



**ULUSAL ELEKTRİK ENERJİSİ TÜKETİM
MİKTARLARININ TAHMİN EDİLMESİNDE
KULLANILAN YÖNTEMLERİN KARŞILAŞTIRILMASI
VE BİRLEŞİK KRALLIK VERİLERİ İLE BİR
UYGULAMA**

Nefise ERTÖY



T.C.
BURSA ULUDAĞ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**ULUSAL ELEKTRİK ENERJİSİ TÜKETİM MİKTARLARININ TAHMİN
EDİLMESİNDE KULLANILAN YÖNTEMLERİN KARŞILAŞTIRILMASI VE
BİRLEŞİK KRALLIK VERİLERİ İLE BİR UYGULAMA**

NEFİSE ERTÖY
0009-0009-9167-8768

Dr. Öğr. Üyesi Mehmet AKANSEL
(Danışman)

YÜKSEK LİSANS TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI

BURSA – 2024
Her Hakkı Saklıdır

ÖZET

Yüksek Lisans Tezi

ULUSAL ELEKTRİK ENERJİSİ TÜKETİM MİKTARLARININ TAHMİN EDİLMESİNDE KULLANILAN YÖNTEMLERİN KARŞILAŞTIRILMASI VE BİRLEŞİK KRALLIK VERİLERİ İLE BİR UYGULAMA

Nefise ERTÖY

Bursa Uludağ Üniversitesi
Fen Bilimleri Enstitüsü
Endüstri Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Mehmet AKANSEL

Elektrik talep tahmini, enerji sistemlerinin güvenli, ekonomik ve çevresel olarak sürdürülebilir bir şekilde işletilmesi için hayati öneme sahiptir. Son zamanlarda, elektrik talep tahmin yöntemleri üzerine yapılan çalışmalar, hem klasik yöntemlerin iyileştirilmesi hem de yeni teknolojilerin entegrasyonu üzerine yoğunlaşmıştır. Bu çalışmada, Birleşik Krallık'taki elektrik talebi veri seti kullanılarak elektrik talebini tahmin etmek amaçlanmıştır. Bu amaç doğrultusunda, Mevsimsel Otoregresif Entegre Hareketli Ortalama (SARIMA), Prophet, Uzun Kısa Süreli Bellek (LSTM), XGBoost ve Lineer Artırma Regresyonu gibi çeşitli tahmin modelleri Python yazılım dili kullanılarak Google Colab ortamında uygulanmıştır. Modellerin performansları, Kök Ortalama Kare Hatası (RMSE) ve Ortalama Mutlak Yüzde Hatası (MAPE) doğruluk ölçüm yöntemleri kullanılarak karşılaştırılmıştır. Sonuçlar, LSTM yönteminin diğer yöntemlere üstünlüğünü göstermektedir. Ayrıca, sonuçlar Prophet ve XGBoost modellerinin de iyi performans sergilediğini göstermiştir.

Anahtar Kelimeler: Elektrik Talebi Tahminlemesi, SARIMA, Prophet, LSTM, XGBoost, Lineer Artırma Regresyonu

2024, ix + 85 sayfa.

ABSTRACT

MSc Thesis

COMPARISON OF METHODS USED IN PREDICTING NATIONAL ELECTRICITY ENERGY CONSUMPTION AMOUNTS AND AN APPLICATION WITH UNITED KINGDOM DATA

Nefise ERTÖY

Bursa Uludağ University
Graduate School of Natural and Applied Sciences
Department of Industrial Engineering

Supervisor: Asst. Prof. Mehmet AKANSEL

Electricity demand forecasting is vital for the safe, economical, and environmentally sustainable operation of energy systems. Recently, studies on electricity demand forecasting methods have focused on both improving classical methods and integrating new technologies. In this study, the aim is to forecast electricity demand using the electricity demand dataset of the United Kingdom. For this purpose, various forecasting models such as Seasonal Autoregressive Integrated Moving Average (SARIMA), Prophet, Long Short-Term Memory (LSTM), and XGBoost, Linear Boost Regression were implemented using the Python programming language in the Google Colab. The performances of the models were compared using accuracy measurement methods such as Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The results show the superiority of the LSTM method over other methods. Additionally, the results indicate that the Prophet and XGBoost models also perform well.

Key words: Electricity Demand Forecasting, SARIMA, Prophet, LSTM, XGBoost, Linear Boost Regression

2024, ix + 85 pages.

TEŞEKKÜR

Tez çalışmamın oluşma sürecinde, Dr. Öğr. Üyesi Mehmet AKANSEL'in yönlendirici, anlayışlı ve teşvik edici yaklaşımıyla verdiği yardımlar için sonsuz teşekkürlerimi sunarım.

Bugünlere gelmemde payı çok büyük olan, beni eğitimimin her aşamasında sonsuz bir özveri ile destekleyen aileme ve eşime sonsuz teşekkürlerimi sunarım.

Son olarak, bu aşamaya gelene kadar üzerimde emeği olan tüm hocalarıma ve arkadaşlarıma teşekkür ederim.

Nefise ERTÖY

.../.../.....



İÇİNDEKİLER

	Sayfa
ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR.....	iii
SİMGELER VE KISALTMALAR DİZİNİ.....	v
ŞEKİLLER DİZİNİ.....	vi
ÇİZELGELER DİZİNİ.....	viii
1. GİRİŞ.....	1
2. KURAMSAL TEMELLER ve KAYNAK ARAŞTIRMASI.....	3
2.1. Elektrik Enerjisi Talep Tahminlerinin Önemi.....	3
2.2. Elektrik Enerjisi Talebini Etkileyen Faktörler.....	4
2.3. Literatür Taraması.....	5
3. MATERYAL ve YÖNTEM.....	8
3.1. Birleşik Krallık'taki Elektrik Tüketim Talebi Verileri.....	8
3.2. Veri Madenciliği.....	11
3.3. Keşifsel Veri Analizi ve Özellik Mühendisliği.....	12
3.4. Elektrik Enerjisi Talep Tahmin Yöntemleri.....	16
3.5. Zaman Serileri.....	22
3.6. XGBoost ve Lineer Artırma Regresyonu	27
3.7. Prophet.....	30
3.8. LSTM (Uzun Kısa Dönemli Hafıza).....	33
4. BULGULAR.....	36
4.1. SARIMA Uygulaması.....	36
4.2. XGBoost Uygulaması.....	48
4.3. Lineer Artırma Regresyonu Uygulaması.....	52
4.4. Prophet Uygulaması.....	57
4.5. LSTM ve Derin LSTM Uygulamaları.....	66
5. TARTIŞMA VE SONUÇ.....	77
KAYNAKLAR.....	80
ÖZGEÇMİŞ.....	85

SİMGELER ve KISALTMALAR DİZİNİ

Kısaltmalar	Açıklama
ADF	Augmented Dickey-Fuller
AE	Ortalama Hata
ANFIS	Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemleri
ANN	Yapay Sinir Ağları
AR	Otoregresif Model
ARCH	Otoregresif Koşullu Değişen Varyans
ARFIMA	Otoregresif Kesirli Entegre Hareketli Ortalama
ARIMA	Otoregresif Entegre Edilmiş Hareketli Ortalama
ARMA	Otoregresif Hareketli Ortalama
DNN	Derin Sinir Ağları
DT	Karar Ağacı
ELM	Aşırı Öğrenme Makinesi
EDA	Keşifsel Veri Analizi
GA	Genetik Algoritmaları
GARCH	Genelleştirilmiş Otoregresif Koşullu Değişen Varyans
GBRT	Gradient Boosted Regresyon Ağacı
GRU	Kapılı Tekrarlayan Üniteler
KPSS	Kwiatkowski, Philips, Schmidt, Shin
LSTM	Uzun Kısa Dönemli Bellek
MA	Hareketli Ortalama Modeli
MAPE	Ortalama Mutlak Yüzde Hata
MSE	Hata Kareler Ortalaması
MTMV	Multi Task- Multi Variable
MW	Megawatt
RMSE	Hata Karelerinin Ortalamasının Karekökü
RNN	Tekrarlayan Sinir Ağı
SARIMA	Mevsimsel Otoregresif Entegre Hareketli Ortalama
SARIMAX	Mevsimsel Otoregresif Entegre Hareketli Ortalama ve Dışsal Değişken
SVM	Destek Vektör Makineleri
VAR	Vektör Otoregresyon
XGBOOST	Ekstrem Gradyan Arttırma
YSA	Yapay Sinir Ağları

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 3.1. Avrupa'daki düşük güç hatları ve hizmet dışı hatlar hariç HVDC bağlantıları.....	9
Şekil 3.2. Veri madenciliği süreci (Fayyad ve diğerleri, 1996).....	11
Şekil 3.3. İletim sistem talebi (tsd) sıfır olan satırların sayısı.....	13
Şekil 3.4. Yıllara göre elektrik tüketimi dağılımı.....	14
Şekil 3.5. 2013 ve 2023 yıllarının elektrik tüketimi kıyası.....	14
Şekil 3.6. Aylara göre elektrik tüketimi dağılımı.....	15
Şekil 3.7. Aylara göre en yüksek hava sıcaklıkları (C°).....	15
Şekil 3.8. Saatlere göre elektrik tüketimi dağılımı.....	16
Şekil 3.9. SARIMA gösterimi (Goswami ve Kandali, 2020).....	26
Şekil 3.10. LSTM ağ yapısı (Zhang ve diğerleri, 2023).....	35
Şekil 4.1. Günlük toplam elektrik talebi ve gecikmeli değerler grafiği...	37
Şekil 4.2. Zaman serisinin orijinali için hareketli ortalama ve standart sapma.....	38
Şekil 4.3. Bir günlük fark alınmış veri seti üzerinde hareketli ortalama ve standart sapma.....	38
Şekil 4.4. Trend, mevsimsellik ve gürültü bileşeni.....	39
Şekil 4.5. Mevsimsel bileşenin hareketli ortalama ve standart sapması.....	40
Şekil 4.6. Bir yıllık fark alınmış veri seti üzerinde hareketli ortalama ve standart sapma.....	40
Şekil 4.7. Orijinal zaman serisinin otokorelasyon ve kısmi otokorelasyon grafikleri.....	41
Şekil 4.8. Mevsimsel bileşenin otokorelasyon ve kısmi otokorelasyon grafikleri.....	42
Şekil 4.9. SARIMA Model 0 grafikleri A) t için standartlaştırılmış artış gösterimi B) Histogram ve tahmin eğrisi C) Normal Q-Q grafiği D) Korelasyon grafiği E) Tahmin grafiği.....	44
Şekil 4.10. SARIMA Model 1 grafikleri A) t için standartlaştırılmış artış gösterimi B) Histogram ve tahmin eğrisi C) Normal Q-Q grafiği D) Korelasyon grafiği E) Tahmin grafiği.....	45
Şekil 4.11. SARIMA Model 2 grafikleri A) t için standartlaştırılmış artış gösterimi B) Histogram ve tahmin eğrisi C) Normal Q-Q grafiği D) Korelasyon grafiği E) Tahmin grafiği.....	46
Şekil 4.12. SARIMA Model 2 gelecek tahminleri.....	47
Şekil 4.13. Test verileri ile tahmin edilen değerlerin karşılaştırması (XGBoost).....	48
Şekil 4.14. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Basit XGBoost).....	49
Şekil 4.15. Tarihsel veriler ve Basit XGBoost, 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	49
Şekil 4.16. Basit XGBoost, 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	49
Şekil 4.17. Test verileri ile tahmin edilen değerlerin karşılaştırması (XGBoost CV&GS).....	50
Şekil 4.18. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki	

	haftalık (XGBoost CV & GS).....	51
Şekil 4.19.	Tarihsel veriler ve XGBoost CV & GS 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	51
Şekil 4.20.	XGBoost CV & GS 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	52
Şekil 4.21.	Test verileri ile tahmin edilen değerlerin karşılaştırması (Basit Lineer Arttırma Regresyonu).....	53
Şekil 4.22.	Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Basit Lineer Arttırma Regresyon).....	53
Şekil 4.23.	Tarihsel veriler ve Basit Lineer Arttırma Regresyonu 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	54
Şekil 4.24.	Basit Lineer Arttırma Regresyonu 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	54
Şekil 4.25.	Test verileri ile tahmin edilen değerlerin karşılaştırması (Lineer Arttırma Regresyonu CV&GS).....	55
Şekil 4.26.	Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık.....	55
Şekil 4.27.	Tarihsel veriler ve Lineer Arttırma Regresyonu CV&GS 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	56
Şekil 4.28.	Lineer Arttırma Regresyonu CV&GS 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	56
Şekil 4.29.	Prophet trend ve mevsimsellik grafikleri.....	57
Şekil 4.30.	Test verileri ile tahmin edilen değerlerin karşılaştırması (Basit Prophet).....	58
Şekil 4.31.	Test verileri ile tahmin edilen değerlerin karşılaştırması (Basit Prophet).....	59
Şekil 4.32.	Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Basit Prophet).....	59
Şekil 4.33.	Tarihsel veriler ve Basit Prophet Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	60
Şekil 4.34.	Basit Prophet Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	60
Şekil 4.35.	Prophet Tatilli Model trend ve mevsimsellik grafikleri.....	61
Şekil 4.36.	Tatillerin dahil edildiği Prophet test verileri ile tahmin edilen değerlerin karşılaştırması.....	61
Şekil 4.37.	Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Tatilli Prophet).....	62
Şekil 4.38.	2019 Noel haftası tatilli Prophet modeli test verisi ve tahmin verisi kıyası.....	62
Şekil 4.39.	2019 Noel haftası tatilsiz Prophet modeli test verisi ve tahmin verisi kıyası.....	63
Şekil 4.40.	Tarihsel veriler ve Tatilli Prophet Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	63
Şekil 4.41.	Tatilli Prophet Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	64
Şekil 4.42.	Prophet (CV & GS) test verileri ile tahmin edilen değerlerin karşılaştırması.....	64

Şekil 4.43.	Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Prophet CV & GS).....	65
Şekil 4.44.	Tarihsel veriler ve Prophet CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	65
Şekil 4.45.	Prophet CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	66
Şekil 4.46.	LSTM kayıp evrimi.....	67
Şekil 4.47.	Test verileri ile tahmin edilen değerlerin karşılaştırması (LSTM).....	68
Şekil 4.48.	Test verileri ile tahmin edilen değerlerin karşılaştırması- İki haftalık (LSTM).....	68
Şekil 4.49.	Tarihsel veriler ve Basit LSTM Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	69
Şekil 4.50.	Basit LSTM Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	69
Şekil 4.51.	Test verileri ile tahmin edilen değerlerin karşılaştırması (LSTM CV & GS).....	70
Şekil 4.52.	Test verileri ile tahmin edilen değerlerin karşılaştırması- İki haftalık (Basit LSTM vs LSTM CV & GS).....	70
Şekil 4.53.	Tarihsel veriler ve LSTM CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	71
Şekil 4.54.	LSTM CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	71
Şekil 4.55.	Derin LSTM kayıp evrimi.....	72
Şekil 4.56.	Test verileri ile tahmin edilen değerlerin karşılaştırması (Derin LSTM).....	73
Şekil 4.57.	Test verileri ile tahmin edilen değerlerin karşılaştırması- İki haftalık (Derin LSTM vs Basit LSTM).....	73
Şekil 4.58.	Tarihsel veriler ve Derin LSTM Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	74
Şekil 4.59.	Derin LSTM Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	74
Şekil 4.60.	Test verileri ile tahmin edilen değerlerin karşılaştırması (Derin LSTM CV & GS).....	75
Şekil 4.61.	Test verileri ile tahmin edilen değerlerin karşılaştırması- İki haftalık (Derin LSTM).....	75
Şekil 4.62.	Tarihsel veriler ve Derin LSTM CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	76
Şekil 4.63.	Derin LSTM CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri.....	76

ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 5.1. Modellerin hata metrik kıyası.....	78



1. GİRİŞ

Elektrik enerjisi, modern dünyanın vazgeçilmez bir unsurudur. Teknolojik ilerlemeler, sanayileşme ve artan nüfus ile birlikte elektrik enerjisine olan talep sürekli olarak artmaktadır. Elektrik enerjisinin üretimi, dağıtımı ve tüketimi, ülkelerin ekonomik ve sosyal gelişimi üzerinde doğrudan etkili olmaktadır. Bu nedenle, elektrik enerjisi talebinin doğru bir şekilde tahmin edilmesi, enerji politikalarının oluşturulması, enerji arz güvenliğinin sağlanması ve sürdürülebilir enerji yönetimi için büyük önem taşımaktadır (Sarwar ve diğerleri, 2023).

Elektrik enerjisi talep tahmini, enerji arz-talep dengesini sağlamak, enerji sistemlerini iyileştirmek ve enerji maliyetlerini düşürmek için kritik bir rol oynamaktadır (Bashir ve diğerleri, 2022). Doğru talep tahminleri, enerji arzının verimli bir şekilde yönetilmesine ve enerji kesintilerinin önlenmesine yardımcı olur (Waheed ve Xu, 2024). Elektrik yükü tahmini modelleri, istatistiksel, makine öğrenmesi, derin öğrenme modelleri ve hibrit olmak üzere kategorilere ayrılabilir. İstatistiksel modeller genellikle doğrusal ilişkiler üzerine kurulu olsa da makine öğrenmesi modelleri doğrusal olmayan etkileşimleri de hesaba katarak daha iyi tahmin sonuçları sağlayabilir. Son yıllarda özellikle derin öğrenme modelleri ve hibrit modellerin elektrik yük tahmininde kullanımı artmıştır (Zhang ve diğerleri, 2022).

Bu çalışmanın amacı, elektrik enerjisi talep tahmini konusunda mevcut literatürü ve kuramsal temelleri inceleyerek bu alandaki bilgi birikimine katkıda bulunmaktır. Ayrıca, elektrik enerjisi talep tahmininde kullanılan yöntemlerin etkinliğini değerlendirmek ve bu yöntemlerin enerji yönetimi açısından sağladığı faydaları ortaya koymaktır. SARIMA, XGBoost, Lineer Arttırma Regresyonu (Linear Boost Regression), Prophet, LSTM ve Derin (Stacked) LSTM gibi farklı model ve yöntemleri kullanarak elektrik enerjisi talep tahmininde performans değerlendirmesi yapılmıştır. Bu çeşitlilik, farklı yöntemlerin avantaj ve dezavantajlarını ortaya koyarak, hangi yöntemlerin hangi koşullarda daha etkili olduğunu göstermiştir. Yapay zekâ ve makine öğrenimi tabanlı yöntemlerin geleneksel istatistiksel yöntemlere göre daha yüksek doğruluk sağladığı görülmüştür.

Özellikle LSTM ve Derin LSTM modellerinin üstün performans sergilemesi, bu tür yöntemlerin elektrik talebi tahmininde etkili olduğunu vurgulamaktadır.

Çalışmada, kullanılan modellerin performansı MAPE ve RMSE gibi hata metrikleri ile değerlendirilmiştir. SARIMA modelleri, zaman serisinin karmaşıklığını yeterli düzeyde tanımlayamamış ve düşük performans sergilemiştir. En iyi SARIMA modelinin MAPE değeri %13,02 olarak bulunmuş, ancak modelin normallik ve otokorelasyon sorunları nedeniyle tahmin güvenilirliği sınırlı kalmıştır. XGBoost modeli, hiperparametre optimizasyonu ile oldukça iyi sonuçlar vermiştir. Doğrulama MAPE değeri %7,29 ve RMSE değeri 2655,08 MW olarak hesaplanmıştır. Elektrik talebi tahmininde yüksek doğruluk sağlamış ve uzun vadeli tahminlerde iyi performans göstermiştir. Lineer Arttırma Regresyonu modeli, MAPE değeri %8,20 ve RMSE değeri 2963,08 MW ile başarılı bir performans sergilemiştir. Hiperparametre optimizasyonu ile MAPE değeri %7,73 ve RMSE değeri 2810 MW olarak elde edilmiştir. Buna karşılık, modelin çalışma süresi uzundur ve optimizasyon gerektirebilecek alanlar bulunmaktadır. Prophet modeli, %9,34'lük MAPE değeri ile oldukça iyi bir performans sergilemiştir. Model, trend, mevsimsellik ve tatil etkilerini dikkate alarak tahmin yapmış ve gelecekteki talep tahminlerinde başarılı olmuştur. Uzun kısa süreli bellek (LSTM) derin öğrenmede en yaygın kullanılan tekniklerden biridir ve son yıllarda güç yükü tahmininde de yaygın olarak kullanılmaktadır. LSTM modeli, %7,44 MAPE ve 2709,97 MW RMSE değeri ile ve derin LSTM modeli ise, %7,16 MAPE ve 2597,09 MW RMSE değeri ile başarılı sonuçlar vermişlerdir. Hiperparametre optimizasyonu uygulanmış LSTM modeli ile MAPE değeri %7,18'ye ve RMSE değeri 2623,11 MW'a iyileştirilmiş, derin LSTM modelinde ise, %7,13 MAPE ve 2618 MW RMSE değerleri ile en iyi performans elde edilmiştir.

Bölüm 2'de, Türkiye'de ve dünyada elektrik enerjisi talep tahmininde kullanılan yöntemler incelenmiştir. Ayrıca, elektrik enerjisi talep tahminlerinin önemi ve talebi etkileyen faktörler hakkında bilgiler sunulmuştur. Bölüm 3'te, bu çalışmada kullanılan veriler ve yöntemler ayrıntılı olarak açıklanmıştır. Bölüm 4'te, deneysel sonuçların analizi ve model performansını değerlendirmek için kullanılan değerlendirme metrikleri sunulmuştur. Bölüm 5, sonuçları ve gelecekteki araştırma yönlerinin tartışmasını sunmaktadır.

2. KURAMSAL TEMELLER ve KAYNAK ARAŞTIRMASI

2.1. Elektrik Enerjisi Talep Tahminlerinin Önemi

Elektrik, atomların içindeki elektronların hareketinden kaynaklanan bir enerji türüdür. M.Ö. 600'lü yıllarda keşfedilen elektrik, günümüzde hayatımızın vazgeçilmez bir parçası haline gelmiştir. 19. yüzyılda yapılan keşiflerle elektrik akımının üretimi, iletimi ve kullanımı gelişmiş, ampul, elektrik motoru, iletişim sistemleri gibi teknolojilerin ortaya çıkmasını sağlamıştır. Günümüzde yenilenebilir enerji kaynaklarının kullanımı artmakta ve akıllı şebekeler ile enerji depolama teknolojileri geliştirilmektedir. Elektrik, modern dünyanın temelini oluşturarak, enerji ihtiyacının karşılanması, endüstriyel üretim, iletişim ve yaşam kalitesinin artması gibi birçok alanda önemli bir rol oynamaktadır (Küçük, 2022).

Elektrik talep tahmini, enerji arzının talebi karşılayacak şekilde planlanmasına yardımcı olur. Elektrik talebinin yüksek doğruluk düzeyinde tahmin edilmesi ise, elektrik kesintilerinin önlenmesi, enerji güvenliğinin sağlanması, enerji üretim ve dağıtım maliyetlerinin optimize edilmesi, enerji altyapısının (örneğin, yeni santraller, iletim hatları, dağıtım ağları) planlanması ve geliştirilmesinde önemli bir rol oynar (Akman ve diğerleri, 2018). Serbestleşmiş enerji piyasalarında, elektrik talep tahmini, piyasa katılımcılarının (üreticiler, dağıtıcılar, tüketiciler) stratejik kararlar almasına yardımcı olur. Talebi doğru tahmin eden şirketler, piyasa fiyatlarını daha iyi öngörebilir ve uygun ticaret stratejileri geliştirebilir. Doğru talep tahmini, enerji üretiminin verimli ve çevre dostu bir şekilde yapılmasına yardımcı olur, böylece karbon ayak izinin azaltılmasına katkıda bulunur (Kılınç, 2023).

Yenilenebilir enerji kaynakları (rüzgâr, güneş) genellikle değişken ve öngörülemez olduğundan, elektrik talep tahmini, bu kaynakların entegrasyonunu ve yönetimini kolaylaştırır. Böylece, yenilenebilir enerji kaynaklarının daha etkin kullanımı sağlanır. Elektrik talep tahmini, acil durumlar (doğal afetler, ani talep artışları) sırasında enerji sistemlerinin esnekliğini ve dayanıklılığını artırır. Bu, kriz anlarında enerji arzının sürekliliğini sağlamaya yardımcı olur (Akman ve diğerleri, 2018).

Elektrik enerjisi talebi bölgesel, mevsimsel ve anlık dalgalanmalar gösterir. Yıl içinde, en düşük ve en yüksek talep arasında % 200'e varan farklar oluşabilir. Elektrik talebinin aşırı değişkenliği ve elektrik enerjisinin depolanamayan bir enerji türü olması nedeniyle, elektrik arzının sürekli ve kesintisiz olması önemlidir. Elektrik enerjisi talebinin yanlış tahmin edilmesi durumunda, sistem tasarımı ve planlamasında sorunlar ortaya çıkabilir. Bu durumda, zorunlu elektrik kesintileri ve tasarruf uygulamaları gerekebilir. Arz kapasitesinin âtıl hale gelmesi, kaynak dağılımının rasyonel yapılamaması ve kaynak israfı gibi olumsuz sonuçlara sebep olabilir (Akgül, 2013).

2.2. Elektrik Enerjisi Talebini Etkileyen Faktörler

Elektrik enerjisi talebini belirleyen birçok unsur bulunmaktadır ve bu unsurlar bir ülkenin veya bölgenin ekonomik, sosyal ve demografik koşullarına bağlı olarak değişiklik gösterebilir. Elektrik enerjisi talebini etkileyen başlıca faktörler şunlardır:

- **Ekonomik Durum:** Ülkenin ekonomik büyüme ve gelişme seviyesi, elektrik talebinin en önemli belirleyicilerindendir. Ekonomik büyüme arttıkça, endüstriyel faaliyetler ve hane halkı tüketimi de artar, bu da elektrik talebinin yükselmesine yol açar.
- **Nüfus:** Hızla artan nüfus, elektrik talebini artırır. Daha fazla insan, daha fazla konut, iş yeri ve altyapı ihtiyacı anlamına gelir ve bu da elektrik talebinin yükselmesine neden olur.
- **Hava Koşulları:** Mevsimsel değişiklikler, sıcaklık, nem ve iklim koşulları elektrik talebini etkiler.
- **Endüstriyel Faaliyetler:** Endüstriyel tesislerin üretim kapasiteleri ve işletmelerin faaliyet durumu, elektrik talebini önemli ölçüde etkiler. Endüstriyel faaliyetlerin yoğun olduğu bölgelerde elektrik talebi genellikle daha yüksektir.
- **Teknolojik Gelişmeler:** Teknolojik ilerlemeler ve enerji verimliliği önlemleri elektrik talebini etkiler. Daha verimli cihazlar ve enerji tasarruflu sistemlerin kullanılması, elektrik tüketimini azaltabilir.

- **Enerji Politikaları:** Enerji politikaları ve teşvikler, elektrik talebinin yönlendirilmesinde önemli rol oynar. Yenilenebilir enerji kaynaklarına verilen teşvikler ve enerji verimliliği programları elektrik talebini etkileyebilir.

Bu faktörler, elektrik enerjisi talebinin artmasına veya azalmasına neden olabilir. Bu nedenle, elektrik talebini etkileyen unsurların dikkatle incelenmesi ve yönetilmesi, enerji sektöründe sürdürülebilir ve dengeli bir yaklaşımın benimsenmesini sağlar (Jin ve diğerleri, 2024; Aldarraji ve diğerleri, 2024).

2.3. Literatür Taraması

Akgül (2013), Türkiye'deki elektrik üretimi ve tüketimini ARIMA, VAR ve YSA yöntemleri ile tahmin etmiştir. Çalışmada, yapay sinir ağlarının üretim ve tüketim verileri için en düşük hata oranına sahip yöntem olduğu belirlenmiştir.

Tarkun (2023), elektrik enerjisi talep tahminini ARCH, VAR, YSA-NAR, YSA-Çoklu model mimarileri ile gerçekleştirmiştir. Eviews 10 ve MATLAB yazılımları kullanılarak yapılan analizlerde, en düşük hata değerine sahip modelin çok değişkenli yapay sinir ağı olduğu belirlenmiştir.

Güven ve Kayalica (2023), Türkiye'deki hane halklarının elektrik tüketimini etkileyen faktörleri Hane Halkı Bütçe Anketi verilerini kullanarak belirlemeyi amaçlamıştır. Bu çalışmada, Karar Ağacı (DT), Rastgele Orman (RF) ve Gradyan Arttırmalı Regresyon Ağacı (GBRT) yöntemleri uygulanmıştır. En düşük Kök Ortalama Kare Hatası (RMSE) değerine sahip olan GBRT yöntemi, elektrik tüketimi üzerindeki değişkenlerin etkisini analiz etmek için tercih edilmiştir. Çalışma makine öğrenimi tabanlı yöntemlerin, elektrik talebi tahmininde yüksek doğruluk sağladığını ve önemli belirleyicilerin etkilerini ortaya koyabildiğini göstermektedir.

Grandón ve diğerleri (2023), ulusal elektrik talebini tahmin etmek için klasik istatistik ve makine öğrenimini birleştiren yenilikçi bir hibrit yaklaşım sunmaktadır. 2013-2020 yılları arasında Ukrayna'nın elektrik tüketimine dair saatlik veriler kullanılarak elektrik

tüketiminin saatlik, günlük ve yıllık zaman serileri analiz edilmiştir. Uzun vadeli model, yıllık trend, makroekonomik regresyon analizi ile değerlendirilmiştir. Orta vadeli model, sıcaklık ve takvim değişkenlerini içeren ve ARIMA ile LSTM tabanlı yaklaşımlarını birleştiren bir yapıyı içermektedir. Kısa vadeli model ise saatlik mevsimselliği, takvim değişkenleri ve kalıntı için birden fazla ARMA modeli ile tespit edilmektedir. Sonuçlar, en iyi tahmin modelinin, birden fazla regresyon modeli ve kalıntı tahmini için LSTM hibrit modelinin kombinasyonu olduğunu göstermektedir.

Zhang ve diğerleri (2023), kısa ve orta vadeli güç yüklerini tahmin etmek amacıyla iki deney gerçekleştirmiştir. Bu deneylerde, önerilen MTMV-CNN-LSTM modeli, diğer modellerle karşılaştırılmış ve daha üstün tahmin sonuçları sağladığı doğrulanmıştır.

Bozlak ve Yaşar (2024), Python programlama dilini kullanarak SARIMAX, LSTM ve CNN-LSTM modellerinin performanslarını karşılaştırmıştır. Almanya elektrik piyasasında gelecekteki gün için elektrik fiyatlarını tahmin etmek amacıyla kullanılan bu modellerin performansları, MAE, RMSE, MAPE ve Diebold-Mariano testi gibi çeşitli metriklerle değerlendirilmiştir. LSTM ve CNN-LSTM modellerinin, daha büyük zaman adımlarına izin verdikleri için SARIMAX' tan daha iyi performans sergilediği ortaya konmuştur.

Khan ve diğerleri (2022), kısa vadeli elektrik yük tahmini (ELF) için yenilikçi bir iki aşamalı çerçeve geliştirmiştir. İlk aşamada, verilerin temizlenmesi için çeşitli işlemler uygulanmıştır. İkinci aşamada ise, derin bir artık Konvolüsyonel Sinir Ağı (CNN) ve yığılmış Uzun Kısa Süreli Bellek (LSTM) mimarisi kullanılarak elektrik tüketiminin mekansal ve zamansal özellikleri modellenmiştir. Önerilen model, Bireysel Hane Elektrik Gücü Tüketimi (IHEPC) ve Pennsylvania–New Jersey–Maryland (PJM) veri setleri üzerinde test edilmiştir. Sonuçlar, modelin hata oranlarını önemli ölçüde azalttığını göstermektedir. Özellikle IHEPC veri seti için, modelin saatlik yükler üzerindeki performansı, Ortalama Mutlak Hata, Ortalama Kare Hata ve Kök Ortalama Kare Hata açısından en düşük hata oranlarına ulaşmıştır. CNN, LSTM, Gated Recurrent Unit (GRU), Multilayer Perceptron (MLP), Destek Vektör Regresyonu (SVR), Lineer Regresyon, Karar Ağacı ve Aşırı Öğrenme Makinesi (ELM) modellerle

karşılaştırıldığında, önerilen modelin hata oranlarını önemli ölçüde azalttığı görülmüştür. PJM veri setinde de benzer sonuçlar elde etmişlerdir.

Ensafi ve diğerleri (2022), bir perakende mağazasının satış geçmişini içeren bir kamu veri setini analiz ederek mobilya satışlarını tahmin etmeyi amaçlamışlardır. Bu doğrultuda, farklı tahmin modelleri uygulanmıştır. İlk olarak, SARIMA ve Üçlü Üstel Düzeltme gibi geleneksel zaman serisi tahmin teknikleri kullanılmıştır. Ardından, Prophet, LSTM ve CNN gibi daha ileri düzey yöntemler devreye sokulmuştur. Modellerin performansları, RMSE ve MAPE gibi çeşitli doğruluk ölçütleriyle değerlendirilmiştir. Sonuçlar, Derin (Stacked) LSTM yönteminin diğer yöntemlere kıyasla üstün olduğunu ortaya koymaktadır. Ayrıca, Prophet ve CNN modellerinin de başarılı performans gösterdiği belirlenmiştir.

Wang ve diğerleri (2021), çalışmalarında, ham yük verilerini bir trend serisi ve dalgalanma alt serilerine ayırmak için VMD ve SampEn tabanlı uyarlanabilir bir ayrıştırma yöntemi (SVMD) önermişlerdir. Bu doğrultuda, trend serisi için doğrusal regresyon modeli ve her dalgalanma alt serisi için XGBoost regresyon modeli kullanılarak tahmin modelleri oluşturulmuş, XGBoost'un hiper-parametreleri, Bayesian optimizasyon algoritması (BOA) ile optimize edilmiştir. Ayrıca, endüstriyel müşterilerin elektrik tüketim davranışını etkileyen faktörler de dikkate alınarak, hibrit yöntemin doğruluğu artırılmıştır. Önerilen yöntem, Çin ve İrlanda'daki farklı endüstriyel müşteriler üzerinde çeşitli senaryolarla test edilmiştir. Sonuçlar, önerilen modelin, mevcut yük tahmin modellerine kıyasla önemli ölçüde daha iyi performans gösterdiğini ortaya koymaktadır.

Ilıç ve diğerleri (2021), çalışmalarında, zaman serileri tahmini için bir Explainable Boosted Linear Regression (EBLR) algoritması geliştirmişlerdir. Bu algoritma, modelin hatalarını regresyon ağaçlarıyla açıklayarak doğrusal olmayan özellikleri dahil etmektedir. Çalışmalarında, EBLR'nin temel model performansını önemli ölçüde artırdığını ve yorumlanabilir sonuçlar sağladığını gözlemlemişlerdir. Bu bulgular, EBLR'nin zaman serileri tahmini için umut verici bir yöntem olduğunu göstermektedir.

3. MATERYAL ve YÖNTEM

Bu çalışmada Birleşik Krallık'ın Ulusal Elektrik Şebekesi National Grid ESO tarafından açık kaynak olarak paylaşılan veri seti kullanılmıştır. Geçmiş iletim sistem talebi verileri dikkate alınarak SARIMA, Prophet, LSTM, Derin LSTM, XGBoost ve Lineer Arttırma Regresyonu kullanılarak hangi yöntemin daha iyi tahmin değerleri sağladığı araştırılmış ve 2024 yılının geri kalanı için talep tahminleri elde edilmiştir. Veri kümesi, Python programlama dilinde çalışan Google Colab üzerinde çeşitli kütüphane ve paketler kullanılarak işlenmiştir. Matematiksel hesaplamalar ve veri analizi için numpy ve pandas, veri görselleştirme için ise matplotlib ve seaborn kullanılmıştır. Zaman serisi modelleme ve istatistiksel analiz için statsmodels kullanılmıştır. Derin öğrenme ve makine öğrenimi modelleri oluşturmak ve eğitmek için keras ve tensorflow kullanılırken, ağaç tabanlı modellerle doğrusal regresyonu birleştiren lineartree ve veri ön işleme, model seçimi ve değerlendirme için de yaygın olarak kullanılan bir makine öğrenimi kütüphanesi olan scikit-learn tercih edilmiştir. Bu kütüphane ve paketler, geçmişte çok çeşitli veri analizi ve makine öğrenimi görevlerinde kullanılmış ve kapsamlı analizlere olanak sağlamıştır.

