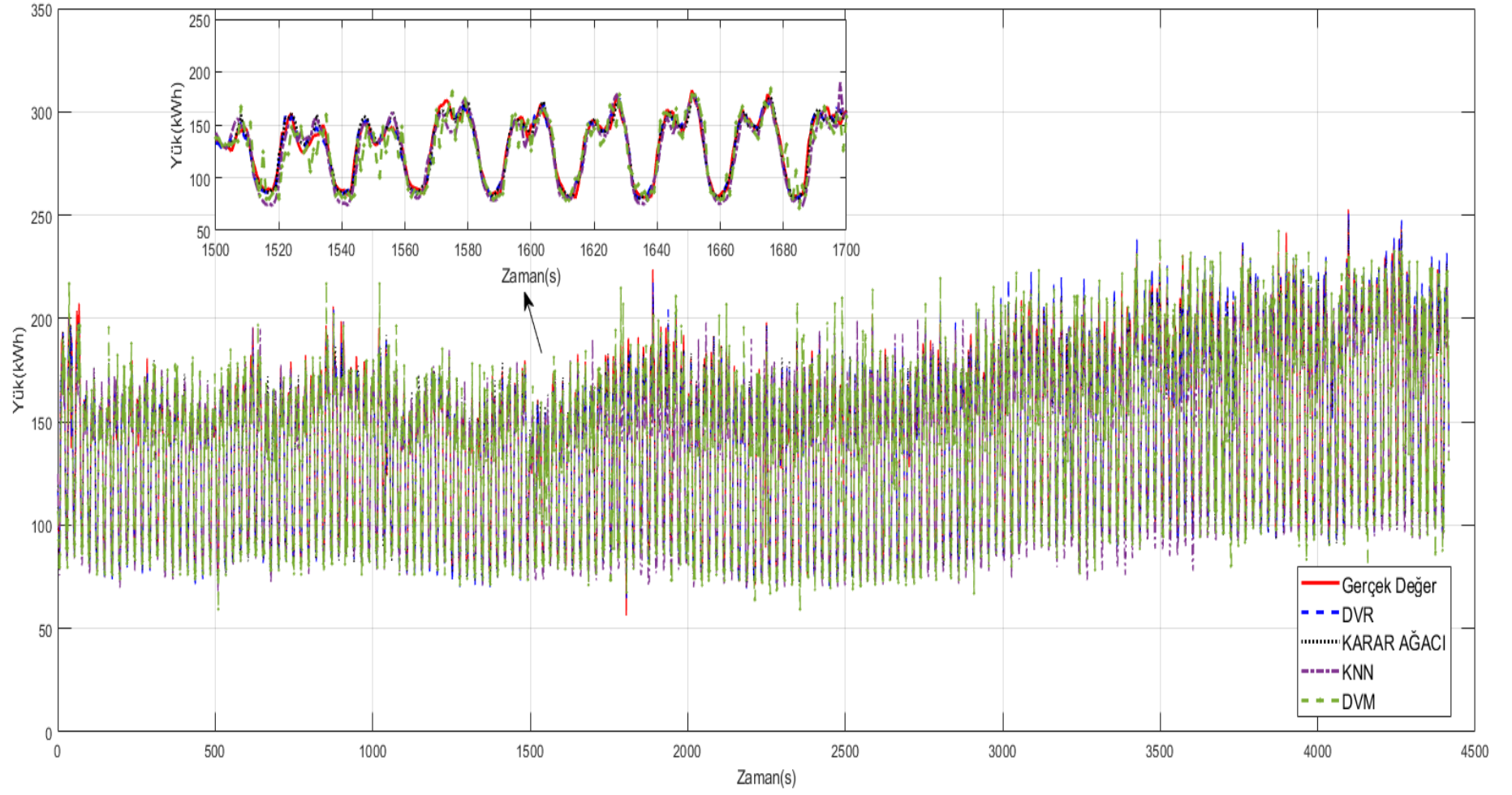
Şekil A.3 : Model-3 tahmin sonuçlarının karşılaştırılması



