

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ARIMA VE XGBOOST MODELLERİ İLE ELEKTRİK
TALEP TAHMİNİ**

YÜKSEK LİSANS TEZİ

Muhammed Can ÖZDEMİR

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
SİSTEMLERİ MÜHENDİSLİĞİ**
Tez Danışmanı : Prof. Dr. Nejat YUMUŞAK

Haziran 2021

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ARIMA VE XGBOOST MODELLERİ İLE ELEKTRİK
TALEP TAHMİNİ**

YÜKSEK LİSANS TEZİ

Muhammed Can ÖZDEMİR

**Enstitü Anabilim Dalı : BİLGİSAYAR VE BİLİŞİM
SİSTEMLERİ MÜHENDİSLİĞİ**

**Bu tez 21/06/2021 tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile
kabul edilmiştir.**

Üye

Jüri Başkanı

Üye

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Muhammed Can ÖZDEMİR

TEŞEKKÜR

Tez çalışmam boyunca desteklerini esirgemeyen ve bilgi ve tecrübeleriyle daima yol gösteren danışman hocam Prof. Dr. Nejat Yumuşak’a teşekkürü bir borç bilirim.



İÇİNDEKİLER

TEŞEKKÜR.....	i
İÇİNDEKİLER	ii
SİMGELER VE KISALTMALAR LİSTESİ	iv
ŞEKİLLER LİSTESİ	v
TABLolar LİSTESİ	vi
ÖZET.....	vii
SUMMARY	viii
 BÖLÜM 1.	
GİRİŞ	1
 BÖLÜM 2.	
LİTERATÜR TARAMASI.....	3
 BÖLÜM 3.	
YÖNTEM.....	5
3.1. Zaman Serileri.....	5
3.1.1. Birleşik otoregresif hareketli ortalama yöntemi (ARIMA).....	7
3.1.1.1. AR modeli	7
3.1.1.2. MA modeli	8
3.1.1.3. ARMA modeli	8
3.1.1.4. ARIMA modeli.....	9
3.2. Regresyon	9
3.2.1. Lineer regresyon.....	10
3.2.1.1. En küçük kareler yöntemi.....	10
3.2.2. Ağaç bazlı regresyon yöntemleri.....	12

3.2.2.1. Karar ağaçları	12
3.2.2.2. Ekstrem gradyan yükseltme (XGBoost).....	13
BÖLÜM 4.	
UYGULAMA	15
4.1. ARIMA Modeli.....	15
4.1.1. Diferansiyel katsayısı (d) hesabı	15
4.1.2. Gecikme derecesi (p) hesabı	18
4.1.3. Hareketli ortalama derecesi (q) hesabı	19
4.1.4. ARIMA modelinin oluşturulması.....	22
4.2. XGBoost Modeli.....	24
4.3. Tutarlılık Analizi.....	28
BÖLÜM 5.	
TARTIŞMA VE SONUÇ	29
KAYNAKLAR	30
ÖZGEÇMİŞ	34

SİMGELER VE KISALTMALAR LİSTESİ

ACF	: Autocorrelation Function (Kendiyle ilgileşim fonksiyonu)
ADF	: Augmented Dickey Fuller (Genişletilmiş Dickey Fuller)
ARIMA	: Autoregressive Integrated Moving Average (Tümleşik kendiyle bağlaşımlı hareketli ortalama)
ARIMAX	: ARIMA with eXogenous variable (Dışsal değişken içerebilen ARIMA)
MAK	: Maksimum ayrılabilir kapasite
MAPE	: Mean Absolute Percent Error (Bağıl mutlak yüzdesel hata)
MSE	: Mean Square Error (Ortalama kare hata)
PACF	: Partial Autocorrelation Function (Kısmi kendiyle ilgileşim fonksiyonu)
RMS	: Reduce Measurement Station (Gaz basıncı düşürme ve ölçüm istasyonu)
SARIMA	: Seasonal Autoregressive Integrated Moving Average (Sezonsal tümleşik kendiyle bağlaşımlı hareketli ortalama)
SVM	: Support Vector Machine (Destek vektör makinası)

ŞEKİLLER LİSTESİ

Şekil 3.1. Trend etkisi	6
Şekil 3.2. Konjonktür etkisi	6
Şekil 3.3. Trend + konjonktür etkisi	6
Şekil 3.4. Trend + düzensiz etkiler	6
Şekil 3.5. Konjonktür + düzensiz etkiler.....	6
Şekil 3.6. Trend + konjonktür + düzensiz etkiler.....	6
Şekil 3.7. Örnek veriler ve örnek tahminleme	11
Şekil 3.8. Örnek tahminleme B için hata gösterimleri	11
Şekil 3.9. Karar ağacı yapısı	13
Şekil 4.1. Eğitim ve test verileri.....	15
Şekil 4.2. Farklandırılmış seriye ait grafikler.....	17
Şekil 4.3. AR parametresi (p) grafiği	19
Şekil 4.4. MA (q) derecesi hesaplanması grafiği	21
Şekil 4.5. Model ve gerçek veri karşılaştırması	23
Şekil 4.6. ARIMA gelecek tahminleri ve gerçek değerler	23
Şekil 4.7. Seçilen parametrelerin önem dereceleri.....	25
Şekil 4.8. XGBoost tahmin sonuçları.....	25
Şekil 4.9. Özel parametrelerin önem tablosu	27
Şekil 4.10. Özel parametreler eklendikten sonra yapılan tahminleme.....	27
Şekil 4.11. Tüm modellerin karşılaştırılması	28

TABLÖLAR LİSTESİ

Tablo 3.1. Zaman serisi bileşenleri [16].....	6
Tablo 4.1. XGBoost verileri.....	24
Tablo 4.2. Özel parametreler tablosu	26
Tablo 4.3. Yöntemlerin tutarlılıkları	28

ÖZET

Anahtar kelimeler: ARIMA, zaman serileri, istatistiksel analiz, XGBoost, Elektrik talep tahmini

Elektrik enerjisi toplumun her alanda ihtiyaç duyduğu ikincil enerji gruplarında bulunan bir enerji türüdür. Elektrik enerjisi, teknolojik gelişmeler doğrultusunda farklı alanlarda daha yaygın bir şekilde kullanılabilir hale gelmektedir. Elektrik enerjisinin artan ihtiyaçları karşılayabilecek şekilde üretilip, doğru zamanda doğru şekilde kullanıcılara aktarılması, oldukça önemli bir problemidir. Bu çalışmada geçmiş elektrik verileri kullanılarak kısa vadeli elektrik talep tahmini yapılmıştır. Tahminlemede ARIMA ve XGBoost modelleri kullanılmıştır. Avustralya'nın Victoria eyaletinin 2000 yılından 2019 yılına kadar alınmış yarım saatlik elektrik talep verileri kullanılmıştır. Sadece ocak ayının ilk haftasının tahminlenmesi amaçlanmıştır. Bu amaçla bu verilerin sadece yılın ilk haftası kullanılmış, bu şekilde tüm modellerin kısa vadeli tahminleme performansı karşılaştırılmıştır. ARIMA modelinin parametreleri ACF ve PACF grafiklerinden faydalanılarak oluşturulmuştur. XGBoost modelinin parametreleri ise karşılaştırılarak seçilmiştir. XGBoost modelinin eğitimi için zaman bilgisi ayrılarak bir parametre seti oluşturulmuş ve sonrasında bu çalışmaya özgü üretilmiş parametreler eklenmiş ve tüm modellerin karşılaştırılması yapılmıştır. Sonuç olarak görülmüştür ki, ARIMA ve XGBoost modelleri ilk kıyaslandığında benzer sonuçlar vermektedir. Ancak eklediğimiz parametrelerin XGBoost modelinin doğrulunu kayda değer şekilde arttırdığı ve istatistiksel yaklaşımdan daha iyi sonuçlar elde ettiği görülmektedir. MAPE değeri 0.10'dan 0.05'e düşürülmüş ve hata azaltılmıştır. Bu yaklaşımla yapay zekâ modellerine, verinin durumu gözetilerek, eklenecek parametreler ile doğruluğunun arttırılabileceği anlaşılmaktadır. Gelecek çalışmalar için zaman serisinden elde edilen parametreler dışında, sıcaklık, nüfus gibi dış etmenlerinde eklenip aynı etmenlerle yapay zekâ ve istatistiksel modellerin karşılaştırılması yapılabilir.

ELECTRICAL DEMAND FORECAST WITH ARIMA AND XGBOOST MODELS

SUMMARY

Keywords: ARIMA, time series, statistical analysis, XGBoost, Electricity demand forecasting

Electric energy is a type of energy found in secondary energy groups that society needs in every field. Electrical energy is becoming more widely available in different fields in line with technological developments. Generating electrical energy in a way to meet the increasing needs and transferring it to the users in the right way at the right time is a very important problem. In case of excessive generation of electrical energy, storage is very inefficient in terms of cost and feasibility. Likewise, if electrical energy cannot meet the need, many problems may arise in terms of industry and social life. Due to the nature of electrical energy, it must be supplied as needed and delivered to industry and society. In this respect, countries or other organizations need to plan their energy policies, investments and operations sufficiently before in order to provide safe and efficient electricity. It may take years to implement a technical or strategic decision made in this area. To realize these processes, a country, organization, companies, etc. It needs to forecast future electricity demand and then plan every detail about it. In this study, a short term electricity demand forecast has been made. ARIMA and XGBoost models were used in the estimation. Australia's state of Victoria demanded nuclear water without question from 2000 to the 2019 market. It is intended to predict only the first week of January. This route is only in the first week of summer, these all models are analyzed for short term forecasting. While applying the ARIMA model, ACF and PACF graphics were used and the ARIMA model was created. The XGBoost model was chosen by comparison. For the training of the XGBoost model, a parameter set was created as time information and then these end parameters were added and all models were compared. As a result, it is seen that ARIMA and XGBoost models give similar results when first compared. It is seen that the XGBoost model increases its value and the parameters it learns give better results. The MAPE value was reduced from 0.10 to 0.05, the error was reduced. It is understood that with this message, artificial intelligence models can be increased with the parameter to be added, taking into account the data. For future research, it can be made to compare real intelligence and similar models with the same factors in external factors such as parameters, temperature, population, obtained from time series.