3.1. Birleşik Krallık'taki Elektrik Tüketim Talebi Verileri

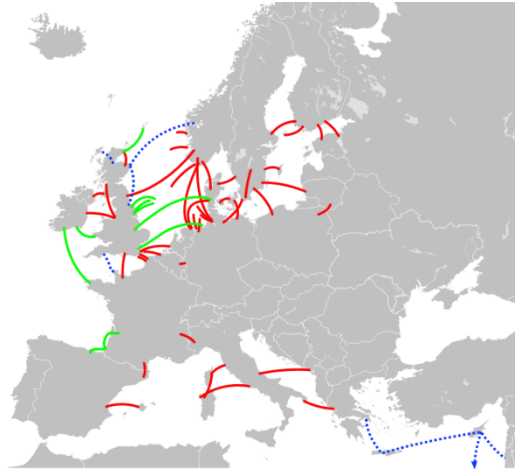
Birleşik Krallık'ın Ulusal Elektrik Şebekesi National Grid ESO tarafından yönetilen bu veri seti, enerji talebi, üretim kapasitesi ve çeşitli enerji kaynaklarına dair kapsamlı bilgiler sunmaktadır. National Grid ESO, Büyük Britanya'nın elektrik sistem operatörü olarak görev yapmaktadır. 2009 yılından bu yana, Büyük Britanya'daki elektrik talebiyle ilgili verileri toplamaktadır. Bu veriler saatte iki kez güncellenmekte olup, günde toplam 48 güncelleme yapılmaktadır. Bu güncellemeler sonucunda elde edilen değerler, takip eden yarım saatlik periyodun ortalama MW değerlerini yansıtmaktadır.

National Grid ESO'nun açık kaynak olarak sunduğu yıllık veri setlerinin birleştirilmiş halini gösteren "Historic_demand_year_2009_2024.csv" dosyası kullanılmıştır. (<https://www.nationalgrideso.com/data-portal/historic-demand-data>)

Veri setinde bulunan sütunlar arasında “settlement_date” (tarih) ve “settlement_period” (yarım saatlik dönemler) gibi önemli bilgiler yer almaktadır. Bu veri setleri, elektrik talebi ve tüketimi ile ilgili detaylı analizler yapmak ve geleceğe yönelik planlamaları desteklemek açısından büyük önem taşımaktadır.

Veri setinin temel bileşenlerinden biri olan Ulusal Talep (nd), istasyon yükü, pompalama depolama ve interkonnektör ihracatını içermeyen National Grid ESO’nun kendi iletim sistemlerine bağlı ölçülen üretimin toplamıdır. Bu değer, MW cinsinden ölçülür ve National Grid ESO tarafından yapılan operasyonel üretim ölçümlerine dayalı olarak hesaplanır. İletim Sistemi Talebi (tsd) ise nd’ye ek olarak, istasyon yükü tahmini, hidro depolama pompaları ve interkonnektör ihracatını içerir ve yine MW cinsinden ölçülür.

İnterkonnektörler, yüksek gerilimli doğru akım iletim hatlarıdır ve farklı elektrik sistemleri arasında elektrik enerjisi alışverişini sağlayarak enerji ticaretine olanak tanır. Bu yapılar genellikle uzun mesafelerde kullanılır, büyük güç kapasitesine sahiptir ve farklı ülkeler veya bölgeler arasında enerji alışverişini kolaylaştırır. Veri seti çeşitli interkonnektör akışlarını da içerir. IFA İnterkonnektör Akışı (ifa_flow), IFA2 İnterkonnektör Akışı (ifa2_flow), Moyle İnterkonnektör Akışı (moyle_flow), Doğu Batı İnterkonnektör Akışı (east_west_flow), Nemo İnterkonnektör Akışı (nemo_flow) ve Kuzey Denizi Bağlantı Elemanı Akışı (nsl_flow) gibi interkonnektörler, ilgili bağlantı elemanındaki akışı gösterir. Bu akışlar, - işareti GB’den çıkış gücünü, + işareti ise GB’ye ithal gücü gösterir ve MW cinsinden ölçülür.



Şekil 3.1. Avrupa'daki düşük güç hatları ve hizmet dışı hatlar hariç HVDC bağlantıları

Bölgesel Talep, İngiltere ve Galler Talebi (england_wales_demand), ND'ye benzer şekilde, İngiltere ve Galler için talebi gösterir. Yerleşik Üretim olarak, Yerleşik Rüzgâr Üretimi (embedded_wind_generation) ve Yerleşik Rüzgâr Kapasitesi (embedded_wind_capacity), İngiltere ve Galler'deki rüzgâr çiftliklerinden tahmin edilen rüzgâr üretimi ve kapasitesini ifade eder. Bu değerler gerçek çıkışı bilinmediği için en iyi modele dayalı bir tahmin sunar. Benzer şekilde, Yerleşik Güneş Üretimi (embedded_solar_generation) ve Yerleşik Güneş Kapasitesi (embedded_solar_capacity), GB'deki güneş panellerinden elde edilen üretim ve kapasiteyi tahmin eder. Rüzgâr ve güneş üretimi, hava verilerine dayalı fiziksel modeller kullanılarak tahmin edilir. Bu modeller, kurulu kapasite ve konumu kullanarak üretim tahminlerini oluşturur.

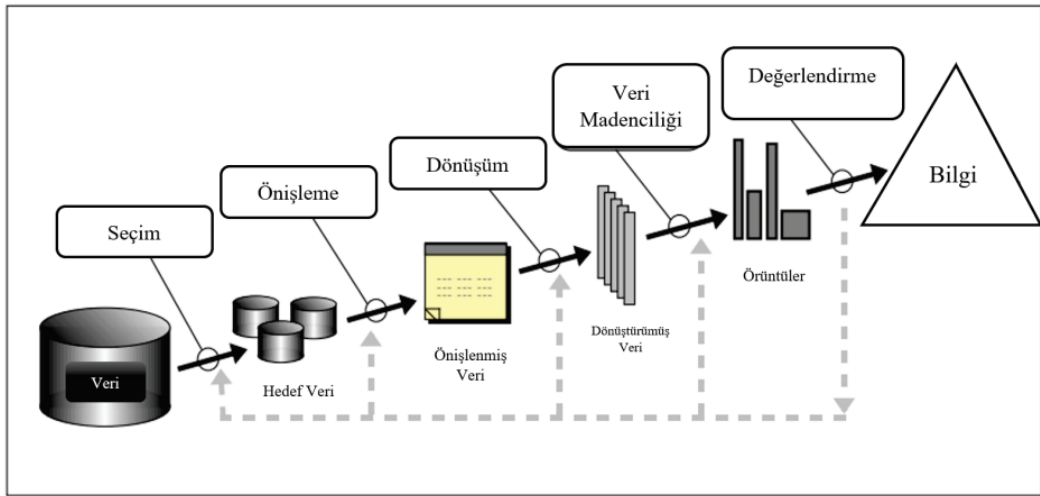
"non_bm_stor" Birleşik Krallık elektrik piyasasında kullanılan bir terimdir. Bu terim, "Balance Mechanism" (Dengeleme Mekanizması) dışında kalan enerji depolama sistemlerini ifade eder. Balance Mechanism, ulusal elektrik şebekesindeki arz ve talebi dengelemek için kullanılan bir sistemdir ve genellikle büyük enerji üreticilerini ve tüketicilerini kapsar. "non_bm_stor" ise bu sistemin dışında kalan, daha küçük ölçekli veya bağımsız enerji depolama tesislerini ifade eder. Bu tür depolama sistemleri, genellikle yenilenebilir enerji kaynaklarıyla entegre çalışır ve enerji piyasasına doğrudan katılmadan yerel ölçekte enerji depolama ve yönetimi sağlar. Bu, bataryalar, pompa depolama sistemleri veya diğer enerji depolama teknolojilerini içerebilir. Pompalama Depolama Talebi (pump_storage_pumping) ise hidro pompalama depolama ünitelerindeki pompalama talebini gösterir.

Bu veri seti, enerji talebi ve üretim kapasitesi hakkında detaylı bilgiler sunarak enerji yönetimi ve planlamasına büyük katkı sağlar. Ulusal Talep ve İletim Sistemi Talebi, bölgesel talep, yerleşik rüzgâr ve güneş üretimi, kısa vadeli işletme rezervi ve interkonnektör akışları gibi çeşitli bileşenler, enerji sisteminin verimliliğini artırmak ve enerji arz güvenliğini sağlamak amacıyla analiz edilebilir. Bu tür veri setleri, enerji sektöründe karar vericilere ve planlayıcılara değerli bilgiler sunarak, sürdürülebilir ve güvenilir bir enerji sistemi oluşturulmasına yardımcı olur.

3.2. Veri Madenciliği

Veri madenciliği, büyük veri kümelerinden bilgi çıkarmak için istatistiksel ve matematiksel tekniklerin kullanıldığı disiplinler arası bir alandır. Veri madenciliği, bilinmeyen, gizli veya faydalı bilgileri keşfetmek, desenleri tanımlamak ve gelecekteki eğilimleri tahmin etmek için veri analizi yapar. Bu sayede karar verme süreçlerini geliştirmek, pazarlama stratejilerini iyileştirmek ve iş süreçlerini iyileştirmek gibi amaçlar için kullanılır.

Veri madenciliği, geleneksel tahmin yöntemlerinin yetersiz kaldığı durumlarda etkili bir çözüm olarak dikkat çekmektedir. Büyük veri yığınlarından ilginç ve faydalı desenler çıkarmayı amaçlayan bu yöntem, betimleyici ve öngörücü olmak üzere iki ana göreve ayrılır. Betimleyici görevler, veri setinin genel özelliklerini ortaya koyarken; öngörücü görevler, mevcut verilerden yola çıkarak geleceğe yönelik tahminler yapmayı hedefler (Puspita ve diğerleri, 2019).



Şekil 3.2. Veri madenciliği süreci (Fayyad ve diğerleri, 1996)

Veri Madenciliği Aşamaları:

- **Veri Toplama:** İlk adım, analiz edilecek verilerin toplanmasıdır. Bu veriler çeşitli kaynaklardan gelebilir, örneğin veri tabanları, veri ambarları, web siteleri, sensörler ve diğer veri toplama araçları.

- **Veri Hazırlama:** Toplanan veriler genellikle ham ve kullanıma hazır değildir. Bu aşamada veriler temizlenir, uygun biçime dönüştürülür ve eksik veriler tamamlanır. Ayrıca, veri setleri arasında birleştirme ve filtreleme işlemleri yapılır.
- **Veri Dönüştürme:** Verilerin analize uygun hale getirilmesi için dönüştürme işlemleri yapılır. Bu aşamada özellik seçimi, özellik çıkarma ve veri normalizasyonu gibi işlemler gerçekleştirilir.
- **Veri Madenciliği:** Bu aşamada, verilerden anlamlı kalıplar ve ilişkiler çıkarmak için çeşitli algoritmalar ve teknikler kullanılır. Bu teknikler arasında sınıflandırma, kümeleme, regresyon, birliktelik kuralları analizi ve zaman serisi analizi bulunur.
- **Sonuçları Yorumlama ve Değerlendirme:** Elde edilen sonuçlar yorumlanır ve değerlendirilir. Bu adım, bulunan kalıpların ve ilişkilerin anlamlı olup olmadığını ve iş gereksinimlerine uygunluğunu kontrol etmeyi içerir.

Elde edilen bilgilerin karar vericilere sunulması ve uygulanabilir sonuçların iş süreçlerine entegrasyonu yapılır (Fayyad ve diğerleri, 1996).

3.3. Keşifsel Veri Analizi ve Özellik Mühendisliği

Veri biliminde, keşifsel veri analizi (EDA) ve özellik mühendisliği, etkili modeller oluşturmak için el ele çalışan iki önemli aşamadır. EDA, verileri görselleştirme ve istatistiksel analizlerle inceleyerek, gizli kalıpları, anomalileri ve ilişkileri ortaya çıkarmayı hedefler. Bu bilgiler, veri setinin daha iyi anlaşılmasına ve sonraki adımların planlanmasına yardımcı olur. Özellik mühendisliği ise, modeller için daha anlamlı ve uygun özellikler oluşturarak ham verileri dönüştürür. Mevcut özellikleri birleştirerek yeni özellikler türetmek, gereksiz veya tekrarlayan özellikleri kaldırmak gibi işlemleri içeren bu süreç, modelin doğruluğunu ve genelleştirilebilirliğini artırır. (<https://medium.com/@betulavcil/ke%C5%9Fi%CC%87fsel-veri%CC%87-anali%CC%87zi%CC%87-eda-exploratory-data-analysis-f620cfd12420>)

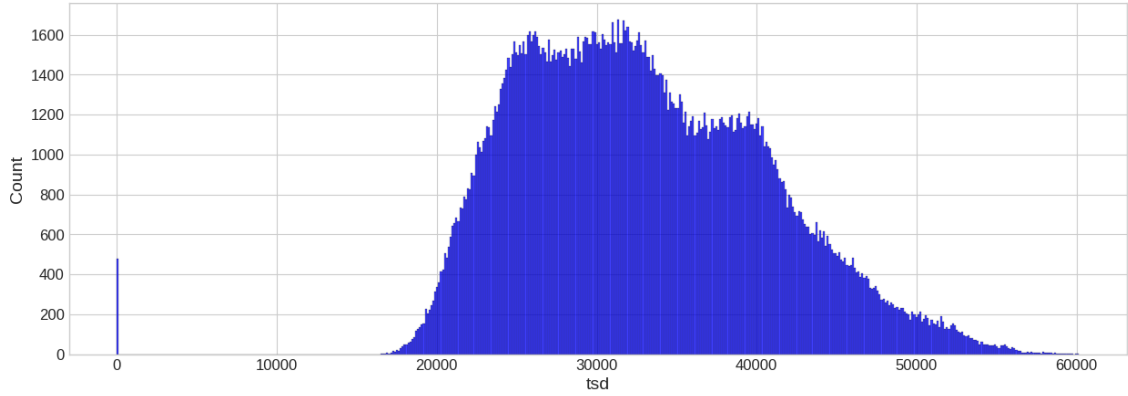
Veri kümesini anlamak ve dönüştürmek amacıyla, öncelikle boş değerli sütunlar incelenmiş, analiz sonucunda dört sütunda boş değerler tespit edilip bu sütunların analiz

sonuçlarını etkilemeyecekleri düşünülerek veri kümesinden çıkarılmıştır (nsl_flow, eleclink_flow, scottish_transfer, viking_flow).

Yerleşim dönemi değerleri, günde kaç numune alındığını ifade eder. Her bir örneğin günün 30 dakikasını temsil ettiği göz önüne alındığında günlük maksimum örnek sayısının 48 olması beklenir. Bu nedenle, 48'den fazla örnek içeren gün satırları temizlenmiştir.

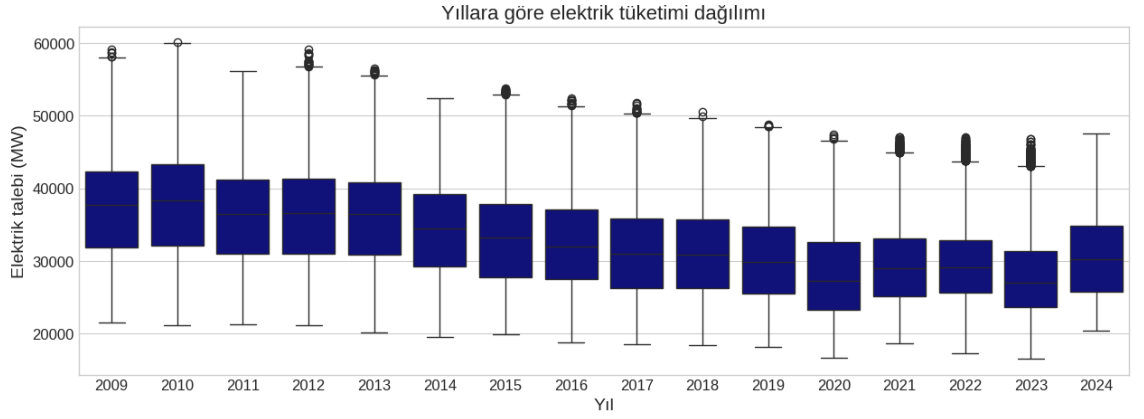
Tarih biçimini saatlik değerleri içerecek şekilde değiştirmek için “period_hour” sütununa yerleşim dönemine bağlı saat değerleri yazılıp ve gerçek tarihle birleştirilmiştir. Zaman serisi verilerinin mevsimselliği göz önüne alınarak, haftanın günü ve yılın günü gibi tarihten yazdırılan yeni özellikler eklenmiştir.

Zaman serisi tahminlerinde resmî tatiller önemli bir rol oynar. Bu nedenle, her günün resmî tatil olup olmadığını belirten yeni bir sütun eklenmiştir. Veri seti İngiltere ve Galler'in elektrik talebini kapsadığından, her iki ülkenin aynı resmî tatil günlerine sahip olup olmadığı da kıyaslanmıştır.



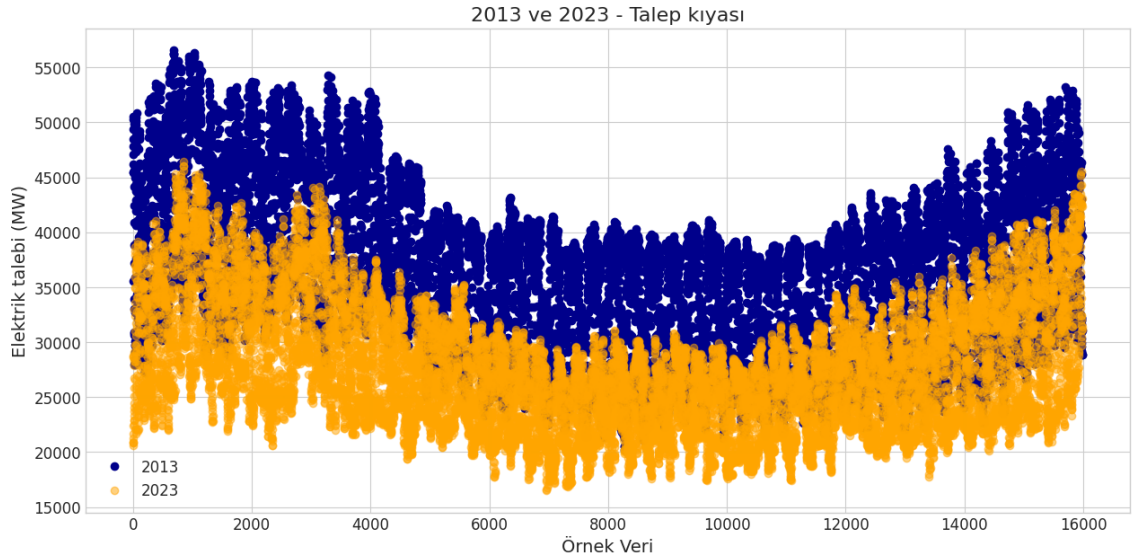
Şekil 3.3. İletim sistem talebi (tsd) sıfır olan satırların sayısı

İletim sistem talebi (tsd) değeri sıfıra eşit olan satırların günleri veri setinden kaldırılmıştır.



Şekil 3.4. Yıllara göre elektrik tüketimi dağılımı

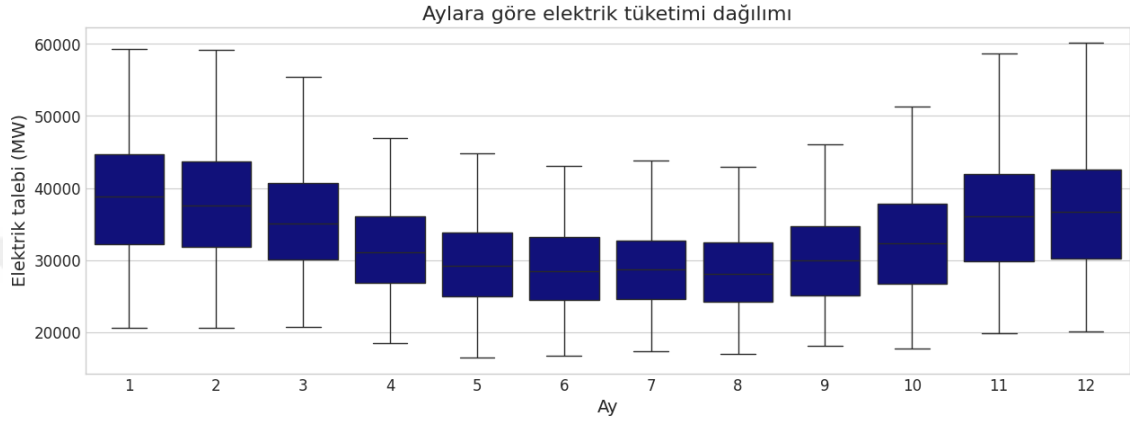
2009-2024 yılları arasında iletim sistemi talebi grafiği incelendiğinde talepte genel bir düşüş gözlenmektedir.



Şekil 3.5. 2013 ve 2023 yıllarının elektrik tüketimi kıyası

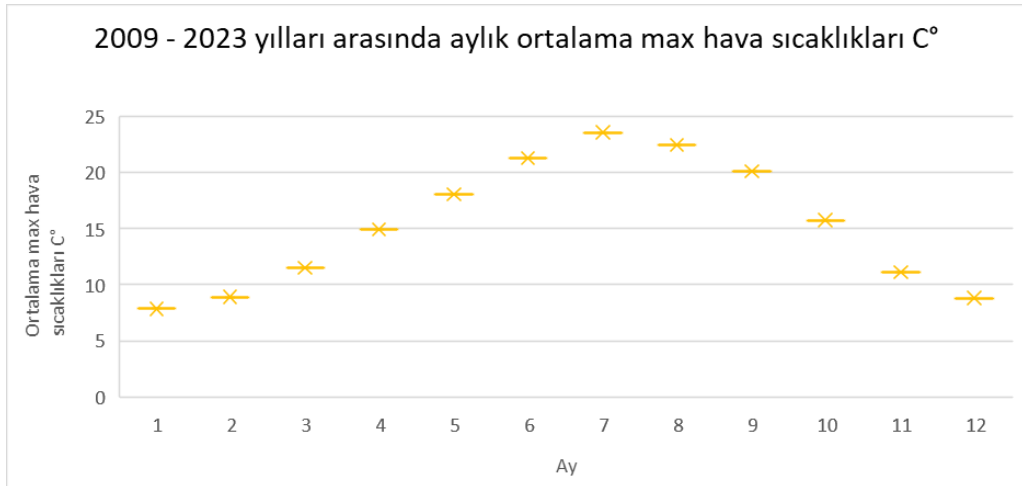
2013 ve 2023 yıllarındaki talep karşılaştırıldığında, ortalama 10 000 MW'lık bir düşüş olduğu görülmektedir. Bu, tüm verilerin bulunduğu görseldeki genel düşüş trendini desteklemektedir. 26 Eylül 2024 tarihli Birleşik Krallık hükümetinin elektrik tüketimi raporuna göre, İngiltere'de nihai enerji tüketimi 2000'li yılların başından 2023 yılı sonuna kadar azalma eğilimindedir. Bu düşüş, sıcak hava koşulları, davranış değişiklikleri ve artan fiyatlar gibi faktörlerle desteklenmiştir. 2023 yılında sanayi sektöründe de rekor düşük enerji tüketimi gözlemlenmiştir. Öte yandan, ulaşım sektöründe enerji tüketimi

artmıştır, özellikle havacılık sektörü Covid-19 pandemisinden toparlanmaya devam etmiştir. Evsel sektörde enerji yoğunluğu, enerji tasarrufu önlemleri sayesinde azalmaya devam etmiştir. Ulaşım sektöründe de enerji yoğunluğu iyileşmiş, seyahat kısıtlamalarının kaldırılmasıyla yolcu yüklemelerinde artış yaşanmıştır. (<https://www.gov.uk/government/statistics/energy-consumption-in-the-uk-2024>)



Şekil 3.6. Aylara göre elektrik tüketimi dağılımı

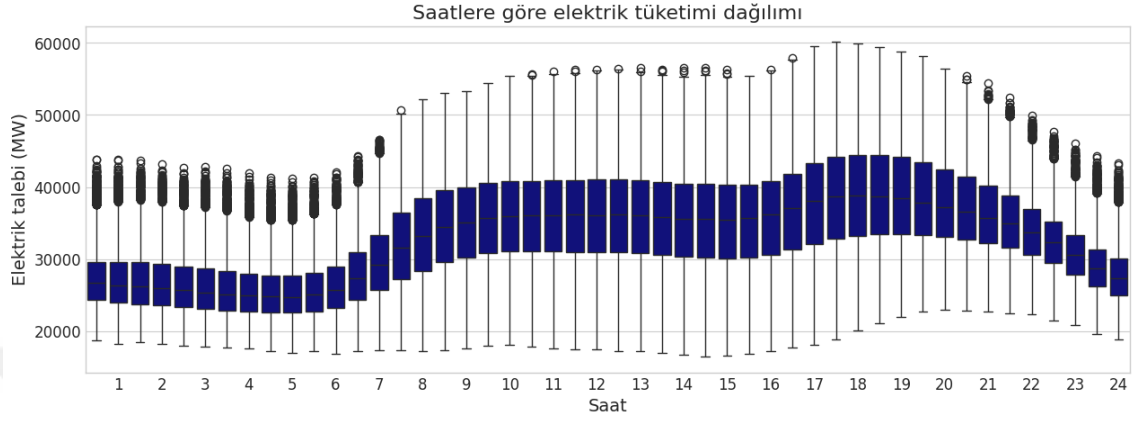
Aylara göre elektrik tüketimi dağılımı incelendiğinde, beklenildiği gibi kış aylarında tüketimin arttığı, yaz aylarında ise en düşük tüketim değerlerinin olduğu gözlenmektedir.



Şekil 3.7. Aylara göre en yüksek hava sıcaklıkları (°C)

Aylık ortalama hava sıcaklığı verisi Met Office resmî web sitesinde açık kaynak olarak sağlanan verilerden Oxford lokasyonu verileri dikkate alınarak kıyaslanmıştır. Met Office, Birleşik Krallık'ın ulusal hava tahmini ve iklim servisi olarak bilinir. 1854 yılında

kurulmuş olan Met Office, dünya çapında tanınan en eski ve en saygın meteoroloji kuruluşlarından biridir. Aylara göre elektrik tüketimi dağılımı havanın sıcaklık değeri ile uyumlu bir grafik sergilemektedir.



Şekil 3.8. Saatlere göre elektrik tüketimi dağılımı

Saatlik talep incelendiğinde, beklenildiği gibi uyku saatleriyle uyumlu olarak 24:00-6:00 saatleri arasında daha düşük olan elektrik talebi değeri, saat 7'den sonra artış göstermekte ve akşam saatlerinde talep artışı gözlemlenmektedir. Birleşik Krallık Hükümeti'nin resmî sitesinde yayınlanan "Elektrik Talebindeki Mevsimsel Değişiklikler" başlıklı makale, saatlere göre elektrik tüketimi dağılımı grafiğini desteklemektedir. (https://assets.publishing.service.gov.uk/media/5a7ba36540f0b645ba3c59b2/Seasonal_variations_in_electricity_demand.pdf)

3.4. Elektrik Enerjisi Talep Tahmin Yöntemleri

Elektrik enerjisi talep tahminlerinde genellikle enerji dağıtım şirketleri veya enerji piyasası işletmecileri tarafından sağlanan geçmiş yıllara ait tarihsel tüketim verileri kullanılır. Sıcaklık, nem gibi meteorolojik veriler ve nüfus, ekonomik büyüme gibi sosyodemografik veriler de elektrik enerjisi talep tahminlerinde önemli rol oynar.

Elektrik enerjisi talep tahminleri kısa, orta ve uzun vadeli olarak yapılmaktadır. Kısa vadeli tahminler genellikle günlük ve haftalık olarak yapılırken, orta vadeli tahminler aylık ve mevsimsel bazda gerçekleştirilir. Uzun vadeli tahminler ise yıllık ve daha uzun dönemleri kapsayan tahminlerdir (Zhang ve diğerleri, 2023).

Elektrik talebini tahmin etmek için çeşitli yöntemler kullanılır ve her birinin kendi güçlü ve zayıf yönleri vardır. Tahmin yöntemleri üç grupta incelenebilir:

1. **Niteliksel Yöntemler:** Subjektif görüşlere dayanarak genellikle nitel verilerin yorumlanmasına dayanır ve gelecekteki trendleri ve olası senaryoları belirlemede kullanılır. Bu yöntemler arasında uzman görüşleri, odak grup görüşmeleri, anketler ve senaryo analizleri gibi teknikler bulunmaktadır.
2. **Niceliksel Yöntemler:** Matematiksel ve istatistiksel verilerin analiz edilmesi ve matematiksel modellemelerin yapılarak gelecekteki trendlerin ve tahminlerin belirlenmesinde kullanılır. Bu yöntemler arasında regresyon analizi, zaman serisi analizi, yapay zekâ modelleri ve makine öğrenmesi algoritmaları gibi teknikler bulunmaktadır (Akgül, 2013).

Zaman serisi analizi, zamanla değişen verilerin analiz edilmesi ve gelecekteki trendlerin tahmin edilmesi için kullanılan klasik bir yöntemdir. Hareketli Ortalama (Moving Average), kısa vadeli dalgalanmaları düzeltmek ve genel eğilimleri belirlemek için kullanılır. Örneğin, 12 aylık hareketli ortalama, her ay için son 12 ayın ortalamasını alarak mevsimsel etkileri düzeltir. Üstel Düzeltme (Exponential Smoothing), geçmiş verilerin ağırlıklı ortalamasını alarak daha güncel verilere daha fazla ağırlık verir. Bu teknik, özellikle ani değişimlere hızlı tepki vermesi gereken tahminler için kullanışlıdır. ARIMA (AutoRegressive Integrated Moving Average), hem otoregresif (AR) hem de hareketli ortalama (MA) bileşenlerini içeren karmaşık bir modeldir. ARIMA, veriyi trend, mevsimsellik ve rastgele bileşenlere ayırarak tahmin yapar. SARIMA (Seasonal ARIMA), ARIMA modelinin mevsimsel bileşenleri içerecek şekilde genişletilmiş halidir. Elektrik talebindeki mevsimsel değişimleri daha iyi tespit etmek için kullanılır. GARCH (Generalized Autoregressive Conditional Heteroskedasticity), zamanla değişen varyansları (heteroskedastisiteyi) modellemek için kullanılır. Elektrik talebindeki volatilitiyi anlamak için yararlıdır. Yük tahminlerinde istatistiksel modellerin kullanımı, temel varsayımların sağlanması için genellikle durağanlık testi ve fark alma gibi birkaç ön işleme adımını gerektirir. Bu durum,

veri hazırlama ve işleme sürecini oldukça zahmetli hale getirebilir. Buna karşılık, istatistiksel modeller verilerin derinlemesine analizini ve uyumunu sağlayarak, güç sistemlerinin işletimi ve planlaması için önemli referanslar ve rehberlik sunar. Buna rağmen, bu modeller yük değişikliklerindeki belirsizlikle başa çıkmada ve güç sistemlerinin karmaşıklığına uyum sağlamada yetersiz kalabilir. Bu sorunu aşmak için bazı araştırmacılar daha gelişmiş yöntemler önermektedir. Örneğin, Goswami ve Kandali (2020) yük verilerinin mevsimselliğini dikkate alan SARIMA modellerinin, geleneksel ARIMA modellerine göre tahmin doğruluğu ve hata kontrolü açısından üstün olduğunu belirtir. Sonuç olarak, istatistiksel modeller elektrik yükü tahmininde belirli avantajlara ve sınırlamalara sahiptir. Geçmiş verileri analiz ederek bazı temel referanslar ve rehberlik sunabilse de yük değişikliklerindeki belirsizlik ve karmaşıklıkla başa çıkmak için daha kapsamlı ve gelişmiş yöntemlerin dikkate alınması gerekmektedir (Gu ve diğerleri, 2024).

Regresyon analizi, değişkenler arasındaki nedenselliği belirlemek, tahmin yapmak ve gelecekteki olayları öngörmek için kullanılan istatistiksel bir yöntemdir. Bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi ifade eden denklemin doğrusal olup olmamasına göre iki türü vardır. En yaygın kullanılan regresyon türü doğrusal regresyondur ve basit doğrusal regresyon ile çoklu doğrusal regresyon olmak üzere iki alt kategoriye ayrılır. Doğrusal Olmayan Regresyonun ise polinom regresyonu, lojistik regresyon, üstel regresyon gibi alt türleri bulunmaktadır.

Makine öğrenimi modelleri, veriye dayalı tahminlerin yapılmasına olanak sağlar. Yapay Sinir Ağları (ANN), veri içindeki karmaşık, doğrusal olmayan ilişkileri tanımlayabilir. Kısa ve uzun vadeli tahminler için kullanışlıdır. Destek Vektör Makineleri (SVM), regresyon görevleri için etkili olup doğrusal olmayan ilişkileri ele alabilir. Rastgele Ormanlar, birden fazla karar ağacını kullanarak doğruluğu artıran topluluk öğrenme yöntemidir (Gülay ve diğerleri, 2024).

Yapay zekâ ve derin öğrenme tabanlı modeller, yapay zekâ teknolojisinin hızlı gelişimi ile sinir ağı algoritmaları yük tahminlerine dahil edilmiştir. Bu modeller,

kısa vadeli elektrik yükünün zaman serisi özelliklerini iyi bir şekilde tahmin edebilir. Buna karşılık, rastgele üretilen ağırlıklar ve önyargılar nedeniyle kolayca etkilenebilirler ve veri ön işleme modülleri genellikle eksiktir. Facebook tarafından geliştirilen Prophet yöntemi, son zamanlarda çeşitli alanlarda başarıyla uygulanmaktadır. Derin Sinir Ağları (DNN), büyük veri setleri ve karmaşık zaman serisi verilerini işleyebilme kapasitesi ile öne çıkmaktadır.

Son yıllarda, bireysel modellerin dezavantajlarını aşmak için hibrit modeller kullanılmıştır. Bu modeller, farklı optimizasyon algoritmaları ve derin öğrenme tekniklerini birleştirir. Hibrit modeller, tekil modellerden daha iyi performans gösterse de gürültü azaltma ve akıllı optimizasyon algoritmalarının doğrudan kullanımı genellikle uygunsuz olabilir ve daha fazla araştırma gerektirir (Xia ve diğerleri, 2023; Jin ve diğerleri, 2024; Ensafi ve diğerleri, 2022).

- 3. Karma Yöntemler:** Nicel ve nitel yöntemlerin birlikte kullanılmasıyla daha kapsamlı ve doğru tahminler elde etmeyi amaçlayan yöntemlerdir.

Matematiksel ve istatistiksel veri modellerinin performansını değerlendirmek için hata analiz metrikleri kullanılır. Ortalama mutlak hata (MAE), ortalama mutlak yüzde hata (MAPE) yük verisi tahmin modelinin genel tahmin yeteneğini tanımlar. Performansın medyan mutlak değeri (MdAPE), istatistiksel yayılımı ölçer. Ortalama hata (AE) ve hata karelerinin ortalamasının karekökü (RMSE) ise tahmin edilen yük ile gerçek yük arasındaki yayılımın derecesini farklı şekillerde yansıtır. Bu metrikler, tahmin edilen yük ile gerçek yük arasındaki farklılıkları ve yayılımları analiz etmeye yardımcı olur (Xia ve diğerleri, 2023).

Bu çalışmada başarı ölçümü olarak kullanılan metriklerden MAPE, aşağıdaki formülle hesaplanır:

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{Gerçek[i] - Tahmin[i]}{Gerçek[i]} \right| \quad (3.1)$$

Düşük MAPE değerleri, modelin gerçek değerlere yakın tahminler yaptığını ve dolayısıyla daha iyi bir performans sergilediğini gösterir. MAPE, modelin hatalarını yüzdesel olarak ifade ettiği için, farklı ölçeklerdeki zaman serileri arasında veya farklı modellerin performanslarını karşılaştırmada faydalı bir ölçüttür. Buna karşılık, gerçek değerler sıfıra yakın olduğunda MAPE'nin tanımsız olabileceği söylenebilir.

Kök Ortalama Kare Sapma (RMSD) veya Kök Ortalama Kare Hata (RMSE), gerçek veya tahmin edilen değerler ile gözlemlenen değerler veya bir tahminci arasındaki farkların ölçülmesinde sıkça kullanılan iki yakından ilişkili ölçüdür (Fan ve diğerleri, 2021; Khan ve diğerleri, 2022; Mosavi ve diğerleri, 2019).