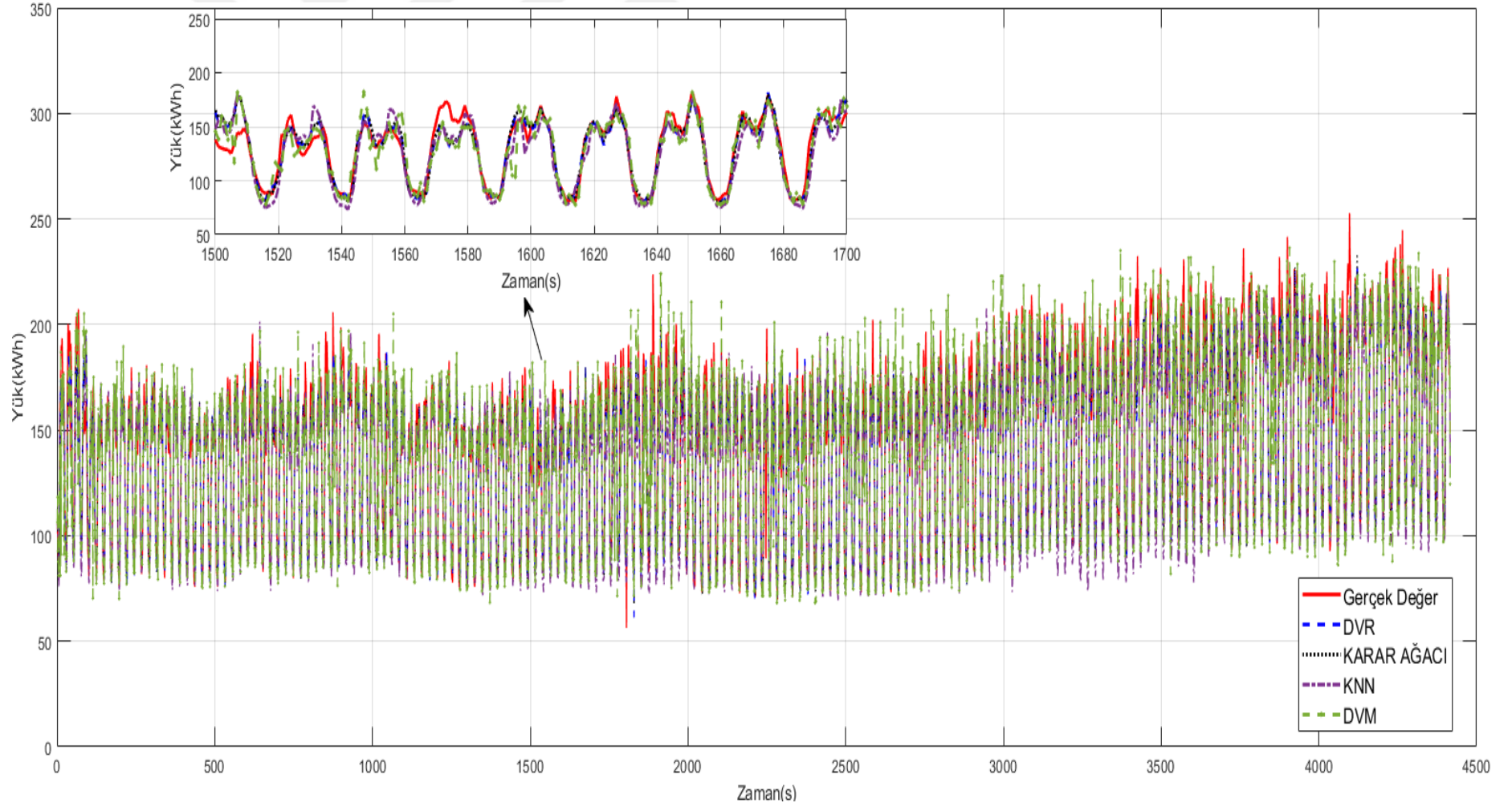
Şekil A.4 : Model-4 tahmin sonuçlarının karşılaştırılması



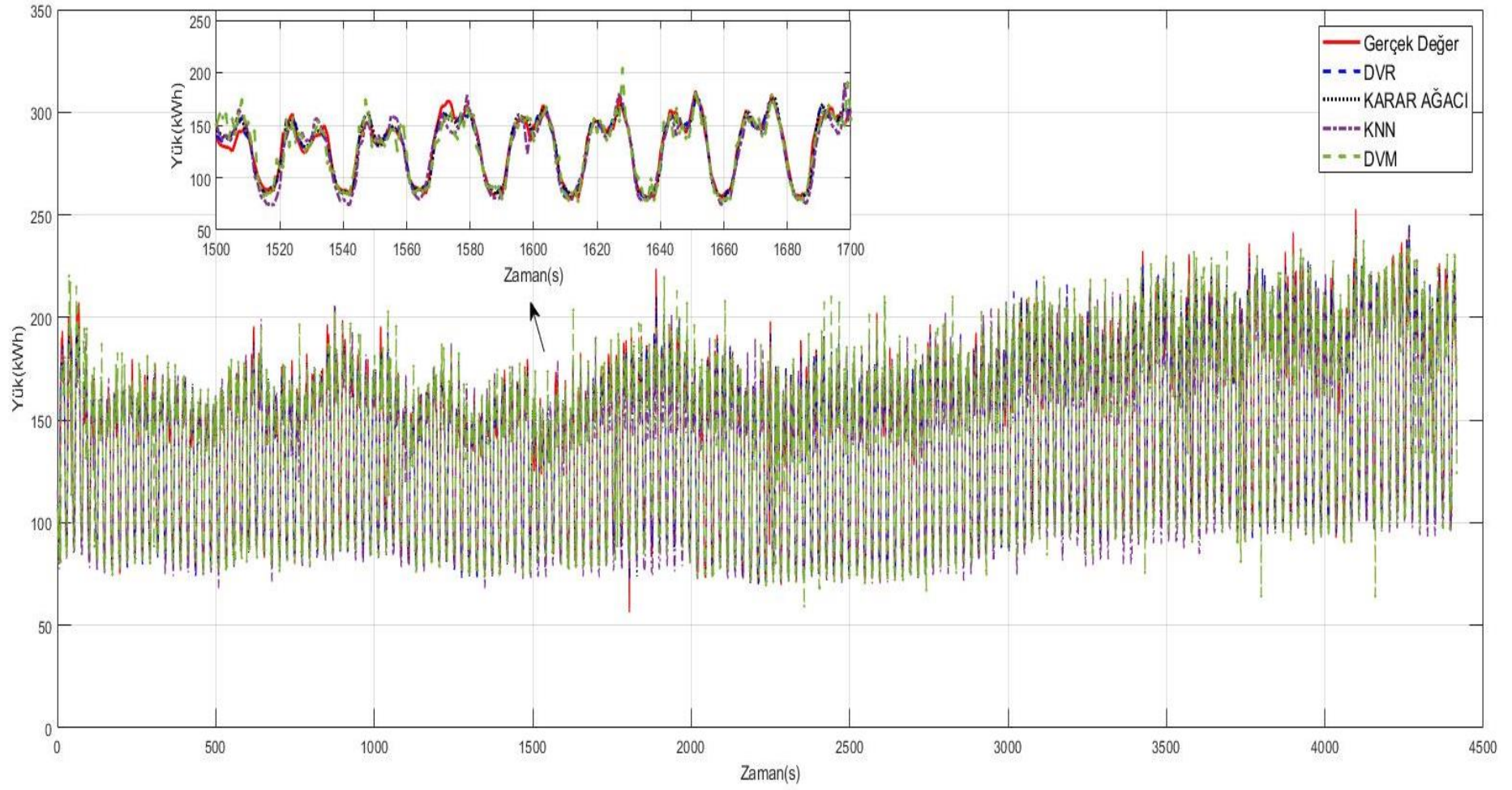
Şekil A.5 : Model-5 tahmin sonuçlarının karşılaştırılması