BÖLÜM 1. GİRİŞ

Elektrik enerjisi toplumun her alanda ihtiyaç duyduğu ikincil enerji gruplarında bulunan bir enerji türüdür. Elektrik enerjisi, teknolojik gelişmeler doğrultusunda farklı alanlarda daha yaygın bir şekilde kullanılabilir hale gelmektedir. Elektrik enerjisinin artan ihtiyaçları karşılayabilecek şekilde üretilip, doğru zamanda doğru şekilde kullanıcılara aktarılması, oldukça önemli bir problemdir. Elektrik enerjisinin fazla üretilmesi durumunda depolanması maliyet ve fizibilite açısından oldukça verimsizdir. Aynı şekilde elektrik enerjisinin ihtiyacı karşılayamaması durumunda sanayi ve toplum yaşamı açısından birçok problem ortaya çıkabilmektedir. Elektrik enerjisinin doğası gereği ihtiyaç kadar arz edilip, sanayi ve topluma iletilmesi gerekmektedir. Bu açıdan güvenli ve verimli elektrik sağlamak için ülkelerin veya diğer kuruluşların enerji politikalarını, yatırımlarını ve operasyonlarını daha önce yeterince planlamaları gerekmektedir. Bu alanda alınmış teknik veya stratejik bir kararın uygulanması yıllar alabilir. Bu süreçleri gerçekleştirmek için bir ülkenin, kuruluşun, şirketlerin vb. gelecekteki elektrik talebini tahmin etmesi ve daha sonra bununla ilgili her ayrıntıyı planlaması gerekmektedir. Gelecekteki elektrik yükünü ihtiyaçlara göre farklı zaman aralıklarında tahminlemek için farklı yöntemler kullanılır. Bu tahminler genellikle zaman aralığına göre 3 farklı bölümde kategorize edilir:

1. Kısa Vadeli Tahminleme (KVT-STLF) : Birkaç dakikadan 24 saat aralığına kadar yapılan tahminlemeler.
2. Orta Vadeli Tahminleme (OVT-MTLF) : Birkaç gün den birkaç aya kadar değişebilen zaman aralığında tahminleme yapılır.
3. Uzun Vadeli tahminleme (UVT-LTLF) : En az bir yıl ve daha fazlası olabilen tahminleme aralığıdır.

Yenilenebilir enerji kaynaklarının doğası, üreticilerin üretimi tam olarak kontrol etmesine izin vermediğinden, en azından üretimi tahmin etmek önemlidir. Güvenilir bir kısa vadeli tahminleme, tesislerin ve enerji sağlayıcıların, yenilenebilir enerjilerin daha çok kullanılması ve gelecekteki akıllı şebekelerin daha karmaşık fiyatlandırma stratejilerinin olmasının getirdiği zorluklarla başa çıkmasına yardımcı olur.

Bu çalışma da öncelikle Bölüm 2.'de konuyla ilgili çalışmalar irdelenmiş ve literatürdeki benzer çalışmalar ve metotlar ortaya konmuştur. Çalışmanın Bölüm 3. kısmı bu çalışmada kullanılmış olan yöntemlerle ilgili teorik bilgileri içermekte olup, ayrıca çalışmada kullanılan veri ile ilgili ön bilgileri içermektedir. Bölüm 4.'de ise çalışma gerçekleştirilme adımları açıklanmıştır. Bölüm 5. ise uygulanan yöntem ile elde edilen sonuçların değerlendirilmesini içermektedir. Çalışmamız; denetimli öğrenme metodu olan XGBoost yöntemi ile, denetimsiz öğrenme metodu olan SARIMA metodunu, zaman serisi analizinde karşılaştırmaktadır. Bu karşılaştırma XGBoost modelinin hata oranını yarı yarıya azaltacak özellikler formüle edilip, model optimize edilerek zenginleştirilmiştir.

BÖLÜM 2. LİTERATÜR TARAMASI

Kısa vadeli tahminleme yaklaşımları arasında ARIMA en temel modeldir. ARIMA temelli modeller normal şartlar altında verimli çalışsa bile, özel günler, mevsimsel sıcaklık değişimleri gibi belirli değişkenlere bağlı oluşan durumları karşılamayabilir. İlgili çalışmada [1] ARIMA tabanlı bir model farklı mevsimler için farklı modelleri baz alarak saatlik yük tahminleri yapmaktadır. Bu şekilde ARIMA modelinin bu zayıflığı, model sayısı artırılarak telafi edilmiştir [1]. Kısa vadeli tahminleme için kullanılabilen diğer bir yöntemde tahmin denklemini düzey, trend ve sezonsallık olarak üç bölüme ayıran Holt-Winters yaklaşımıdır. İlgili çalışmada [2] ARIMA modeli ve Holt-Winter yöntemlerini kısa vadeli tahminleme yaklaşımı olarak ele alıp karşılaştırmıştır. İlgili çalışmada on farklı Avrupa ülkesinden yük verileri kullanarak ilgili verinin mevsimselliği Holt-Winters yöntemiyle düzeltilmesinin ARIMA modelinin doğruluğunu arttırdığı gözlemlenmiştir. Tahminlemede sıcaklık, nem, mevsim, özel gün gibi farklı parametreler kullanılarak geliştirilen, en küçük kareler bazlı regresyon teknikleri kısa vadeli tahminleme yaklaşımları arasında oldukça popülerdir. Çalışma [3]'de farklı çoklu lineer regresyon temelli farklı yaklaşımlar incelenmiştir. Çalışmada yaz ve kış mevsimi için farklı modeller kullanılmıştır ve koşullara bağlı, farklı tekniklerin sonuçları incelenmiştir. Bir teknik ile seçmeden önce verilerin kapsamlı analizinin yapılması gerektiği sonucuna varılmıştır.

Son zamanlarda yapılan çalışmalarda ise yapay zeka temelli algoritmalar sıklıkla kullanılmaktadır [4], [5], [6]. Yapay sinir ağları temelli algoritmalar istatistiksel işlemler gerektirmeden tahminleme yapabilen bir teknik olarak kendini kanıtlamıştır. Herhangi bir modeli kabul edip, o modele ait parametreleri tahmin etmeye ihtiyaç duyulmadığı için, veriye dayalı bir yöntemdir. Yapay sinir ağları sürekli ve doğrusal olmayan herhangi bir veri için tahminleme yapabilir. Kısa vadeli tahminleme için bazı yapay zekâ yöntemleri uygulanmıştır. Bunlar Bulanık mantık, adaptif yapay sinir

ağları, Bayesian yapay sinir ağları ve geri beslemeli yapay sinir ağlarıdır. Doğruluğu arttırmak için yapay sinir ağları ile farklı yöntemler beraber kullanılmıştır, örneğin gradient metodu [7], bulanık uzman sistem [8], wavelet dönüşümü [9][17], bulanık mantık [10], genetik algoritma [11], destek vektör makinesi [12], optimizasyon modeli [5], tavlama benzetimi [13] ve diğer farklı metotlarda enerji tahminlemesine uygulanmıştır [14], [15], [16]. Yapay zekâ modelleri kullanılarak yapılmış olan çalışmalarda, modeli iyileştirmek ekstra çevresel bir değişken ya da bilinen istatistiksel parametrelerin, eğitim verisine yeni bir özellik eklenmesi ile yapılmaktadır. Bizim çalışmamız ise, veri setine özel olarak ayarlanıp, modelin eksik kaldığı yönlerde gelişmesini sağlayan, özel, sadece zaman bağlı yeni istatistiksel parametreler önermektedir.

BÖLÜM 3. YÖNTEM

Kısa vadeli talep tahmini için literatürde farklı yöntemler kullanılmaktadır. Elektrik enerji talep verisi, zaman serisi olmasından dolayı zaman serisi metotlarıyla analiz edilebilmektedir. Elektrik talep yükü farklı faktörlerden etkilenebilir. Regresyon metodu bağımlı ve bağımsız değişkenler arasında bir ilişki kurar, bu sebeple elektrik enerji talep verileri için kullanmaya oldukça uygundur. Bunun yanında benzer şekilde bağımlı ve bağımsız değişkenler arasında ilişki kuran diğer metotlarda kullanılabilir.

3.1. Zaman Serileri

Elektrik talep verileri, zaman serisi tipinde veri olduğu için, zaman serisi analiz metotlarıyla incelenebilir. Literatürde en çok kabul görmüş yöntemler arasında, ARIMA, üstel düzgünleştirme yöntemi sayılabilir. Zaman serileri, doğası gereği farklı bileşenleri içerisinde barındırır. Bu bileşenler matematiksel yöntemlerle ayrıştırılıp, tahminleme için kullanılacak olan verinin daha temiz olmasını ve tahminlemenin daha iyi olmasını sağlar. Zaman serilerini etkileyen bu bileşenler aşağıdaki gibi listelenebilir:

- Trend Etkisi(T) : Trend etkisi serinin uzun dönem davranışını etkileyen bileşendir. Serinin yönünde belirli bir değişikliğin gözlemlendiği durumlarda trend etkisinin varlığından söz edilebilir.
 - Doğrusal trend eğilimi; $T = a + bt + e$ ifade edilir. İfade de a sabit katsayı, b eğim, e gerçekleşme ile tahmin arasında kalan kalıntılardır.
- Konjonktür Etkisi (C): Tekrarlayan ancak belli bir periyodu olmayan bileşenlerdir.
 - Konjonktür etkisi hareketli ortalamaların trende bölünmesiyle bulunur.

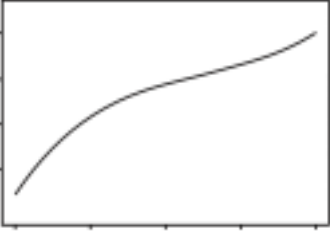
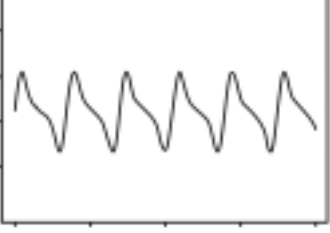
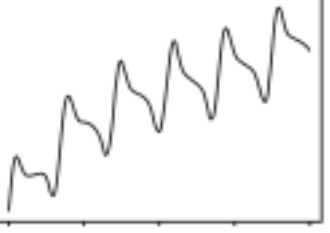
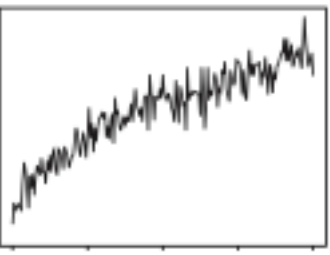
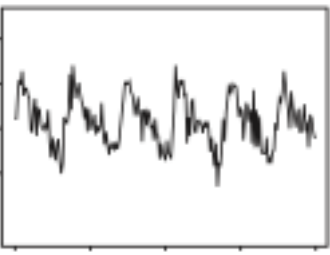
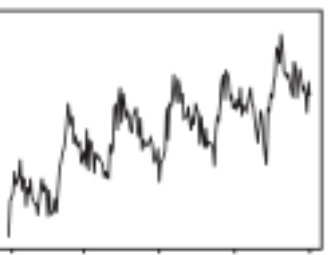
- Sezonallık Etkisi (S) : Zaman serisinin, dönemsel faktörlerden etkilendiği durumlarda varlığından söz edilebilir. Belirli veya bilinen bir zaman aralığında tekrarlayan bir etki gösterir.
 - Sezonallık; gerçek serinin, hareketli ortalamalar serisine bölümüyle bulunmaktadır.
- Düzensiz Hareketler (I): Rastgele ve düzensiz etkiler olarak açıklanır. Tüm diğer etkiler arındırıldıktan sonra kalan ve düzensiz, rastgele olan etkidir.