RMSE, aşağıdaki formülle hesaplanır:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Tahmin[i] - Gerçek[i])^2}{N}} \quad (3.2)$$

Farklı modellerin kıyaslamasında CPU zamanı ve duvar saati zamanı (wall time) da önemli değerlendirme kriterleri olarak kullanılabilir. CPU Zamanı, bir işlemin veya modelin yürütülmesi sırasında işlemcinin ne kadar süre aktif olarak çalıştığını ifade eder. Bu, işlemcinin verimli kullanılıp kullanılmadığını anlamak için kritik bir göstergedir. Duvar Saati Zamanı ise bir işlemin veya modelin başlangıcından bitişine kadar geçen gerçek süreyi temsil eder ve bu süre, kullanıcı perspektifinden işlemin ne kadar sürede tamamlandığını gösterir. Bu süreler, bir modelin seçiminde önemli kriterlerden biri olabilir. Örneğin, belirli bir uygulama için farklı algoritmalar veya modeller denendiğinde, daha düşük CPU zamanı ve duvar saati zamanı gerektiren bir model tercih edilebilir. Bu, özellikle kaynakların sınırlı olduğu veya hızlı geri dönüş süresinin kritik olduğu durumlarda önem kazanır. Dolayısıyla, CPU ve duvar saati zamanları, modellerin performansını kıyaslamak ve hangi modelin belirli bir senaryo için daha uygun olduğunu belirlemek için kullanılabilecek değerli ölçütlerdir.

Model doğrulama, bir tahmin veya makine öğrenmesi modelinin performansını değerlendirmek ve doğruluğunu test etmek için kullanılan süreçtir. Amaç, modelin yeni ve görülmemiş verilere nasıl tepki vereceğini ölçmektir. Model doğrulama yöntemi

olarak bu çalışmada kullanılan Çapraz Doğrulama (Cross Validation), bir modelin ne kadar iyi çalıştığını ve ne kadar genelleştirebildiğini kontrol etmek için kullanılan bir yöntemdir. Bu yöntem, veriyi farklı parçalara böler ve modelin her parçada nasıl çalıştığını test eder. K-Katlı Çapraz Doğrulama (K-Fold Cross Validation) yönteminde, veri seti K eşit parçaya bölünür. Her seferinde bir parça test verisi olarak kullanılırken, geri kalan K-1 parça eğitim verisi olarak kullanılır. Bu işlem K kez tekrarlanır ve her bir parçanın performansı hesaplanır. Sonuçlar, ortalama performans olarak raporlanır. Tek Bırakma Çapraz Doğrulama (Leave-One-Out Cross Validation) yönteminde, veri setindeki her bir gözlem tek başına test verisi olarak kullanılırken, diğer gözlemler eğitim verisi olarak kullanılır. Bu işlem veri setindeki gözlem sayısı kadar tekrarlanır ve sonuçlar ortalama performans olarak raporlanır. Katmanlı K-Katlı Çapraz Doğrulama (Stratified K-Fold Cross Validation) yönteminde, K-Katlı Çapraz Doğrulama uygulanırken, her parçanın veri setindeki sınıf dağılımını koruması sağlanır. Çapraz Doğrulama, modelin aşırı öğrenme (overfitting) veya eksik öğrenme (underfitting) yapıp yapmadığını anlamak için kullanılır. (<https://miracozturk.com/capraz-dogrulama-teknikleri-cross-validation/>)

Matematiksel ve istatistiksel veri modellerinde doğru hiperparametrelerin seçimi, modelin başarımı üzerinde büyük bir etkiye sahiptir. Hiperparametrelerin doğru seçimi için çeşitli teknikler ve stratejiler kullanılır. Bu çalışmada kullanılan Izgara Arama, bu tekniklerdendir. Izgara Arama (Grid Search), modelin hiperparametrelerini optimize etmek amacıyla kullanılır. En iyi model performansını elde etmek için farklı hiperparametre kombinasyonlarını sistematik bir şekilde dener. Öncelikle, optimize edilmek istenen hiperparametreler ve bu hiperparametreler için denenecek değerler belirlenir. Örneğin, bir karar ağacı modelinde max_depth ve min_samples_split hiperparametreleri optimize edilecekse, bu hiperparametreler için denenecek değerler bir ızgara (grid) oluşturur. Örneğin, max_depth için [3, 5, 7] ve min_samples_split için [2, 5, 10] değerleri denenecekse, toplamda $3 \times 3 = 9$ farklı hiperparametre kombinasyonu denenir. Her bir hiperparametre kombinasyonu için model eğitilir ve çapraz doğrulama kullanılarak performansı değerlendirilir. Izgara Arama, her kombinasyonun performansını ölçer ve en iyi performansı sağlayan kombinasyonu seçer. En iyi hiperparametre kombinasyonu seçildikten sonra, bu kombinasyon kullanılarak model yeniden eğitilir ve nihai model elde edilir (Su ve diğerleri, 2012).

3.5. Zaman Serileri

Zaman serileri, belirli zaman aralıklarında toplanan verilerin zamanla nasıl değiştiğini analiz etmek için kullanılan bir veri analizi yöntemidir. Zaman serileri genellikle trendler, mevsimsel etkiler ve rasgele değişimler gibi bileşenleri içerir. Sıkça kullanılan bazı temel terimler şunlardır:

Trend (T): Zaman serisinde uzun vadeli yönü gösteren düzenli artış veya azalışı ifade eder. Trend artan, azalan ve yatay yönlü olabilir.

Mevsimsellik (S): Belirli zaman aralıklarında tekrarlayan bir model gösteren zaman serisine mevsimsel model veya mevsimsellik adı verilir. Mevsimsel kalıplar, birçok zaman serisi türünde gözlenebilir. Örneğin, yaz aylarında ısıtma maliyetleri düşer, kış aylarında artar.

Durağanlık (Stationarity): Zaman serisinin istatistiksel özelliklerinin zamanla değişmediği durumu ifade eder. Zaman serilerinin modellenmesi için öncelikle serinin durağan olup olmadığı belirlenmelidir (Akgül, 2013).

Durağanlık, serinin ortalama, varyans, kovaryans ve birleşik dağılım fonksiyonunun zamana göre sabit kalması anlamına gelir. Tam durağanlık, bu dört bileşenin de sabit kalmasını gerektirirken, zayıf durağanlık sadece ortalama, varyans ve kovaryansın sabit kalmasını gerektirir. Tam durağanlık kavramı uygulamada yaygın rastlanan bir durum değildir. Bu sebeple, zaman serisi analizlerinde genellikle zayıf durağanlık düzeyinde inceleme yapılır (Tarkun, 2023; Sevüktekin ve Çınar, 2007; Işığışık, 1994; Gülay ve diğerleri, 2024).

Augmented Dickey–Fuller testi (ADF), bir zaman serisinin durağanlık özelliğini belirlemek için kullanılır. Test istatistiği negatif olduğunda ve ele alınan zaman serisi durağanlık özelliğine sahip olduğunda, daha negatif bir değer, daha güçlü bir durağanlık göstergesi olarak kabul edilir. ADF testinde, sıfır hipotezi serinin durağan olmadığını veya birim kök içerdiğini ifade eder ($H_0: \alpha = 0$). Alternatif hipotez ise, serinin durağan

olduğunu veya birim kök içermediğini belirtir ($H_0: \alpha < 0$). ADF testi için hesaplanan p değeri test istatistiğinin istatistiksel olarak ne kadar anlamlı olduğunu gösterir. p değeri, sıfır hipotezin doğru olduğu varsayımı altında, gözlemlenen test istatistiği kadar aşırı veya daha aşırı bir sonuç elde etme olasılığıdır. p değeri, bir hipotezi reddetmek veya kabul etmek için bir eşiğe (genellikle 0,05) kıyasla kullanılır. Düşük bir p değeri (örneğin, 0,05'ten küçük), sıfır hipotezin reddedilmesi gerektiğini gösterir. Genellikle p değeri, belirli bir güven seviyesinde (genellikle %95) hipotezin reddedilip reddedilmeyeceğini belirlemek için kullanılır. Düşük p -değeri ($<0,05$), test istatistiğinin istatistiksel olarak anlamlı olduğunu gösterir. Kritik değerler, test istatistiğinin belirli bir güven seviyesindeki (genellikle %1, %5, %10) anlamlılığını belirlemek için kullanılır. Test istatistiği, kritik değerlerden daha düşükse, hipotez reddedilir ve serinin durağan olduğu kabul edilir (Akgül, 2013; Fan ve diğerleri, 2021; Arıkök, 2024).

ADF testi, zaman serisinin birinci farklarını kullanarak otoregresif bir model kurar ve bu modeldeki katsayıların anlamlılığını test eder. Genel formülasyon şu şekildedir:

$$\Delta Y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_p \Delta y_{t-p} + \epsilon_t$$

Burada, α zaman serisinin dalgalandığı seviyeyi ayarlayan sabit terimdir. Δy_t zaman serisi verilerindeki farkı ve βt trend faktörünü ifade eder. γy_{t-1} , mevcut fark (Δy_t) ile zaman serisinin önceki seviyesi (y_{t-1}) arasındaki ilişkiyi yakalar. $\delta_1, \delta_2, \dots, \delta_p$, zaman serisinin gecikmeli farkları için otokorelasyonları hesaba katan katsayılardır. $\Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p}$ ise zaman serisinin gecikmeli farklarıdır. ϵ_t , zaman serisindeki açıklanamayan varyasyonu yakalayan hata terimidir. (<https://unofficed.com/courses/newton-week-3/lessons/augmented-dickey-fuller-test/>)

Zaman serilerinin modellenmesinde AR (otoregresif), MA (hareketli ortalama) ve ARMA (otoregresif hareketli ortalama) gibi durağan süreçler kullanılır. Buna karşılık, zaman serileri genellikle durağan değildir ve bu nedenle serinin farkları alınarak durağanlık sağlanır. Bu tür seriler ARIMA (otoregresif bütünleştirilmiş hareketli ortalama) süreci olarak adlandırılır (Tarkun, 2023).

Otoregresif Model (AR)

Bir zaman serisi kendi gecikmeli değerlerinin bir fonksiyonu şeklinde ifade ediliyorsa otoregresif süreç olarak tanımlanmaktadır. AR(p) modeli, bağımlı değişkeni Y_t geçmiş p dönemdeki bağımsız değişkenleri $(Y_{t-1}), (Y_{t-2}), \dots, (Y_{t-p})$ olan çoklu doğrusal regresyon modelidir. “ p ” değeri o anki gözlem değerini tahmin edebilmek için sondan itibaren kaç tane geçmiş gözleme bakılacağını ifade eder. Bu yüzden adı geçen modele kendi kendine regresyon anlamında otoregresif model adı verilmektedir (Cárdenas ve diğerleri, 2016). “ p .” dereceden bir AR(p) modelinin yapısı:

$$Y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + u_t \quad (3.3)$$

olarak ifade edilir. Burada Y_t , tahmin edilen değeri, $(Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$ geçmiş gözlem değerlerini, $(\phi_1, \phi_2, \dots, \phi_p)$ geçmiş gözlem değerleri için oto-regresif parametreleri, μ sabit terim ve u_t hata terimini ifade eder. AR modeli, geçmiş değerlerin gelecekteki değerler üzerindeki etkisini modellemek için kullanılır (Tarkun, 2023).

Hareketli Ortalama Modeli (MA)

Bir değişkenin t dönemindeki değeri aynı dönemdeki hata terimi u_t ve hata teriminin önceki dönemlerine ait gecikmeli değerleri ile belirleniyorsa $(u_{t-1}, \dots, u_{t-q})$ bu süreç hareketli ortalama süreci adı verilmektedir. MA(q) modelleri, içerdikleri geçmiş dönem hata terimi sayısına göre birinci, ikinci ve genel olarak q 'inci dereceden MA modelleri olarak isimlendirilir. Buradaki q değeri o anki gözlem değerini tahmin edebilmek için sondan itibaren kaç tane hata terimlerinin ağırlıklı ortalamasının kullanılacağını ifade eder. “ q .” dereceden bir MA(q) modelinin yapısı:

$$Y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \quad (3.4)$$

olarak ifade edilir. Burada Y_t , tahmin edilen değeri, $(u_{t-1}, \dots, u_{t-q})$ hata terimlerini, $(\theta_1, \theta_2, \dots, \theta_q)$ hata terimleri ile ilgili hareketli ortalama parametreleri, μ sürecin ortalaması olan

sabit bir sayıyı, u_t hata terimini ifade etmektedir. MA modeli, hata terimlerinin geçmiş değerlerinin etkisini modellemek için kullanılır (Tarkun, 2023).

Otoregresif Hareketli Ortalama Model (ARMA)

AR ve MA modellerinin birleşimi olan ARMA modeli hem geçmiş değerlerin hem de hata terimlerinin bugünkü değeri tahmin etmek için kullanılır. ARMA (p, q) modeli, bağımlı değişkeni (Y_t) geçmiş p dönemdeki bağımsız değişkenler ve q dönemdeki hata terimleri ile açıklar. Bu modelin temel denklemi şu şekildedir:

$$Y_t = \mu + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \quad (3.5)$$

Bu denklem, AR ve MA modellerinin bir kombinasyonunu temsil eder ve daha karmaşık zaman serilerini modellemek için kullanılır. ARMA, verilerdeki eğilimleri işleyebilir ancak mevsimsel bileşenleri doğrudan modelleyemez. Bunun için ARMA'nın genişletilmiş versiyonları olan SARIMA (Mevsimsel ARIMA), ARIMAX ve SARIMAX gibi modeller geliştirilmiştir (Tarkun, 2023; Mateo-Barcos ve diğerleri, 2024).

ARIMA Modeli

ARIMA (Autoregressive Integrated Moving Average) modeli, zaman serisi analizinde yaygın olarak kullanılan istatistiksel yöntemlerden biridir. Bu modelin popülerliği, Box ve Jenkins'in çalışmalarıyla artmış olup, zaman serisi verilerinin analizinde önemli bir araç olarak kabul edilir (Goswami ve Kandali, 2020).

ARIMA modeli, iki farklı doğrusal regresyon formunun AR ve MA modellerinin entegrasyonu ile geliştirilmiş ve fark alma işlemi ile verilerin durağan hale getirilmesini sağlayan bir modeldir. Buradaki Integrated (I) bileşeni, zaman serilerinde görülen trend ve mevsimsellik gibi durağan olmayan durumları ortadan kaldırmak için kullanılan fark alma işlemi temsil eder. ARIMA (p, d, q) modeli, bağımlı değişkeni (Y_t) geçmiş p dönemdeki bağımsız değişkenler, q dönemdeki hata terimleri ve d derecesinde fark alınmış veriler ile açıklar. Bu modelin temel denklemi şu şekildedir:

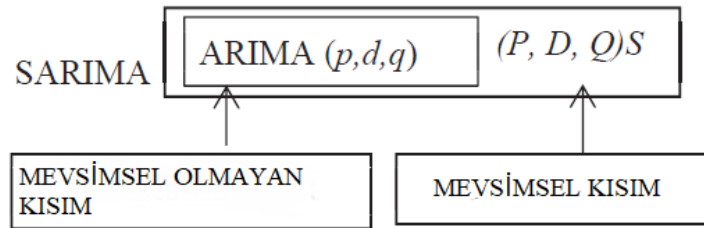
$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) \Delta^d Y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q} \quad (3.6)$$

AR (p) geçmiş değerlerin ağırlıklı toplamını içerir. " p " parametresi, o anki gözlem değerini tahmin edebilmek için sondan itibaren kaç tane geçmiş gözleme bakılacağını belirler. I(d), d değeri serinin durağanlaştırılması için kaçınıcı farkın alınacağını belirler. MA (q), modeldeki hataların önceki hataların ağırlıklı toplamı olduğunu ifade eder. " q " parametresi, sondan itibaren kaç tane hata terimlerinin ağırlıklı ortalamasının kullanılacağını belirler (Tarkun, 2023).

SARIMA Modeli

SARIMA (Seasonal Autoregressive Integrated Moving Average), mevsimsel bileşenleri ve dışsal değişkenleri modele dahil etmek için ARIMA modelinin genişletilmiş versiyonudur. Mevsimsel oto-regresif ve hareketli ortalama bileşenleri ekleyerek mevsimselliği modelleme yeteneğine sahiptir.

Elektrik talebi verileri, mevsimsel değişikliklerden kaynaklanan belirgin periyodik dalgalanmalar gösterir. Bu mevsimsel değişiklikler, SARIMA modeli kullanılarak ele alınabilir.



Şekil 3.9. SARIMA gösterimi (Goswami ve Kandali, 2020)

Genel formuyla SARIMA modeli, SARIMA (p,d,q)(P,D,Q)S olarak ifade edilir. Bu ifadede p , oto-regresif derecesini, P ise o anki gözlem değerini tahmin edebilmek için sondan itibaren S periyodunda geriye dönerek kaç tane geçmiş gözleme bakılacağını belirten mevsimsel oto-regresif derecesini ifade eder. d , fark alma derecesini, D ise serinin durağanlaştırılması için S periyodunda kaçınıcı farkın alınacağını ifade eden mevsimsel

fark alma derecesini temsil eder. q , MA değerini ve Q değeri ise sondan itibaren S periyodunda geriye dönerek kaç tane hata terimlerinin ağırlıklı ortalamasının kullanılacağını ifade eder. (S) mevsimsel periyodun uzunluğunu ifade eder. Modelin mevsimsel kısmı, korelasyon fonksiyonu grafiklerinin mevsimsel gecikmeleri ile gösterilir (Goswami ve Kandali, 2020; Gülay ve diğerleri, 2024; Ensafi ve diğerleri, 2022; Galit ve diğerleri, 2016; Godoy, 2022).

3.6. XGBoost ve Lineer Artırma Regresyonu

XGBoost (Ekstrem Gradyan Artırma)

XGBoost, Gradyan Artırmalı Karar Ağaçları (GBDT) yönteminin geliştirilmiş bir versiyonudur. GBDT'nin avantajlarını taşıırken, kayıp fonksiyonu üzerinde ikinci dereceden Taylor açılımı kullanarak optimizasyonu hızlandırır ve aşırı uyumu önlemek amacıyla düzenleme terimine ağaç modelinin karmaşıklığını ekler. Paralel işlemeyi desteklemesi, GBDT'nin paralel hesaplama yapamama sorununu çözer.

XGBoost, verimliliği, esnekliği ile bilinir ve sınıflandırma, regresyon ve sıralama gibi görevlerde diğer algoritmaları geride bıraktığı gösterilmiştir. Algoritma, birden fazla zayıf öğreniciyi bir araya getirerek güçlü bir öğrenici oluşturur ve her zayıf öğrenici, verinin bir alt kümesi üzerinde eğitilir. Karar ağaçları eğiterek ve bunların tahminlerini birleştirerek nihai tahmini yapar. Bu nedenle, XGBoost'un yük tahmini için kullanılması, endüstriyel müşterilerin yük tahmin doğruluğunu artırabilir (Aldarraji ve diğerleri, 2024).

XGBoost algoritması, zayıf öğrenici olarak Sınıflandırma ve Regresyon Ağacı (CART) kullanır. Her bir yinelemede, önceki zayıf öğrenicileri değiştirmek için artık değer kullanılır. Bu süreç, CART'nin bir topluluk modeli olarak değerlendirilebilir.

$$\hat{y}_{xi} = \sum_{m=1}^M (f_m(X_i)) , f_m \in F \quad (3.7)$$

y_{xi} , i 'nci örneğin tahmin edilen değerini, M modeldeki CART sayısını, $f_m(x_i)$ ise m 'nci ağaçtaki i 'nci örneğin tahmin edilen değerini ifade eder, F ise CART'nin fonksiyon

uzayıdır. XGBoost'un amaç fonksiyonu, kayıp fonksiyonu ve düzenleme terimini içerir (Wang ve diğerleri, 2021).

$$Obj = \sum_{i=1}^n l(y_{xi}, \hat{y}_{xi}) + \sum_{m=1}^M \Omega(f_m)$$

$$\Omega(f_m) = \gamma T + \frac{1}{2} \beta \sum_{j=1}^T w_j^2 \quad (3.8)$$

η örnek sayısını, l ise gerçek değer y_{xi} ile tahmin edilen değer $\hat{y}_{xi}$ arasındaki farkı ölçen ikinci dereceden türevlenebilir bir kayıp fonksiyonunu ifade eder. $\Omega(f_m)$ düzenleme terimini temsil eder. T , ağaçtaki yaprak düğümlerinin sayısıdır, w_j ise yaprak düğümlerinin skorudur, γ ve β ise ağacın karmaşıklığını kontrol eden parametrelerdir (Wang ve diğerleri, 2021).

Amaç fonksiyonunun hedefi, CART'nin yapısını ve w_j 'yi belirlemektir. (En iyi bölünme özelliği ve en iyi bölünme noktası) Amaç fonksiyonu, ikinci dereceden Taylor genişlemesi ile w_j hakkında bir birim ikinci dereceden denkleme basitleştirilebilir (Chen ve Guestrin, 2016) (Wang ve diğerleri, 2021).

$$Obj = \sum_{j=1}^T (\sum_{i \in I_j} g_i w_j + 1/2 (\sum_{i \in I_j} h_i + \beta) w_j^2) + \gamma T \quad (3.9)$$

I_j , j yaprak düğümündeki tüm veri örneklerini temsil eder, g_i ve h_i kayıp fonksiyonunun birinci ve ikinci türevlerini ifade eder. G_j ve H_j şu şekilde tanımlanabilir:

$$G_j = \sum_{i \in I_j} g_i, H_j = \sum_{i \in I_j} h_i \quad (3.10)$$

Yaprak düğümünün optimal skoru w_j^* ve buna karşılık gelen amaç fonksiyonunun optimal değeri, CART'nin yapısının belirlenmiş olduğu durumda yukarıdaki birim ikinci dereceden denklemin çözülmesiyle elde edilir.

$$w_j^* = \frac{-G_j}{H_j + \beta}$$

$$Obj = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \beta} + \gamma T \quad (3.11)$$

Amaç fonksiyonunun değeri ne kadar küçükse, CART'nin yapısı o kadar iyidir. XGBoost, tüm bölünme noktalarını dolaşmak için açgözlü bir algoritma uygular ve nihayetinde bölme işleminden sonra amaç fonksiyonunun en küçük değerine sahip olan bölünme noktasını seçer. Bu aşağıdaki denklemde gösterilmiştir:

$$Gain = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \beta} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \beta} + \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \beta} \right] - \gamma \quad (3.12)$$

Diğer makine öğrenimi algoritmalarına kıyasla, XGBoost üstün sonuçlar sunmasıyla ünlüdür. XGBoost sadece elektrik talebi değil, birden fazla özelliği de kullanabilir (Ayele ve diğerleri, 2023; Abbasimehr ve diğerleri, 2023; Deng ve diğerleri, 2022; Zhang ve Janosik, 2024; Wang ve diğerleri, 2021).

Lineer Artırma Regresyonu (Linear Boost Regression)

Lineer Artırma Regresyon Ağaçları, bir karar ağacı algoritması türüdür ve geleneksel karar ağaçlarından farklı olarak her yaprak düğümünde doğrusal regresyon modeli kullanılır. Bu yaklaşım, karar ağaçlarının esnekliği ile lineer regresyonun doğrusallık avantajını birleştirir.

Lineer Artırma, iki aşamalı bir öğrenme sürecidir. İlk olarak, başlangıç veri seti üzerinde bir lineer model eğitilerek tahminler elde edilir. Ardından, önceki adımın artık değerleri mevcut tüm özellikler kullanılarak bir karar ağacı ile modellenir. Ağaç, en yüksek hataya yol açan yolu (yani en kötü yaprağı) belirler. Hatalara en fazla katkıda bulunan yaprak, ilk aşamada kullanılmak üzere yeni bir ikili özellik oluşturmak için kullanılır. Bu yinelemeler, belirli bir durdurma kriterine ulaşıncaya kadar devam eder (<https://pypi.org/project/linear-tree/>).

Regresyon analizi, bir veya daha fazla değişken arasındaki ilişkinin gücünün belirlenmesini sağlayan istatistiksel bir yaklaşımdır. Bu ilişkiler temelinde bilinmeyen değerlerin tahmin edilmesine yardımcı olabilir. Basit doğrusal regresyon, bağımlı bir değişkenle doğrusal bir ilişki modellemek için bir bağımsız değişkenin kullanılmasını

içerir. Çoklu doğrusal regresyon ise, bağımlı değişkeni tahmin etmek için birden fazla bağımsız değişkenin kullanılmasını kapsar (Aldarraji ve diğerleri, 2024).

Verilen bir veri setini doğrusal bir modele uydurmak için, (c) bir değer, (W) gerçek sayılardan oluşan bir satır vektörü, (X) bağımsız değişkenlerin sütun vektörü ve $\hat{y}$ bağımlı bir değişkenin tahminini ifade edecek şekilde, $y = WX + c$ doğrusal regresyon modelini,

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.13)$$

ifadesini en aza indirecek şekilde ağırlıkları (W) ayarlayarak veri setine uydurur; burada (y_i) , i veri noktası için bağımlı değişkenin gerçek değeridir.

Lineer Artırma Regresyonu (Linear Boost Regression) yapısal karar ağaçları ve doğrusal modellerin güçlü yönlerini birleştirerek karmaşık veri yapıları için etkili çözümler sunar. Hataları ardışık olarak düzeltir bu sebeple yüksek doğruluk sağlar. Farklı zayıf öğrenicilerle kullanılabilir ve genelleme yeteneği yüksektir ancak hesaplama maliyeti yüksektir ve doğru hiperparametreleri bulmak zaman alıcı olabilir (Nofal, 2023; Su ve diğerleri, 2012; Ilıç ve diğerleri 2021).

3.7. Prophet

Prophet, Facebook ekibi tarafından tanıtılan, zaman serilerini trend, mevsimsellik ve tatiller kullanarak modelleyen açık kaynaklı bir zaman serisi analizi kütüphanesidir (Abbasimehr ve diğerleri, 2024, Taylor ve Letham 2018).

Prophet, kullanıcı dostu arayüzü ve sezgisel parametreleri ile zaman serisi tahmininde kolaylık sağlar (Waheed ve Xu 2024). Bayesian tabanlı eğri uydurma yaklaşımı kullanarak zaman serisi verilerini yumuşatır ve tahmin eder. Bu, onu Holt-Winters ve ARIMA gibi standart modellerden ayıran belirgin bir özelliktir (Abbasimehr ve diğerleri, 2024). Eksik veriler ve aykırı değerler gibi karmaşık veri sorunlarıyla etkili bir şekilde başa çıkabilir. Bu, çeşitli veri setlerinde güvenilir tahminler yapmasını sağlar (Zaraket ve

diğerleri, 2024). Tatiller ve özel olayların zaman serisine etkilerini dikkate alarak tahmin doğruluğunu artırır (Bashir ve diğerleri, 2022).

Modelin gösterimi şu şekildedir:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (3.14)$$

Burada $g(t)$ trend bileşenini, yani zaman serisi değerlerindeki periyodik olmayan değişiklikleri ifade eder; $s(t)$ zaman serisinin mevsimsellik unsurunu (örneğin, haftalık ve yıllık) temsil eder; $h(t)$ tatillerin etkilerini belirtir ve ε_t modelin hesaba katmadığı olağandışı değişiklikleri yansıtır (Ensafi ve diğerleri, 2022; Guo ve diğerleri, 2021).

Prophet modelinde mantıksal büyüme ve parça parça doğrusal büyüme olmak üzere iki tür trend bileşeni vardır. Parça parça doğrusal büyüme modeli şu şekilde ifade edilir:

$$g(t) = (k + a(t)^T \delta)t - (m + a(t)^T \gamma) \quad (3.15)$$

Burada t verilerin zaman indeksini temsil eden zaman değişkenidir, k büyüme oranı, $a(t)$ ayar parametrelerinin tasarım matrisidir ve T matrisin devriğini temsil eder, δ büyüme oranının oran ayarlamaları matrisidir, m bir ofset parametresidir ve γ değişim noktalarında düzeltme ayarlamalarının bir matrisidir. Bu katsayılar, model uyum süreci sırasında öğrenilir (Abbasimehr ve diğerleri, 2024).

Mantıksal büyüme modeli parçalı lojistik büyüme modeli kullanılarak formüle edilir. Burada $C(t)$ zamanla değişen kapasitedir.

$$g(t) = \frac{C(t)}{(1 + \exp(-(k + a(t)^T \delta)(t - (m + a(t)^T \gamma)))} \quad (3.16)$$

Prophet modeli, mevsimsellik öğelerini Fourier serisi ile modellemektedir. Bu, zaman serisinin periyodik özelliklerini doğru bir şekilde yansıtmayı sağlar.

$$s(t) = \sum_{n=1}^N \left(a_n \cos \frac{2\pi n t}{P} + b_n \sin \frac{2\pi n t}{P} \right) \quad (3.17)$$

Burada P zaman serisinin beklenen düzenli periyodunu gösterir ve t bir tarihi temsil eder. Bu nedenle, mevsimsel modeli oluşturmanın temeli, N değerini belirtmek ve ardından $2 \times N$ parametreyi belirlemektir, yani $\beta = [a_1, b_1, \dots, a_N, b_N]^T$. Yıllık ve haftalık mevsimsellik söz konusu olduğunda $N = 10$ ve $N = 3$ 'ün çoğu durumda iyi çalıştığı keşfedilmiştir (Abbasimehr ve diğerleri, 2024).

Tatiller ve büyük olaylar zaman serisinde öngörülemeyen dalgalanmalara yol açabilir. Prophet, kullanıcıların bu tarihleri belirli bir matris olarak sağlamasına olanak tanır ve bu etkiler modelde dikkate alınır.

$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_L)] \text{ ve } h(t) = Z(t)\kappa_i \quad (3.18)$$

Her tatil i için önceki ve gelecekteki tarihlerin D_i listesi oluşturulur ve belirli bir zaman noktasının t bir tatil i içerisinde olup olmadığını belirlemek için bir gösterge fonksiyonu kullanılır. $Z(t)$ 'deki L , belirli bir günde bir tatil varsa 1 değerini alır ve aksi takdirde 0 değerini alır. Ardından, her tatilin etkisi, κ_i katsayıları vektörü tarafından bulunur. Bu katsayılar, karşılık gelen tatil gününün varlığı nedeniyle zaman serisindeki beklenen değişimi temsil eder. κ , ortalaması 0 ve standart sapması ν olan normal bir dağılımı izleyen rastgele bir değişkendir. ν parametresi, modeldeki tatil etkisinin gücünü kontrol eder ve ν 'nun daha küçük bir değeri daha zayıf bir tatil etkisini ifade ederken, ν 'nun daha büyük bir değeri daha güçlü bir tatil etkisini ifade eder. Örneğin, Noel Günü'nde normal bir güne göre satışların %50 daha yüksek olması bekleniyorsa, Noel tatili için κ_i 'deki karşılık gelen giriş 0,5 olacaktır (Abbasimehr ve diğerleri, 2024).

$Z(t)$ matrisi, dikkate alınan toplam tatil sayısı olan L uzunluğundadır. $Z(t)$ 'nin her sütunundaki girişler, tatilin belirli bir günde olup olmadığını gösterir, 1 tatilin mevcut olduğunu ve 0 olmadığını gösterir. Tatil bileşeni daha sonra $Z(t)$ 'yi karşılık gelen κ_i katsayıları vektörüyle çarparak genel tahmin modeline dahil edilir. Bu, her girişin zaman serisindeki tatil günlerinin belirli bir zaman noktasındaki beklenen etkisini temsil eden

yeni bir zaman serisi $h(t)$ ile sonuçlanır (Abbasimehr ve diğerleri, 2024; Huang ve diğerleri, 2022).

3.8. LSTM (Uzun Kısa Dönemli Hafıza)

Uzun Kısa Dönemli Hafıza (LSTM) algoritması, bireysel veri noktalarını veya bir dizi veri noktasını işleyebilen ve tahmin amaçlı derin öğrenme alanında kullanılan bir yapay yinelemeli sinir ağı (RNN) mimarisidir. Bu algoritma, zaman serisi verilerini tahmin etmek, işlemek ve sınıflandırmak için uzun süre bilgi tutma kapasitesine sahiptir. Böylece, yakın tarihli bilgilere dayanarak doğru tahminler yapılabilir (Aldarraji ve diğerleri, 2024; Arıkök, 2024).

LSTM, kaybolan gradyanlar sorununu çözmek için geliştirilmiş ileri düzey bir modifikasyon olarak kabul edilir. Standart RNN birimlerinde bulunan nöron katmanları, LSTM'de dört ek nöron katmanı ile değiştirilmiştir. Bu katmanlar, geleneksel iç durumdan farklı olarak, ana kanal olan başka bir durumu içerir. Bu durum, veri setini kapsayan hücre veya ortak durumdur ve uzun vadede bellek paylaşımı için kullanılır. Hücre durumu, yeni veriler eklenerek veya güncelliğini yitirmiş veriler silinerek güncellenir ve genellikle mesafelerinden bağımsız olarak birden çok LSTM birimiyle ilişkilidir (Rick ve Berton, 2022).

LSTM yapısında, hücre durumu ve mevcut girişin sağladığı değerlere çeşitli aktivasyon fonksiyonları uygulanarak hücre durumu değiştirilir. Bu yapı, kaybolan gradyanlar sorununa karşı direnç göstermek için nöron katmanlarının düzenlenmesiyle değiştirilmiştir. LSTM'nin belirli bir yapısı vardır ve bu yapı, daha uzun dizilerin izlenmesi gereken uygulamaların analizinde faydalıdır (Rezaei ve diğerleri, 2022).

LSTM algoritmasında, hücre durumunun yönetimi için aşağıdaki yapı kullanılır:

Unutma Bloğu: Bu blok, değerleri 0 ile 1 arasında olan ve sigmoid fonksiyonu içeren bir nöron katmanını dikkate alır. Önceki hücreye h_{t-1} ait değerler ile giriş değerleri x_t arasındaki hesaplama göz önünde bulundurulur. Nöronlar ve hücre durumu çarpılarak

elde edilen sonuç 0'a yaklaşıyorsa, veriler hücre durumundan silinir. Sonuç 1'e yaklaşıyorsa, veriler kaydedilir (Rick ve Berton, 2022).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.19)$$

Giriş Geçidi: Mevcut girdi yeni bilgiler içerdiğinde, giriş kapısı bu bilgilerin hücre durumu için önemli olup olmadığını veya göz ardı edilip edilmeyeceğini belirler. Giriş kapısında, hiperbolik tanjant katmanı olarak işaretlenen bir aktivasyon fonksiyonu ile birlikte bir sigmoid katmanı bulunur. Bilgiler, aktivasyon katmanı tarafından üretilir ve daha sonra hiperbolik tanjant, hangi verilerin hücre durumuna iletmeyi hak ettiğini belirler. Hesaplama fonksiyonu önceki adımla (unutma geçidi) aynıdır, ancak yeniden kullanılır (Rick ve Berton, 2022).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.20)$$

Hücre Durumu: Hücre durumunun bu bileşeninde, hesaplanan öğeler arasında en yüksek ilgisine sahip güncellenmiş bilgileri korumaya odaklanılır. Ayrıca, gradyanlar stabilize edilir, böylece daha az gelecek hesaplama yapılır (Rick ve Berton, 2022).