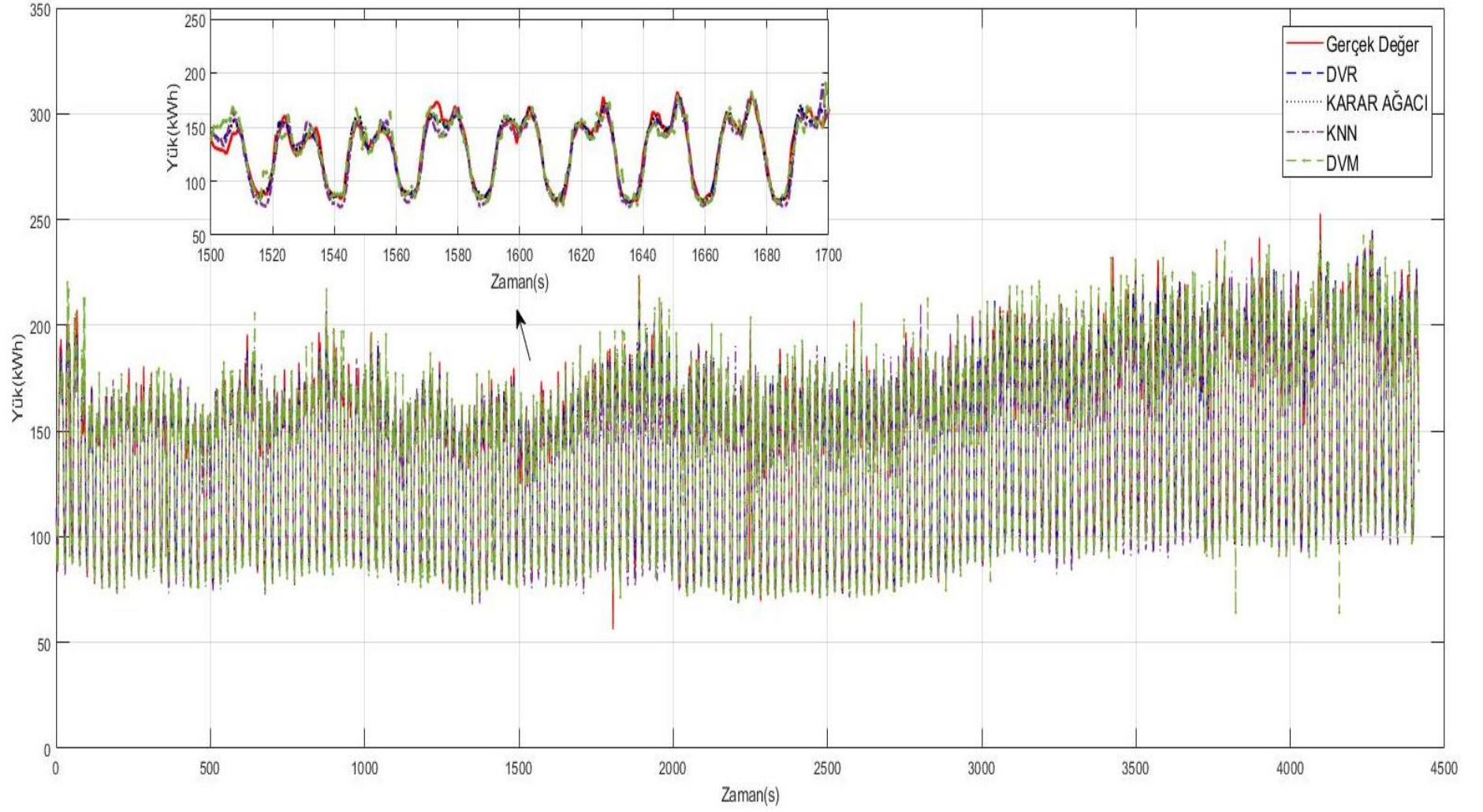
Şekil A.6 : Model-6 tahmin sonuçlarının karşılaştırılması