Bu etkiler; T trend, C konjonktür, S sezonallık, I düzensiz hareketler olmak üzere, iki farklı modelle ifade edilebilir:

Toplamsal model: $Y=T+C+S+I$

Çarpımsal model : $Y=TxCxSxI$

Tablo 3.1. Zaman serisi bileşenleri [16]

 Şekil 3.1. Trend etkisi	 Şekil 3.2. Konjonktür etkisi	 Şekil 3.3. Trend + konjonktür etkisi
 Şekil 3.4. Trend + düzensiz etkiler	 Şekil 3.5. Konjonktür + düzensiz etkiler	 Şekil 3.6. Trend + konjonktür + düzensiz etkiler

Tablo 3.1.'de zaman serilerini oluşturan bileşenler verilmiştir. Bu bileşenler zaman serisinin durağanlığını ve lineerliğini bozmakta olup, tüm tahminleme yöntemlerinin

temel amacı seriyi bu bileşenlerden ayırıp daha lineer ve tahminlenebilir bir hale getirmektir.

Trendin varyasyonları, zaman serisinin seviyesine göre değişmediğinde toplamsal model kullanılırken trend, zaman serisinin seviyesiyle orantılıysa çarpımsal model daha uygun olmaktadır [17].

3.1.1. Birleşik otoregresif hareketli ortalama yöntemi (ARIMA)

Otoregresif hareketli ortalama (ARMA) modelleri olarak adlandırılan genel ve kullanışlı bir zaman serisi modelleri oluşturmak için otoregresif (AR) modelleri hareketli ortalama (MA) modelleri ile etkili bir şekilde birleştirilebilir. Ancak, yalnızca veriler durağanken kullanılabilirler. Bu model sınıfı, veri serilerinin farklılaşmasına izin verilerek durağan olmayan serilere genişletilebilir. Bunlara otoregresif entegre hareketli ortalama (ARIMA) modelleri denir. Box ve Jenkins (1970) ARIMA modellerini popülerleştirmiştir.

Çok çeşitli ARIMA modelleri vardır. Genel mevsimsel olmayan model ARIMA (p, d, q) olarak bilinir:

AR: p = otoregresif parçanın sayısı

I: d = ilk farkın derecesi

MA: q = hareketli ortalama parçanın sayısı.

3.1.1.1. AR modeli

AR modelleri, geçmiş dönem verilerini içerir ve içerdikleri geçmiş dönem verisi sayısına göre isimlendirilir. İçerdiği geçmiş dönem sayısı p ile gösterilir ve AR modelinin derecesi olarak ifade edilir [18]. Genel AR(p) modelinin ifadesi, x gözlem değerleri, ϕ modelin parametreleri olmak üzere şu şekildedir:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + a_t \quad (3.1)$$

Örnek olarak birinci dereceden AR(1) aşağıdaki gibi ifade edilebilir:

$$x_t = \phi_1 x_{t-1} + a_t \quad (3.2)$$

3.1.1.2. MA modeli

MA modelleri, uzunluğu verilen dönemlerin ortalamasından tahmin çıkaran bir yöntemdir. İçerdiği geçmiş dönem sayısı q ile gösterilir ve MA modelinin derecesi olarak ifade edilir. Genel MA(q) modelinin ifadesi, x gözlem değerleri, θ modelin parametreleri olmak üzere şu şekildedir:

$$x_t = \theta_0 a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} \quad (3.3)$$

Örnek olarak birinci dereceden MA(1) aşağıdaki gibi ifade edilebilir ($\theta_0=1$):

$$x_t = a_t - \theta_1 a_{t-1} \quad (3.4)$$

3.1.1.3. ARMA modeli

ARMA; AR ve MA modellerinin bir kombinasyonudur. p terimli AR(p) ve q terimli MA(q) modelinden oluşan bir ARMA modelinde $p+q$ adet terim vardır ve ARMA(p,q) şeklinde gösterilir. Genel ARMA(p,q) modelinin ifadesi Denklem 3.5.'te verilmiştir.

$$x_t = (\phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p}) - (\theta_0 a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}) \quad (3.5)$$

Örnek olarak birinci dereceden ARMA(1,1) aşağıdaki gibi ifade edilebilir ($\theta_0=1$):

$$x_t = \phi_1 x_{t-1} + a_t + \theta_1 a_{t-1} \quad (3.6)$$

3.1.1.4. ARIMA modeli

AR, MA ve ARMA modelleri durağan serilere uygulanabilen modellerdir. Durağan olmayan seriler için ise ARIMA modeli $I(d)$ parametresi ile serinin durağanlaştırma işlemini yapıp, modellerin seriye uygulanabilmesini sağlamaktadır. Bir zaman serisi ortalama değeri etrafında durağan değilse serinin uygun derecede farkları alınarak durağanlık sağlanır bu değer modelin derecesi d olur. Genel ARIMA(p,q,d) modelinin ifadesi:

$$w_t = (\phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p}) - (\theta_0 a_t + \dots + \theta_q a_{t-q}) \quad (3.7)$$

Denklem 3.7.'yi $w_t = \nabla^d a_t$ şeklinde ifade edecek olursak, ∇ ; fark alma operatörü, d ; fark alma derecesi, w_t ; farkı alınmış seri, B ; backshift operatörü olmak üzere, $d=1$ için fark operatörü:

$$\nabla x_t = w_t = x_t - x_{t-1} \quad (3.8)$$

Dolayısıyla ARIMA(1,1,1) modelinin eşitliği de aşağıdaki gibi ifade edilebilir:

$$\nabla x_t - \nabla x_{t-1} = a_t - \theta_1 a_{t-1} \quad (3.9)$$

veya;

$$(1 - \phi_1 B) \nabla x_t = (1 - \theta_1 B) a_t \quad (3.10)$$

şeklinde ifade edilebilir.

3.2. Regresyon

Farklı tahminleme yöntemleriyle karşılaştırıldığında, regresyon modellerinin geliştirilmesi, kullanımı ve yorumlanması basittir, bu nedenle diğer daha karmaşık tekniklere kıyasla, yük tahmini için yaygın olarak kullanılmaktadır [19]–[22].

Regresyon modelleri, enerji tüketimini, elde bulunan ilişkili diğer parametrelerle (örneğin dış hava ve iç bina parametreleri) ilişkilendirmektedir. Bu modeller gerçek geçmiş yük verileri [23], [24] veya benzetilmiş yük verileri [25]–[27] kullanılarak geliştirilebilir. Ayrıca, regresyon modelleri uygulama alanları, yaygın olarak kullanılan girdi parametreleri, model performansını artırma yöntemleri ve diğer modellerle karşılaştırma gibi farklı kategoriler altında sınıflandırılmaktadır.

3.2.1. Lineer regresyon

Bu modelde bir bağımlı ve bir bağımsız değişken arasındaki lineer bağıntı, farklı matematiksel yöntemler kullanılarak ifade edilerek, tahminleme modeli oluşturulmaktadır. Örneğin;

$$i = 1,2,3,4 \dots, n \text{ için } (X_i, Y_i) \quad (3.11)$$

Bağımlı ve bağımsız değişkenlerin birbiriyle olan ilişkisi Denklem 3.12.'deki bağıntıdaki gibi verilebilir.

$$Y = a + bX + e \quad (3.12)$$

Denklemdeki, a değeri kesim noktası, b değeri eğimi, e değeri de hata payını temsil etmektedir. Bu değerlerin hesaplanması için farklı yöntemler vardır.

3.2.1.1. En küçük kareler yöntemi

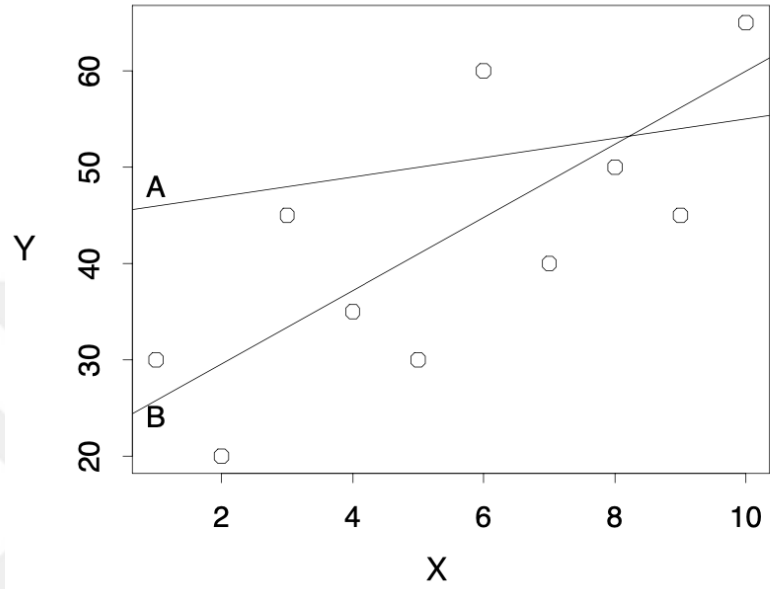
$\hat{Y}$ tahmin değeri olmak üzere, Denklem 3.13.'deki a ve b değerlerini hesaplayıp en az hata ile denklem oluşturulur.

$$\hat{Y} = a + bX + e \quad (3.13)$$

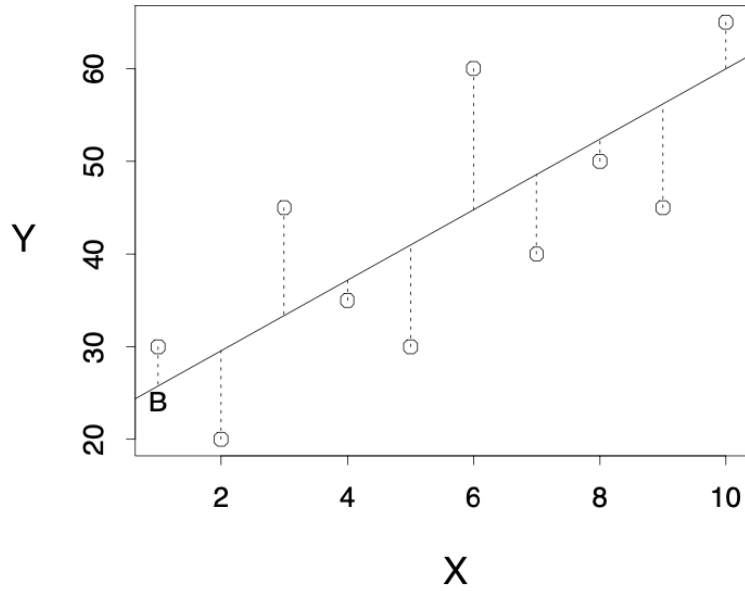
Şekil 3.7.'de gösterilen A ve B doğruları Denklem 3.13.'deki herhangi a ve b değerleri için oluşan tahminlemeleri göstermektedir. Bu tahminleme hataları minimize edilerek

en doğru tahminleme bulunması hedeflenmektedir. Denklem 3.14.'de tahminleme hatası verilmiştir.