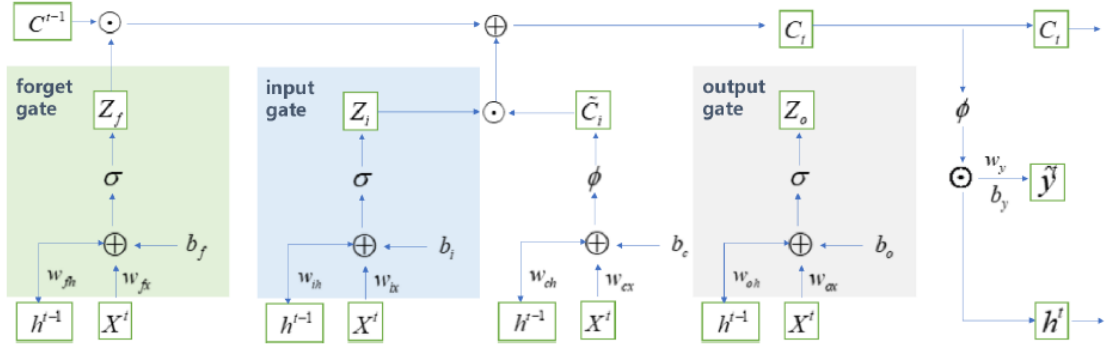
$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.21)$$

Çıkış Geçidi: Hücre durumu değiştirilir ve ardından yeni bir iç durum ve çıktı, yeni bir katman içeren çıkış kapısı tarafından üretilir. Bu sefer, aynı katmandaki $(t + 1)$ 'deki bir sonraki hücreye yönlendirme için giriş geçidinde ve unutma geçidinde zaten gerçekleştirilen benzer işlemler gerçekleştirilir (Rick ve Berton, 2022).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.22)$$

Hücre (Son) Geçidi: Son olarak, bir sonraki katmana (yeni bir LSTM veya başka bir tür) giriş olacak, çıkış geçidinden işlenmiş veriler elde edilir (Rick ve Berton, 2022; Zhang ve diğerleri, 2020; Gülay ve diğerleri, 2024).

$$h_t = o_t \odot \tanh W_C \quad (3.23)$$



Şekil 3.10. LSTM ağ yapısı (Zhang ve diğerleri, 2023)

LSTM, standart ileri beslemeli sinir ağlarından farklı olarak, bilgi akışını geçmiş birkaç örnekten şimdiki zamana aktarmak için geri besleme bağlantılarına sahiptir. Bu özellik, LSTM'nin zaman serisi verilerini işlemek ve tahmin etmek için çok uygun olmasını sağlar. Örneğin, LSTM bölümlenmemiş, bağlı el yazısı tanıma, konuşma tanıma ve ağ trafiğinde anomali tespiti gibi görevler için kullanılır. Ayrıca, LSTM'nin bellek hücresi yapısı, enerji tüketim verilerinin zaman serisi tahmininde mükemmel performans gösterir (Hasim ve diğerleri, 2014; Xiangang ve diğerleri, 2014; Lu ve diğerleri, 2023; Houdt ve diğerleri, 2020; Le ve diğerleri, 2019).

Sonuç olarak, LSTM, RNN'lerin karşılaştığı gradyan kaybolması ve gradyan patlaması gibi sorunları aşmak için geliştirilmiştir. Unutma kapısının varlığı ve hücre durumu gradyanlarının toplama özelliği, LSTM'nin bağlantı ağırlıklarını güncelleyerek gradyan kaybolması problemini önemli ölçüde azaltmasına olanak tanır ancak LSTM'nin tahmin performansı, toplam hücre veya nöron sayısı ve toplam eğitim döngüsü sayısı (epoch) gibi ayar parametrelerine bağlıdır (Chaturvedi ve diğerleri, 2022; El-Azab ve diğerleri, 2024; Huang ve diğerleri, 2022).

4. BULGULAR

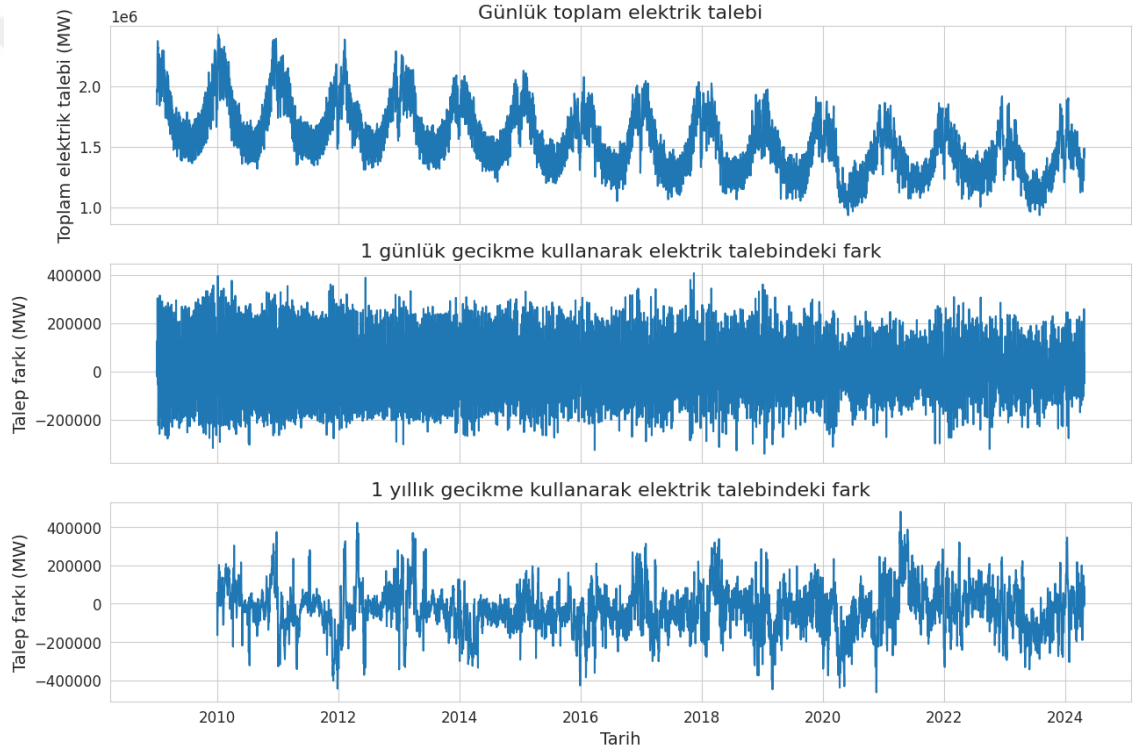
4.1. SARIMA Uygulaması

Günlük elektrik talebi verilerine uygulanan bir SARIMA (Sezonsal Otokorelasyonlu Entegre Hareketli Ortalama) modeli oluşturmak ve değerlendirmek için, zaman serisi analizinin temel adımları takip edilerek, veri hazırlığı, model parametrelerinin belirlenmesi, modelin oluşturulması, değerlendirilmesi ve tahmin yapma aşamaları izlenmiştir. Veriyi düzleştirmek ve uzun bir zaman serisinde temelde "gürültü" olan şeyleri kaldırmak için veri frekansı yarım saatte bir kezden günlük toplam veriye düşürülmüştür.

Geçmiş değerler ve gecikmeler kullanılarak zaman serisi bileşenleri oluşturulmuştur. Bu işlem, verileri modelleme için daha uygun hale getirmek amacıyla yapılır. Gecikme işlemi, zaman serisindeki belirli bir dönemin (örneğin, bir önceki ayın) değerlerini kullanarak mevcut dönemi modellemeye çalışır. Bu, zaman serisinde otokorelasyonları yakalamak için kullanılır. Gecikmeler (lags), birbirinden k zaman birimi uzaklıktaki değerler arasındaki korelasyonu temsil eder. Add_lags işlevi, veri kümesine önceki yıllardan gelen bilgileri içeren gecikmeler ekler. Bu işlevlerin amacı, veri çerçevesine zaman serisi özellikleri eklemek ve önceki dönemlere ait bilgileri kullanarak gecikmeler oluşturarak makine öğrenimi modelleri için daha zengin bir veri kümesi oluşturmaktır. SARIMA modelinde, hem otoregresif (AR) hem de hareketli ortalama (MA) bileşenleri gecikmeli değerleri kullanır. AR(p) kısmında p , kaç gecikmeli değer kullanılacağını belirtirken, MA(q) kısmında q , hata terimlerinin kaç gecikmeli değeriyle çalışılacağını ifade eder.

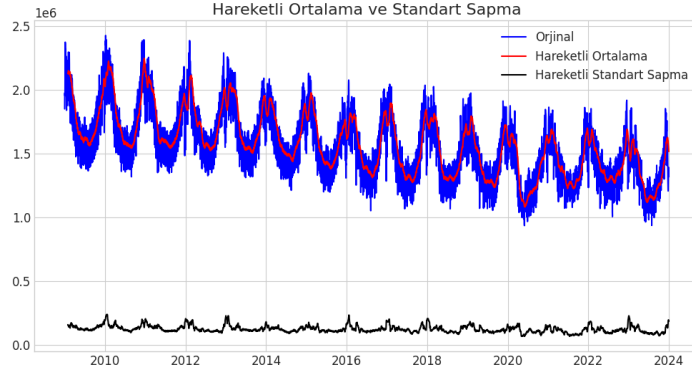
Fark alma işlemi, serideki trendi veya mevsimsel etkileri ortadan kaldırmak için kullanılır. Zaman serisinin durağan (stationary) hale getirilmesi amacıyla yapılır. Fark alma, zaman serisindeki her bir veri noktasından bir önceki veri noktasının çıkarılmasıyla yapılır. Bu işlem, trendin etkilerini azaltır veya ortadan kaldırır.

Zaman serisi modellemesi için gereken otokorelasyonları tespit etmek için gecikmeli değerler grafikleri oluşturulmuştur. Şekil 4.1’teki ilk grafik, günlük toplam elektrik talebini göstermektedir. Talebin zaman içinde dalgalandığını ve belirgin bir mevsimsellik gösterdiğini ortaya koymaktadır. İkinci grafik, günlük toplam elektrik talebinin bir günlük gecikmeli versiyonuyla arasındaki farkı göstermektedir. Bu grafik, talebin günlük olarak nasıl değiştiğini göstererek, bir gün öncesine göre nasıl arttığını veya azaldığını göstermektedir. Üçüncü grafik günlük toplam elektrik talebinin bir yıllık gecikmeli versiyonuyla arasındaki farkı göstermektedir. Talebin bir yıl öncesine göre nasıl arttığını veya azaldığını göstermektedir.



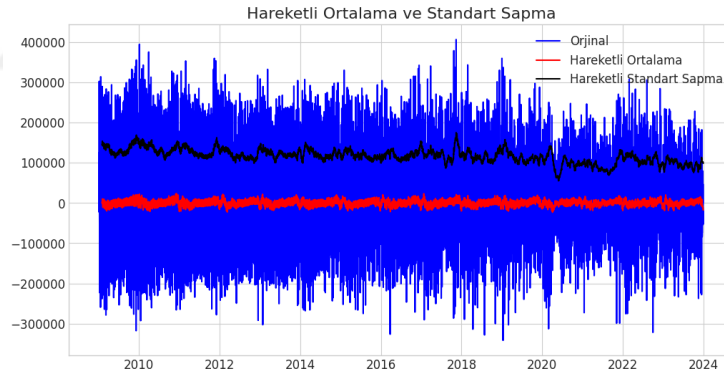
Şekil 4.1. Günlük toplam elektrik talebi ve gecikmeli değerler grafiği

Veri hazırlığının ardından, Augmented Dickey-Fuller (ADF) testi ile verinin durağanlığı kontrol edilmiştir. Durağanlık, zaman serisi verilerinin zaman içinde sabit bir ortalama ve varyansa sahip olmasını ifade eder.



Şekil 4.2. Zaman serisinin orijinali için hareketli ortalama ve standart sapma

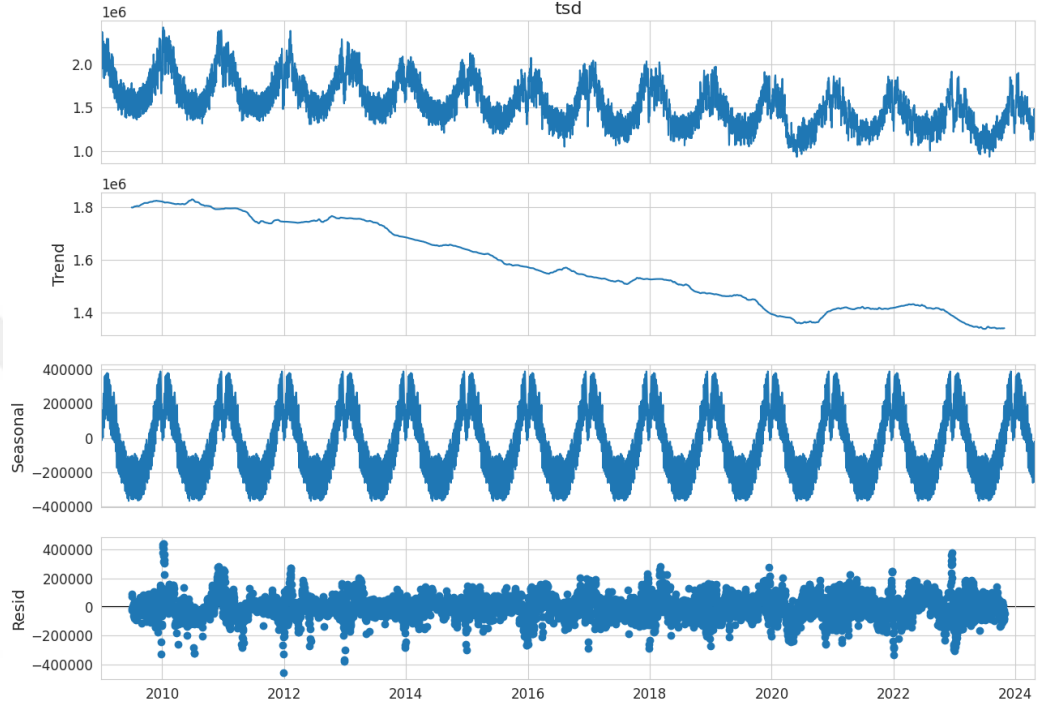
30 günlük pencere üzerinde yapılan ADF testi istatistiği $(-3,811209)$, %1 anlamlılık seviyesindeki kritik değerden $(-3,431527)$ daha düşüktür. p -değeri $(0,002795)$, %1 anlamlılık seviyesinden $(0,01)$ daha küçüktür. Test istatistiklerine ve p değerine göre zaman serisinin birim köke sahip olma olasılığının düşük olduğu ve durağan olma olasılığının yüksek olduğu görülmüştür ancak serinin mevsimsellik ve azalan bir eğilim göstermesi nedeniyle durağan bir serinin gereklerini karşılamadığı gözlenmiştir.



Şekil 4.3. Bir günlük fark alınmış veri seti üzerinde hareketli ortalama ve standart sapma

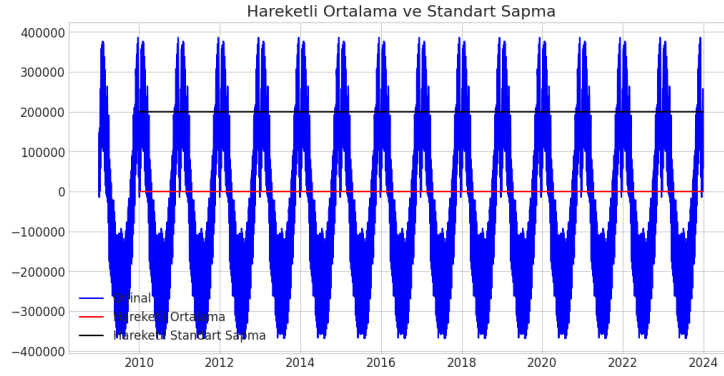
Bir günlük fark alma işlemi uygulandıktan sonra test istatistiği $(-1,780990e+01)$, %1, %5 ve %10 anlamlılık seviyelerindeki kritik değerlerden (sırasıyla $-3,431527$, $-2,862060$ ve $-2,567047$) çok daha düşüktür ve p değeri $(3,192405e-30)$, tüm anlamlılık seviyelerinden çok daha küçüktür. Zaman serisinin birim köke sahip olma olasılığının çok düşük olduğu ve durağan olma olasılığının çok yüksek olduğu yorumlanır. Veriler artık durağan olduğundan modelde d , 1 olarak kullanılabilir.

D 'yi belirlemek için serinin mevsimsel bileşeninin yalıtılması gerekmektedir. Bu, zaman serisi verilerinin yapısını anlamak ve uygun bir zaman serisi modeli seçmek için önemlidir. Veriler mevsimsel ayrışmaya tabi tutulur. Ayrışma, verideki trend, mevsimsellik ve gürültü bileşenlerini ayırır.



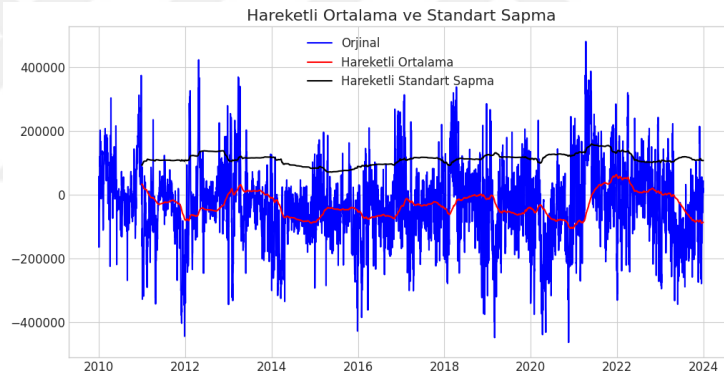
Şekil 4.4. Trend, mevsimsellik ve gürültü bileşeni

Şekil 4.4'teki ilk grafik, orijinal zaman serisinin grafiğini göstermektedir. Zaman serisinin zaman içinde dalgalandığı ve belirgin bir mevsimsellik gösterdiği anlaşılmaktadır. İkinci grafik, zaman serisinin trend bileşenini göstermektedir. Zaman serisinin genel olarak azalan bir trend eğilimi gösterdiği fark edilmektedir. Üçüncü grafik, zaman serisinin mevsimsel bileşenini göstermektedir. Zaman serisinin yıllık bir mevsimsellik gösterdiği belirlenmiştir. Alt grafik, zaman serisinin kalıntı bileşenini göstermektedir. Kalıntılar, trend ve mevsimsellik bileşenlerinin çıkarılmasıyla kalan kısım temsil eder. Kalıntılar, zaman serisinin rastgele varyasyonlarını gösterir.



Şekil 4.5. Mevsimsel bileşenin hareketli ortalama ve standart sapması

Zaman serisi verilerinin mevsimsel bileşeni, ADF testi sonucunda durağan olarak bulunmuştur. Test istatistiği $(-5,397225)$, %1, %5 ve %10 anlamlılık seviyelerindeki kritik değerlerden (sırasıyla $-3,431527$, $-2,862060$ ve $-2,567047$) çok daha düşüktür. p -değeri $(0,000003)$, tüm anlamlılık seviyelerinden çok daha küçüktür.



Şekil 4.6. Bir yıllık fark alınmış veri seti üzerinde hareketli ortalama ve standart sapma

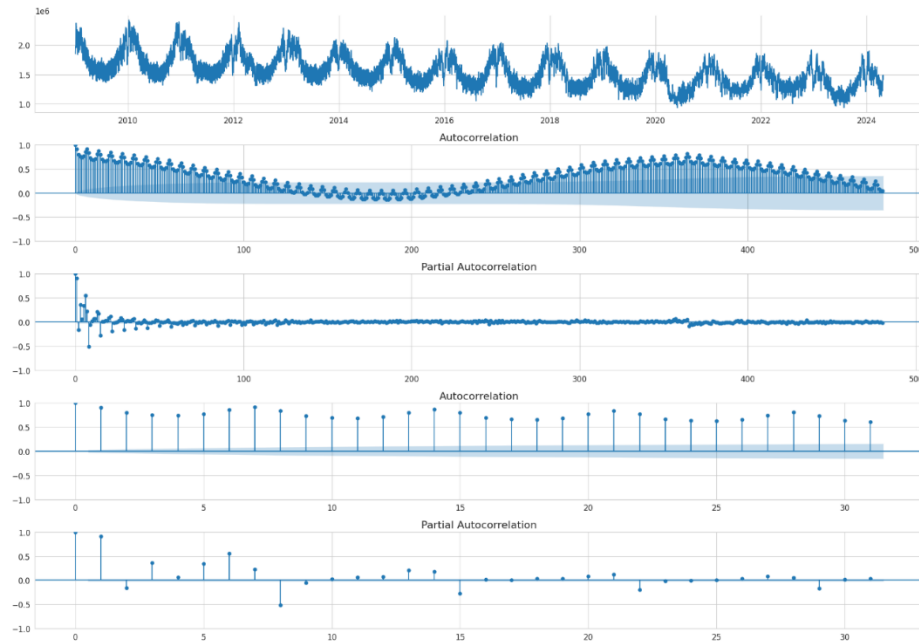
Bir yıllık fark alma işlemi uygulandıktan sonra test istatistiği $(-8,066354)$, %1, %5 ve %10 anlamlılık seviyelerindeki kritik değerlerden (sırasıyla $-3,431607$, $-2,862096$ ve $-2,567066$) çok daha düşüktür. Bu, zaman serisinin birim köke sahip olma olasılığının çok düşük olduğunu gösterir. p -değeri $(1,592428e-12)$, tüm anlamlılık seviyelerinden çok daha küçüktür. Veri setinin zamansal olarak durağan olduğu ve modelin güvenilir olabileceği sonucuna varılabilir. 1 yıllık fark verileri durağan olduğundan D , 1 alınabilir.

“ p ” ve “ q ”yu belirlemek için orijinal verilerin otokorelasyon grafikleri analiz edilebilir. Otokorelasyon Fonksiyonu (ACF Autocorrelation function), herhangi bir serinin gecikmeli değerleriyle olan otomatik korelasyon değerlerini gösteren tam bir otomatik

korelasyon fonksiyonudur. Serinin şimdiki değerlerinin geçmişteki değerleriyle ne kadar iyi ilişkili olduğunu açıklar. Bir zaman serisi; trend, mevsimsellik, döngüsellik ve kalıntı gibi bileşenlerden oluşabilir. ACF, bu bileşenlerin hepsini dikkate alarak korelasyonları belirler, bu nedenle ‘tam otomatik korelasyon grafiği’ olarak adlandırılır. Grafiklerde mavi çubuklar korelasyon katsayılarını, gölgeli alan ise güven aralığını temsil eder.

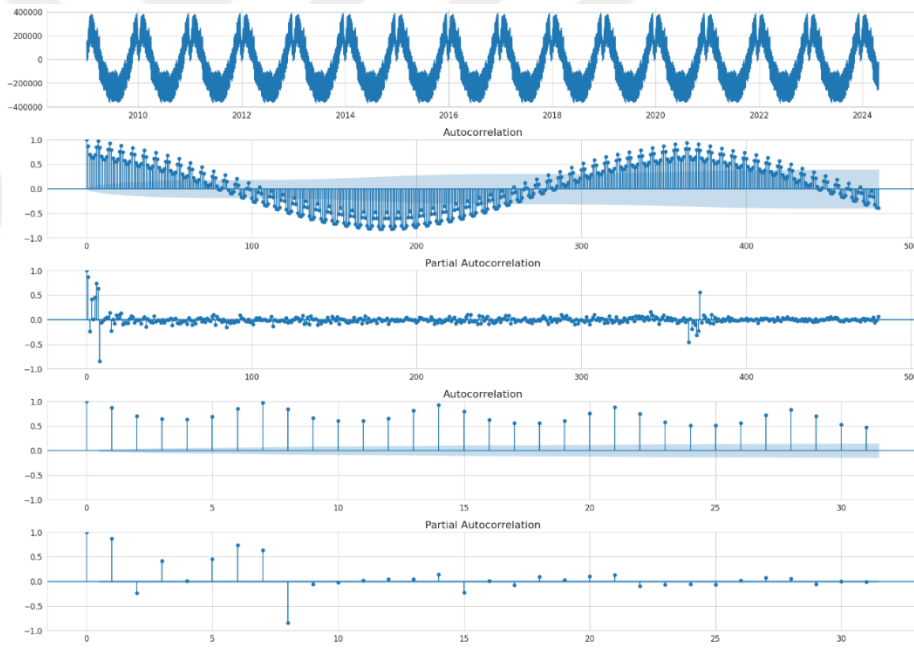
Kısmi Otokorelasyon Fonksiyonu (PACF Partial Autocorrelation Function), PACF, ACF'nin bir alt kümesidir. PACF, diğer veri noktalarının etkilerini hesaba katarak iki zaman noktasındaki gözlemler arasındaki korelasyonu ifade eder. PACF, bir AR modelinde kullanılacak optimal terim sayısını belirlemek için kullanılabilir. Genel olarak, ACF ve PACF grafikleri Otopregresif (AR) ve Hareketli Ortalama (MA) modelleri için p, q değerlerini bulmada çok önemli olan grafikler üretir (Fan ve diğerleri, 2021; Ghimire ve diğerleri, 2024; Arıkök, 2024).

Zaman serisi verilerinin otokorelasyon ve kısmi otokorelasyonları iki farklı gecikme değeri kullanarak görselleştirilmiştir. lags_1=480, ilk otokorelasyon ve kısmi otokorelasyon grafiği için gecikme sayısını, lags_2=31, ikinci otokorelasyon ve kısmi otokorelasyon grafiği için gecikme sayısını belirtir.



Şekil 4.7. Orijinal zaman serisinin otokorelasyon ve kısmi otokorelasyon grafikleri

İlk grafik ham zaman serisi verilerini, ikinci ve üçüncü grafikler zaman serisinin 480 gecikmeye kadar olan otokorelasyonunu ve kısmi otokorelasyonunu göstermektedir. Dördüncü ve beşinci grafikler ise ikinci gecikme değeri ($\text{lags}_2=31$) için otokorelasyon ve kısmi otokorelasyon grafiğini gösterir. Yıllık bir mevsimselliğin açıkça görüldüğü, ancak haftalık bir mevsimselliğin de olduğu anlaşılabılır. Bu grafiklerde, p ve q değerleri, otokorelasyon ve kısmi otokorelasyon fonksiyonlarına göre belirlenebilir. p , otokorelasyon fonksiyonunda, ilk önemli gecikme, güvenlik sınırlarının dışında kalan ilk gecikmedir. q , kısmi otokorelasyon fonksiyonunda, ilk önemli gecikme, güvenlik sınırlarının dışında kalan ilk gecikmedir. Ne yazık ki, otokorelasyon değerleri yaklaşık 120 gecikmeye kadar güven aralığına düşmez ve kısmi otokorelasyon değerleri sürekli olarak güven aralığına ulaşmaz ve birden fazla sivri uç vardır. Başlangıç için p değeri 7 ve q değeri olarak 2 seçilebilir.



Şekil 4.8. Mevsimsel bileşenin otokorelasyon ve kısmi otokorelasyon grafikleri

Mevsimsel parametreleri tahmin etmek için 1 yıllık fark alınmış değerler kullanılmıştır. İlk grafikte, mevsimsel bileşenin ham verilerinin yıllık veya iki yılda bir olabilecek tekrar eden bir desenle güçlü bir mevsimsel bileşene sahip olduğu gözlenmiştir. Bu grafikler, değerlerin güven aralığına düşmesi için birçok gecikmeye ihtiyaç duyduğunu göstermektedir. Grafikler, hem uzun dönem (yıllık) hem de kısa dönem (aylık veya haftalık) mevsimsel yapıların varlığını gösterir. SARIMA modeli gibi mevsimsel

modellerin kullanılması gerektiği sonucuna ulaşılabilir. Başlangıç tahmin olarak, $P=3$, $Q=2$ ve $M=12$ seçilmiştir.

Veri seti, 2019 Haziran öncesi eğitim seti ve sonrası test seti olarak ayrılmıştır. Eğitim verisi, modelin eğitilmesi için kullanılırken, test verisi modelin performansının değerlendirilmesi için kullanılmıştır. SARIMA modeli, statsmodels ve pmdarima kütüphaneleri kullanılarak oluşturulmuş ve farklı parametre kombinasyonları ile eğitilmiştir. Model parametreleri, order ve seasonal_order olmak üzere iki kısımda belirlenmiştir. Verilerin belirli bir dağılım, model veya hipotezle uyumlu olup olmadığını değerlendirmek için bazı istatistiksel testler kullanılmıştır:

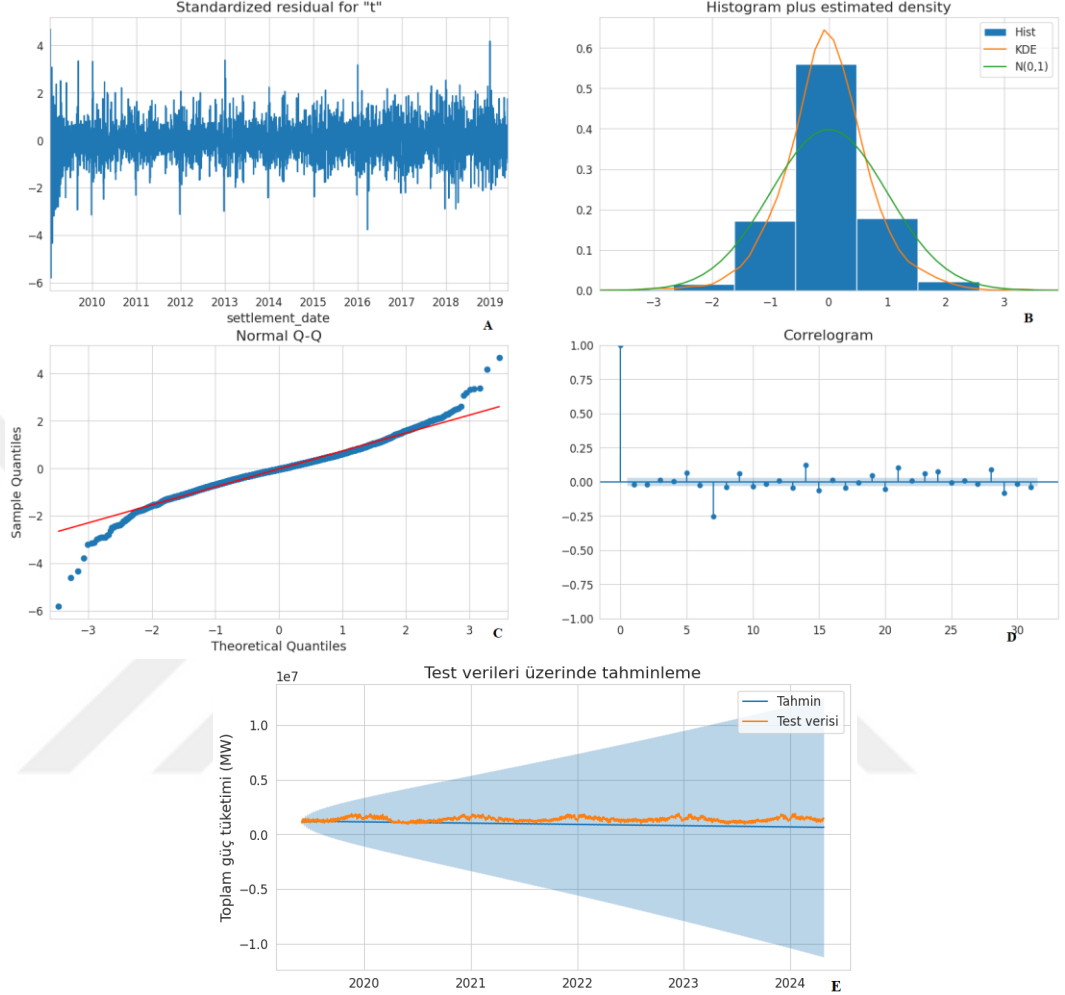
Normality Testi (Jarque-Bera Testi): Jarque-Bera testi ile modelin kalıntılarının normal dağılıma uygun olup olmadığı test edilmiştir. Eğer p değeri belirli bir anlamlılık seviyesinden (örneğin 0,05) küçükse, veri setinin normal dağılım göstermediği sonucuna varılır.

Seri Korelasyon Testi (Ljung-Box Testi): Ljung-Box testi ile modelin hata terimlerinin otokorelasyon içerip içermediğini kontrol eder. Yani, kalıntıların seri bağımlı olup olmadığını test eder. Eğer p değeri belirli bir anlamlılık seviyesinden küçükse, kalıntılarda otokorelasyon olduğu sonucuna varılır.

Heteroskedastisite Testi (Break-Variance Testi): Break-Variance testi ile kalıntıların varyansının sabit olup olmadığı kontrol edilmiştir. Hata terimlerinin sabit bir varyansa sahip olup olmadığını kontrol eder. Heteroskedastisite, hata terimlerinin varyansının zaman içinde değişmesi durumudur. Eğer p değeri belirli bir anlamlılık seviyesinden küçükse, heteroskedastisite olduğu sonucuna varılır (Mikayilov ve diğerleri, 2023).

Durbin-Watson Testi: Hata terimlerinin otokorelasyon içerip içermediğini kontrol eder. Durbin-Watson istatistiği 0 ile 4 arasında bir değer alır. 2'ye yakın bir değer, otokorelasyon olmadığını gösterir. 2'den küçük değerler pozitif otokorelasyonu, 2'den büyük değerler ise negatif otokorelasyonu gösterir (Su ve diğerleri, 2012).

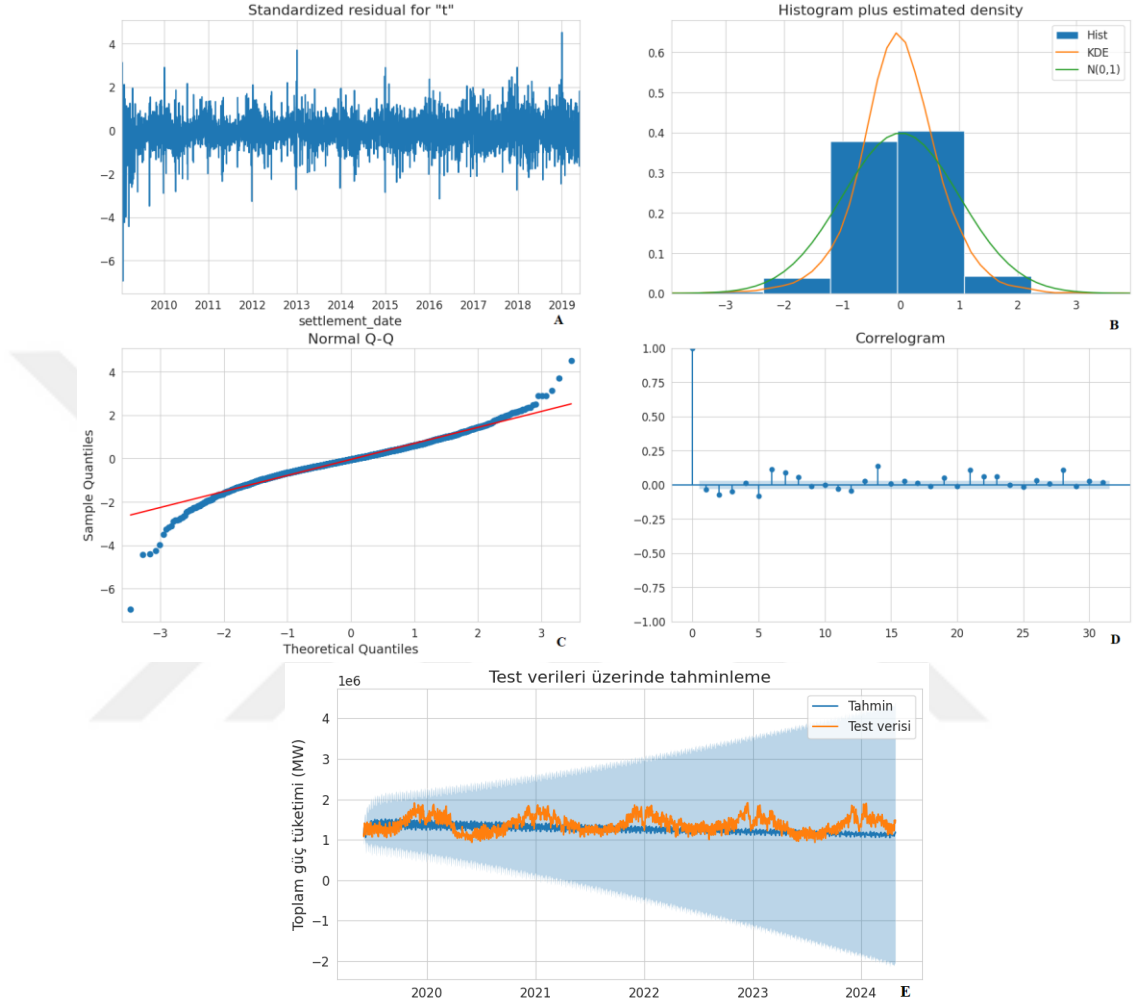
Başlangıç için secilen Model parametreleri ile kurulan SARIMA modelinin parametrelerini (7, 1, 2) (p, d, q) ve (3, 1, 2, 12) (P, D, Q, s), önceden belirlenen tahmin süresini (50) ve gecikme sayısını (31) alan bir model (Model 0) tasarlanmıştır.