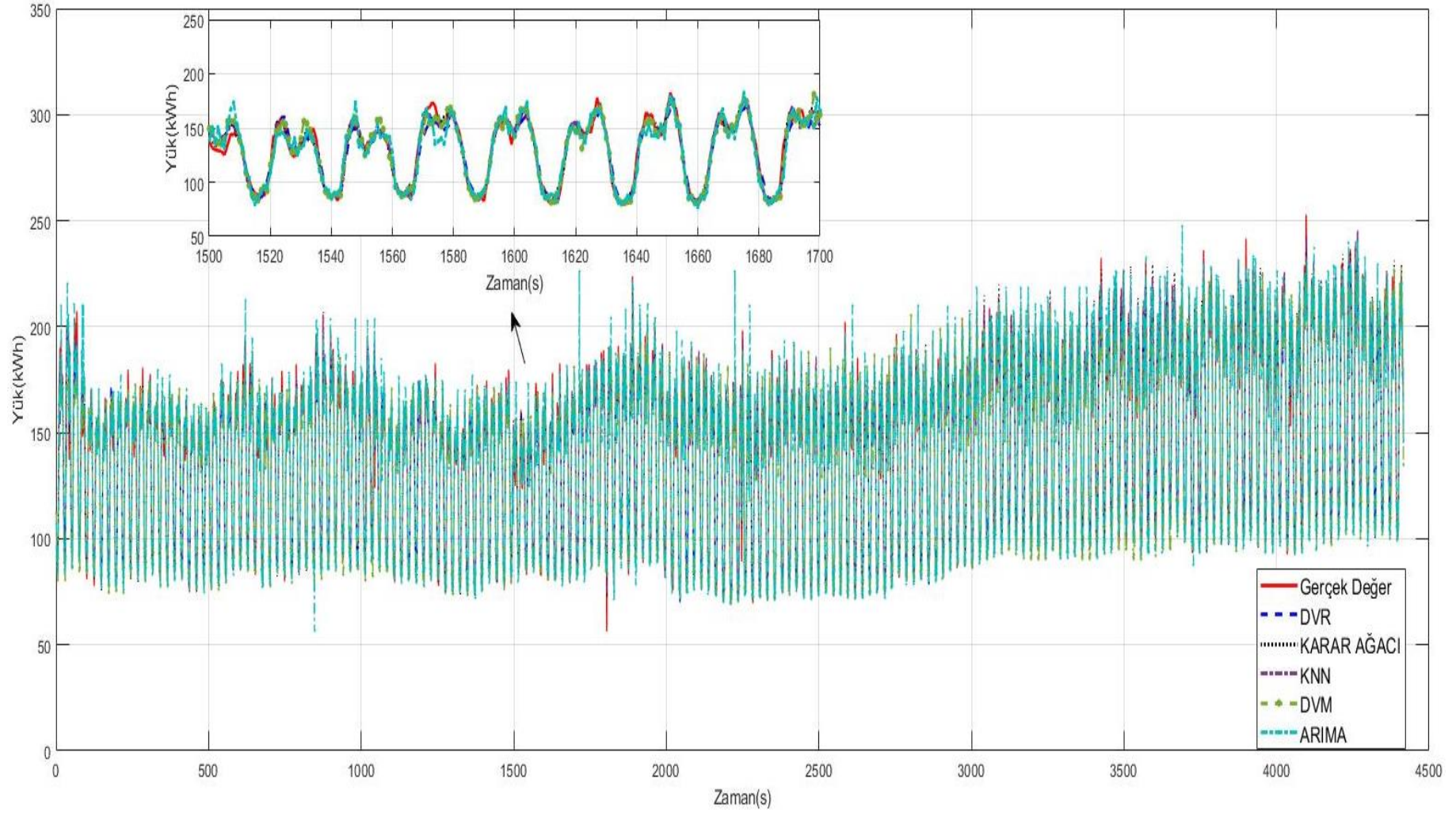
Şekil A.7 : Model-7 tahmin sonuçlarının karşılaştırılması