$$e_i = Y_i - \hat{Y}_i \quad (3.14)$$



Şekil 3.7. Örnek veriler ve örnek tahminleme



Şekil 3.8. Örnek tahminleme B için hata gösterimleri

Denklem 3.14.'de formülü verilen hata değeri minimize edilmelidir. Tahminlemenin tüm datamız için doğruluğunu arttırıp, hata değerlerini azaltmak için bu hata değerini her bir veri noktamız için hesaplamalıyız. Hataların negatif ve pozitif değerlerinin birbirini sıfırlamaması için karelerini almak zorundayız. Hesaplamalarımızı tüm verileri işleme katarak yapmalıyız.

$$\text{Hata Kareleri Toplamı} = e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2 = \sum_{i=1}^n e_i^2 \quad (3.15)$$

Oluşturacağımız tahminlemenin amacı Denklem 3.15.'deki değeri minimize etmek olacaktır. Bunun için, hata değerini Denklem 3.13. ve 3.14.'den çekip yerine yazarsak;

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - a - bX_i)^2 \quad (3.16)$$

Son olarak elde ettiğimiz denklemden a ve b değerlerini çektiğimiz zaman denklemimizi oluşturmuş olacağız.

$$b = \frac{\sum_{i=1}^n (X_i - \hat{X}_i) (Y_i - \hat{Y}_i)}{\sum_{i=1}^n (X_i - \hat{X}_i)^2} \quad (3.17)$$

$$a = \hat{Y} - b\hat{X} \quad (3.18)$$

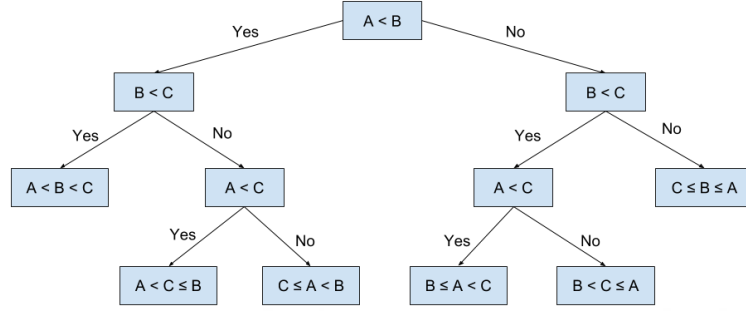
İfadesi ile denklemimiz elde edilmiş olur. Bu denklem doğrultusunda yapılan minimizasyon süreci Şekil 3.7. ve Şekil 3.8. ile ifade edilebilir.

3.2.2. Ağaç bazlı regresyon yöntemleri

3.2.2.1. Karar ağaçları

Karar ağaçları, hedef değeri tahmin etmek için veri kümesini daha küçük alt kümelerle nasıl bölmesi gerektiğini öğrenir. Koşul veya test “yaprak” (düğüm) ve olası sonuçlar

“dallar” (kenarlar) olarak temsil edilir. Bu ayırma işlemi, başka kazanç sağlanamayana veya önceden ayarlanmış bir kural karşılanıncaya kadar devam eder. Bu yöntemi ifade eden karar ağacı yapısı bir Şekil 3.9.’da gösterilmiştir.



Şekil 3.9. Karar ağacı yapısı

$$\hat{y} = \hat{f}(x) = \sum_{m=1}^M c_m I\{x \in R_m\} \quad (3.19)$$

Her örnek bir yaprak düğüme denk gelir. $I\{x \in R_m\}$, x , R_m alt kümesindeyse 1, aksi takdirde 0 döndüren özdeşlik fonksiyonudur. Bir örnek, yaprak düğüm R_l 'e düşerse, tahmin edilen sonuç $\hat{y} = c_l$ olur, burada c_l , yaprak düğüm R_l 'deki tüm eğitim örneklerinin ortalamasıdır.

3.2.2.2. Ekstrem gradyan yükseltme (XGBoost)

XGBoost, 2016 yılında Chen tarafından önerilmiştir [28]. Sınıflandırma ve Regresyon Ağacına (CART) dayanır, bölüm özniteliklerini yeniden tanımlar ve bölüm özniteliklerini belirlemek için kayıp işlevinin en aza indirilmesini kullanır. Bu yöntem, elektrik yükünü tahmin etmek [29], petrol fiyatını tahmin etmek [30], kazayı tahmin etmek [31] için başarılı bir biçimde kullanılmıştır. Doğruluğu diğer modellere göre daha yüksek ve eğitim maliyeti de diğer modellerden daha azdır. XGBoost, kayıp işlevini en aza indirmek için en uygun özelliği belirleyin, kaybı en aza indirir, Denklem 3.20. ile tanımlanır:

$$Obj(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (3.20)$$

$L(y_i, \hat{y}_i)$ i-inci numunenin eğitim hatası anlamına gelir, T yaprak düğümlerin sayısı, w yaprak düğümlerin puanı ve γ, λ katsayıdır.

K'inci ağaç eklendiğinde, önceki ağaç eğitilmiştir. Eğitim hatası ve önceki ağacın normal süresi sabit hale gelir. Kayıp işlevi aşağıdaki gibi yazılır:

$$Obj(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 + C \quad (3.21)$$

Kayıp fonksiyonunun Ortalama Kare Hata fonksiyonu olduğunu ve Denklem 3.20.'nin ikinci dereceden Taylor açılımı olduğunu varsayarsak sonuç, Denklem 3.22. ile ifade edilebilir.

$$Obj^t(\theta) \approx \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (3.22)$$

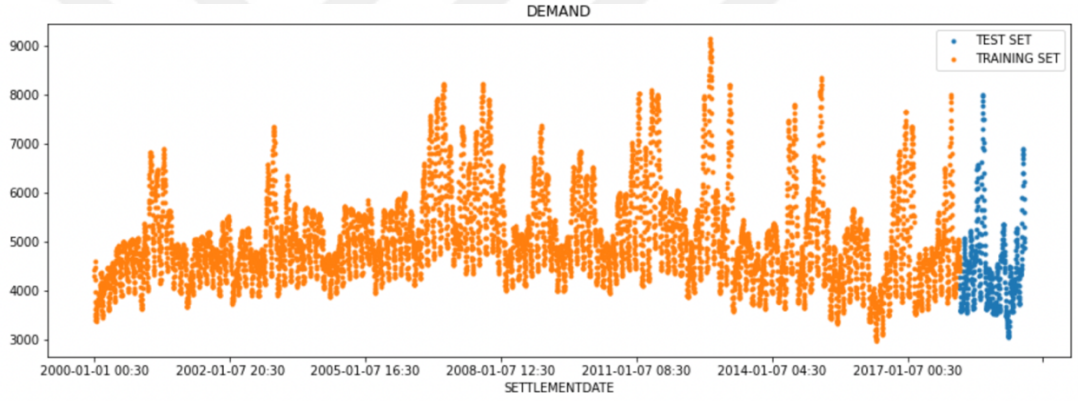
Değerleri yerine koyarsak, $g_i = \partial_{\hat{y}^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)})$, $h_i = \frac{1}{2} \partial_{\hat{y}^{(t-1)}}^2 L(y_i, \hat{y}_i^{(t-1)})$

$$Obj^t(\theta) = \sum_{j=1}^T \left[\left(\sum_{i \in I} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I} h_i + \lambda \right) w_j^2 \right] + \gamma T \quad (3.23)$$

w_j , bu yaprak düğümünün j-inci puanıdır. Denklem 3.23. için w_j ile ilgili kısmi türevi hesaplanarak, en uygun yaprak düğüm skorunu w_j^* ve minimum kayıp değeri Obj^* elde edilebilir.

BÖLÜM 4. UYGULAMA

Elimizde Avustralya'nın Victoria eyaletine ait 2000-2019 yıllarına ait yarım saatlik elektrik talep verileri bulunmaktadır. Bu çalışmada ocak ayının ilk haftası tahminlemesi hedeflenmekte olduğu için, her yılın sadece ocak ayının ilk haftası model oluşturulurken kullanılmıştır. Her iki modelde de 2019 yılına kadar olan veriler eğitim verisi, kalan veri ise test verisi olarak kullanılmıştır.



Şekil 4.1. Eğitim ve test verileri

4.1. ARIMA Modeli

4.1.1. Diferansiyel katsayısı (d) hesabı

Seriye diferansiyel işlem yapılmasının sebebi seriyi durağanlaştırmaktır. Serimizin ortalaması, varyansı ve kovaryansı zaman boyunca sabit kalıyorsa, durağan olduğu söylenebilir. Ancak diferansiyel katsayısı diğer model parametrelerini de etkileyebileceği için dikkatli seçilmelidir. Doğru seçilmiş bir diferansiyel katsayısı oldukça küçük olup, ortalama değer etrafında değer alır ve oto korelasyon fonksiyonunu olabildiğince sıfır etrafında tutmalıdır. Dolayısıyla ilk olarak, statmodels paketinden Augmented Dickey Fuller testi (`adfuller()`) kullanarak serinin

durağan olup olmadığı kontrol edilecektir. Çünkü, sadece seri durağan değilse diferansiyele ihtiyacınız vardır. Aksi takdirde, diferansiyel gerekmez, yani $d = 0$ olur. ADF testinin p değeri, zaman serilerinin durağan olup olmadığının tespitinde kullanılır. Testin p değeri önem seviyesinden (0.05) düşükse, sıfır hipotezini reddedersiniz ve zaman serisinin gerçekten durağan olduğunu çıkarırsınız. Yani, bizim durumumuzda, $P \text{ Değeri} > 0.05$ olduğunda diferansiyel katsayısını bulmamız gerekir. Bunun için verimiz örneklenip, adfuller fonksiyonu uygulanmıştır:

Kod :

```
from statsmodels.tsa.stattools import adfuller
import pandas as pd
from numpy import log
from google.colab import files
from google.colab import drive
drive.mount('/content/gdrive')
data = pd.read_csv('/content/gdrive/My
Drive/demand/AllData.csv', index_col=0)
data.index = pd.to_datetime(data.index)
result=adfuller(data['TOTALDEMAND'].dropna().resample('MS').mean(
))
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
```