Şekil 4.9. SARIMA Model 0 grafikleri **A)** t için standartlaştırılmış artık gösterimi **B)** Histogram ve tahmin eğrisi **C)** Normal Q-Q grafiği **D)** Korelasyon grafiği **E)** Tahmin grafiği

Model 0, MAPE değeri olarak %32,32 elde edilmiştir. İstatistiksel test sonuçları, kalıntıların normal dağılımdan sapma gösterdiğini ($p < 0,05$) ve otokorelasyon içerdiğini ($p < 0,05$) göstermektedir. Heteroskedastisite testi, varyansın sabit olduğunu göstermiştir ($p > 0,05$). Durbin-Watson değeri, 2,04 ile kalıntılar arasında otokorelasyon olmadığını işaret etmektedir. Model 0'ın eğitimi ve testi toplamda yaklaşık 6 dakika 12 saniye sürmüştür. Modelin karmaşıklığı ve veri setinin büyüklüğü göz önüne alındığında süre makul olsa da, hedeflenen düşük MAPE değerine ulaşamadığı için modelin performansı yetersiz bulunmuştur.

SARIMA modelinin parametrelerini $(7, 1, 7)$ (p, d, q) ve $(3, 1, 2, 12)$ (P, D, Q, S), önceden belirlenen tahmin süresini (50) ve gecikme sayısını (31) alan bir model (Model 1) tasarlanmıştır.

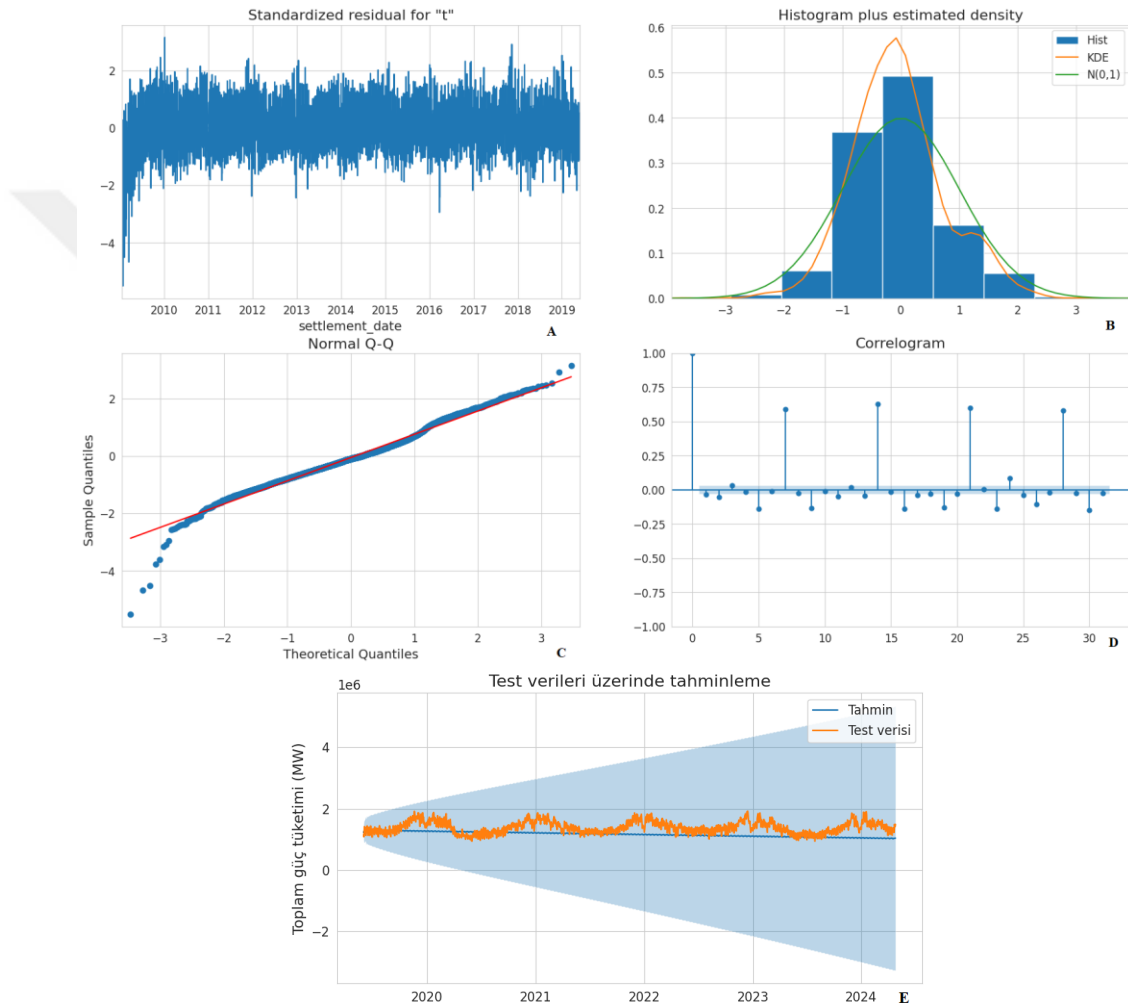


Şekil 4.10. SARIMA Model 1 grafikleri **A)** t için standartlaştırılmış artık gösterimi **B)** Histogram ve tahmin eğrisi **C)** Normal Q-Q grafiği **D)** Korelasyon grafiği **E)** Tahmin grafiği

Model 1, en iyi performansı göstermiştir ve MAPE değeri olarak 13,02 elde edilmiştir. Bu, tahminlerin gerçek değerlere oldukça yakın olduğunu gösterir. Normality ve Ljung-Box testleri, kalıntıların normal dağılımdan sapma ve otokorelasyon içerdiğini göstermektedir ($p < 0,05$). Ancak, heteroskedastisite testi, varyansın sabit olduğunu ($p > 0,05$) ortaya koymuştur. Durbin-Watson değeri 2,07 olarak hesaplanmış ve otokorelasyon olmadığını göstermiştir. Bu modelin eğitimi ve tahmini 8 dakika 21 saniye sürmüştür. Daha karmaşık bir model olmasına rağmen, test sonuçları göz önüne alındığında bu süre,

elde edilen yüksek doğruluk ile kıyaslandığında kabul edilebilir bir süre olarak değerlendirilmiştir.

Model parametreleri yeniden değiştirilmiştir. SARIMA modelinin parametrelerini (3, 1, 2) (p, d, q) ve (0, 1, 3, 12) (P, D, Q, S), önceden belirlenen tahmin süresini (50) ve gecikme sayısını (31) alan bir model (Model 2) tasarlanmıştır.

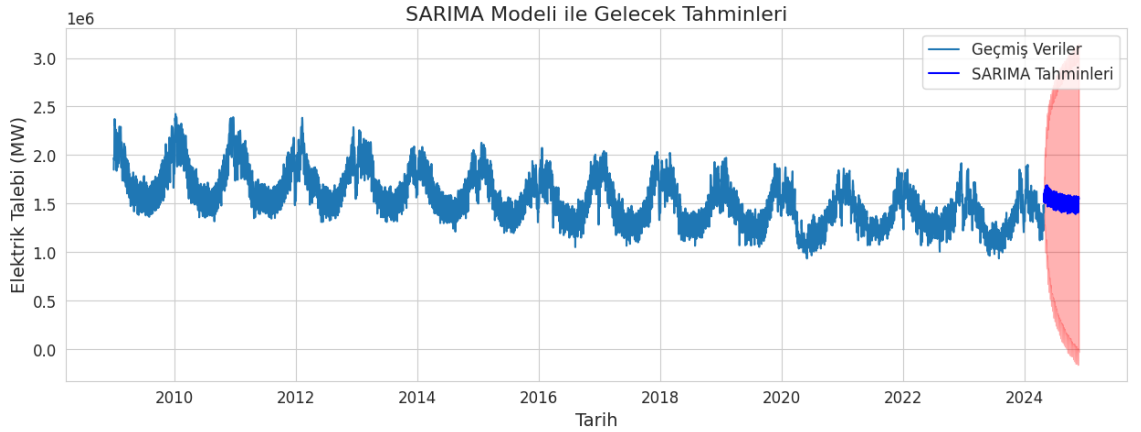


Şekil 4.11. SARIMA Model 2 grafikleri **A)** t için standartlaştırılmış artık gösterimi **B)** Histogram ve tahmin eğrisi **C)** Normal Q-Q grafiği **D)** Korelasyon grafiği **E)** Tahmin grafiği

Model 2, MAPE değeri %17,59 olarak elde edilmiştir. Modelin tahminleri, Model 1 kadar olmasa da yine de gerçek değerlere yakın sonuçlar vermiştir. İstatistiksel testler, kalıntıların normal dağılımdan sapma gösterdiğini ($p < 0,05$) ve otokorelasyon içerdiğini

($p < 0,05$) göstermektedir. Heteroskedastisite testi, varyansın sabit olmadığını ($p < 0,05$) ortaya koymuştur. Durbin-Watson değeri 2,06 ile otokorelasyon olmadığını işaret etmektedir. Bu model, en hızlı çalışan modeldir ve toplamda 1 dakika 50 saniyede tamamlanmıştır. Daha basit bir model yapısına sahip olması, daha hızlı hesaplama süreleri sağlarken, doğruluk açısından Model 1'in gerisinde kalmıştır. Bu nedenle, zaman kısıtlamalarının önemli olduğu durumlarda tercih edilebilir.

Üç farklı SARIMA modeli üzerinde yapılan testler, modellerin tahmin doğruluğu ve performans süreleri açısından önemli farklılıklar göstermiştir. Model 0, %32,32 MAPE değeri ile diğer modellere kıyasla daha düşük bir tahmin doğruluğu sergilemiştir. Model 1, %13,02 MAPE değeri ile en iyi performansı göstermiş ve tahminlerin gerçek değerlere oldukça yakın olduğunu ortaya koymuştur. Bu model, daha karmaşık yapısına rağmen kabul edilebilir bir süre içinde sonuç vermiştir. Model 2 ise %17,59 MAPE değeri ile orta düzeyde bir doğruluk sağlamış, ancak daha hızlı bir hesaplama süresi sunmuştur. Bu sonuçlar, Model 1'in hem doğruluk hem de güvenilirlik açısından en uygun model olduğunu göstermektedir, ancak belirli durumlarda Model 2'nin hızlı sonuç verme avantajı da dikkate alınabilir.



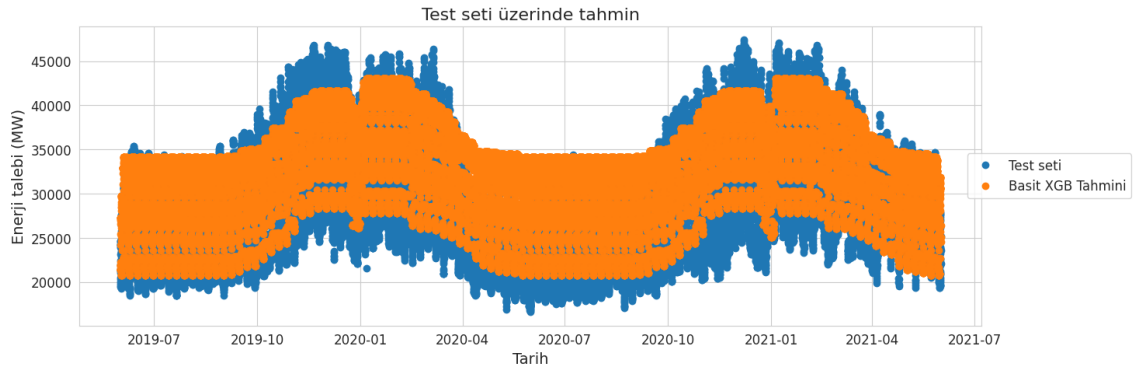
Şekil 4.12. SARIMA Model 2 gelecek tahminleri

SARIMA modeli günlük elektrik talebini tahmin etmek için etkili bir araçtır ancak, normallik ve otokorelasyon sorunları nedeniyle tahmin güvenilirliği sınırlı kalmıştır. Modelin performansı kullanılan model parametreleri ve veri setinin özellikleri gibi faktörlere bağlı olarak geliştirilebilir.

4.2. XGBoost Uygulaması

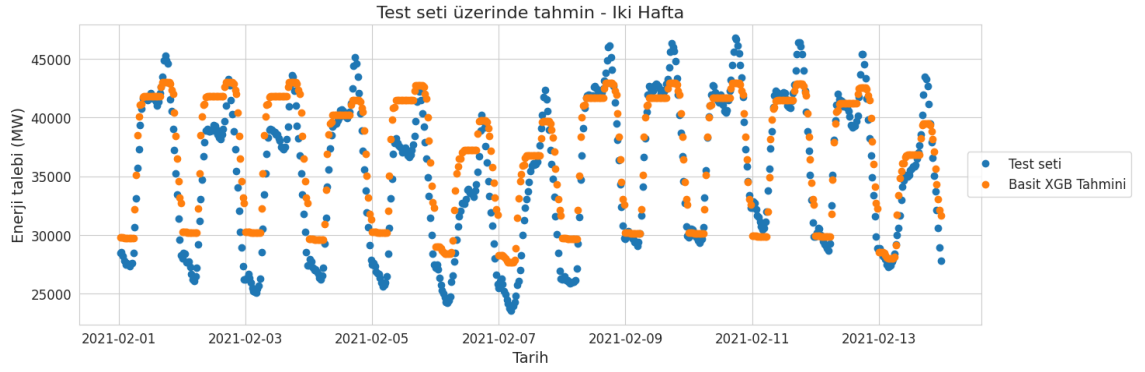
XGBoost, esnekliği ve yüksek performansı sayesinde günümüzde yaygın olarak kullanılan güçlü bir algoritmadır. XGBoost, paralel hesaplama ve verimli bellek kullanımı sayesinde çok yüksek performans sunar. Bu, büyük veri setleri üzerinde hızlı eğitim sağlar. Algoritmanın performansı, hiperparametrelerin dikkatli bir şekilde ayarlanmasına bağlıdır, bu da zaman alıcı olabilir.

Veriler, 2019 Haziran öncesi eğitim seti, 2019 Haziran ile 2021 Haziran arası test seti ve 2021 Haziran sonrası dışarıda tutma seti olarak ayrılmıştır. 2021 sonrası dışarıda tutma seti, model eğitilirken bağımsız bir değerlendirme yapılabilmesi amacıyla kullanılmıştır.



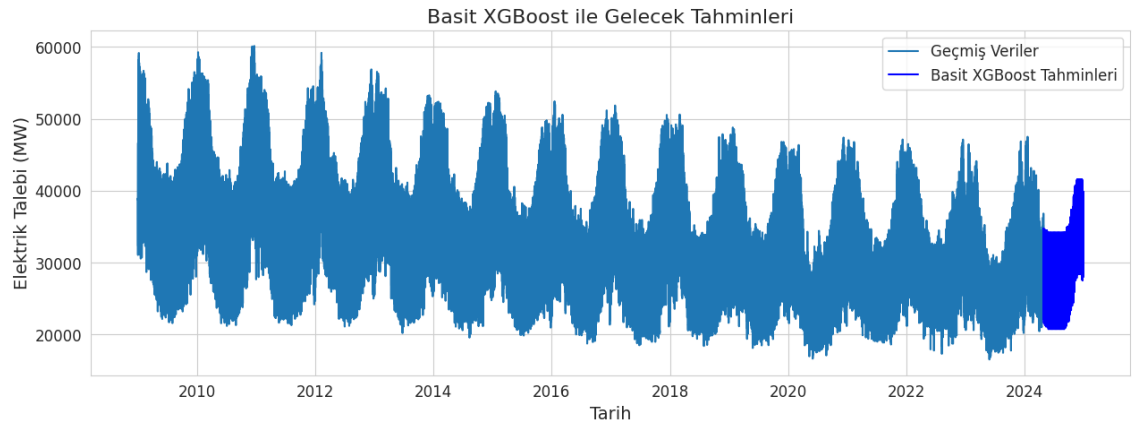
Şekil 4.13. Test verileri ile tahmin edilen değerlerin karşılaştırması (Basit XGBoost)

XGBoost modeli, zirveleri ve vadileri tespit etmekte zorlanmaktadır. Basit modelin Ortalama Mutlak Yüzde Hatası (MAPE) %11,29, Kök Ortalama Kare Hatası (RMSE) ise 3786,88 MW olarak hesaplanmıştır. Bu sonuçlar, modelin belirli dönemlerdeki hassasiyetini artırmak için daha fazla iyileştirmeye ihtiyaç duyabileceğini göstermektedir.

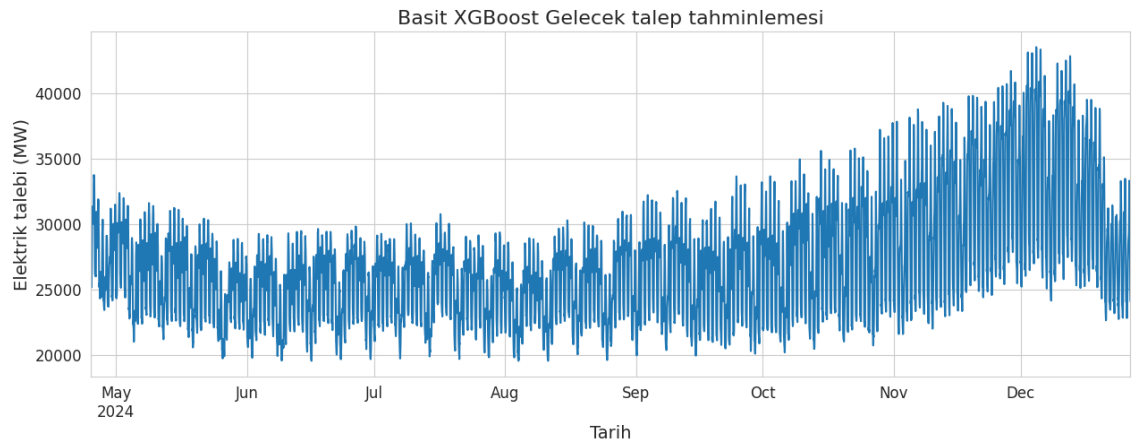


Şekil 4.14. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Basit XGBoost)

İki haftalık bir örneklem desenine bakıldığında, tahmin deseninin genel deseni tanımlayabildiği fakat uç noktaları temsil eden verilere uzak kaldığı fark edilmiştir. Modelin gelecek talep tahminleri aşağıdaki şekillerde görselleştirilmiştir.



Şekil 4.15. Tarihsel veriler ve Basit XGBoost, 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

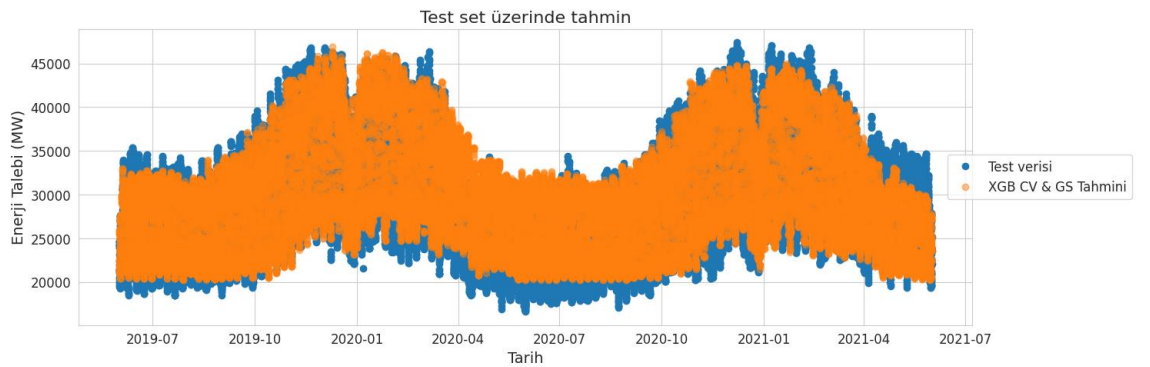


Şekil 4.16. Basit XGBoost, 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

Genel olarak, en iyi tahmin performansını elde etmek için hiperparametre ayarı esastır. Model hiperparametrelerini ayarlayarak daha iyi bir tahmine yol açıp açmadığını görmek için grid araması ve çapraz doğrulama kullanılmıştır. XGBoost, regresyon modeli oluşturma sürecinde manuel olarak ayarlanması gereken bir düzineden fazla hiperparametreye sahiptir. XGBoost'un hiperparametreleri genel parametreler, artırıcı parametreler ve görev parametrelerinden oluşur. Artırıcı parametreler performans üzerinde en büyük etkiye sahiptir.

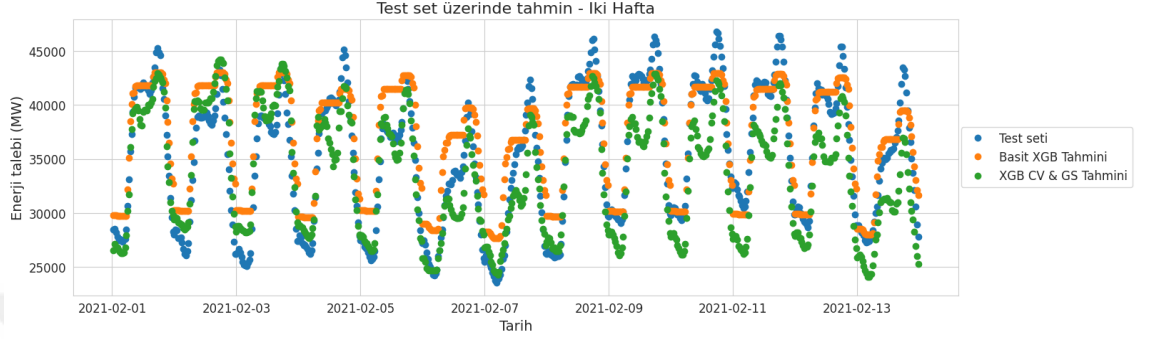
TimeSeriesSplit kullanılarak dört katmanlı bir çapraz doğrulama stratejisi uygulanmıştır. Her bir katmanda eğitim ve test veri setleri oluşturulmuş, her test seti için bir yıllık veri (48 saatlik dönemlerle çarpılmış 365 gün) ayrılmıştır. Bu yaklaşım, modelin genelleme yeteneğini değerlendirmek ve zaman içerisindeki değişikliklere nasıl tepki verdiğini anlamak için önemli bir adımdır.

XGBoost regresyon modeli için hiperparametre optimizasyonu gerçekleştirilmiştir. Modelin performansını artırmak amacıyla, max_depth, n_estimators ve subsample gibi hiperparametreler üzerinde bir arama yapılmıştır. max_depth, bir ağacın alabileceği maksimum derinliği belirlerken; n_estimators, modelde kullanılacak toplam ağaç sayısını ifade etmektedir. subsample ise her bir ağacın eğitilmesi için kullanılan veri alt kümesinin oranını belirlemektedir. GridSearchCV kullanılarak yapılan bu titiz arama süreci sonucunda, en iyi parametreler {'max_depth': 5, 'n_estimators': 500, 'subsample': 0,95} olarak bulunmuştur. Bu parametreler, modelin genel doğruluğunu artırmak ve aşırı öğrenmeyi önlemek için dikkatlice seçilmiştir.



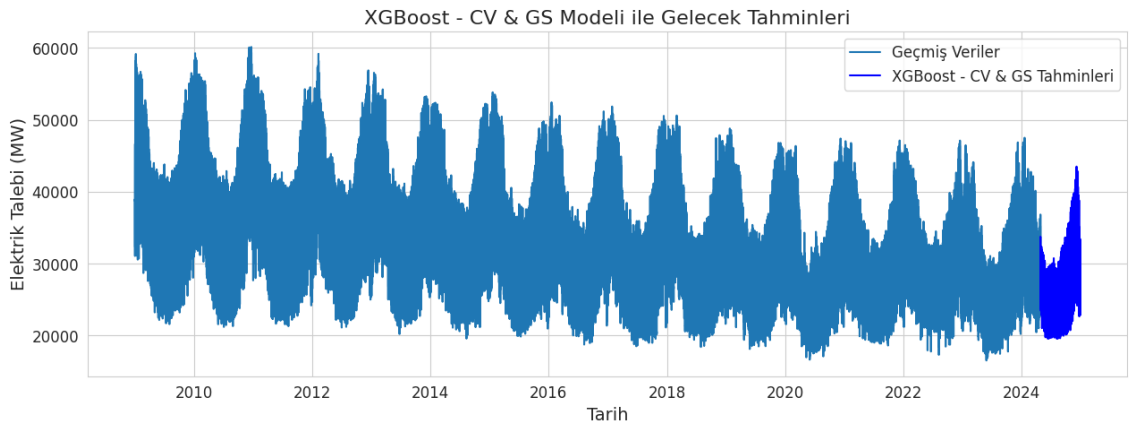
Şekil 4.17. Test verileri ile tahmin edilen değerlerin karşılaştırması (XGBoost CV&GS)

Çapraz doğrulama ve ızgara arama uygulanmış XGBoost modelinin test seti üzerindeki tahminlerini karşılaştırdığımızda, hiperparametre optimizasyonu yapılmış model tahminlerinin daha iyi olduğu görülmektedir. İyileştirilmiş modelin MAPE sonucu %7,29, RMSE değeri 2655,08 MW'tır.

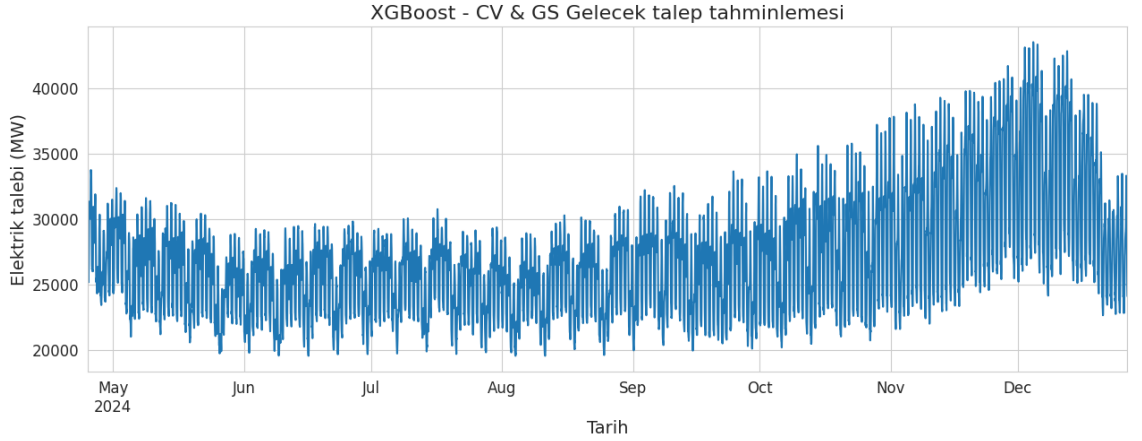


Şekil 4.18. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (XGBoost CV & GS)

İki haftalık bir örneklem desenine bakıldığında, çapraz doğrulama ve ızgara arama ile iyileştirilmiş modelin dip noktaları daha iyi temsil ettiği genel desene uyumuna rağmen aykırı dalgalanmalarda günlük mevsimselliği ve dalgalanmaları iyi yansıtmadığı görülmüştür. İyileştirilmiş modelin gelecek talep tahminleri aşağıdaki şekillerde görselleştirilmiştir.



Şekil 4.19. Tarihsel veriler ve XGBoost CV & GS 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

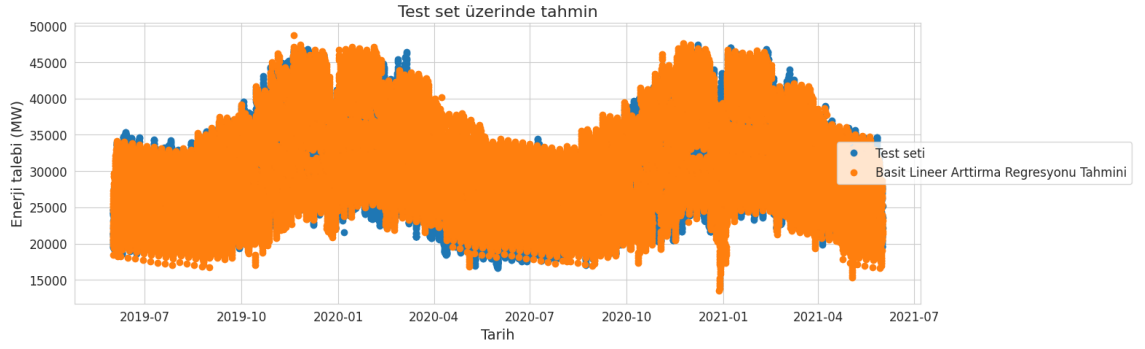


Şekil 4.20. XGBoost CV & GS 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

Basit XGBoost modeli için MAPE değeri %11,29 iken, çapraz doğrulama ve ızgara arama (CV & GS) uygulanmış modelde bu değer %7,29'a düşmüştür. Basit model için RMSE 3786,88 iken, optimize edilmiş modelde 2655,08'e düşmüştür. Basit modelin CPU zamanı 19,7 saniye iken, optimize edilmiş modelde bu süre 13 dakika 51 saniyedir. Hiperparametre optimizasyonu daha fazla hesaplama gerektirdiğinden, işlem süresi artmıştır. Optimize edilmiş modelin daha iyi tahmin doğruluğu sağladığı ancak bu doğruluğu elde etmek için daha fazla hesaplama süresi gerektirdiği görülmektedir. Daha doğru tahminler isteyen uygulamalar için optimize edilmiş modelin tercih edilebileceğini, ancak hesaplama kaynakları ve zamanın sınırlı olduğu durumlarda basit modelin yeterli olabileceği söylenebilir.

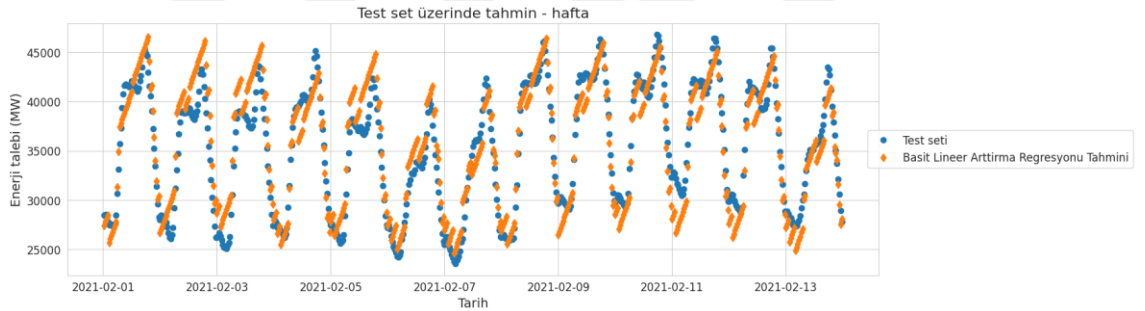
4.3. Lineer Arttırma Regresyonu Uygulaması

Enerji talebini tahmin etmek için öncelikle bir Lineer Regresyon modeli ile başlayan ve daha sonra Lineer Arttırma Regresyon sınıfı ile güçlendirilen bir model oluşturulmuştur. Model 350 temel tahmin edici (modelin oluşturacağı ağaç sayısı) kullanarak eğitilmiş, her bir temel tahmin edicinin en fazla 6 seviyeli bir karar ağacı olabileceği maksimum derinlik belirlenmiştir. Eğitilmiş model kullanarak test verileri tahmin edilmiştir.



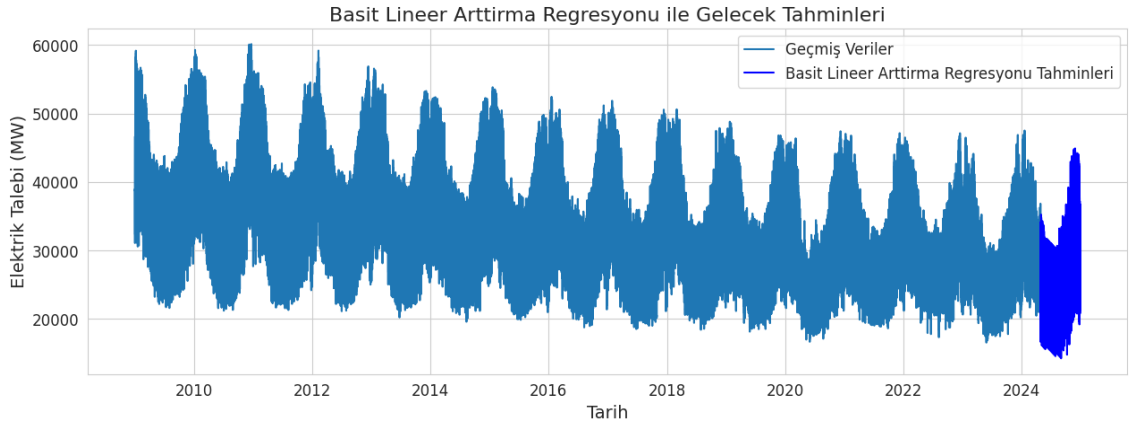
Şekil 4.21. Test verileri ile tahmin edilen değerlerin karşılaştırması (Basit Lineer Arttırma Regresyonu)

Gerçek değerler ile tahmin değerlerini gösteren grafikler incelendiğinde modelin başarılı bir tahminleme yaptığı ancak bazı sapmalar ve süreksizlikler olduğu görülmektedir. Ortalama Yüzde Mutlak Hata (MAPE) değeri %8,20 ve RMSE değeri 2963,08 MW olarak hesaplanmıştır. Bu değerler, modelin enerji talebini tahmin etme konusunda başarılı olduğunu göstermektedir.

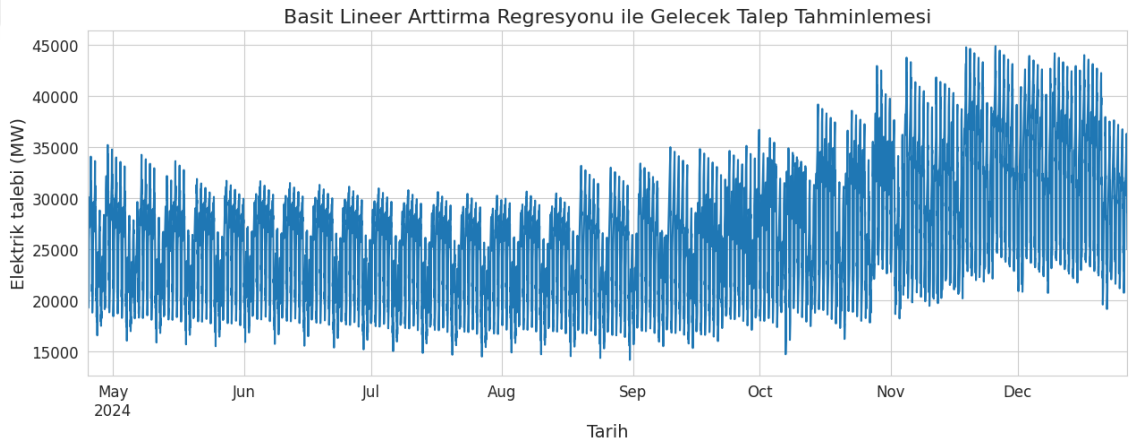


Şekil 4.22. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Basit Lineer Arttırma Regresyon)

29 dakika 53 saniyelik CPU zamanı, modelin karmaşıklığı göz önüne alındığında, oldukça uzun bir süredir. Bu, modelin daha hızlı eğitilmesi için bazı optimizasyon tekniklerinin kullanılabileceğini gösterir. Örneğin, daha az sayıda temel tahmin edici kullanmak, daha küçük bir maksimum derinlik seçmek veya daha verimli bir temel tahmin edici kullanmak gibi yöntemler denenebilir fakat bu durumda performans düşüşü gözlemlenebilir. Modelin gelecek talep tahminleri aşağıdaki grafiklerdeki gibi çizdirilmiştir.



Şekil 4.23. Tarihsel veriler ve Basit Lineer Arttırma Regresyonu 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

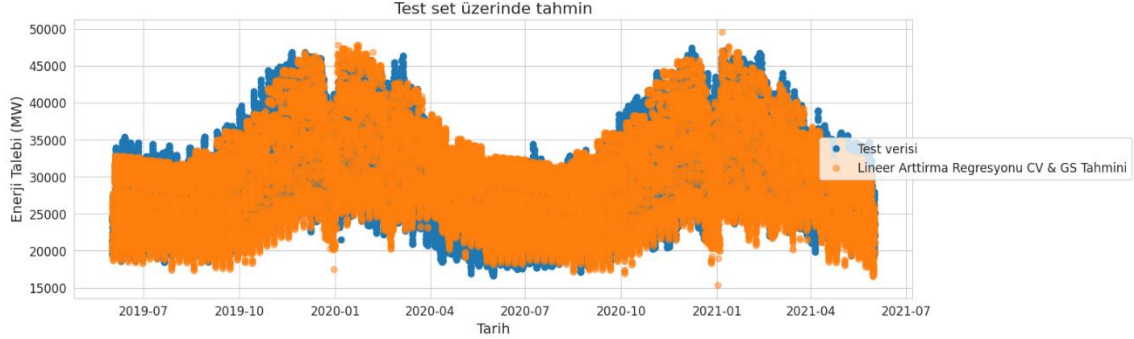


Şekil 4.24. Basit Lineer Arttırma Regresyonu 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

Lineer Arttırma Regresyon algoritmasının enerji talebini tahmin etmek için kullanılabileceği ve MAPE ve RMSE değerleri dikkate alındığında, modelin tahminlerde iyi bir performans gösterdiği gözlenmiştir.