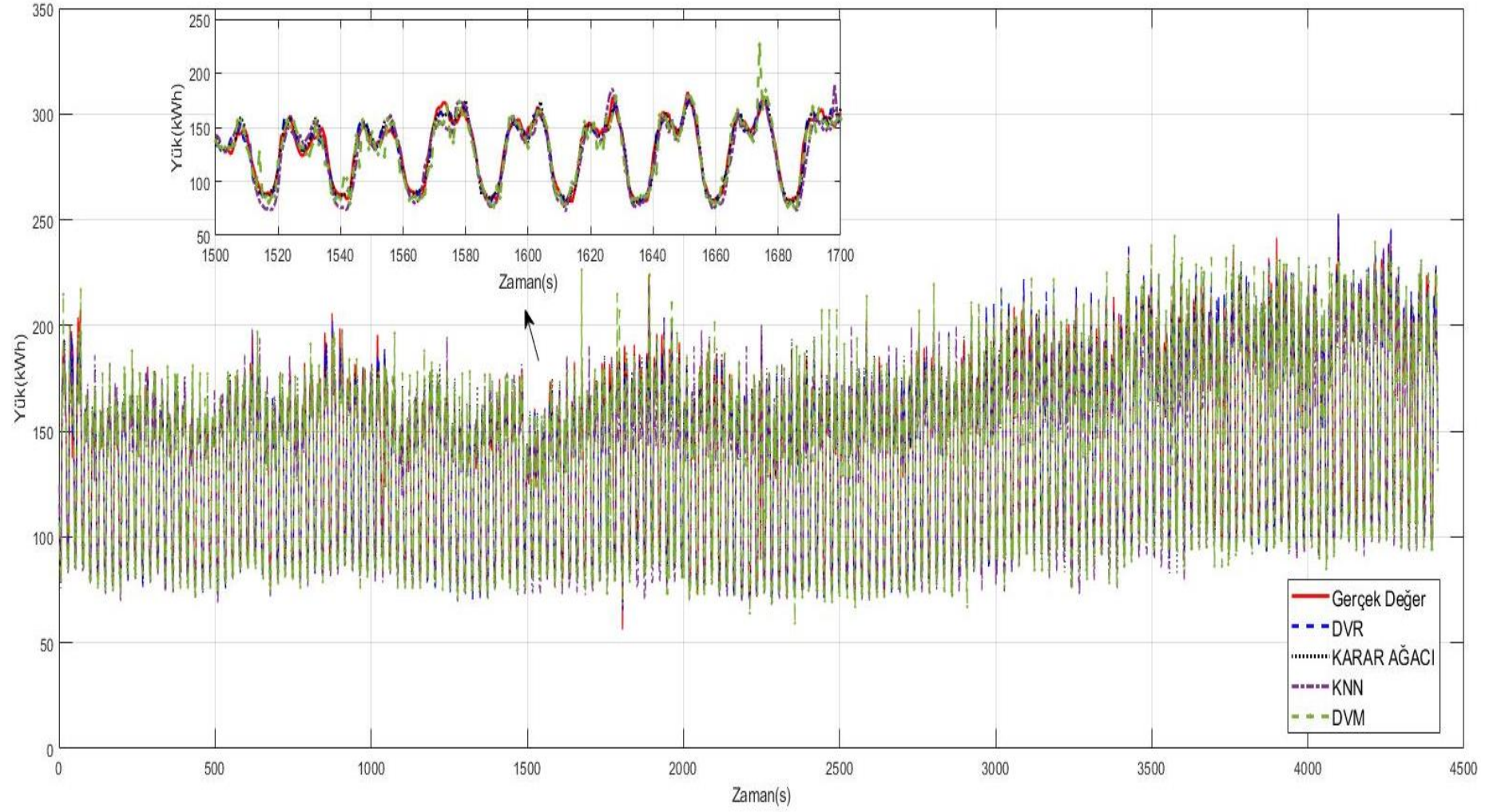
Şekil A.8 : Model-8 tahmin sonuçlarının karşılaştırılması



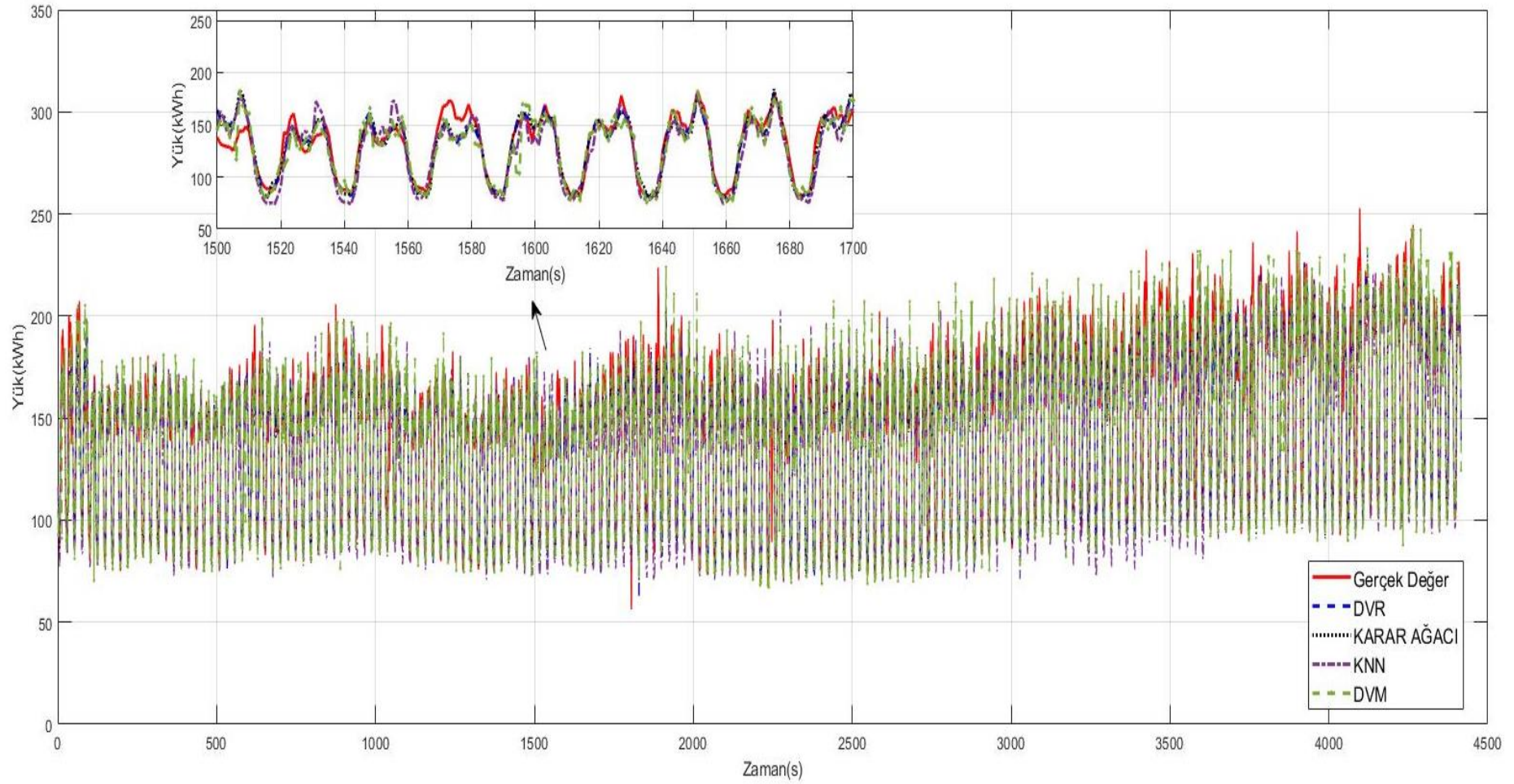
Şekil A.9 : Model-9 tahmin sonuçlarının karşılaştırılması