Çıktı:

```
ADF Statistic: 0.804343
p-value: 0.991706
```

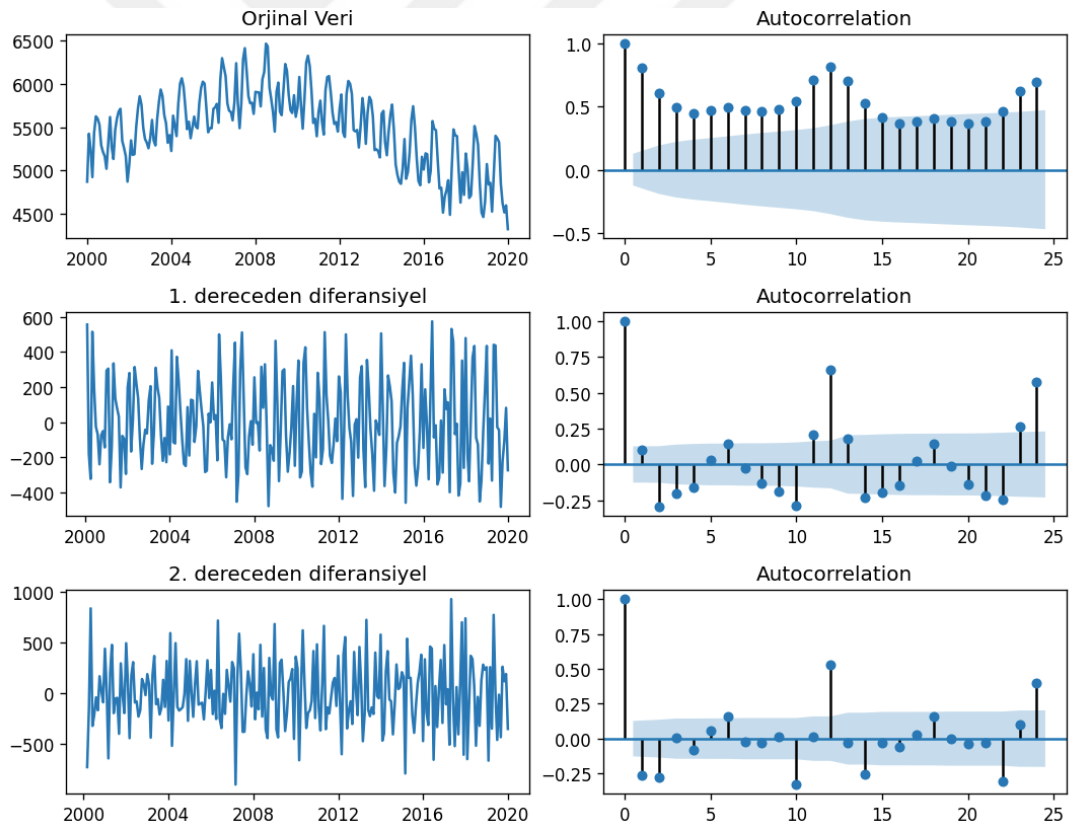
P-değeri önem seviyesinden daha büyük olduğundan, diziyi diferansiyel almaya devam edilir ve otokorelasyon grafiği elde edilir. Aşağıdaki kod bloğunda serimizin diferansiyellerini ve otokorelasyon grafiklerini çizdiriyoruz.

```
import numpy as np, pandas as pd
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import matplotlib.pyplot as plt
plt.rcParams.update({'figure.figsize':(9,7), 'figure.dpi':120})
```

```

df = data['TOTALDEMAND'].dropna().resample('MS').mean()
df.index =
pd.to_datetime(data.dropna().resample('MS').mean().index)
fig, axes = plt.subplots(3, 2)
plot_acf(df, ax=axes[0,1])
axes[0, 0].plot(df); axes[0, 0].set_title('Orjinal Veri')
plot_acf(df.diff().dropna(), ax=axes[1,1])
axes[1, 0].plot(df.diff().dropna()); axes[1, 0].set_title('1.
dereceden diferansiyel')
plot_acf(df.diff().diff().dropna(), ax=axes[2,1])
axes[2, 0].plot(df.diff().diff().dropna()); axes[2,
0].set_title('2. dereceden diferansiyel')
fig.tight_layout()
plt.show()

```



Şekil 4.2. Farklandırılmış seriye ait grafikler

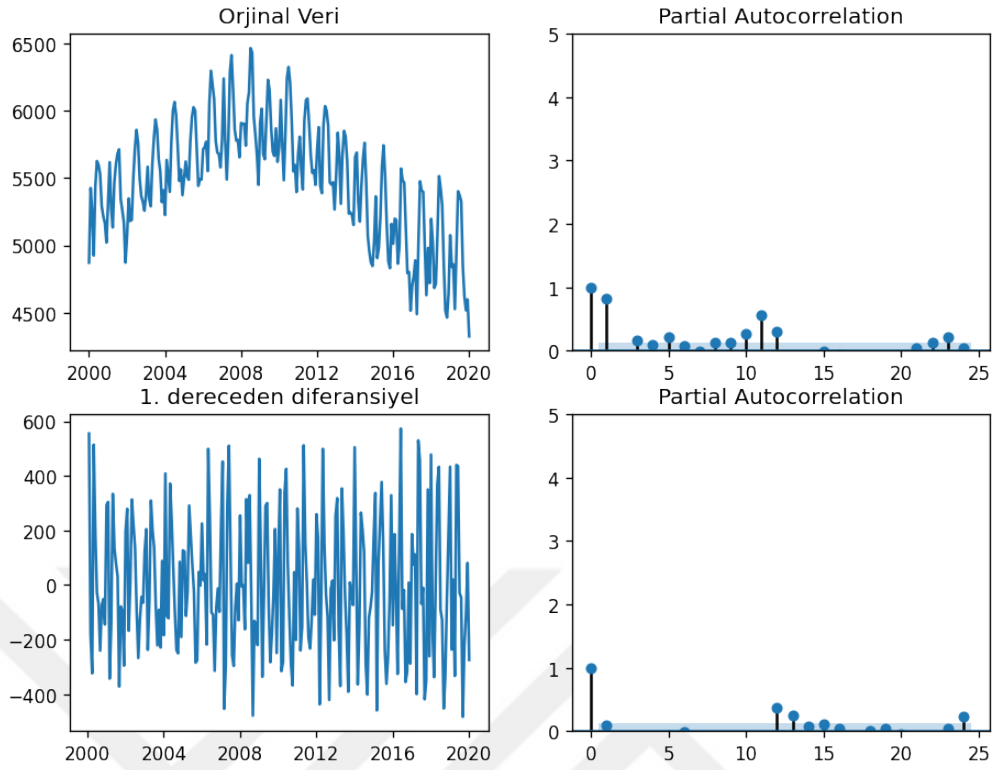
Serimiz Şekil 4.2.'de görüldüğü gibi 2.dereceden diferansiyeli alınınca durağanlaştı ancak, oto-korelasyon grafiğinde, peş peşe olan noktaların, aniden sıfıra inmesi serinin

aşırı farklılaştırılmış olduğunu göstermektedir. Bu sebeple serinin 1. dereceden farklılaştırılmış modeli tercih edilir.

4.1.2. Gecikme derecesi (p) hesabı

Modelimize kaçınıcı dereceden bir gecikme eklememiz gerektiğini bulmak için, parçalı oto korelasyon yöntemini kullanmamız gerekmektedir [32]. Parçalı oto korelasyon, zaman serisi ve zaman serisinin gecikme eklenmiş hali arasındaki korelasyon olarak ifade edilebilir. Bu sonuca bakarak p kat sayısının modele olacak olan katkısı gözlemlenebilir.

```
df = data['TOTALDEMAND'].dropna().resample('MS').mean()
df.index=
pd.to_datetime(data.dropna().resample('MS').mean().index)
plt.rcParams.update({'figure.figsize':(9,7), 'figure.dpi':120})
fig, axes = plt.subplots(2, 2)
axes[0, 0].plot(df); axes[0, 0].set_title('Orjinal Veri')
axes[0, 1].set(ylim=(0,5))
plot_pacf(df.dropna(), ax=axes[0, 1])
axes[1, 0].plot(df.diff()); axes[1, 0].set_title('1. dereceden
diferansiyel')
axes[1, 1].set(ylim=(0,5))
plot_pacf(df.diff().dropna(), ax=axes[1, 1])
plt.show()
```



Şekil 4.3. AR parametresi (p) grafiği

Şekil 4.3.'den anlaşılacağı üzere, mavi bölge yani güven aralığındaki değişim göz önüne alınırsa, birinci dereceden almak yeterli bir düzeltme sağlıyor.

4.1.3. Hareketli ortalama derecesi (q) hesabı

AR(p) teriminin derecesi için PACF grafiğine baktığımız gibi, MA(q) terimlerinin derecesi için ACF grafiğine bakabiliriz[32]. MA(q) terimi teknik olarak gecikmeli tahminin hatasıdır. ACF, durağanlaştırılmış serilerdeki herhangi bir otokorelasyonu kaldırmak için kaçınıcı dereceden MA(q) terimine ihtiyaç olduğunu söyler.