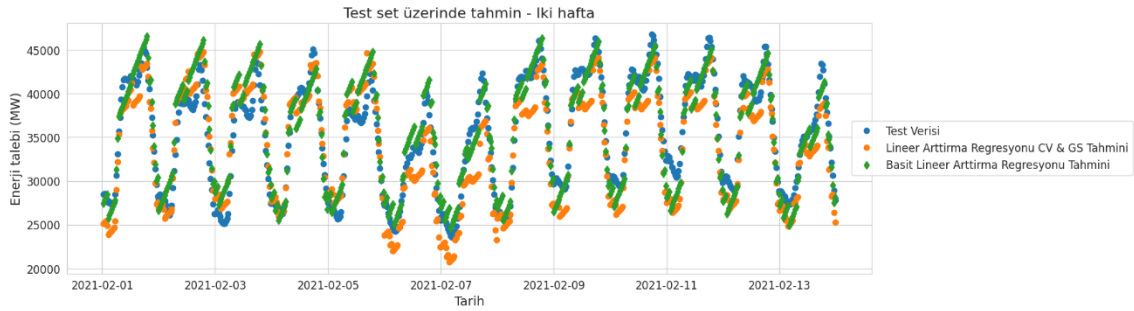
Lineer Arttırma Regresyon modeli için hiperparametre optimizasyonu yapılmıştır. Modelin temel taşıyıcısı olarak LinearRegression kullanılmış ve max_depth ile n_estimators gibi hiperparametreler optimize edilmiştir. GridSearchCV kullanılarak yapılan bu kapsamlı optimizasyon süreci, negatif kök ortalama kare hatası (neg_root_mean_squared_error) metriği kullanılarak değerlendirilmiştir. Yapılan optimizasyon sonucunda, en iyi parametreler {'max_depth': 3, 'n_estimators': 200} olarak

tespit edilmiştir. Bu parametreler, modelin doğruluğunu artırmak ve genelleme yeteneğini geliştirmek amacıyla seçilmiştir.



Şekil 4.25. Test verileri ile tahmin edilen değerlerin karşılaştırması (Lineer Arttırma Regresyonu CV&GS)

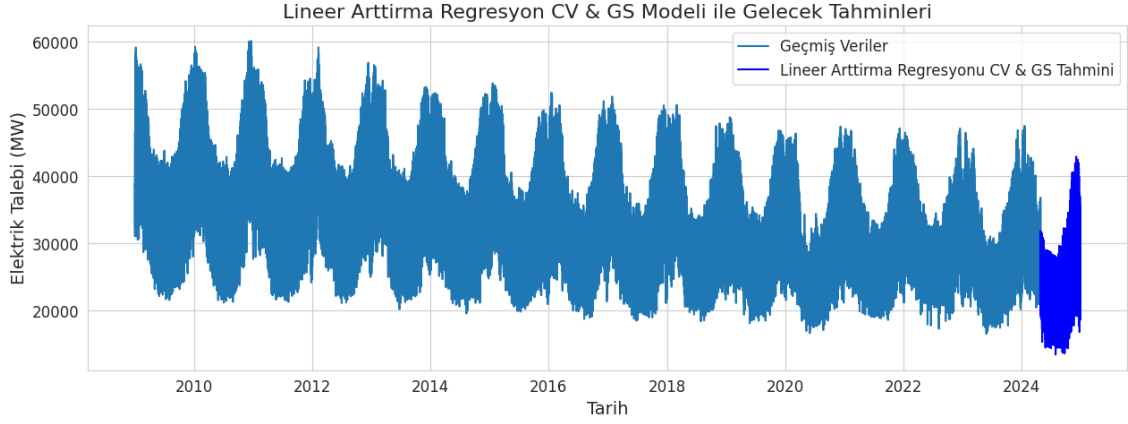
Test verileri ile modelin tahmin ettiği değerlerin karşılaştırıldığı grafikte, modelin genel noktaları iyi temsil ettiği fakat modelin belirli veri noktalarında beklenmeyen tahminler ürettiği gözlenmiştir. Modelin değerlendirilmesi sonucunda, Ortalama Mutlak Yüzde Hatası (MAPE) %7,73 ve Kök Ortalama Kare Hatası (RMSE) 2810 MW olarak bulunmuştur. Ancak, modelin eğitilmesi ve çalıştırılması için gereken CPU süresi 1 saat 22 dakika 51 saniye gibi uzun bir süreye ulaşmıştır.



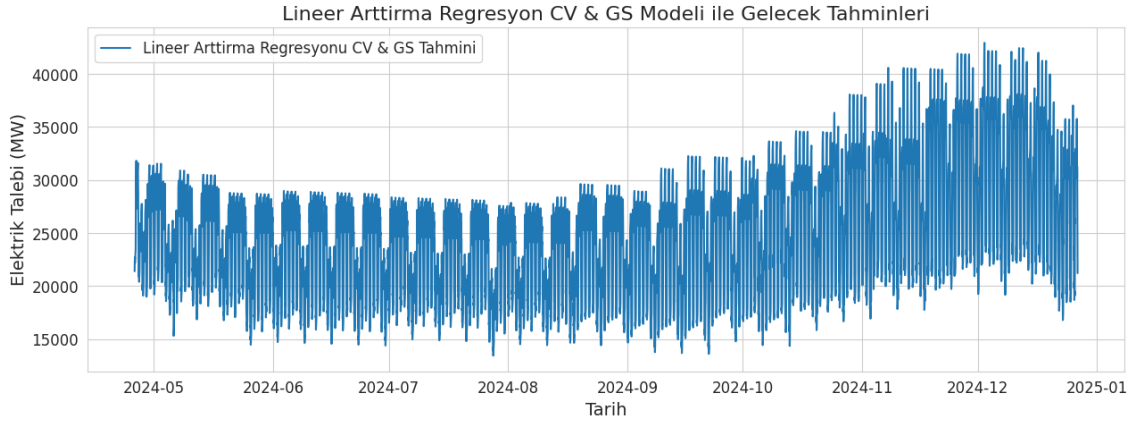
Şekil 4.26. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık

Şekil 4.26’da iki modelin tahmin performansları karşılaştırılmıştır. Her iki model de genel eğilimi yakalamış fakat Lineer Arttırma Regresyonu CV & GS gerçek değerlere daha yakın tahminlerde bulunarak daha iyi bir performans sergilemiştir. Hiperparametre optimizasyonu yapılmış olan CV&GS modeli, genel performans metriklerinde daha iyi sonuçlar verse de, aykırı noktalar üzerindeki duyarlılığı artırmak için ek iyileştirmelere

ihtiyaç duyabilir. Modelin gelecek talep tahminleri aşağıdaki grafiklerdeki gibi çizdirilmiştir.



Şekil 4.27. Tarihsel veriler ve Lineer Arttırma Regresyonu CV&GS 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri



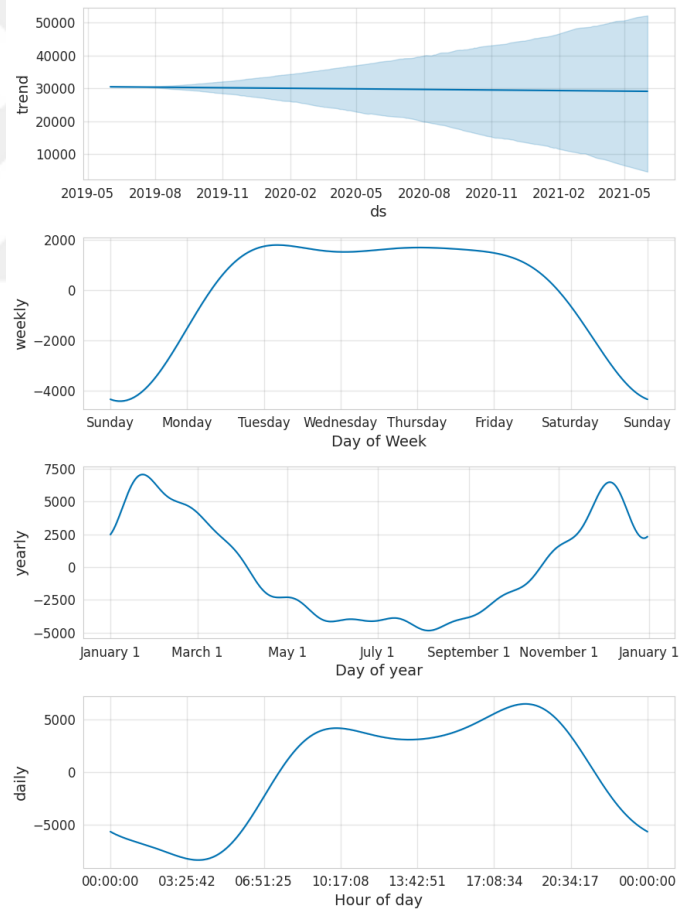
Şekil 4.28. Lineer Arttırma Regresyonu CV&GS 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

Basit model için MAPE değeri %8,2 iken, optimize edilmiş modelde bu değer %7,73'e düşmüştür. Basit modelin RMSE değeri 2963,08 iken, optimize edilmiş modelde 2810'a düşmüştür. İyileştirilmiş MAPE ve RMSE degerleri CV & GS modelinin daha hassas ve doğru tahminler ürettiğini göstermektedir. Basit modelin CPU zamanı 29 dakika 53 saniye, optimize edilmiş modelin ise 1 saat 22 dakika 51 saniyedir. Optimize edilmiş modelin daha fazla hesaplama gerektirdiği açıktır. CV & GS modeli, daha iyi performans sağlasa da, daha yüksek hesaplama maliyeti ve zaman gereksinimi ile gelir. Bu sonuçlar, doğruluk ile işlem süresi arasında bir denge kurma gerekliliğini vurgulamaktadır.

4.4. Prophet Uygulaması

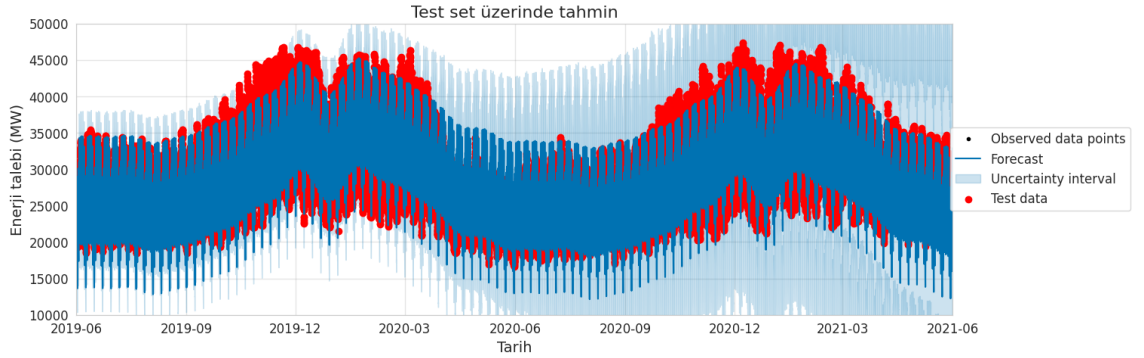
Prophet modeli, zaman serisi tahmininde güçlü ve esnek bir araç olarak öne çıkmaktadır. Trend, mevsimsellik ve tatil etkilerini ayrı ayrı modelleyerek, karmaşık zaman serisi verilerinde yüksek doğruluk sağlar. Kullanım kolaylığı, esneklik ve farklı alanlarda başarılı uygulama örnekleri, Prophet'i zaman serisi tahmininde tercih edilen bir araç haline getirmektedir.

Prophet, güven aralıklarıyla birlikte tahmin bileşeninin (trendler ve mevsimsellikler) kolayca görselleştirilmesine olanak tanıyan bir fonksiyon içerir. Trend ve mevsimsellik aşağıdaki grafiklerdeki gibi görselleştirilmiştir.



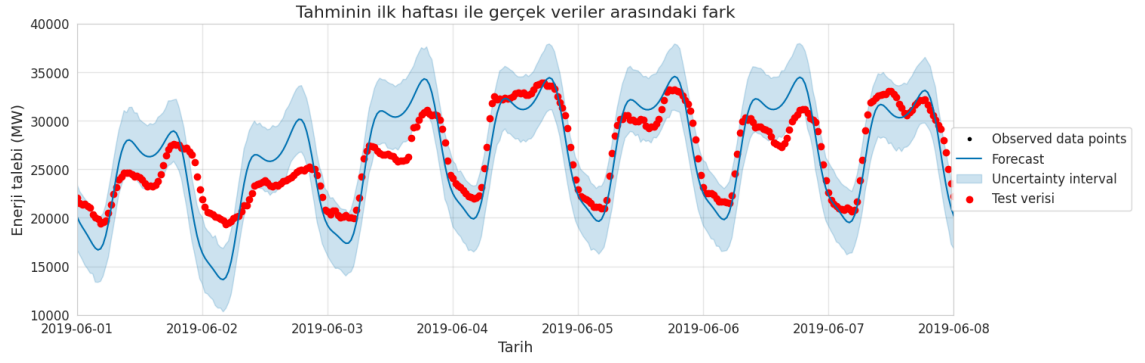
Şekil 4.29. Prophet trend ve mevsimsellik grafikleri

Trend çizgisi, 2019 Haziran ile 2021 Haziran arasındaki dönemde zaman serisinde hafif bir yükseliş trendi göstermektedir. Mavi gölgeli alan, trend için güven aralığını temsil etmektedir. Haftalık desen, zaman serisi değerinin Pazar günü en düşük, Salı ve Çarşamba günleri zirveye ulaştığını ve ardından Pazar gününe kadar kademeli olarak azaldığını göstermektedir. Yıllık desen, Ocak ayında bir mevsimsel zirve, ardından Mayıs ayına kadar kademeli bir düşüş göstermektedir. Zaman serisi değeri daha sonra tekrar artmaya başlar, Kasım ayında başka bir zirveye ulaşır ve ardından Ocak ayına kadar azalır. Günlük desende zaman serisi değeri sabah 03:00 civarında en düşük değerini görür, saat 10:00 civarında bir zirve yapar, ardından saat 20:00 civarına kadar belirli azalma ve dalgalanmalarla devam eder. Saat 20:00 civarında yeni bir zirveye ulaştıktan sonra, sabah 03:00'e kadar düşüş gösterir. Genel olarak, grafikler zaman serisinin hafif bir yükseliş trendine, güçlü bir haftalık yıllık ve günlük mevsimselliğe sahip olduğunu göstermektedir. Bu, zaman serisinin doğasına bağlı olarak çeşitli faktörlerle ilgili olabilir.



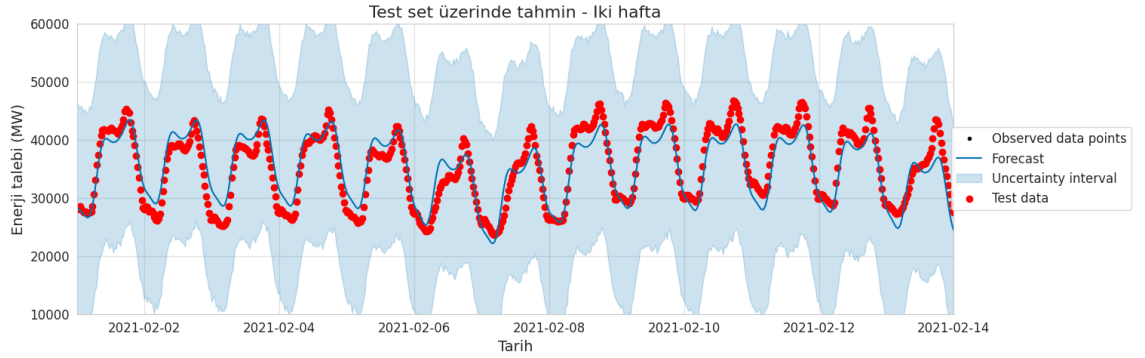
Şekil 4.30. Test verileri ile tahmin edilen değerlerin karşılaştırması (Basit Prophet)

Gerçek değerler ile tahmin değerlerini gösteren grafik incelendiğinde tahmin edilen veriler gerçek verileri oldukça iyi bir şekilde temsil etmektedir. Buna karşılık, tahmin aralığı oldukça geniştir, bu da tahminin doğruluğu konusunda bazı belirsizlikler olduğunu ve modelin daha fazla iyileştirmeye ihtiyaç duyabileceğini göstermektedir.



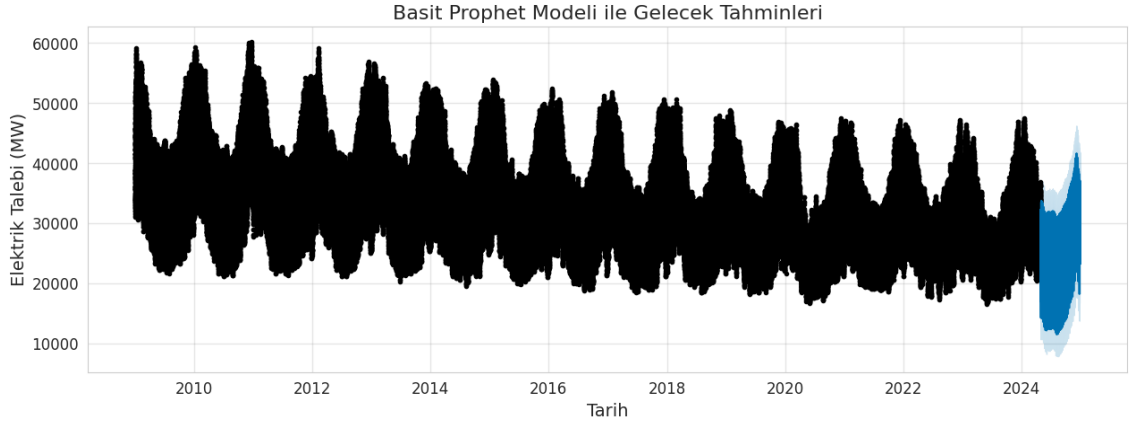
Şekil 4.31. Test verileri ile tahmin edilen değerlerin karşılaştırması (Basit Prophet)

Tahmin edilen ilk haftanın tahmin ve test verileri detay kıyası incelendiğinde zaman serisi modelinin gerçek verileri oldukça iyi bir şekilde tahmin edebildiğini göstermekle birlikte, bazı sapmalar da görülmektedir. Özellikle, 03/06/2019 ve 07/06/2019 tarihlerinde tahmin gerçek verilerin biraz altında kalmıştır.

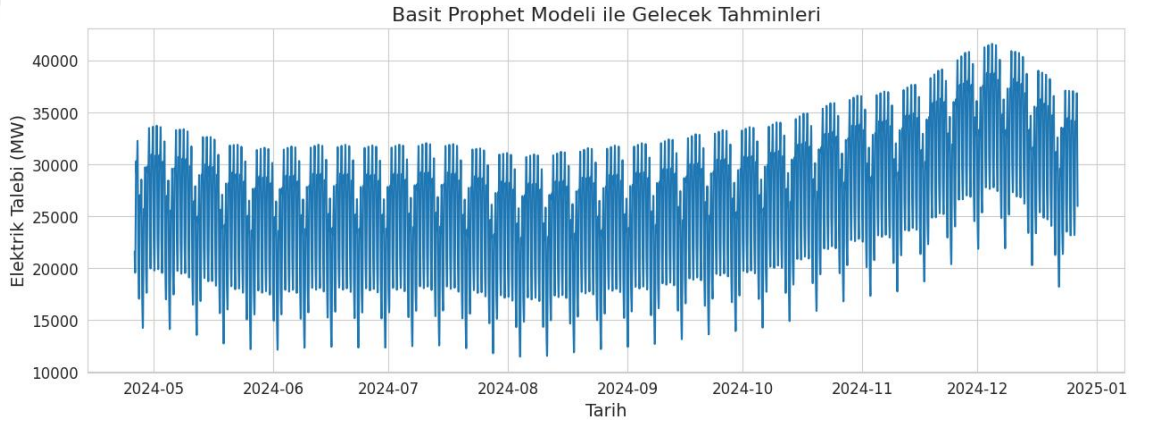


Şekil 4.32. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Basit Prophet)

İki haftalık bir örneklem desenine bakıldığında, modelin mevsimselliği genel olarak iyi tespit ettiği gözlemlenmiştir. MAPE değeri %9,37 ve RMSE değeri 3259,96 MW olarak hesaplanmıştır. Bu değerler, modelin enerji talebini tahmin etme konusunda oldukça başarılı olduğunu göstermektedir. Modelin geleceğe yönelik tahminleri ise sonraki iki şekilde ayrıntılı olarak sunulmuştur.

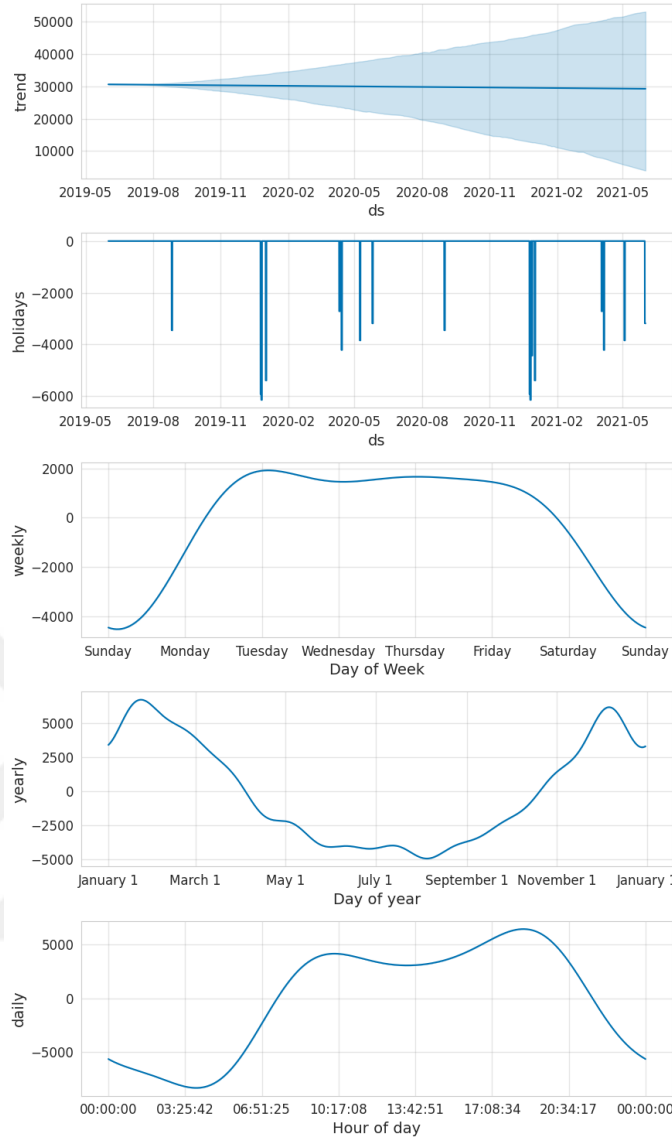


Şekil 4.33. Tarihsel veriler ve Basit Prophet Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

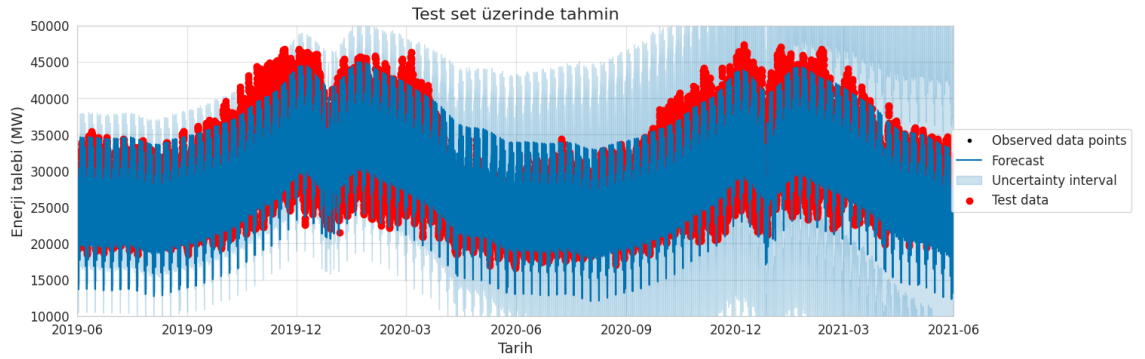


Şekil 4.34. Basit Prophet Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

Tatillerin dikkate alındığı Prophet modeli için trend ve mevsimsellik grafikleri incelendiğinde her bir tatil günü için zaman serisi değerinde önemli bir düşüş gözlemlenmektedir. Haftalık, yıllık ve günlük desen grafikleri tatilsiz modelle aynı gerçekleşmiştir.

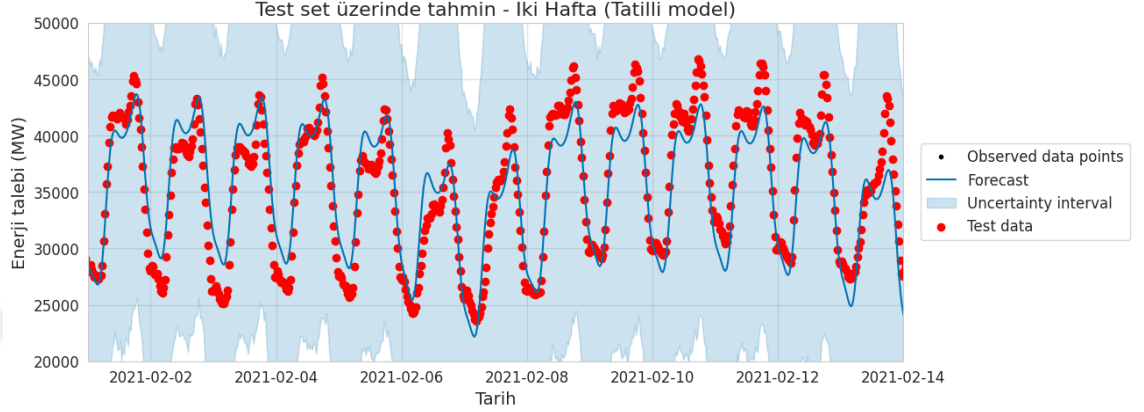


Şekil 4.35. Prophet Tatilli Model trend ve mevsimsellik grafikleri



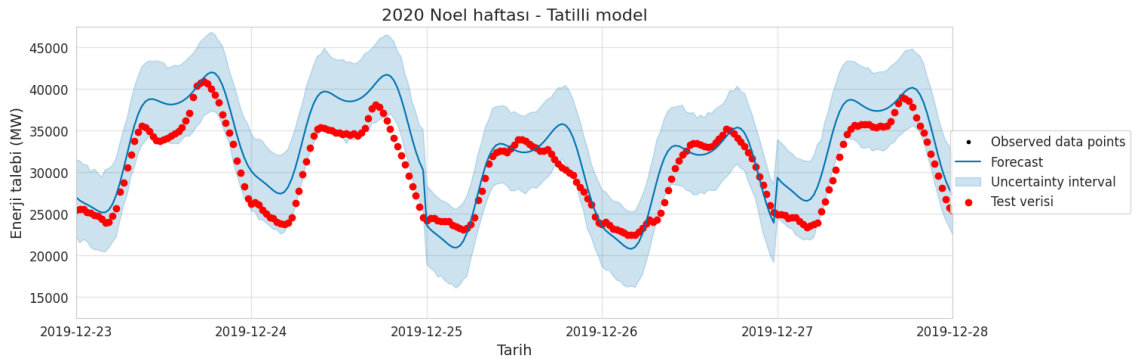
Şekil 4.36. Tatillerin dahil edildiği Prophet test verileri ile tahmin edilen değerlerin karşılaştırması

Tatilli Prophet modelinin gerçek değerler ile tahmin değerlerini gösteren grafik incelendiğinde basit modele çok benzer bir desen elde edilmiştir. Tahmin edilen veriler gerçek verileri oldukça iyi bir şekilde temsil etmektedir. Buna karşılık, tahmin aralığı oldukça geniştir.

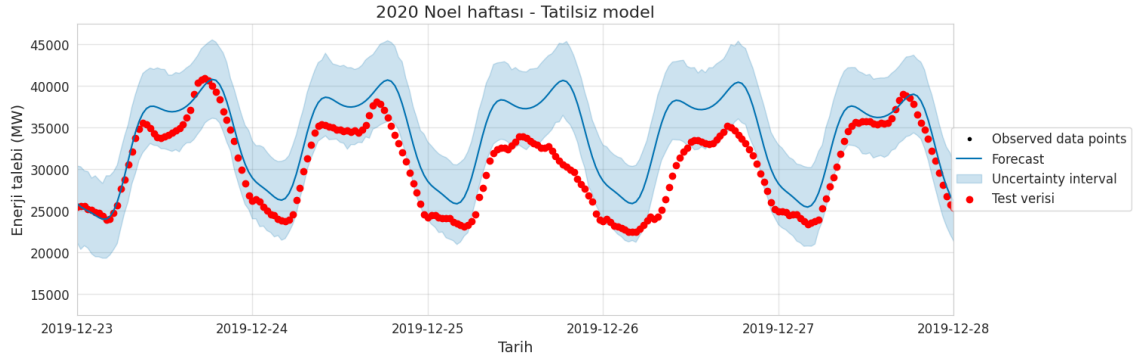


Şekil 4.37. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Tatilli Prophet)

Tatil etkilerinin dikkate alındığı tahmin modeli, gerçek verileri iyi bir şekilde takip etmektedir, ancak bazı sapmalar mevcuttur. Dalgalanmalardaki belirli zirveler veya çukurlar tatil günleri ile açıklanabilir. Prophet modeli, tatilleri dikkate alarak tahmin yaparken %9,34'lük bir MAPE ile oldukça iyi bir performans sergilemektedir. MAPE değeri %9,37'den %9,34'ye, RMSE değeri 3259,96 MW'dan 3236,53 MW'a iyileşmiştir. Modelin tahminlerinin genellikle gerçek değerlere yakın olduğu söylenebilir.

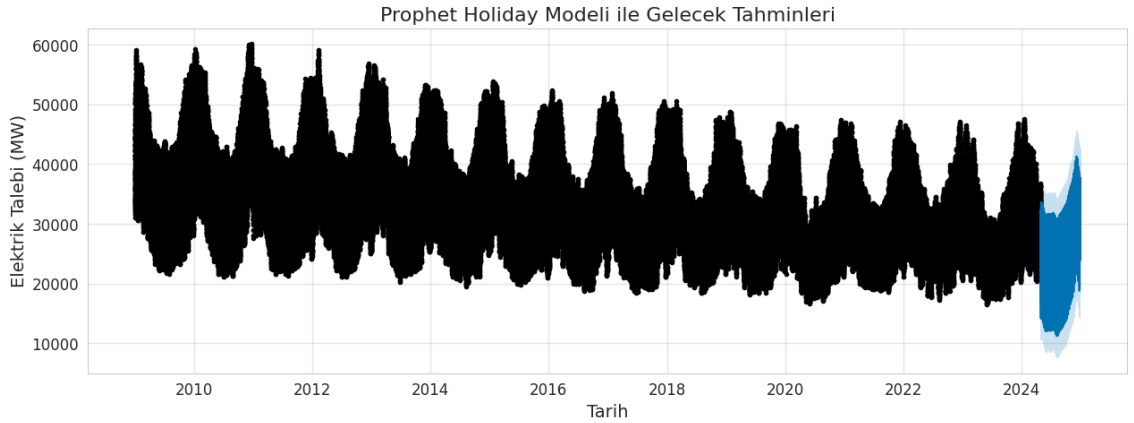


Şekil 4.38. 2019 Noel haftası tatilli Prophet modeli test verisi ve tahmin verisi kıyası

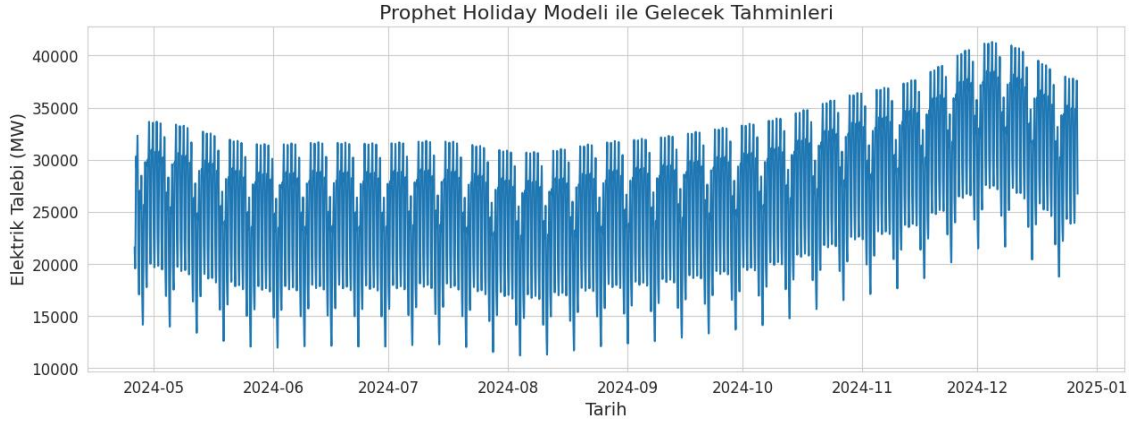


Şekil 4.39. 2019 Noel haftası tatilsiz Prophet modeli test verisi ve tahmin verisi kıyası

2019 Noel haftası için enerji talebi karşılaştırıldığında tatillerin dikkate alındığı model, Noel tatili gibi özel günlerdeki enerji talebini daha doğru bir şekilde tahmin etmektedir, ancak bu modelin güven aralığı biraz daha geniştir. Tatillerin dikkate alınmadığı model, daha genel bir tahmin sunar ve güven aralığı biraz daha dardır. Hangi modelin daha iyi olduğu, uygulamaya özel olarak değerlendirilmelidir. Eğer Noel tatili gibi özel günlerde enerji talebi daha doğru bir şekilde tahmin etmek isteniyorsa, tatillerin dikkate alındığı model tercih edilebilir. Ancak, daha genel bir tahmin ve daha dar bir güven aralığı isteniyorsa, tatillerin dikkate alınmadığı model tercih edilebilir. Modelin geleceğe yönelik tahminleri ise Şekil 4.40 ve Şekil 4.41’de ayrıntılı olarak sunulmuştur.