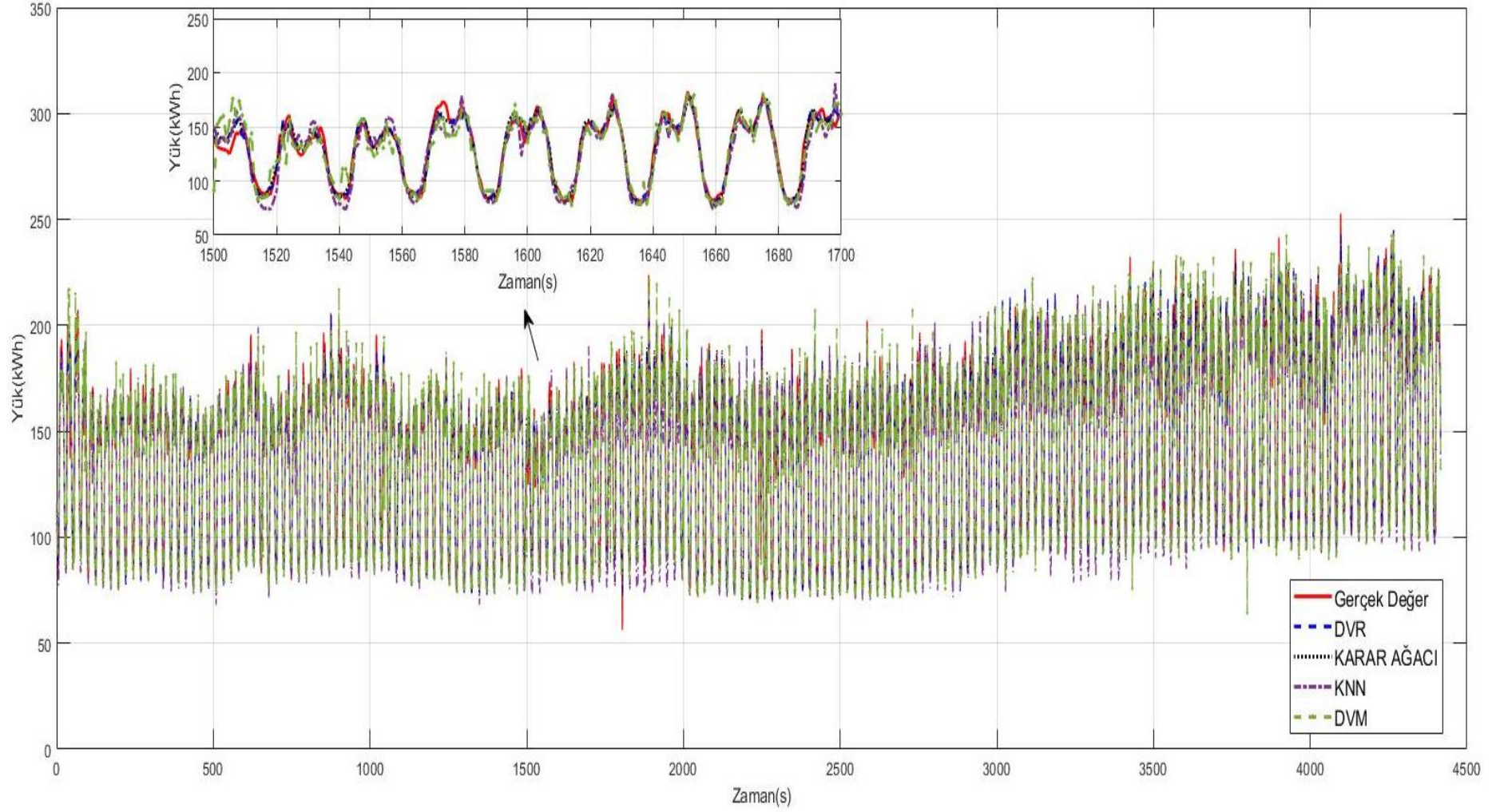
Şekil A.10 : Model-10 tahmin sonuçlarının karşılaştırılması