Farklı serilerin otokorelasyon grafiğini görelim, bunu için verinin örneklem aralığını belirleyip, diferansiyelini aldık:

```
import pandas as pd
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import matplotlib.pyplot as plt
plt.rcParams.update({'figure.figsize':(9,12), 'figure.dpi':120})
```

```

df = data['TOTALDEMAND'].dropna().resample('MS').mean()
df.index =
pd.to_datetime(data.dropna().resample('MS').mean().index)

fig, axes = plt.subplots(3, 2)

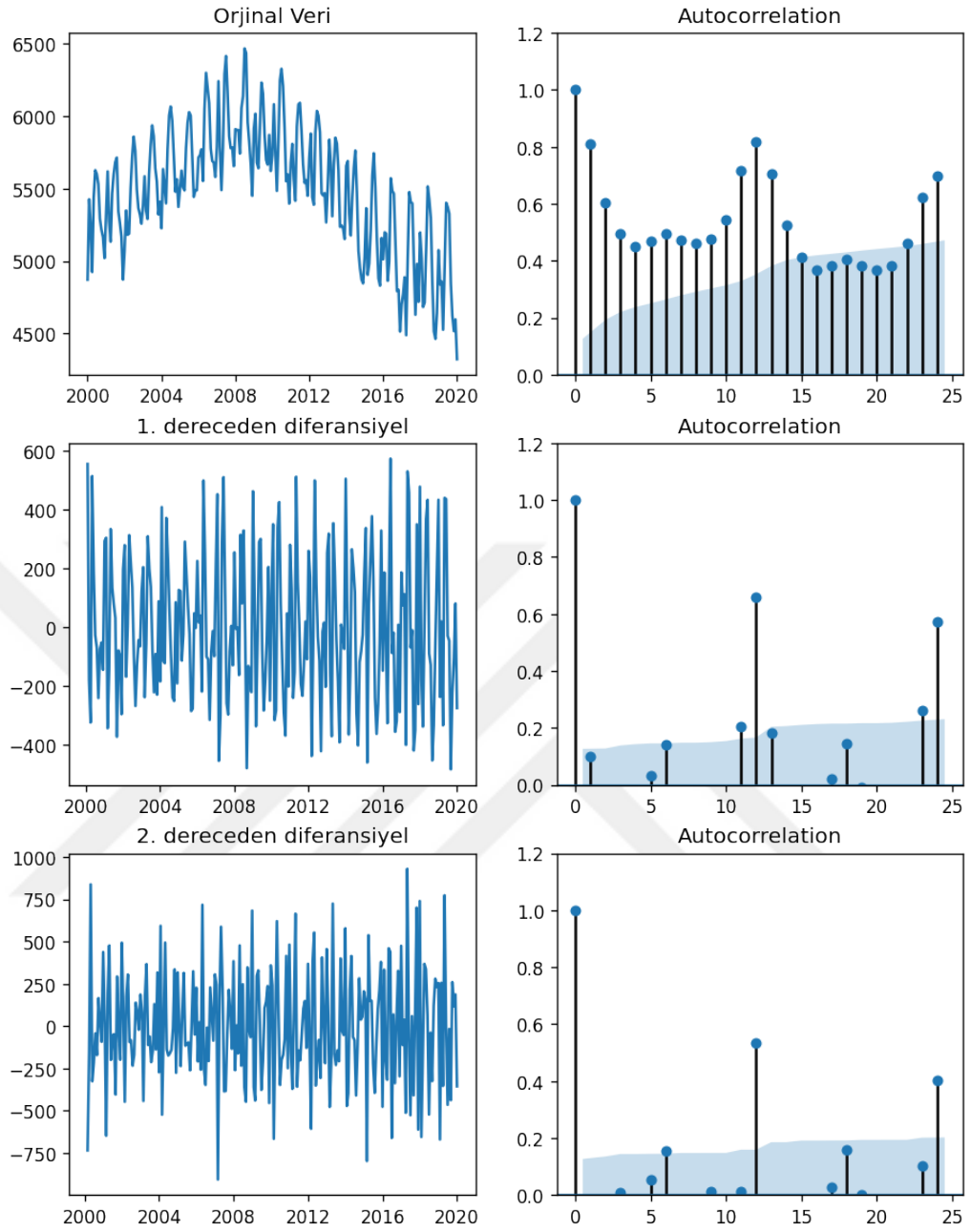
axes[0,0].plot(df); axes[0,0].set_title('Orjinal Veri')
axes[0,1].set(ylim=(0,1.2))
plot_acf(df.dropna(), ax=axes[0,1])

axes[1,0].plot(df.diff()); axes[1,0].set_title('1. dereceden
diferansiyel')
axes[1,1].set(ylim=(0,1.2))
plot_acf(df.diff().dropna(), ax=axes[1,1])

axes[2,0].plot(df.diff().diff()); axes[2,0].set_title('2.
dereceden diferansiyel')
axes[2,1].set(ylim=(0,1.2))
plot_acf(df.diff().diff().dropna(), ax=axes[2,1])

plt.show()

```



Şekil 4.4. MA (q) derecesi hesaplanması grafiği

Şekil 4.4.'te görüldüğü üzere, 2. Dereceden oto korelasyon grafiğinin güven aralığındaki değerleri daha fazla olduğu için 2. Derece MA(q) tercih edilebilir.

4.1.4. ARIMA modelinin oluşturulması

Artık p, d ve q değerlerini belirlediğimize göre, ARIMA modelini oluşturabiliriz. Modelimizi oluşturmak için Statisticsmodels paketinde ARIMA() uygulamasını kullanacağız.

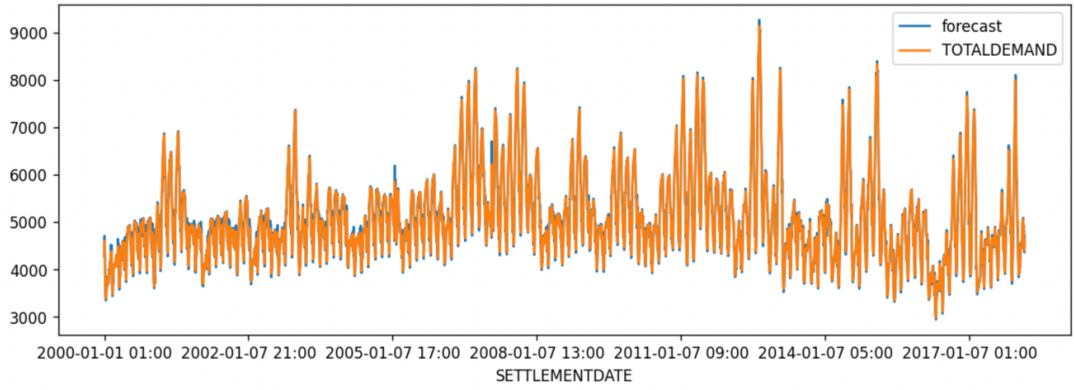
```
# p=1,d=1,q=2 ARIMA Model
model = ARIMA(data_train, order=(1,1,2))
model_fit = model.fit()
print(model_fit.summary())
```

Çıktı:

ARIMA Model Results						
Dep. Variable:	D.TOTALDEMAND	No. Observations:	6383			
Model:	ARIMA(2, 1, 1)	Log Likelihood	-39740.105			
Method:	css-mle	S.D. of innovations	122.360			
Date:	Sun, 04 Apr 2021	AIC	79490.211			
Time:	20:25:23	BIC	79524.017			
Sample:	1	HQIC	79501.914			
	coef	std err	z	P> z	[0.025	0.975]
const	-0.0546	4.167	-0.013	0.990	-8.221	8.112
ar.L1.D.TOTALDEMAND	1.0665	0.027	38.932	0.000	1.013	1.120
ar.L2.D.TOTALDEMAND	-0.1859	0.019	-9.984	0.000	-0.222	-0.149
ma.L1.D.TOTALDEMAND	-0.6752	0.024	-28.690	0.000	-0.721	-0.629
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	1.1804	+0.0000j	1.1804	0.0000		
AR.2	4.5580	+0.0000j	4.5580	0.0000		
MA.1	1.4811	+0.0000j	1.4811	0.0000		

Modelimizin AIC ve Log Likelihood değerleri modelimizin doğruluğu hakkında bilgi vermektedir. AIC değerinin yüksek olması ve Log Likelihood değerinin düşük olması istenmektedir.

```
model_fit.plot_predict(dynamic=False)
plt.show()
```

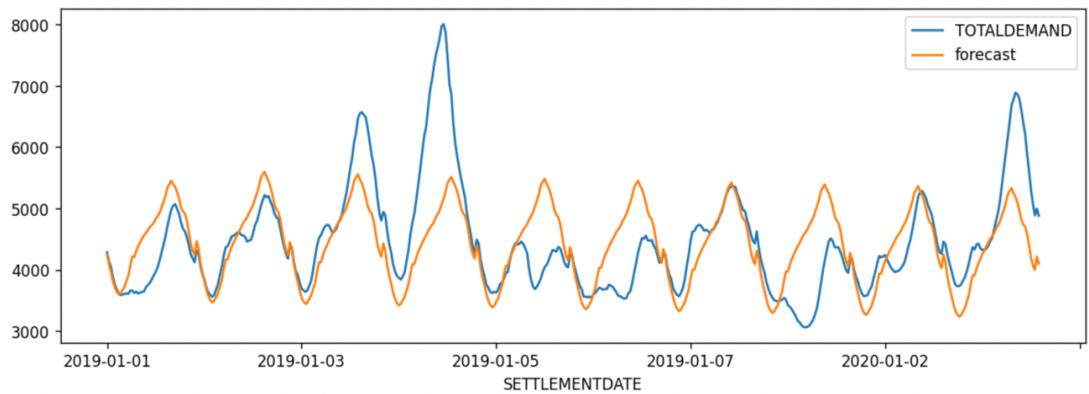


Şekil 4.5. Model ve gerçek veri karşılaştırması

Oluşturmuş olduğumuz model 2019 yılına kadarki verilerle eğitilmiştir. Sonrasında ise sezonsallık parametreleri de eklenerek tahminleme yapılmıştır.

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
mod = SARIMAX([int(m) for m in data_train],
               order=(1, 1, 2),
               seasonal_order=(1, 1, 2, 48),
               enforce_stationarity=False,
               enforce_invertibility=False)