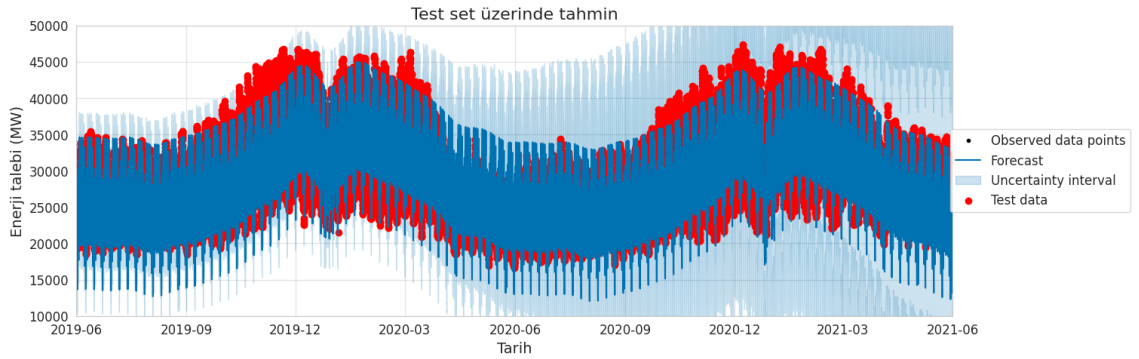


Şekil 4.40. Tarihsel veriler ve Tatilli Prophet Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri



Şekil 4.41. Tatilli Prophet Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

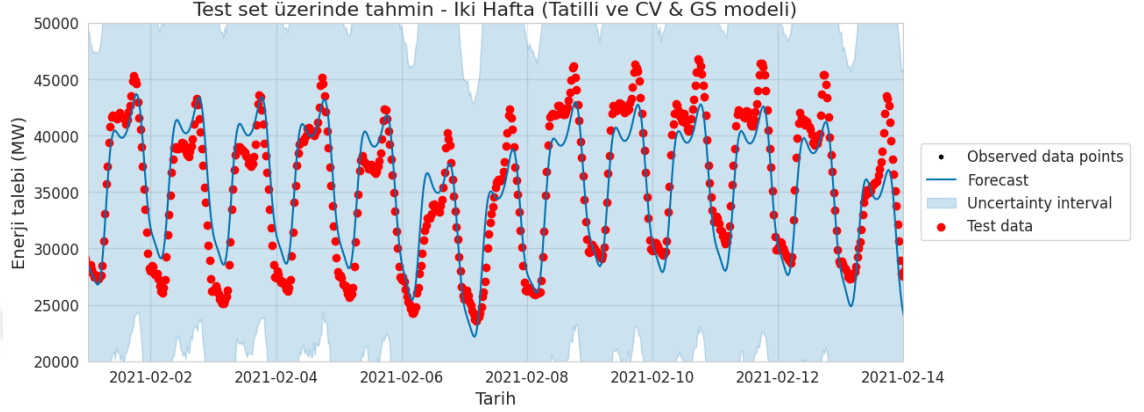
Çapraz doğrulama ve ızgara arama kullanılarak en iyi parametre seti bulunduğunda, tüm veri kümesi kullanılarak model eğitilip ve ardından gelecek tahminleri yapılmıştır. TimeSeriesSplit ile zaman serisi doğrulama yöntemi uygulanmıştır. Hiperparametre optimizasyonu için, changepoint_prior_scale ve seasonality_prior_scale parametreleri üzerinde bir ızgara arama (grid search) yapılmıştır. Bu parametreler, modelin değişim noktalarına ve mevsimselliğe ne kadar esnek olacağını belirler. Parametre kombinasyonlarının her biri için, model eğitilmiş ve test verileri üzerinde tahminler yapılmıştır. Her bir modelin performansı, Kök Ortalama Kare Hatası (RMSE) metriği kullanılarak değerlendirilmiştir. Sonuçlar, changepoint_prior_scale için 0,05 ve seasonality_prior_scale için 2,5 değerlerinin en iyi performansı sağladığını göstermiştir.



Şekil 4.42. Prophet (CV & GS) test verileri ile tahmin edilen değerlerin karşılaştırması

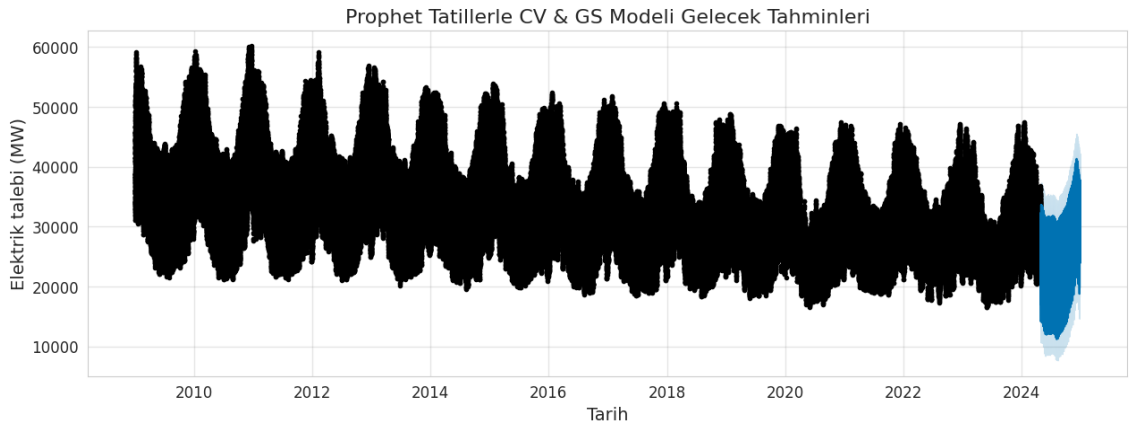
Tatilli hiperparametre optimizasyonu yapılmış Prophet modelinin gerçek değerler ile tahmin değerlerini gösteren grafik incelendiğinde diğer modellere çok benzer bir desen

elde edilmiştir. Optimizasyon sonrası en iyi parametrelerle eğitilen Prophet modeli, test verileri üzerinde değerlendirilmiş ve yeni modelin MAPE değeri %9,34, RMSE değeri ise 3235,91 MW olarak gerçekleşmiştir. MAPE değeri tatil içeren Prophet modeli ile aynı kalırken, RMSE değeri bir miktar iyileşme göstermiştir.

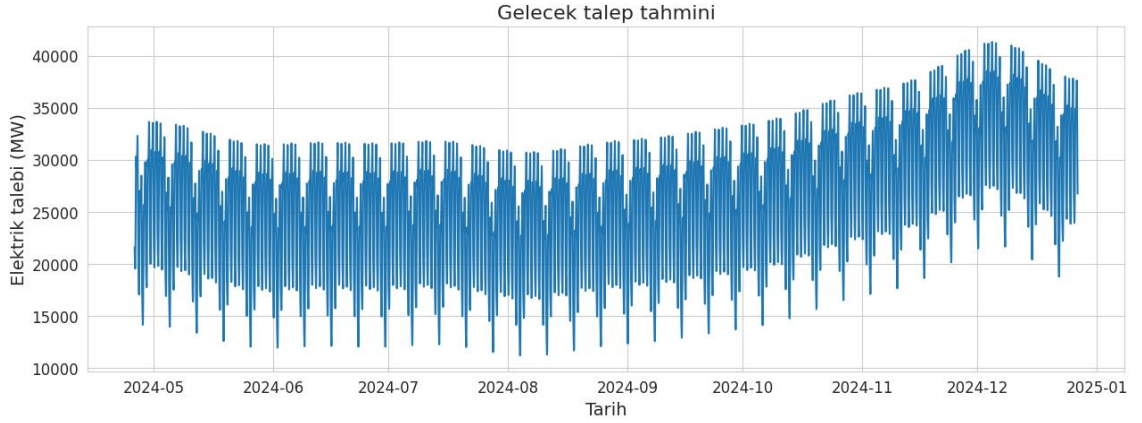


Şekil 4.43. Test verileri ile tahmin edilen değerlerin karşılaştırması - İki haftalık (Prophet CV & GS)

Tatil etkileri ile birlikte çapraz doğrulama ve ızgara arama (grid search) yöntemlerinin kullanıldığı modelin iki haftalık deseni incelendiğinde gerçek verileri iyi bir şekilde takip ettiği, ancak bazı sapmalar olduğu gözlemlenmiştir. Bu modelin gelecek talep tahminlemesi aşağıdaki grafikteki gibi tahminlenmiştir.



Şekil 4.44. Tarihsel veriler ve Prophet CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri



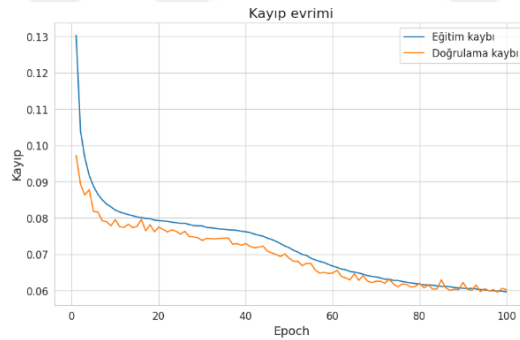
Şekil 4.45. Prophet CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

Basit Prophet için MAPE değeri %9,37 iken, tatil etkisi eklenmiş ve optimize edilmiş modellerde bu değer %9,34'tür. MAPE açısından, tatil etkisi eklenmesi ve optimize edilmesi modellerin performansını az bir oranda iyileştirmiştir. Basit Prophet RMSE değeri 3259,96, tatil etkisi eklenmiş modelde 3236,53, ve optimize edilmiş modelde 3235,91'dir. RMSE değerlerinde, tatil etkisi eklenmesi ve optimize edilmesi ile hafif bir iyileşme gözlemlenmiştir. Basit Prophet modeli için CPU zamanı 30,5 saniye, tatil etkisi eklenmiş modelde 37,3 saniye, ve optimize edilmiş modelde 3 dakika 32 saniyedir. Modellerin hesaplama süreleri oldukça iyidir. Modelin doğruluğunu artırmak ve zaman serisi tahminlerinde daha güvenilir sonuçlar elde etmek adına yapılan hiperparametre optimizasyonunun etkili olduğu görülmüştür.

4.5. LSTM ve Derin LSTM Uygulaması

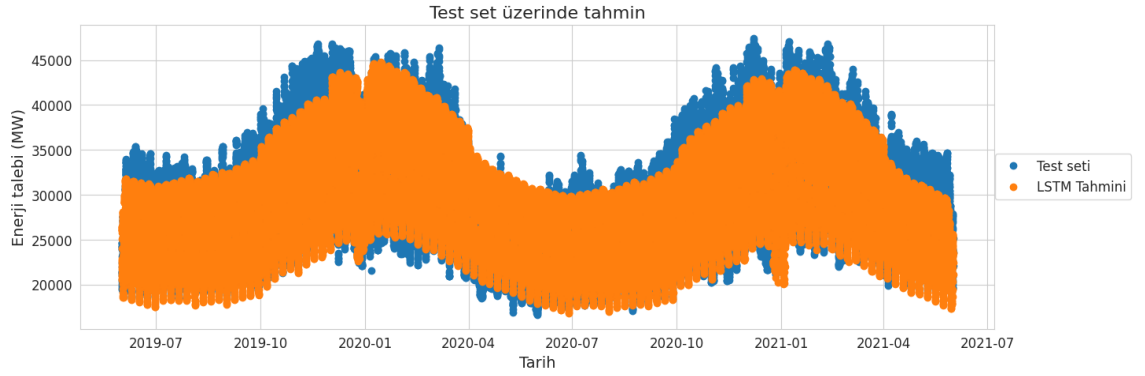
LSTM, tekrarlayan sinir ağlarının (RNN) bir türüdür ve zaman serisi verilerinin uzun vadeli bağımlılıklarını öğrenmede etkilidir. Özellikle geçmiş verilerin gelecekteki tahminleri etkileyebileceği durumlar için uygundur (El-Azab ve diğerleri, 2024). Derin LSTM, birden çok LSTM katmanının üst üste yerleştirildiği derin bir sinir ağıdır. Derin yapısı sayesinde hem kısa vadeli hem de uzun vadeli bağımlılıkları daha iyi öğrenme kapasitesine sahiptir. LSTM ve Derin LSTM, zaman serisi verilerindeki karmaşık bağımlılıkları öğrenme yetenekleri sayesinde, elektrik talep tahmininde yüksek doğruluk sağlar. Özellikle uzun vadeli bağımlılıkları temsil etme kapasiteleri, bu algoritmaları elektrik talep tahmini gibi uygulamalarda güçlü kılar.

Zaman serisi verilerini modellemek için bir LSTM (Uzun Kısa Süreli Bellek) tabanlı derin öğrenme ağı oluşturulmuştur. Eğitim, test ve hold-out veri setleri, zaman aralıklarına göre belirlenmiştir: Eğitim verileri 1 Haziran 2019 tarihinden önceki verileri, test verileri ise 1 Haziran 2019 ile 1 Haziran 2021 tarihleri arasındaki verileri kapsamaktadır. 1 Haziran 2021 sonrası veriler ise hold-out veri seti olarak ayrılmıştır. Modelin girdileri, tatil durumu, günün belirli özellikleri gibi zamanla ilgili değişkenler içeren özellikler setine dayanmaktadır. Bu veriler, MinMaxScaler ile ölçeklendirilerek modelin daha iyi öğrenmesi sağlanmıştır. Modelin yapısı, 256 birimlik bir LSTM katmanı, ardından %50 oranında dropout ve bir çıkış katmanından oluşmaktadır. Kodda zaman adımı 1 olarak belirlenmiştir. Yani, model her seferinde sadece bir zaman adımına bakarak tahmin yapar. Daha fazla zaman adımı kullanmak, modelin daha karmaşık kalıpları öğrenmesine yardımcı olabilir ve tahmin doğruluğunu artırabilir. Kök Ortalama Kare Hatası (RMSE) kayıp fonksiyonu olarak belirlenmiş ve "adam" optimizasyon algoritması ile model derlenmiştir. Eğitim sürecinde erken durdurma, model kontrol noktası kaydetme ve öğrenme oranı azaltma gibi geri çağırımlar kullanılarak modelin performansı optimize edilmiştir.



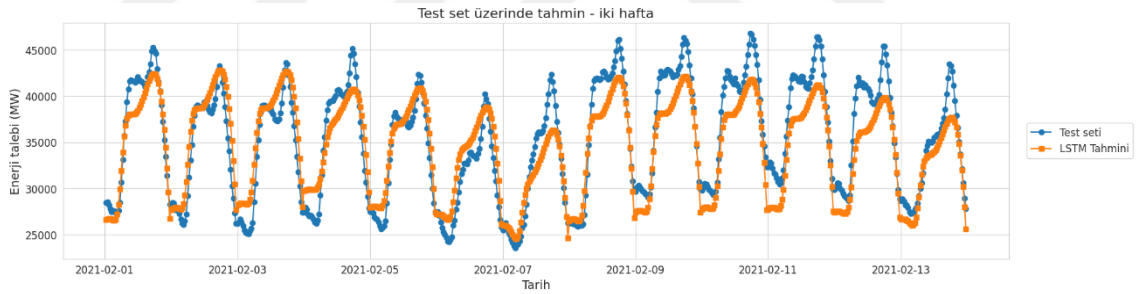
Şekil 4.46. LSTM kayıp evrimi

Bir katmanlı LSTM modelinin ayar parametrelerine bakıldığında eğitim süreci boyunca kayıp ve doğrulama kaybı değerlerinde azalma gözlemlenmiştir, modelin öğrenmeye devam ettiğini ve performansının arttığı söylenebilir. Kayıp ve doğrulama kaybı değerleri yaklaşık 0,060 düzeyinde kararlı hale gelmiştir ve modelin tahminlerinin genellikle doğru olduğu ve kabul edilebilir bir hata oranına sahip olduğu görülmektedir.



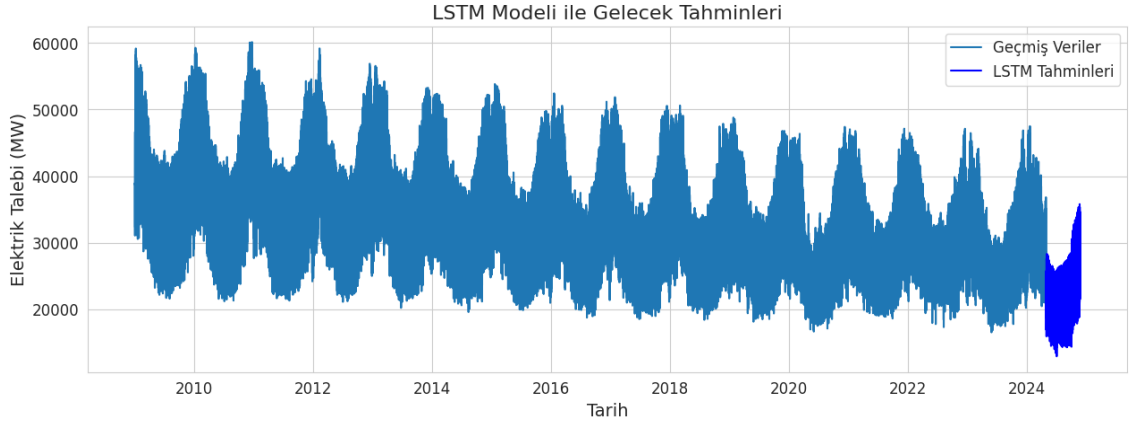
Şekil 4.47. Test verileri ile tahmin edilen değerlerin karşılaştırması (LSTM)

Tahminler ve gerçek değerler arasında belirgin bir uyum görülmekte, ancak bazı sapmalar mevcuttur. Modelin Ortalama Mutlak Yüzde Hatası (MAPE) %7,44 ve Kök Ortalama Kare Hatası (RMSE) 2709,97 MW olarak hesaplanmıştır. Modelin eğitimi toplamda 43 dakika 22 saniye sürmüştür, bu süreçte modelin doğruluğu ve genelleme yeteneği artırılmıştır. Sonuçlar, LSTM modelinin zaman serisi tahminlerinde etkili bir araç olduğunu göstermektedir.

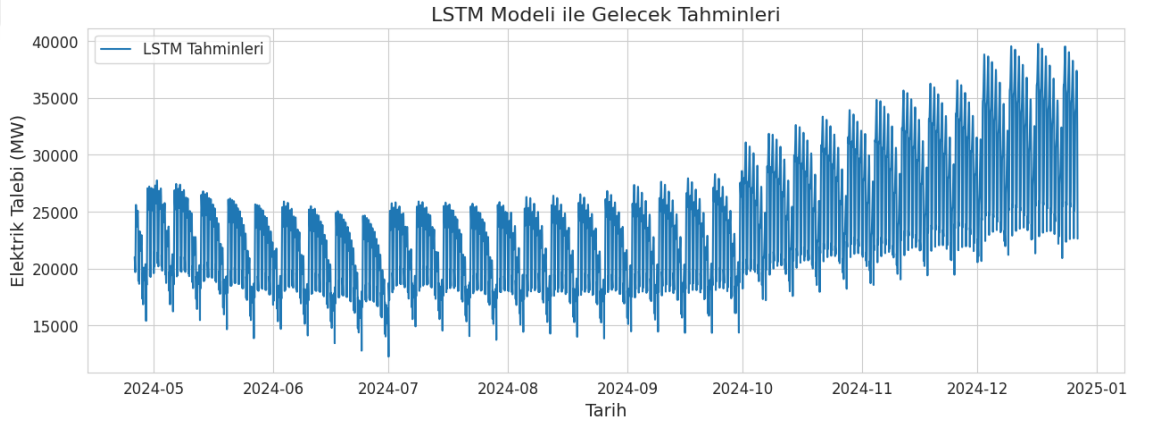


Şekil 4.48. Test verileri ile tahmin edilen değerlerin karşılaştırması- İki haftalık (LSTM)

İki haftalık bir örneklem desenine bakıldığında, tahmin deseninin gerçekleşen genel deseni tanımlayabildiği fakat tepe noktaları temsil eden verilere uzak kaldığı fark edilmiştir. Modelin gelecek talep tahmin grafikleri Şekil 4.49 ve Şekil 4.50'deki gibi gerçekleşmiştir.

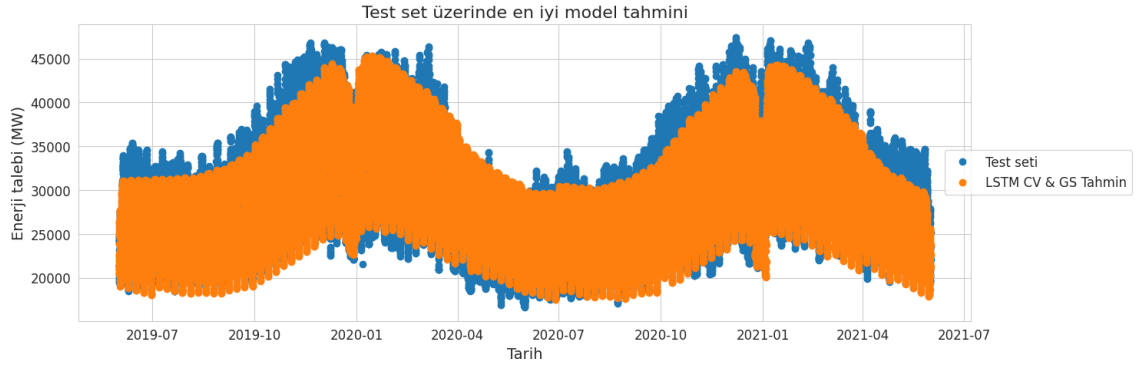


Şekil 4.49. Tarihsel veriler ve Basit LSTM Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri



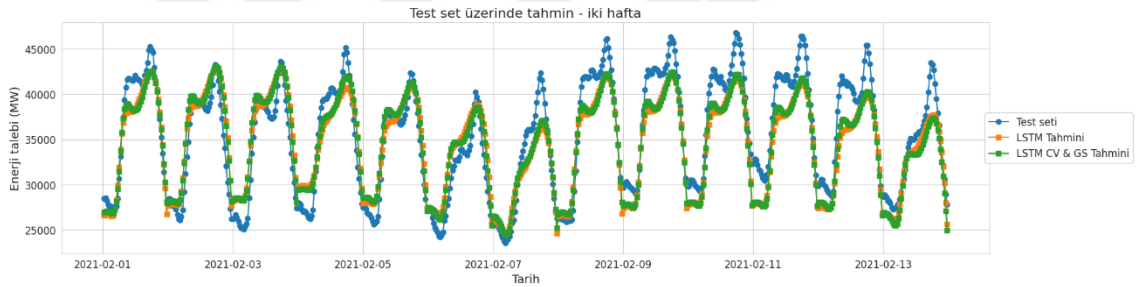
Şekil 4.50. Basit LSTM Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

LSTM tabanlı derin öğrenme modeli için hiperparametre optimizasyonu gerçekleştirilmiştir. Model, Keras kullanılarak oluşturulmuş ve Scikit-learn kütüphanesinin GridSearchCV fonksiyonu ile optimize edilmiştir. Hiperparametre aramaları, `batch_size`, `epochs`, `model__dropout_rate` ve `optimizer` gibi parametreler üzerinde gerçekleştirilmiştir. LSTM modeli, 256 hücreli bir katman, dropout katmanı ve bir çıkış katmanı ile tasarlanmıştır. Çapraz doğrulama için `TimeSeriesSplit` kullanılarak iki dilim oluşturulmuş ve her bir hiperparametre kombinasyonu üzerinde eğitim yapılarak en iyi model belirlenmiştir. En iyi sonuçlar `batch_size` 72, `epochs` 150, `model__dropout_rate` 0.5 ve `optimizer` olarak 'adam' kullanıldığında elde edilmiştir.



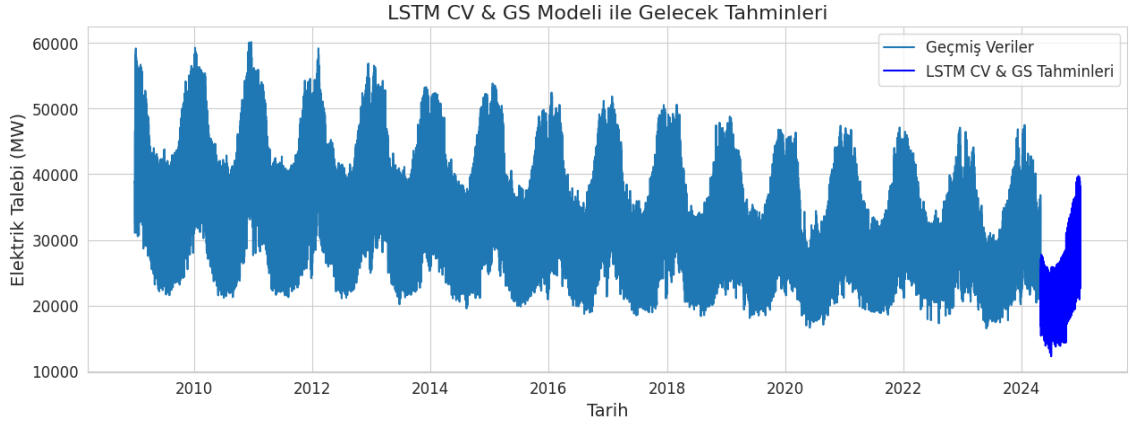
Şekil 4.51. Test verileri ile tahmin edilen değerlerin karşılaştırması (LSTM CV & GS)

Tahminler ve gerçek değerler arasındaki uyum basit modele göre bir miktar artmıştır. Modelin eğitimi ve optimizasyonu toplamda 2 saat 51 dakika 55 saniye sürmüştür, bu sürecin sonunda elde edilen modelin Ortalama Mutlak Yüzde Hatası (MAPE) %7,18 ve Kök Ortalama Kare Hatası (RMSE) 2623,11 MW olarak ölçülmüştür.

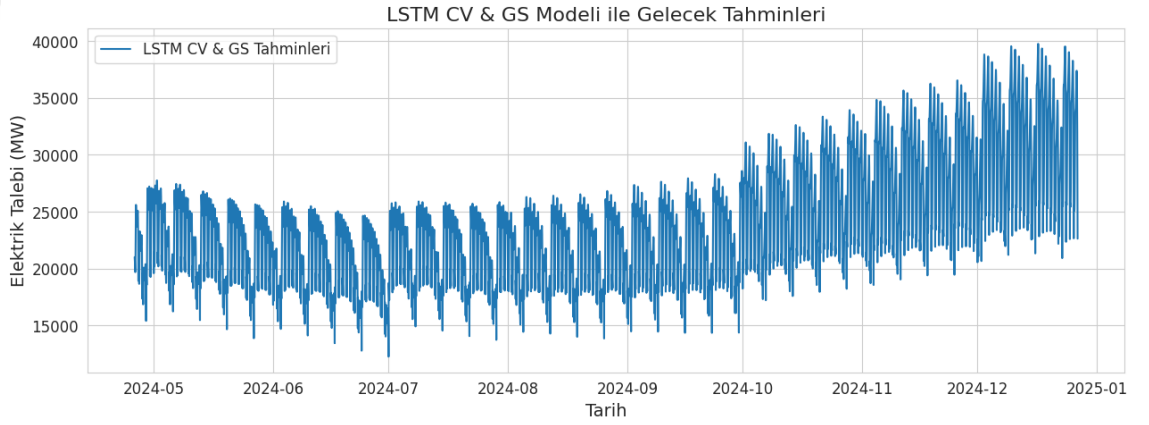


Şekil 4.52. Test verileri ile tahmin edilen değerlerin karşılaştırması- İki haftalık (Basit LSTM vs LSTM CV & GS)

Grafikte, iyileştirilmiş modelin, gerçek verilere daha yakın tahminler yaptığı ve dalgalanmaları daha iyi yakaladığı görülmüştür. Basit model için MAPE değeri %7,44 iken, iyileştirilmiş modelde bu değer %7,18'e düşmüştür. Basit modelin RMSE değeri 2709,97 MW iken, iyileştirilmiş modelde bu değer 2623,11 MW'a düşmüştür. CPU zamanı 43 dakika 22 saniyeden, 2 saat 51 dakika 55 saniyeye çıkmıştır. İyileştirilmiş model daha fazla hesaplama gerektirdiği için bu süre artmıştır. Modelin gelecek talep tahmin grafikleri aşağıdaki gibi gerçekleşmiştir.



Şekil 4.53. Tarihsel veriler ve LSTM CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri



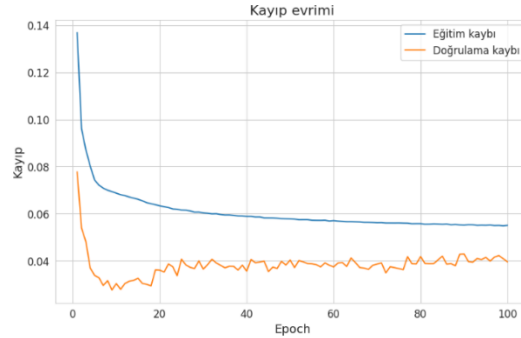
Şekil 4.54. LSTM CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

Sonuçlar, iyileştirilmiş LSTM modelinin enerji talebi tahminlerinde yüksek doğruluk sağladığını göstermektedir. Ancak, bu doğruluğu elde etmek için çok daha fazla hesaplama süresi ve kaynağa ihtiyaç duyulmaktadır. Bu çalışma, hiperparametre optimizasyonunun derin öğrenme modellerinde doğruluğu artırmada kritik bir rol oynadığını ve bu süreçte dikkatli bir dengeleme yapılması gerektiğini ortaya koymaktadır. Daha yüksek doğruluk gerektiren uygulamalar için iyileştirilmiş model tercih edilebilirken, daha hızlı sonuçların gerektiği durumlarda basit model yeterli olabilir.

Derin LSTM, daha iyi tahminler üretmek için birden fazla LSTM biriminin katmanlar halinde üst üste konulduğu bir LSTM varyasyonudur. Derin LSTM, çok katmanlı yapısı

sayesinde hem kısa vadeli hem de uzun vadeli bağımlılıkları daha iyi öğrenme kapasitesine sahiptir. Bu sayede, zaman serisi verilerindeki karmaşık bağımlılıkları daha etkili bir şekilde modelleyebilir ve daha yüksek doğruluk sağlayabilir. Buna karşılık, tepe noktalarının doğru bir şekilde tanımlanabilmesi için modelin daha fazla iyileştirilmesi gerekebilir. Bu iyileştirmeler, hiperparametre optimizasyonu, ek özelliklerin eklenmesi veya daha fazla eğitim verisi kullanılması gibi yöntemlerle gerçekleştirilebilir.

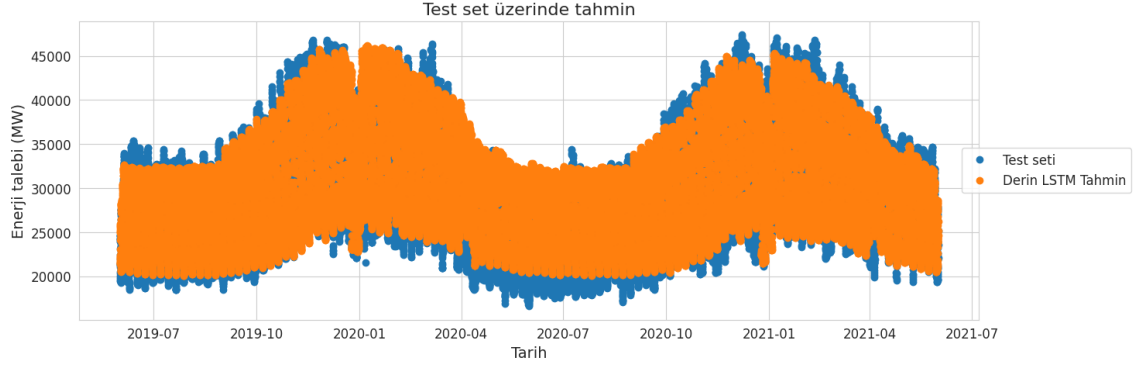
Derin bir LSTM (Uzun Kısa Süreli Bellek) modeli geliştirilmiştir. Eğitim verileri 1 Haziran 2019 tarihinden önceki, test verileri ise 1 Haziran 2019 ile 1 Haziran 2021 tarihleri arasındaki verileri kapsamaktadır. 1 Haziran 2021 sonrası veriler hold-out seti olarak kullanılmıştır. Model, üç katmanlı LSTM yapısından oluşmaktadır: İlk katmanda 256 birim, ikinci katmanda 128 birim ve üçüncü katmanda 32 birim bulunmaktadır. Her bir LSTM katmanından sonra %50 oranında dropout uygulanarak modelin aşırı öğrenmesi engellenmiştir. Modelin çıkış katmanı, tek bir nörondan oluşmaktadır. Kayıp fonksiyonu olarak Kök Ortalama Kare Hatası (RMSE) kullanılmış ve "adam" optimizasyon algoritması ile derlenmiştir. Eğitim sürecinde erken durdurma, model kontrol noktası kaydetme ve öğrenme oranı azaltma gibi geri çağırımlar uygulanmıştır.



Şekil 4.55. Derin LSTM kayıp evrimi

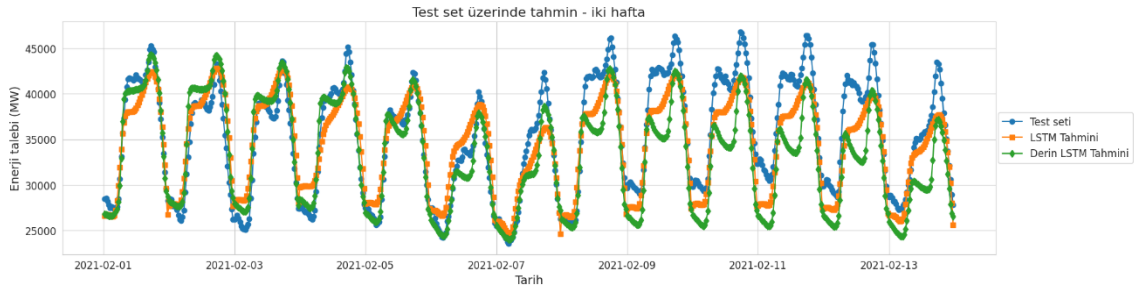
Çok Katmanlı LSTM modelinin eğitim süresi sabit kalmış ve modelin verileri aynı hızda işleyebildiği görülmüştür. Öğrenme hızı 0,0010 olan model, düşük eğitim ve doğrulama kayıpları ile iyi bir genelleme yeteneği sergilemiştir. Eğitim kaybı epoch 10'da 0,1369'dan 0,0687'ye, doğrulama kaybı ise 0,0778'den 0,0304'e düşmüştür. Doğrulama kaybındaki dalgalanmalar, modelin genelleme yeteneğini yansıtır. Doğrulama kaybı artmaya

başladığında, en iyi genelleme yeteneği korunarak erken durdurma stratejisi uygulanmıştır.



Şekil 4.56. Test verileri ile tahmin edilen değerlerin karşılaştırması (Derin LSTM)

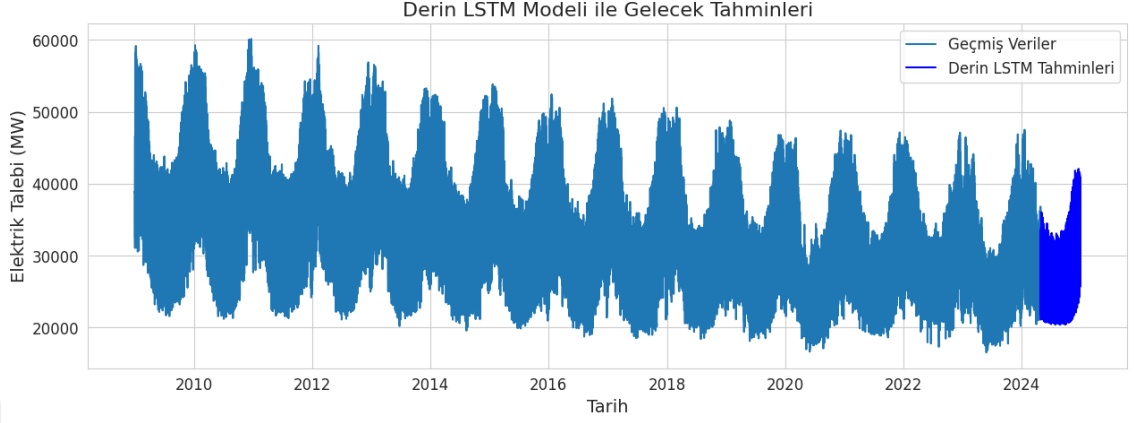
Test ve tahmin deseni incelendiğinde tahminlerde büyük bir iyileşmenin olduğu fakat yine tahminlerde tepe noktaların temsil edilmediği fark edilmiştir. Eğitim sonunda, modelin Ortalama Mutlak Yüzde Hatası (MAPE) %7,16 ve Kök Ortalama Kare Hatası (RMSE) 2597,09 MW olarak hesaplanmıştır. Modelin eğitimi 1 saat 53 dakika 41 saniye sürmüştür.