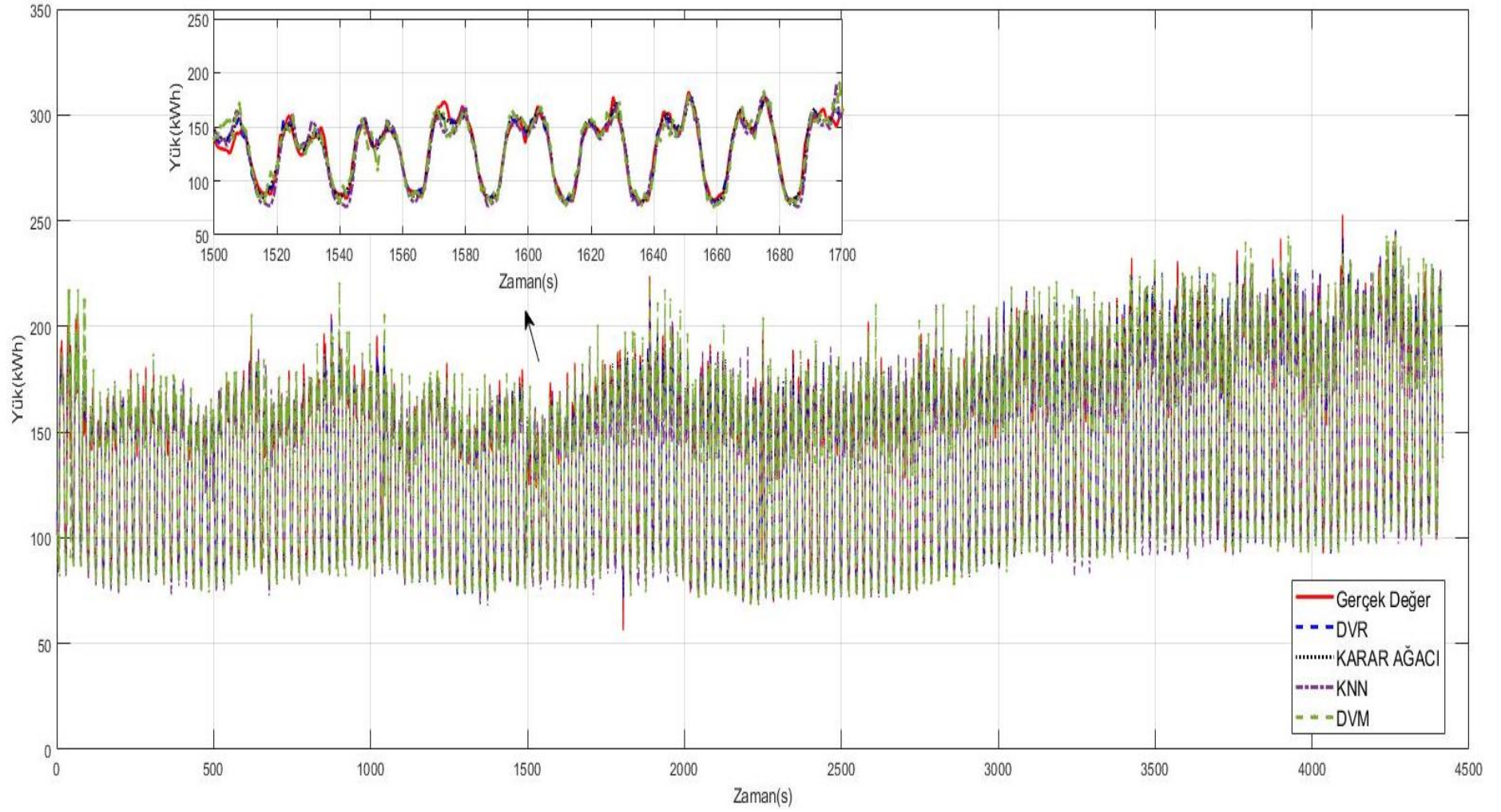
Şekil A.11 : Model-11 tahmin sonuçlarının karşılaştırılması