results = mod.fit(dispatch=False)
sarima = results.forecast(steps=len(data_test), alpha=0.05)
```



Şekil 4.6. ARIMA gelecek tahminleri ve gerçek değerler

Şekil 4.6.'da eğitilmiş olan ve sezonsallıkta eklenmiş olan ARIMA modelimizin 2019 ve 2020 yıllarına ait ocak ayının ilk haftasını tahminlenmesi grafiği verilmiştir.

4.2. XGBoost Modeli

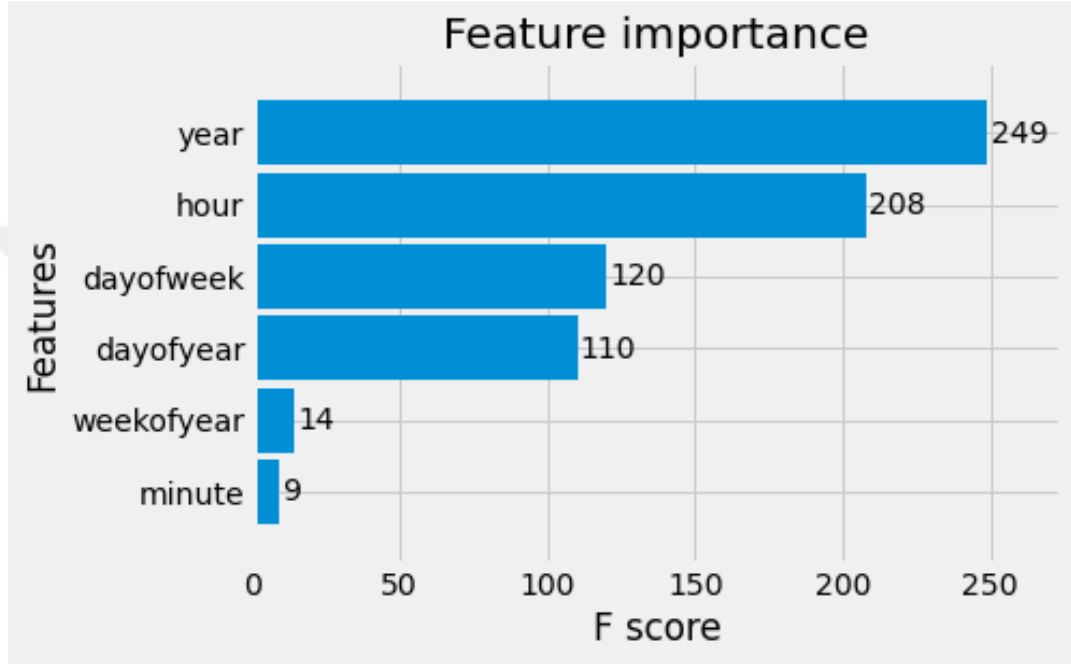
XGBoost yöntemi, gözetimli öğrenme metodudur, verimizin gözetimli modellerde kullanılabilir olmasını sağlamak için veriye, sonuç verisi dışında, sonuç verisi ile ilişkili parametreler eklenmesi gerekmektedir. Bizim verimiz zaman serisi olduğu için verimize ekleyeceğimiz parametreleri tarih bilgisinden alabiliriz [33]. Tarih parametreleri eklendikten sonra, veri kümemiz Tablo 4.1.'deki hali almaktadır.

Tablo 4.1. XGBoost verileri

DATE	DEMAD	hour	min	dayof week	quarter	month	year	dayof year	dayof month	week ofyear
2000-01-01	4419,037	0	30	5	1	1	2000	1	1	52
2000-01-01	3718,437	14	0	5	1	1	2000	1	1	52
2000-01-01	3694,442	14	30	5	1	1	2000	1	1	52
2000-01-01	3707,042	15	0	5	1	1	2000	1	1	52
2000-01-01	3734,84	15	30	5	1	1	2000	1	1	52
2000-01-01	3823,265	16	0	5	1	1	2000	1	1	52
2000-01-01	3857,703	16	30	5	1	1	2000	1	1	52
2000-01-01	3957,37	17	0	5	1	1	2000	1	1	52
2000-01-01	3938,97	17	30	5	1	1	2000	1	1	52
2000-01-01	3939,78	18	0	5	1	1	2000	1	1	52
2000-01-01	3724,937	13	30	5	1	1	2000	1	1	52
2000-01-01	3931,095	18	30	5	1	1	2000	1	1	52

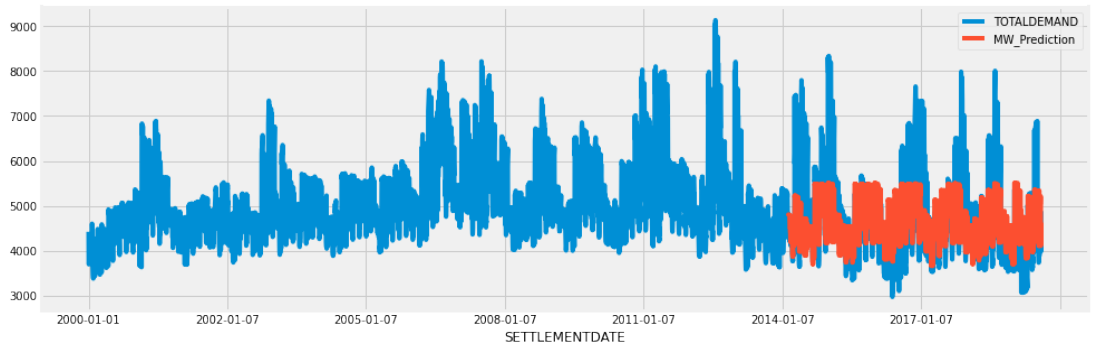
Kod:

```
reg = xgb.XGBRegressor(n_estimators=1000)
reg.fit(X_train, y_train,
        eval_set=[(X_train, y_train), (X_test, y_test)],
        early_stopping_rounds=50,
        verbose=False)
```



Şekil 4.7. Seçilen parametrelerin önem dereceleri

Şekil 4.7. modelde kullanılan parametrelerin model sonucunda ne kadar etkili olduğunu göstermektedir.



Şekil 4.8. XGBoost tahmin sonuçları

Şekil 4.8.'de görüleceği gibi XGBoost modeli pik noktaları bulmakta oldukça zayıf kalmıştır. Modeli bu alanda güçlendirmek için modele yeni parametreler eklenmiştir. Bu parametreler min, max, peak, step olarak isimlendirilmiş olup formül haline getirilmesi aşağıdaki gibidir:

Ismin parametresi hesabı:

$$(\mu - x_t) > (\mu - \min) / 2 \rightarrow 1 \quad (4.1)$$

Ismax parametresi hesabı:

$$(x_t - \mu) > (\mu - \min) / 20 \rightarrow 1 \quad (4.2)$$

Isstep parametresi hesabı:

$$(x_t - x_{t+1}) > ((\mu - \min) + (\max - \mu)) / 2 \rightarrow 1 \quad (4.3)$$

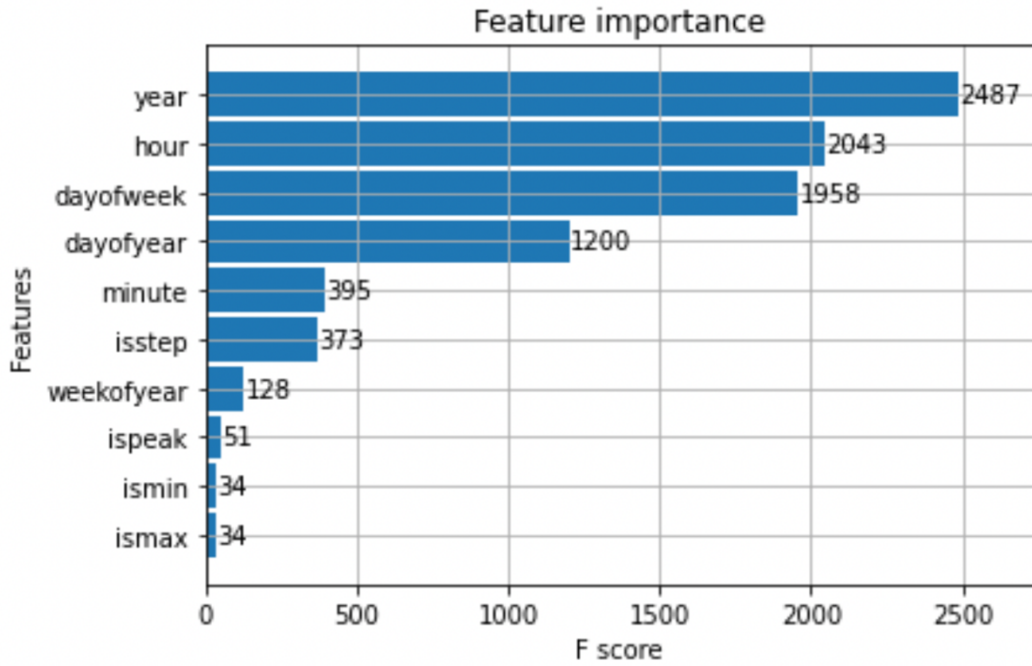
Ispeak parametresi hesabı:

$$(x_t) > \mu * 1.5 \rightarrow 1 \quad (4.4)$$

Bu hesaplamalar her bir değer için yapılmış olup verimize, aşağıdaki Tablo 4.2.'te eklenmiştir.

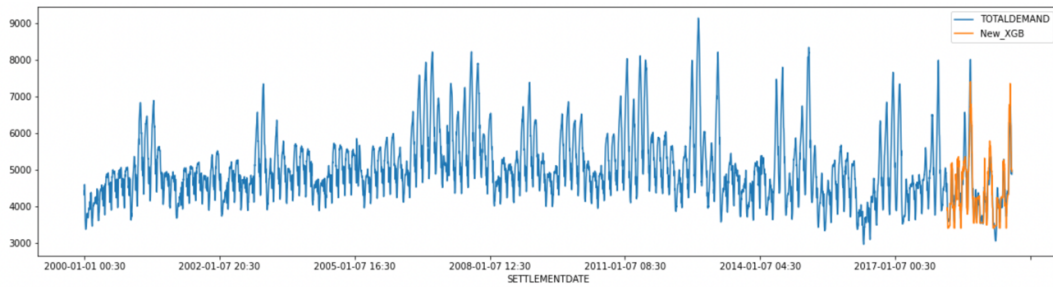
Tablo 4.2. Özel parametreler tablosu

SETTLEMENTDATE	ispeak	ismin	isstep	ismax
1.01.2000 00:30	0	0	0	0
1.01.2000 01:00	0	0	1	0
1.01.2000 01:30	0	0	0	0
1.01.2000 02:00	0	0	1	0
1.01.2000 02:30	0	0	1	0



Şekil 4.9. Özel parametrelerin önem tablosu

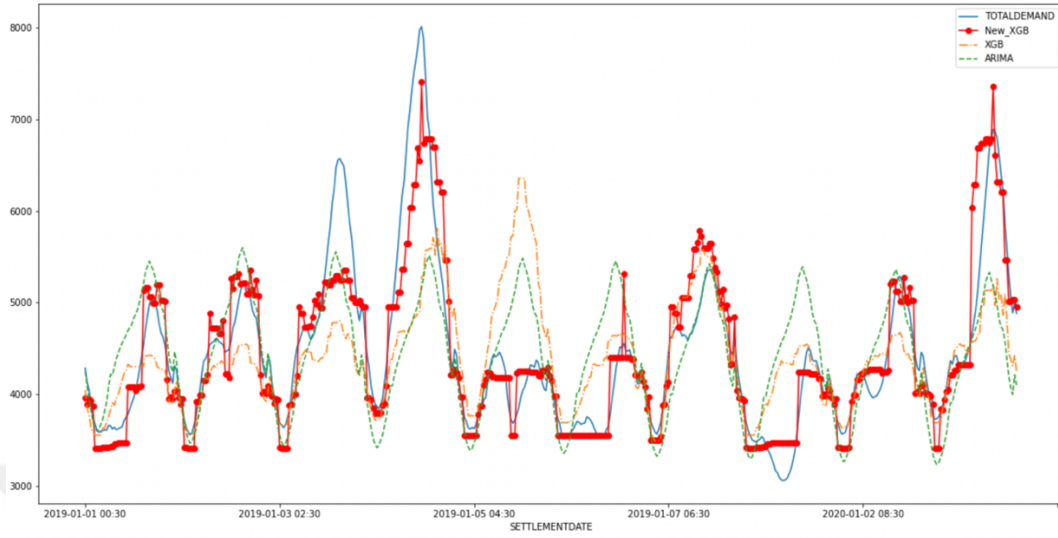
Şekil 4.9.'de eklemiş olduğumuz ve diğer özelliklerin önem aralıkları verilmiştir. Bu parametreler eklendikten sonra modelimiz tekrar oluşturulup, tekrar tahminleme yapılmıştır.



Şekil 4.10. Özel parametreler eklendikten sonra yapılan tahminleme

Şekil 4.10.'da ise XGBoost modelimizin geliştirmiş olduğumuz parametreler eklendikten sonraki tahminleme grafiği verilmiştir. Tahminleme performansında gözle görülür bir gelişme mevcuttur.

4.3. Tutarlılık Analizi



Şekil 4.11. Tüm modellerin karşılaştırılması

Şekil 4.11.'de aynı dataset kullanılarak yapılan tahminleme sonuçları gösterilmektedir.

Tablo 4.3. Yöntemlerin tutarlılıkları

Method	New XGB	XGB	ARIMA
mape	0.055055	0.103977	0.113040
me	-16.193.963	-59.667.597	-33.931.905
mae	261.903.310	489.718.882	511.977.434
mpe	-0.001189	0.006968	0.011368
rmse	385.687.234	751.521.645	740.843.444
corr	0.902601	0.537140	0.560828

Tablo 4.3.'de görüldüğü üzere XGBoost ve ARIMA modellerinin mape değerleri yaklaşık olarak aynıdır. Belirlediğimiz parametreler eklendikten sonra ise XGBoost modelinin mape oranı yarı yarıya azalmaktadır.

BÖLÜM 5. TARTIŞMA VE SONUÇ

Bu çalışmanın amacı istatistiksel bir yöntem olan ARIMA ile bir yapay zekâ algoritması olan karar ağacının gelişmiş bir versiyonu diyebileceğimiz XGBoost modelinin bir elektrik tüketim veri kümesi üzerindeki başarımını karşılaştırmaktır.

Bu çalışmada kısa vadeli elektrik talep tahmini yapmak için Avustralya'nın Victoria eyaletinin 2000 yılından 2019 yılına kadar alınmış yarım saatlik elektrik talep verileri kullanılmıştır. Sadece ocak ayının ilk haftasının tahminlenmesi için bu verilerin sadece yılın ilk haftası kullanılmış, bu şekilde tüm modellerin kısa vadeli tahminleme performansı karşılaştırılmıştır. ARIMA modelinin parametreleri belirlenirken ACF ve PACF grafiklerinden faydalanılmış ve ARIMA modeli oluşturulmuştur. XGBoost modelinin parametreleri ise karşılaştırılarak seçilmiştir. XGBoost modelinin eğitimi için zaman bilgisi ayrılarak bir parametre seti oluşturulmuş ve sonrasında bu çalışmaya özgü üretilmiş parametreler eklenmiş ve tüm modellerin karşılaştırılması yapılmıştır. Sonuç olarak görülmüştür ki, ARIMA ve XGBoost modelleri ilk kıyaslandığında benzer sonuçlar vermektedir. Ancak eklediğimiz parametrelerin XGBoost modelinin doğrulunu kayda değer şekilde arttırdığı ve istatistiksel yaklaşımdan daha iyi sonuçlar aldığı görülmektedir. MAPE değeri 0.10'dan 0.05'e düşürülmüş hata azaltılmıştır. Bu yaklaşımla yapay zekâ modellerine, verinin durumu gözetilerek, eklenecek parametreler ile doğruluğunun arttırılabileceği anlaşılmaktadır. Gelecek çalışmalar için zaman serisinden elde edilen parametreler dışında, sıcaklık, nüfus gibi dış etmenlerinde eklenip aynı etmenlerle yapay zekâ ve istatistiksel modellerin karşılaştırılması yapılabilir.

KAYNAKLAR

- [1] HAGAN and M., "Identification Techniques of Box and Jenkins Applied to the Problem of Short-Term Load Forecasting," *IEEE PES SM, Mex. City, Mex.*, 1977.
- [2] J. W. Taylor and P. E. McSharry, "Short-term load forecasting methods: An evaluation based on European data," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 2213–2219, 2007, doi: 10.1109/TPWRS.2007.907583.
- [3] I. Moghram and S. Rahman, "Analysis and evaluation of five short-term load forecasting techniques," *IEEE Trans. Power Syst.*, vol. 4, no. 4, pp. 1484–1491, 1989, doi: 10.1109/59.41700.
- [4] H. Chen, C. A. Cañizares, and A. Singh, "ANN-based short-term load forecasting in electricity markets," *Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf.*, vol. 2, no. WINTER MEETING, pp. 411–415, 2001, doi: 10.1109/PESW.2001.916876.
- [5] H. Mao, X. J. Zeng, G. Leng, Y. J. Zhai, and J. A. Keane, "Short-term and midterm load forecasting using a bilevel optimization model," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 1080–1090, 2009, doi: 10.1109/TPWRS.2009.2016609.
- [6] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2013, doi: 10.1561/20000000039.
- [7] L. M. Saini and M. K. Soni, "Artificial Neural Network-Based Peak Load Forecasting Using Conjugate Gradient Methods," *IEEE Power Eng. Rev.*, vol. 22, no. 7, p. 59, 2002, doi: 10.1109/MPER.2002.4312410.
- [8] D. Srinivasan, "Parallel neural network-fuzzy expert system strategy for short-term load forecasting : System implementation and performance evaluation," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1100–1106, 1999, doi: 10.1109/59.780934.
- [9] Z. A. Bashir and M. E. El-Hawary, "Applying wavelets to short-term load forecasting using PSO-based neural networks," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 20–27, 2009, doi: 10.1109/TPWRS.2008.2008606.

- [10] S. Chenthur Pandian, K. Duraiswamy, C. C. A. Rajan, and N. Kanagaraj, "Fuzzy approach for short term load forecasting," *Electr. Power Syst. Res.*, vol. 76, no. 6–7, pp. 541–548, 2006, doi: 10.1016/j.epsr.2005.09.018.
- [11] D. E. Goldberg and J. H. Holland, "Genetic Algorithms and Machine Learning," *Machine Learning*, vol. 3, no. 2. Springer, pp. 95–99, 1988, doi: 10.1023/A:1022602019183.
- [12] B. J. Chen, M. W. Chang, and C. J. Lin, "Load forecasting using support vector machines: A study on EUNITE Competition 2001," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821–1830, 2004, doi: 10.1109/TPWRS.2004.835679.
- [13] P. F. Pai and W. C. Hong, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electr. Power Syst. Res.*, vol. 74, no. 3, pp. 417–425, 2005, doi: 10.1016/j.epsr.2005.01.006.
- [14] L. B. Rasmussen, P. Bacher, H. Madsen, H. A. Nielsen, C. Heerup, and T. Green, "Load forecasting of supermarket refrigeration," *Appl. Energy*, vol. 163, pp. 32–40, 2016, doi: 10.1016/j.apenergy.2015.10.046.
- [15] A. N. Adamakos and M. K. Titsias, "Short-term load forecasting using a cluster of neural networks for the Greek energy market," *ACM Int. Conf. Proceeding Ser.*, vol. 18-20-May-, 2016, doi: 10.1145/2903220.2903222.
- [16] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis, 2nd Edition*. Wiley, 2003.
- [17] "6.1 Time series components | Forecasting: Principles and Practice." <https://otexts.com/fpp2/components.html> (accessed May 05, 2020).
- [18] T. H. Naylor, T. G. Seaks, and D. W. Wichern, "Box-Jenkins Methods: An Alternative to Econometric Models," *Int. Stat. Rev. / Rev. Int. Stat.*, vol. 40, no. 2, p. 123, Aug. 1972, doi: 10.2307/1402755.
- [19] H. X. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6. Pergamon, pp. 3586–3592, Aug. 01, 2012, doi: 10.1016/j.rser.2012.02.049.
- [20] H. K. Alfares and M. Nazeeruddin, "Electric load forecasting: Literature survey and classification of methods," *Int. J. Syst. Sci.*, vol. 33, no. 1, pp. 23–34, Jan. 2002, doi: 10.1080/00207720110067421.
- [21] A. Fouquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," *Renewable and Sustainable Energy Reviews*, vol. 23. Pergamon, pp. 272–288, Jul. 01, 2013, doi: 10.1016/j.rser.2013.03.004.

- [22] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting - A review," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 2. Pergamon, pp. 1223–1240, Feb. 01, 2012, doi: 10.1016/j.rser.2011.08.014.
- [23] N. Nassif, "Single and multivariate regression models for estimating monthly energy consumption in schools in hot and humid climates," *Energy Engineering: Journal of the Association of Energy Engineering*, vol. 110, no. 5. Taylor & Francis, pp. 33–54, Aug. 01, 2013, doi: 10.1080/01998595.2013.10750242.
- [24] R. Ramanathan, R. Engle, C. W. J. Granger, F. Vahid-Araghi, and C. Brace, "Short-run forecasts of electricity loads and peaks," *Int. J. Forecast.*, vol. 13, no. 2, pp. 161–174, Jun. 1997, doi: 10.1016/S0169-2070(97)00015-0.
- [25] J. C. Lam, K. K. W. Wan, D. Liu, and C. L. Tsang, "Multiple regression models for energy use in air-conditioned office buildings in different climates," *Energy Convers. Manag.*, vol. 51, no. 12, pp. 2692–2697, Dec. 2010, doi: 10.1016/j.enconman.2010.06.004.
- [26] L. Yang, J. C. Lam, J. Liu, and C. L. Tsang, "Building energy simulation using multi-years and typical meteorological years in different climates," *Energy Convers. Manag.*, vol. 49, no. 1, pp. 113–124, Jan. 2008, doi: 10.1016/j.enconman.2007.05.004.
- [27] I. Turiel, R. Boschen, M. Seedall, and M. Levine, "Simplified energy analysis methodology for commercial buildings," *Energy Build.*, vol. 6, no. 1, pp. 67–83, Jan. 1984, doi: 10.1016/0378-7788(84)90008-2.
- [28] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-August-2016, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
- [29] H. Zheng, J. Yuan, and L. Chen, "Short-Term Load Forecasting Using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation," *Energies*, vol. 10, no. 8, 2017, doi: 10.3390/en10081168.
- [30] M. Gumus and M. S. Kiran, "Crude oil price forecasting using XGBoost," *2nd Int. Conf. Comput. Sci. Eng. UBMK 2017*, pp. 1100–1103, 2017, doi: 10.1109/UBMK.2017.8093500.
- [31] W. Yucong and W. Bo, "Research on EA-Xgboost Hybrid Model for Building Energy Prediction," *J. Phys. Conf. Ser.*, vol. 1518, no. 1, 2020, doi: 10.1088/1742-6596/1518/1/012082.
- [32] A. Inoue, "AR and MA representation of partial autocorrelation functions, with applications," *Probab. Theory Relat. Fields*, vol. 140, no. 3–4, pp. 523–551, 2008, doi: 10.1007/s00440-007-0074-1.

- [33] S. Singh and A. Yassine, “Big data mining of energy time series for behavioral analytics and energy consumption forecasting,” *Energies*, vol. 11, no. 2, 2018, doi: 10.3390/en11020452.



ÖZGEÇMİŞ

Adı Soyadı

: Muhammed Can ÖZDEMİR

ÖĞRENİM DURUMU

Derece	Eğitim Birimi	Mezuniyet Yılı
Yüksek Lisans	Sakarya Üniversitesi / Bilgisayar Mühendisliği	Devam ediyor
Lisans	Sakarya Üniversitesi / Elektrik Elektronik Mühendisliği	2015
Lise	Akyazı Anadolu Lisesi	2011