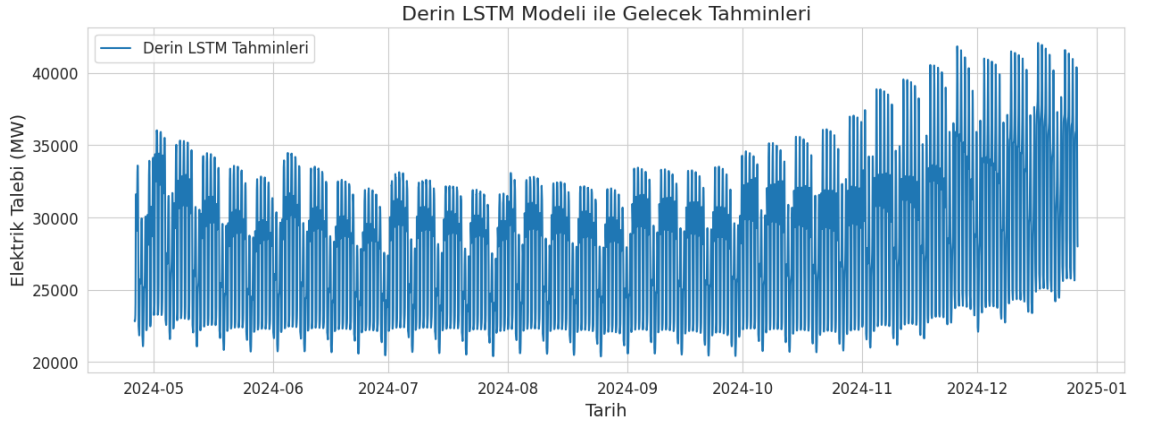
Şekil 4.57. Test verileri ile tahmin edilen değerlerin karşılaştırması- İki haftalık (Derin LSTM vs Basit LSTM)

İki haftalık örnek dönem için, derin LSTM ağının basit LSTM ağı deseni incelendiğinde derin LSTM modelinin daha iyi performans gösterdiği görülür. MAPE değeri %7,44'ten %7,16'ya, RMSE değeri 2709,97'den 2597,09 MW'a iyileşmiştir. Bu derin LSTM modeli, basit LSTM uygulamasına kıyasla daha iyi bir doğruluk ve genelleme yeteneği sunarak, zaman serisi tahminlerinde etkili bir araç olduğunu kanıtlamıştır. Ancak, hesaplama süresi ve maliyetleri göz önünde bulundurularak bu tür derin modellerin

kullanımı dikkatli bir şekilde değerlendirilmelidir. Derin LSTM modelinin gelecek talep tahmin grafikleri Şekil 4.58 ve Şekil 4.59'daki gibi gerçekleşmiştir.



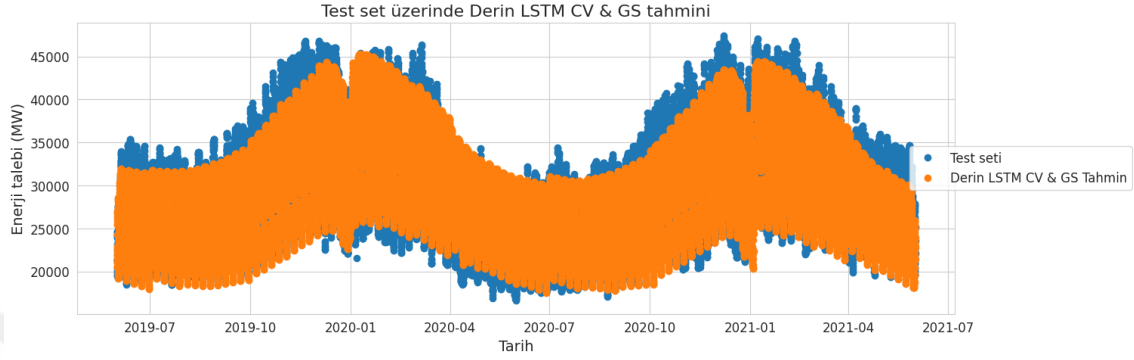
Şekil 4.58. Tarihsel veriler ve Derin LSTM Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri



Şekil 4.59. Derin LSTM Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

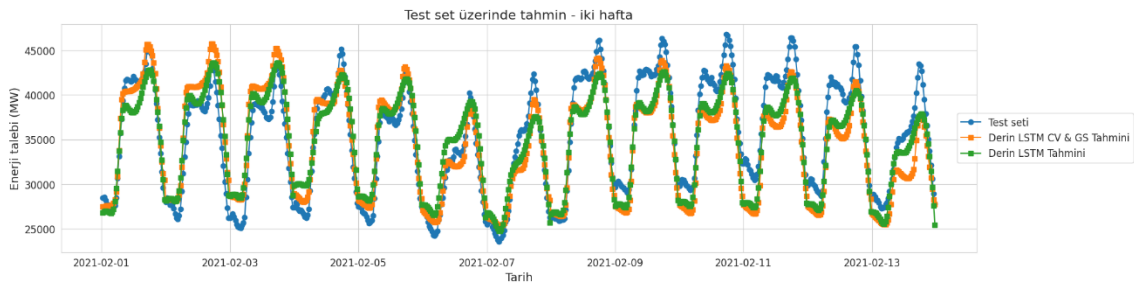
Derin LSTM modeli için hiperparametre optimizasyonu gerçekleştirilmiştir. Model, Scikit-learn'ün GridSearchCV fonksiyonu kullanılarak optimize edilmiştir ve KerasRegressor ile sarılmıştır. Modelin yapısı üç katmanlı bir LSTM mimarisi üzerine kurulmuştur: İlk katmanda 256 birim, ikinci katmanda 128 birim ve son katmanda 32 birim bulunmaktadır. Her bir LSTM katmanından sonra dropout uygulanarak modelin aşırı öğrenmesi engellenmiştir. Modelin kayıp fonksiyonu olarak Kök Ortalama Kare Hatası (RMSE) kullanılmış ve "adam" ve "rmsprop" optimizasyon algoritmaları test edilmiştir. Hiperparametre optimizasyonu için, model__dropout_rate, batch_size, epochs

ve optimizasyon parametreleri üzerinde bir arama gerçekleştirilmiştir. Çapraz doğrulama için TimeSeriesSplit kullanılarak iki dilim oluşturulmuştur. En iyi sonuçlar, batch_size 72, epochs 150, model__dropout_rate 0.5 ve optimizasyon olarak 'adam' kullanıldığında elde edilmiştir.



Şekil 4.60. Test verileri ile tahmin edilen değerlerin karşılaştırması (Derin LSTM CV & GS)

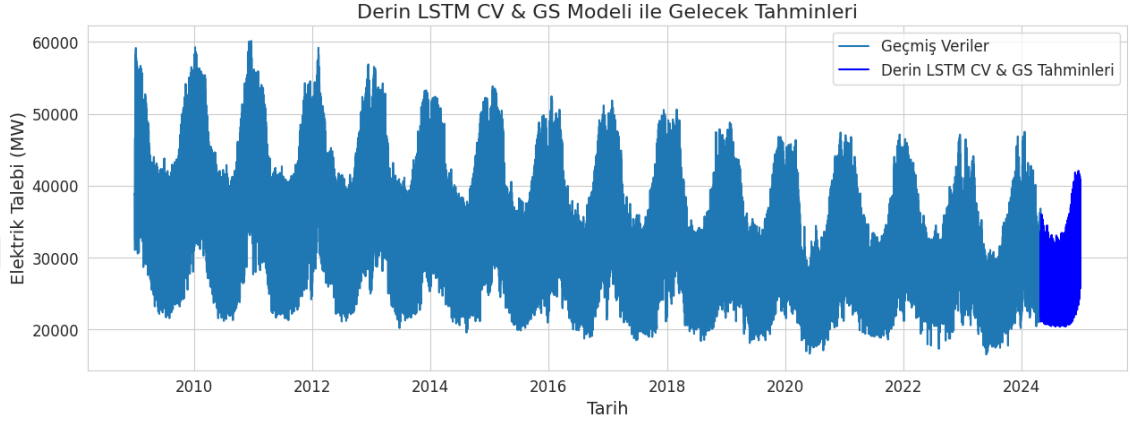
Tahminler genel olarak gerçek değerlere oldukça yakın ve tahminlerin varyansı daha düşük görünmekte, sıkı bir uyum gözlenmekte; bu da basit Derin LSTM modeline göre daha stabil bir model performansı olduğunu göstermektedir. En iyi parametrelerle eğitilen modelin Ortalama Mutlak Yüzde Hatası (MAPE) %7,13 ve Kök Ortalama Kare Hatası (RMSE) 2618.00 MW olarak ölçülmüştür. Modelin eğitimi ve optimizasyonu toplamda 4 saat 55 dakika 19 saniye sürmüştür.



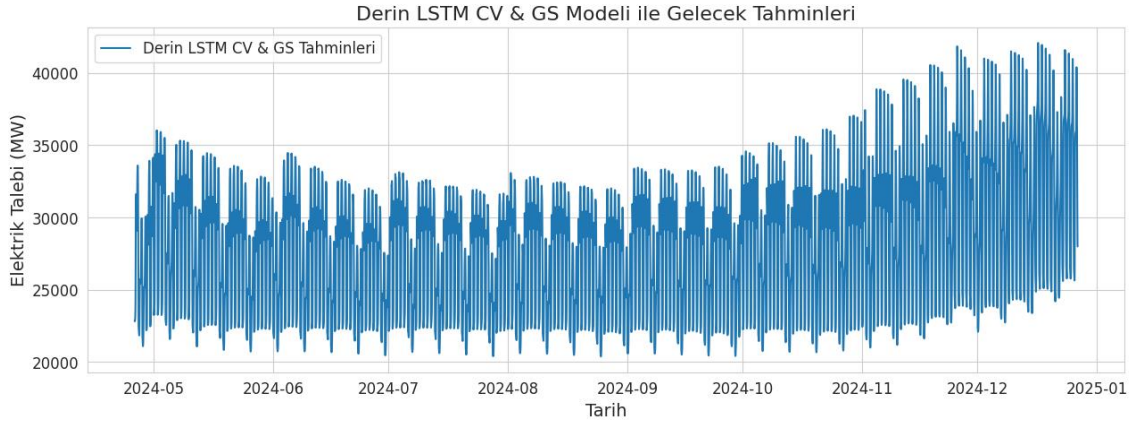
Şekil 4.61. Test verileri ile tahmin edilen değerlerin karşılaştırması- İki haftalık (Derin LSTM)

Derin LSTM basit modelinin MAPE değeri %7,16 iken, optimize edilmiş modelde bu değer %7,13'e düşmüştür, ancak fark oldukça küçüktür. RMSE değeri de basit modelde 2597,09 MW iken optimize edilmiş modelde 2618 MW olmuştur. Basit modelin CPU

süresi 1 saat 53 dakika 41 saniye iken, optimize edilmiş modelin süresi 4 saat 55 dakika 19 saniyedir. Optimize edilmiş model, çok küçük bir MAPE avantajı sağlarken, daha fazla hesaplama süresi ve kaynak gerektirmiştir. Bu nedenle, benzer performansla daha az kaynak gerektiren basit model tercih edilebilir veya farklı hiperparametre aramaları yapılabilir. Optimize edilmiş modelin tahmin grafikleri Şekil 4.62 ve Şekil 4.63'te gösterilmiştir.



Şekil 4.62. Tarihsel veriler ve Derin LSTM CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri



Şekil 4.63. Derin LSTM CV & GS Modeli 2024 yılı Mayıs-Aralık dönemi için gelecek talep tahminleri

Genel olarak, LSTM modelleri, enerji tüketimi tahminlerinde makul ve güvenilir sonuçlar üretebilmektedir. Buna karşılık, daha düşük hata oranları elde etmek için modelin ve hiperparametrelerin iyileştirilmesi veya daha karmaşık modellerin değerlendirilmesi düşünülebilir.

5. TARTIŞMA VE SONUÇ

Bu tez çalışmasında, Birleşik Krallık'ın National Grid ESO enerji sistem operatörü tarafından sağlanan ve 2009-2024 yılları arasındaki yarım saatlik ortalama MW tüketim verilerini içeren bir zaman serisi analizi gerçekleştirilmiştir. Çalışmanın amacı, 2024 yılı Mayıs-Aralık ayları arasındaki enerji tüketimini tahmin etmek, farklı makine öğrenimi ve zaman serisi modellerinin performanslarını karşılaştırarak en uygun modeli belirlemek ve iyileştirme önerileri sunmaktır. Doğru tahminler, enerji üretim ve dağıtımını en verimli şekilde planlamak, enerji israfını önlemek, maliyetleri azaltmak ve enerji güvenliğini sağlamak açısından kritik öneme sahiptir. Bu nedenle, çeşitli algoritmaların karşılaştırılması, tahmin doğruluğunu artırmak ve bu algoritmaların güçlü ve zayıf yönlerini belirlemek açısından önem taşımaktadır. Bu süreç, kullanıcıların ihtiyaçlarına göre en uygun modeli seçmelerine olanak tanır.

Çalışmada SARIMA, Prophet, XGBoost, Lineer Arttırma Regresyonu ve LSTM gibi modeller kullanılmış ve bu modeller, Ortalama Mutlak Yüzde Hatası (MAPE), Kök Ortalama Kare Hatası (RMSE), CPU zamanı ve duvar saati zamanı gibi metriklerle değerlendirilmiştir. Elde edilen sonuçlar, her modelin doğruluk, hız ve hesaplama maliyeti açısından farklı avantajlar sunduğunu ortaya koymuştur. Python programlama dili ve Google Colab platformu kullanılarak yapılan uygulamalar, derin öğrenme tabanlı yöntemlerin daha yüksek doğruluk sağladığını göstermiştir.

Veri hazırlama aşamasında, veri seti temizlenmiş ve yeni özellikler eklenmiştir. Model eğitimi aşamasında, geçmiş tüketim verileri kullanılarak modeller eğitilmiş ve gelecekteki elektrik talebini tahmin etmeleri sağlanmıştır. Tahmin aşamasında, eğitilen modeller belirli bir dönem için elektrik talebini tahmin etmek amacıyla kullanılmış ve tahminler oluşturulmuştur. Model değerlendirme aşamasında, modellerin performansı MAPE ve RMSE başarı metrikleri ile ölçülmüştür. Çalışmanın sonuçları, 11. Nesil Intel(R) Core(TM) i7-11850H @ 2.50GHz işlemci ve 32.0 GB RAM özelliklerine sahip bir bilgisayar kullanılarak elde edilmiştir.

Çalışmada kullanılan modellerin MAPE ve RMSE performans karşılaştırması aşağıdaki Çizelge 5.1’de verilmektedir:

Çizelge 5.1. Modellerin hata ölçü karşılaştırması

Metric	LSTM Basit	LSTM CV & GS	Derin LSTM	Derin LSTM CV & GS	XGBoost Basit	XGBoost CV & GS	Lineer Arttırma Regresyonu Basit	Lineer Arttırma Regresyonu CV & GS	Prophet Basit	Prophet Tatilli	Prophet CV & GS	SARIMA Model0	SARIMA ModelI	SARIMA Model2
MAPE	7.44	7.18	7.16	7.13	11.29	7.29	8.2	7.73	9.37	9.34	9.34	32.32	13.02	17.59
RMSE	2709.97	2623.11	2597.09	2618	3786.88	2655.08	2963.08	2810	3259.96	3236.53	3235.91			
CPU Time	43dk 22sn	2s 51dk 55sn	1s 53dk 41sn	4s 55dk 19sn	19.7 sn	13dk 51sn	29dk 53sn	1s 22dk 51sn	30.5 sn	37.3 sn	3dk 32sn	6dk 12sn	8dk 21sn	1dk 50sn
Walltime	32dk 34sn	2s 38dk 2sn	1s 45dk 29sn	5s 2dk 18sn	12 sn	8dk 7sn	19dk 20sn	1s 5dk 18sn	4dk 12sn	5dk 3sn	22dk 40sn	4dk 48sn	6dk 18sn	1dk 20sn

Çizelge 5.1'de görüldüğü üzere, LSTM ve Derin LSTM modelleri, en düşük MAPE ve RMSE değerleriyle zaman serisi tahmini için en uygun seçenekler olarak öne çıkmaktadır.

Derin öğrenme tabanlı LSTM modelleri, özellikle %7,16 MAPE ve 2597,09 MW RMSE ile Derin LSTM modeli ve %7,13 MAPE ve 2618 MW RMSE ile Derin LSTM CV & GS modeli, en yüksek doğruluk oranlarına sahiptir. Ancak, bu modellerin uzun eğitim süreleri ve yüksek hesaplama kaynakları gereksinimleri, kaynak yönetiminin dikkatlice yapılmasını zorunlu kılmaktadır. Tek katmanlı LSTM modeli de oldukça iyi bir performans sergilemiş olup, MAPE değeri %7,44 ve RMSE değeri 2709,97 MW olarak ölçülmüştür. Hiperparametre optimizasyonu yapılan modelde ise MAPE değeri %7,18 ve RMSE değeri 2623,11 MW olarak gerçekleşmiştir.

Basit XGBoost modeli, en kısa CPU ve işlem süresine sahip olmasına rağmen, %11,29 MAPE ve 3786,88 MW RMSE ile yüksek hata oranları göstermiştir. XGBoost CV & GS modeli ise %7,29 MAPE ve 2655,08 RMSE ile iyi bir denge sağlamaktadır. Lineer Arttırma Regresyonu Basit modeli, %8,2 MAPE ile diğer modellere göre daha düşük doğruluk sergilemiştir. Ancak, hiperparametre optimizasyonu yapılmış model %7,73 MAPE değeri ile iyileşme göstermiştir ve RMSE değeri de 2963,08 MW'tan 2810 MW'a düşmüştür.

SARIMA modelleri, zaman serisi karmaşıklığı ile mücadelede zorlanmış ve güvenilirlik sorunları yaşamıştır; SARIMA Model0 %32,32 MAPE ile en yüksek hata oranına

sahiptir. Prophet modeli ise tatil etkilerini hesaba katarak nispeten daha iyi bir performans göstermiş, ancak doğruluk oranları diğer modellere göre daha düşük kalmıştır. Prophet modelinin performansı, daha fazla parametre eklenerek geliştirilebilir. Prophet Basit, Prophet Tatilli ve "Prophet CV & GS" modelleri sırasıyla %9,37, %9,34 ve %9,34 MAPE değerleriyle daha düşük hata oranları sunmuştur. RMSE açısından Prophet modelleri 3259,96 MW, 3236,53 MW ve 3235,91 MW değerleriyle performans göstermiştir. CPU ve işlem süreleri açısından Prophet modelleri XGBoost'tan sonra en hızlı sonuçları sunarken, SARIMA modelleri daha uzun işlem süreleri gerektirmiştir (Trull ve diğerleri, 2024, Wang ve diğerleri, 2023).

Sonuç olarak, model seçimi uygulamanın gereksinimlerine bağlı olarak yapılmalıdır. Yüksek doğruluk talep eden uygulamalar için daha uzun eğitim sürelerine katlanmak gerekebilir; bu durumda derin öğrenme modelleri tercih edilebilir. Öte yandan, hızlı sonuç gerektiren durumlarda hızlı modeller uygun olabilir. Veri yapısı ve kaynak kısıtlamaları göz önünde bulundurularak model seçimi dikkatli bir şekilde yapılmalıdır.

Çalışmaya, veri ön işleme ve özellik mühendisliği konularında katkılar sağlanabilir; diğer özelliklerin talep tahminindeki etkileri test edilebilir ve hava durumu, piyasa koşulları gibi dış faktörler modele eklenebilir. Model optimizasyonu için ızgara arama yerine Random Search, Bayesian Optimization gibi diğer teknikler denenebilir. Modellerin hiperparametreleri daha geniş bir aralıkta ve detaylı bir şekilde optimize edilebilir. Mevsimsellik ve trendlerin daha iyi anlaşılması için SARIMA, LSTM ve Prophet modellerinin bir arada kullanıldığı hibrit modeller geliştirilebilir. LSTM modellerinde daha fazla zaman adımı kullanarak modelin daha karmaşık kalıpları öğrenmesi sağlanabilir ve bu da tahmin doğruluğunu artırabilir.

Birleşik Krallık'taki elektrik talebi tahminlemesi ile ilgili olarak, bu çalışmanın sonuçları, enerji talebi tahmininde kullanılan modellerin performansını anlamak ve iyileştirmek için değerli bilgiler sunmaktadır. Özellikle LSTM ve Derin LSTM modelleri, yüksek doğruluk oranları ile öne çıkarken, XGBoost ve Prophet modelleri de dikkate değer performans göstermiştir. Sonuç olarak, bu çalışma, elektrik talebi tahmininde daha doğru ve güvenilir modeller geliştirmek için önemli bir rehber sunmaktadır.

KAYNAKLAR

- Abbasimehr, H., Behboodi, A., Bahrini, A., (2024). A novel hybrid model to forecast seasonal and chaotic time series. *Expert Systems with Applications*, 239: 122461. ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2023.122461>.
- Abbasimehr, H., Paki, R., Bahrini, A., (2023). A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models. *Sustainable Computing: Informatics and Systems*, 38: 100863. ISSN 2210-5379, <https://doi.org/10.1016/j.suscom.2023.100863>.
- Akgül, S. (2013). Yapay Sinir Ağları ile Türkiye Elektrik Enerjisi Talep Tahmini ve Diğer Yöntemlerle Karşılaştırmalı Analizi.[Yayınlanmamış Doktora Tezi]. Atatürk Üniversitesi.
<https://atauni.edu.tr/yuklemeler/f7ecbba8b3a5e802f31e56621fa3f7af.pdf>
- Akman, T., Yılmaz, C., Sönmez, Y. (2018). Elektrik Yüğü Tahmin Yöntemlerinin Analizi. *Gazi Mühendislik Bilimleri Dergisi*, 4(3): 168-175. <https://doi.org/10.30855/GJES.2018.04.03.003>
- Aldarraj, M., Vega-Márquez, B., Pontes, B., Mahmood, B., Riquelme, J.C., (2024). Addressing energy challenges in Iraq: Forecasting power supply and demand using artificial intelligence models. *Heliyon*, 10(4): e25821. ISSN 2405-8440, <https://doi.org/10.1016/j.heliyon.2024.e25821>.
- Arıkök, Y.A. (2024). A Comparative Study of ARIMA and LSTM for Two-Days Ahead Forecasting of Electricity Demand in İzmir-Manisa Region in Turkey. [Yayınlanmamış Yüksek Lisans Tezi]. Yaşar University
- Ayele, S.T., Ageze, M.B., Zeleke, M.A., Miliket, T.A., (2023). Adama II wind farm long-term power generation forecasting based on machine learning models. *Scientific African*, 21: e01831. ISSN 2468-2276, <https://doi.org/10.1016/j.sciaf.2023.e01831>.
- Bashir, T., Haoyong, C., Tahir, M. F., Liqiang, Z. (2022). Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. *Energy Reports*, 8: 1678-1686. ISSN 2352-4847, <https://doi.org/10.1016/j.egyr.2021.12.067>.
- Bashir, T., Haoyong, C., Tahir, M.F., Liqiang, Z., (2022) Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN, *Energy Reports*, Volume 8, Pages 1678-1686, ISSN 2352-4847, <https://doi.org/10.1016/j.egyr.2021.12.067>.
- Box, G.E.P., Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. *Journal of the American Statistical Association*, 68(342): 199-201. ISBN 978-1-118-67502-1
- Bozlak, Ç.B., Yaşar, C.F. (2024). An optimized deep learning approach for forecasting day-ahead electricity prices. *Electric Power Systems Research*, 229: 110129. <https://doi.org/10.1016/j.epsr.2024.110129>
- Chaturvedi, S., Rajasekar, E., Natarajan, S., McCullen, N., (2022). A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. *Energy Policy*, 168: 113097. ISSN 0301-4215, <https://doi.org/10.1016/j.enpol.2022.113097>.
- Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In: *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. <https://doi.org/10.1145/2939672.293978>

- Deng, X., Ye, A., Zhong, J., Xu, D., Yang, W., Song, Z., Zhang, Z., Guo, J., Wang, T., Tian, Y., Pan, H., Zhang, Z., Wang, H., Wu, C., Shao, J., Chen, X., (2022). Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model. *Energy Reports*, 8: 8661-8674. ISSN 2352-4847, <https://doi.org/10.1016/j.egyr.2022.06.072>.
- El-Azab, H.I., Swief, R.A., El-Amary, N.H., Temraz, H.K. (2024). Machine and deep learning approaches for forecasting electricity price and energy load assessment on real datasets. *Ain Shams Engineering Journal*, 15(4): 102613. <https://doi.org/10.1016/j.asej.2023.102613>
- Ensafi, Y., Amin, S.H., Zhang, G., Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*, 2(1): 100058. <https://doi.org/10.1016/j.jjimei.2022.100058>
- Fan, G.F., Yu, M., Dong, S.Q., Yeh, Y.H., Hong, W.C. (2021). Forecasting short-term electricity load using hybrid support vector regression with grey catastrophe and random forest modeling. *Utilities Policy*, 73: 101294. <https://doi.org/10.1016/j.jup.2021.101294>
- Galit, S., Kenneth, C.L.Jr., (2016), *Practical Time Series Forecasting with R*. 2016
- Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S. (2024). Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting. *Applied Energy*, 353(Part A): 122059. <https://doi.org/10.1016/j.apenergy.2023.122059>
- Ghimire, S., Deo, R.C., Casillas-Pérez, D., Salcedo-Sanz, S., Pourmousavi, S.A., Acharya, U.R. (2024). Probabilistic-based electricity demand forecasting with hybrid convolutional neural network-extreme learning machine model. *Engineering Applications of Artificial Intelligence*, 132: 107918. <https://doi.org/10.1016/j.engappai.2024.107918>
- González Grandón, T., Schwenzer, J., Steens, T., Breuing, J. (2024). Electricity demand forecasting with hybrid classical statistical and machine learning algorithms: Case study of Ukraine. *Applied Energy*, 355: 122249. <https://doi.org/10.1016/j.apenergy.2023.122249>
- Goswami, K., Kandali, A.B. (2020). Electricity Demand Prediction using Data Driven Forecasting Scheme: ARIMA and SARIMA for Real-Time Load Data of Assam. Dibrugarh University. <https://doi.org/10.1109/ComPE49325.2020.9200031>
- Gu, L., Wang, J., Liu, J. (2024). A combined system based on data preprocessing and optimization algorithm for electricity load forecasting. *Computers & Industrial Engineering*. <https://doi.org/10.1016/j.cie.2024.110114>
- Gülay, E., Sen, M., Akgün, Ö.B. (2024). Forecasting electricity production from various energy sources in Türkiye: A predictive analysis of time series, deep learning, and hybrid models. *Energy*, 286: 129566. <https://doi.org/10.1016/j.energy.2023.129566>
- Guo, L., Fang, W., Zhao, Q., Wang, X., (2021). The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality. *Computers & Industrial Engineering*, 161: 107598. ISSN 0360-8352, <https://doi.org/10.1016/j.cie.2021.107598>.
- Güven, D., Kayalica, M.Ö. (2023). Analysing the determinants of the Turkish household electricity consumption using gradient boosting regression tree. *Energy for Sustainable Development*, 77: 101312. <https://doi.org/10.1016/j.esd.2023.101312>

- Huang, Y., Bai, Y., Yu, Q., Ding, L., Ma, Y., (2022). Application of a hybrid model based on the Prophet model, ICEEMDAN and multi-model optimization error correction in metal price prediction. *Resources Policy*, 79: 102969. ISSN 0301-4207, <https://doi.org/10.1016/j.resourpol.2022.102969>.
- Hussein, A., Awad, M. (2024). Time series forecasting of electricity consumption using hybrid model of recurrent neural networks and genetic algorithms. *Measurement: Energy*, 2: 100004. <https://doi.org/10.1016/j.meaeene.2024.100004>
- Jin, H., Guo, J., Tang, L., Du, P. (2024). Long-term electricity demand forecasting under low-carbon energy transition: Based on the bidirectional feedback between power demand and generation mix. *Energy*, 286: 129435. <https://doi.org/10.1016/j.energy.2023.129435>
- Khan, Z. A., Ullah, A., Haq, I. U., Hamdy, M., Mauro, G. M., Khan, M., Hijji, M., Baik, S. W. (2022). Efficient Short-Term Electricity Load Forecasting for Effective Energy Management. *Sustainable Energy Technologies and Assessments*, 53(Part A): 102337. ISSN 2213-1388, <https://doi.org/10.1016/j.seta.2022.102337>.
- Kılınc, E.C. (2023). Yenilenebilir Enerji ve Fosil Yakıt Tüketiminin Ekolojik Ayak İzi Üzerindeki Etkisi: Türkiye Örneği. *FESA Dergisi*, 8(3): 731-749. DOI: 10.29106/fesa.1307807
- Küçük, S. (2022). Elektriğin Tarihi. *Sosyal Bilimler Dergisi / The Journal of Social Sciences, Akademik Sosyal Araştırmalar Dergisi*, 9(56): 38-55. ISSN: 2149-0821. <http://dx.doi.org/10.29228/SOBIDER.57290>.
- Le, X.H., Ho, H.V., Lee, G., Jung, S. (2019). Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water*, 11: 1387. <https://doi.org/10.3390/w11071387>
- Loizidis, S., Kyprianou, A., Georghiou, G.E. (2024). Electricity market price forecasting using ELM and Bootstrap analysis: A case study of the German and Finnish Day-Ahead markets. *Applied Energy*, 363: 123058. <https://doi.org/10.1016/j.apenergy.2024.123058>
- Lu, Y., Sheng, B., Fu, G., Luo, R., Chen, G., Huang, Y., (2023). Prophet-EEMD-LSTM based method for predicting energy consumption in the paint workshop. *Applied Soft Computing*, 143: 110447. ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2023.110447>.
- Mateo-Barcos, S., Ribó-Pérez, D., Rodríguez-García, J., Alcázar-Ortega, M. (2024). Forecasting electricity demand of municipalities through artificial neural networks and metered supply point classification. *Energy Reports*, 11: 3533-3549. <https://doi.org/10.1016/j.egyr.2024.03.023>
- Mikayilov, J.I., Alyamani, R., Darandary, A., Javid, M., Hasanov, F.J. (2023). Modeling and forecasting industrial electricity demand for Saudi Arabia: Uncovering regional characteristics. *The Electricity Journal*, 36(8): 107331. <https://doi.org/10.1016/j.tej.2023.107331>
- Mosavi, A., Ardabili, F., Shamshirband, S. (2019). Demand Prediction with Machine Learning Models; State of the Art and a Systematic Review of Advances. <http://dx.doi.org/10.20944/preprints201905.0175.v1>
- Nofal, S., (2023). Forecasting next-hour electricity demand in small-scale territories: Evidence from Jordan. *Heliyon*, 9(9): e19790. ISSN 2405-8440, <https://doi.org/10.1016/j.heliyon.2023.e19790>.
- Rao, C., Zhang, Y., Wen, J., Xiao, X., Goh, M. (2023). Energy demand forecasting in China: A support vector regression-compositional data second exponential

- smoothing model. *Energy*, 263(Part C): 125955. ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2022.125955>.
- Rezaei, N., Rajabi, R., Estebarsari, A. (2022). Electricity Price Forecasting Model based on Gated Recurrent Units. 2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe, IEEEIC / I and CPS Europe 2022. <https://doi.org/10.1109/IEEEIC/ICPSEurope54979.2022.9854746>
- Rick, R., Berton, L., (2022). Energy forecasting model based on CNN-LSTM-AE for many time series with unequal lengths. *Engineering Applications of Artificial Intelligence*, 113: 104998. ISSN 0952-1976, <https://doi.org/10.1016/j.engappai.2022.104998>.
- Sarwar, S., Aziz, G., Tiwari, A.K. (2023). Implication of machine learning techniques to forecast the electricity price and carbon emission: Evidence from a hot region. *Geoscience Frontiers*, 2023: 101647. <https://doi.org/10.1016/j.gsf.2023.101647>
- Su, X., Yan, X., Tsai, C-L. (2012). Linear regression. *WIREs Comput Stat*, 4: 275-294. <https://doi.org/10.1002/wics.1198>.
- Tarkun, S. (2023). Elektrik Talebinin Zaman Serisi Analizi, Yapay Sinir Ağları ve Hibrit Yöntem ile Tahmini. [Yayınlanmamış Doktora Tezi]. Bursa Uludağ Üniversitesi
- Trull, O., García-Díaz, J. C., Peiró-Signes, A., (2024). msf, a forecasting library to predict short-term electricity demand based on multiple seasonal time series. *Journal of Computational Science*, 78: 102280. ISSN 1877-7503, <https://doi.org/10.1016/j.jocs.2024.102280>.
- Valbuena Godoy, J.N. (2022). Demand forecasting of Fast-Moving Consumer Goods based on modeling of time series and deep learning methods. [Unpublished master dissertation], Universidad Central/Universidad Jorge Tadeo Lozano.
- Valencia-Cárdenas, M., Díaz-Serna, F.J., Correa-Morales, J.C. (2016). Multi-product Inventory Modeling with Demand Forecasting and Bayesian Optimization. *DYNA*, 83(198): 97-104. <https://doi.org/10.15446/dyna.v83n198.51310>
- Waheed, W., Xu, Q., (2024). The COVID-19 pandemic's impact on user consumption: A shift towards very efficient load forecasting. *Sustainable Energy, Grids and Networks*, 39: 101426. ISSN 2352-4677, <https://doi.org/10.1016/j.segan.2024.101426>.
- Wang, C., Li, Z., Ni, X., Shi, W., Zhang, J., Bian, J., Liu, Y., (2023). Residential water and energy consumption prediction at hourly resolution based on a hybrid machine learning approach. *Water Research*, 246: 120733. ISSN 0043-1354, <https://doi.org/10.1016/j.watres.2023.120733>.
- Wang, Y., Sun, S., Chen, X., Zeng, X., Kong, Y., Chen, J., Guo, Y., Wang, T., (2021). Short-term load forecasting of industrial customers based on SVM and XGBoost. *International Journal of Electrical Power & Energy Systems*, 129: 106830. ISSN 0142-0615, <https://doi.org/10.1016/j.ijepes.2021.106830>.
- Xia, Y., Wang, J., Wei, D., Zhang, Z. (2023). Combined framework based on data preprocessing and multi-objective optimizer for electricity load forecasting. *Engineering Applications of Artificial Intelligence*, 119: 105776. <https://doi.org/10.1016/j.engappai.2022.105776>
- Zaraket, K., Harb, H., Bennis, I., Jaber, A., Abouaissa, A., (2024). Hyper-Flophet: A neural Prophet-based model for traffic flow forecasting in transportation systems. *Simulation Modelling Practice and Theory*, 134: 102954. ISSN 1569-190X, <https://doi.org/10.1016/j.simpat.2024.102954>.

- Zhang, B., Li, N., Shi, F., Law, R. (2020). A deep learning approach for daily tourist flow forecasting with consumer search data. *Asia Pacific Journal of Tourism Research*, 25(3): 323-339. <https://doi.org/10.1080/10941665.2019.1709876>
- Zhang, C., Ma, L., Luo, Z., Han, X., Zhao, T., (2024). Forecasting building plug load electricity consumption employing occupant-building interaction input features and bidirectional LSTM with improved swarm intelligent algorithms. *Energy*, 288: 129651. ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2023.129651>.
- Zhang, L., Janosik, D., (2024). Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches. *Expert Systems with Applications*, 241: 122686. ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2023.122686>.
- Zhang, S., Chen, R., Cao, J., Tan, J. (2023). A CNN and LSTM-based multi-task learning architecture for short and medium-term electricity load forecasting. <https://doi.org/10.1016/j.epsr.2023.109507>
- Vidal, A. (2023, Kasım 12) <https://www.kaggle.com/code/albertovidalrod/uk-electricity-consumption-prediction-time-series/notebook>
- Ozturk, M. (2024, 03 Ağustos). <https://miracozturk.com/capraz-dogrulama-teknikleri-cross-validation/>
- Avcil, B. (2024, Nisan 6). <https://medium.com/@betulavcil/ke%C5%9Fi%CC%87fsel-veri%CC%87-anali%CC%87zi%CC%87-eda-exploratory-data-analysis-f620cfd12420>
- Artley, B. (2024, Temmuz 12). <https://towardsdatascience.com/time-series-forecasting-with-arima-sarima-and-sarimax-ee61099e78f6>
- Independent Statistics and Analysis. (2024, Nisan). <https://www.iea.org/energy-system/electricity>
- National Energy System Operator. (2024, Nisan). <https://www.nationalgrideso.com/data-portal/daily-demand-update>
- Met Office. (2024, Temmuz 2024) <https://www.metoffice.gov.uk/research/climate/maps-and-data/historic-station-data>
- Duca, A. (2024, Mayıs 11). <https://towardsdatascience.com/4-different-approaches-for-time-series-analysis-7e2364fadcb9>
- Ilıç, I.,Görgülü, B., Cevik, M., Baydoğan, M.G., (2021). Explainable boosted linear regression for time series forecasting, <https://doi.org/10.1016/j.patcog.2021.108144>.
- Tasasiz, E. (2024, Ağustos 3). <https://medium.com/@edatasasiz/k-fold-cross-validation-nedir-ab2107fce616>
- (2024, Ekim 23). <https://www.gov.uk/government/statistics/energy-consumption-in-the-uk-2024>
- Gavin,C.(2024,Ekim23).https://assets.publishing.service.gov.uk/media/5a7ba36540f0b645ba3c59b2/Seasonal_variations_in_electricity_demand.pdf
- Unofficed.com (2024, Aralık 6) <https://unofficed.com/courses/newton-week-3/lessons/augmented-dickey-fuller-test/>
- Cerliani, M. (2024, Aralık 6) <https://pypi.org/project/linear-tree/>