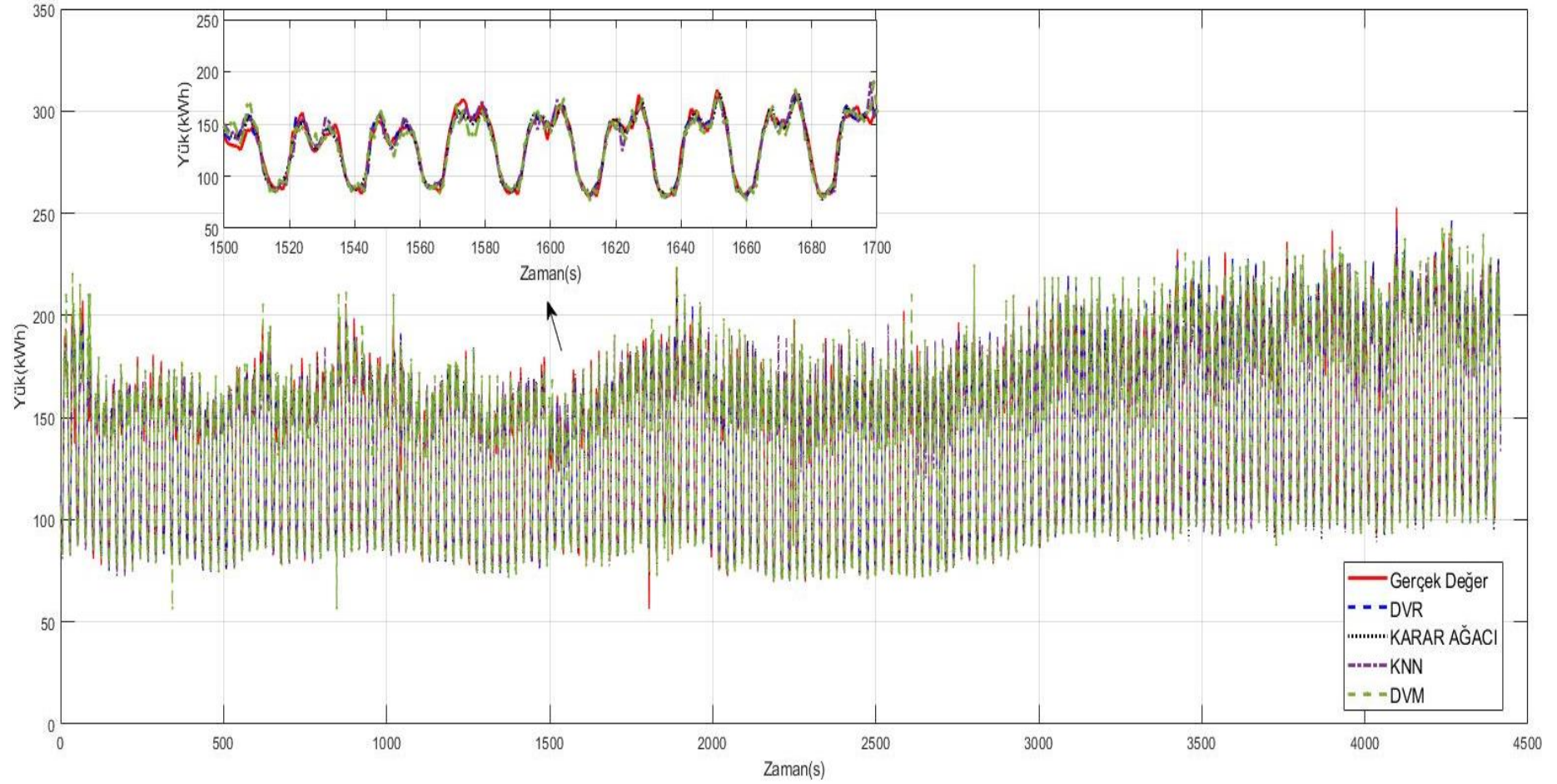
Şekil A.12 : Model-12 tahmin sonuçlarının karşılaştırılması



Şekil A.13 : Model-13 tahmin sonuçlarının karşılaştırılması



Şekil A.14 : Model-14 tahmin sonuçlarının karşılaştırılması



Şekil A.15 : Model-15 tahmin sonuçlarının karşılaştırılması



ÖZGEÇMİŞ

Ad-Soyad : Mert SAÇLI
Doğum Tarihi ve Yeri : 1993 /Adana
E-posta : saclim@itu.edu.tr



ÖĞRENİM DURUMU:

- **Lisans** : -2017, İstanbul Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Elektrik Mühendisliği
-2018, İstanbul Teknik Üniversitesi, Makine Fakültesi, Mekatronik (Yandal)
- **Yüksek lisans** : -2018, İstanbul Medeniyet Üniversitesi, İşletme Anabilim Dalı, İşletme

MESLEKİ DENEYİM VE ÖDÜLLER:

- Türk Hava Yolları Teknik A.Ş- Süreç ve İş Geliştirme Mühendisi

DİĞER YAYINLAR, SUNUMLAR VE PATENTLER:

- Saçlı, M., Ayan, O., Silsüpür, M., & Türkay, B. E. (2017). “Investigation of power quality analysis of three-phase inverter topologies for renewable energy systems”,10th International Conference on Electrical and Electronics Engineering (ELECO) (pp. 1445-1449). IEEE.
- Saçlı, M., Ayan, O., Silsüpür, M., & Türkay, B. E. (2019). “Power quality comparison of single-phase and three-phase inverter topologies for renewable energy systems”. International Journal of Smart Grid and Clean Energy, 8(2), 174-183.
- Saçlı, M., Ayan, O., Silsüpür, M., & Türkay, B. E. 2017. “Investigation of Power Quality Analysis of Single-Phase Inverter Topologies for Solar Energy Systems”, International Conference on Condition Monitoring, Diagnosis and Maintenance (CMDM 2017),88-96.
- Ayan, O, Saçlı, M, Türkay, B. E. (2017). “Residential Load Management through Smart Plugs in Smart Grids”, International Conference on Condition Monitoring, Diagnosis and Maintenance (CMDM 2017),78-87.