



Hava durumu verilerinin kısa dönem elektrik yük tahmini üzerine etkileri The effects of weather data on short-term electricity load forecasting

Murat Ünlü^{1*} , Habil Ergen¹ 

¹Elektrik Mühendisliği Bölümü, Mühendislik Fakültesi, Kocaeli Üniversitesi, Kocaeli, Türkiye.
muratunlu@kocaeli.edu.tr, habilergen@gmail.com

Geliş Tarihi/Received: 24.11.2024
Kabul Tarihi/Accepted: 05.06.2025

Düzeltilme Tarihi/Revision: 19.04.2025

doi: 10.5505/pajes.2025.38921
Araştırma Makalesi/Research Article

Öz

Elektrik tüketim talebinin doğru bir şekilde tahmin edilmesi, enerji şirketleri için planlama ve kaynak yönetimi açısından büyük bir öneme sahiptir ve bu nedenle, hava durumunun elektrik yük tahminine olan etkilerinin tam olarak belirlenmesi gerekmektedir. Bu çalışmada, hava durumunun kısa dönem elektrik yük tahmini üzerindeki etkileri incelenmiştir. Catboost, LightGBM ve XGBoost gibi güçlü makine öğrenmesi modelleri ile geçmiş yük verileri, gün özellikleri ve tatil günleri gibi bağımsız değişkenlerle birleştirilerek elektrik yük tahmini yapılmıştır. İlk olarak hava durumu verileri hariç tutularak tahmin yapılmış, ardından hava durumu verileri eklenerek modellerin performansları karşılaştırılmıştır. Modellerin performans değerlendirilmesi için ortalama mutlak yüzdesel hata (MAPE), ortalama karesel hata (MSE) ve ortalama mutlak hata (MAE) kullanılmıştır. Bir yıl boyunca modellerin günlük performansı ölçülmüştür. Hava durumu verilerinin veri setine eklenmesiyle birlikte MAPE değerlerinde LightGBM modelinde %3.213, XGBoost modelinde %3.404 ve CatBoost modelinde %6.671 performans iyileşmesi olduğu tespit edilmiştir.

Anahtar kelimeler: Kısa dönem yük tahmini, Hava durumu, Makine öğrenmesi, MAPE.

Abstract

Accurate estimation of electricity consumption demand is of great importance for energy companies in terms of planning and resource management, and therefore, it is necessary to fully determine the effects of weather conditions on electricity load estimation. In this study, the effects of weather conditions on short-term electricity load estimation were investigated. Electric load estimation was performed by combining historical load data with independent variables such as day characteristics and holidays with a powerful machine learning model such as CatBoost, LightGBM and XGBoost. First, estimation was performed by excluding weather data, and then the performances of the models were compared by adding weather data. Mean absolute percentage error (MAPE), mean square error (MSE) and mean absolute error (MAE) were used to evaluate the performance of the models. Daily performance of the models was measured for a year. With the addition of weather data to the dataset, MAPE values were found to improve by 3.213% for the LightGBM model, 3.404% for the XGBoost model and 6.671% for the CatBoost model.

Keywords: Short-term load estimation, Weather, Machine learning, MAPE.

1 Giriş

Elektrik tüketim tahmini, enerji talebinin belirlenmesi, elektrik tedarikinin planlanması ve enerji kaynaklarının optimize edilmesi gibi stratejik konularda son derece kritik bir rol oynamaktadır. Doğru ve güvenilir tahminler, enerji şirketlerinin sürdürülebilirlik hedeflerine ulaşmasına yardımcı olmakla kalmaz, aynı zamanda enerji kaynaklarının verimli bir şekilde kullanılmasını da sağlamaktadır [1]. Bu nedenle, elektrik tüketim tahmini, enerji sektöründe karar alma süreçlerinde vazgeçilmez bir araç olarak kabul edilmektedir. Enerji talebindeki değişkenlikler, hava koşulları, tüketici alışkanlıkları, ekonomik faktörler ve diğer etkenler tarafından etkilendiğinden, doğru tahminler yapmak oldukça zor olabilmektedir. Ancak, gelişmiş veri analitiği teknikleri ve yapay zekâ uygulamaları kullanılarak bu tahminler daha doğru ve güvenilir hale getirilebilir. Büyük veri analizi, öngörü modelleri ve simülasyon yöntemleri gibi araçlar, elektrik tüketim tahmininde önemli bir rol oynamaktadır [2].

Elektrik tüketim tahmini, enerji tedarikinin planlanması açısından da büyük bir öneme sahiptir. Elektrik üretim tesislerinin kapasite planlaması ve dağıtım ağına optimize edilmesi, enerji şirketlerinin verimli bir şekilde enerji üretmelerini ve dağıtmalarını sağlamaktadır. Bu da enerji

arzında herhangi bir aksama veya kaynak israfının önlenmesine yardımcı olmaktadır [1].

Elektrik talebinin tahmin edilmesi süreci, geçmiş elektrik tüketim verilerinin analizi, gün içindeki özelliklerin dikkate alınması, tatil günleri ve hava durumu gibi faktörlerin değerlendirilmesini içermektedir. Elektrik tüketiminin doğru bir şekilde tahmin edilmesi, enerji şirketlerinin verimli planlama yapabilmeleri ve kaynaklarını etkin bir şekilde yönetebilmeleri için son derece kritik bir rol oynamaktadır. Bu nedenle, elektrik yük tahmini sürecinin sürekli olarak geliştirilmesi ve iyileştirilmesi, enerji sektörünün sürdürülebilirlik ve verimlilik hedeflerine ulaşmasında kritik bir öneme sahiptir.

Elektrik üretim kaynaklarının tüketici taleplerine göre ayarlanması, sürdürülebilirlik düzeyinin yükseltilmesi ve kaynak tüketimi açısından oldukça önemlidir [3]. Elektrik depolanmasındaki fiziksel sınırlamalar nedeniyle, elektrik üretiminin gerçek zamanlı olarak tüketim talebiyle birlikte ayarlanması gerekmektedir. Akıllı şebeke teknolojilerinin ilerlemesi ve başta fotovoltaik sistemler olmak üzere yenilenebilir enerjinin etkisinin artmasıyla birlikte, elektrik güç sistemlerinde kısa vadeli operasyonel planlama ile ilgili karmaşıklık artmış ve elektrik dağıtım şirketlerinin güç şebekelerini güvenilir bir şekilde işletmesi için önemli zorluklar ortaya çıkmıştır [1].

*Yazışılan yazar/Corresponding author

Elektrikli araçların, bölgesel düzeyde dağıtık yenilenebilir üretimin ve enerji depolamanın yaygınlaşmasında beklenen artış, kısa vadeli operasyonlarda arz ve enerji kontrol edilebilirliğindeki belirsizlikleri daha da artırmaktadır. Bu sebeple, doğru talep tahminleri, elektrik güç sistemlerinde planlama ve operasyonel kararları desteklemek için sistem operatörleri ve analistler için hayati önem taşımaktadır [2],[4].

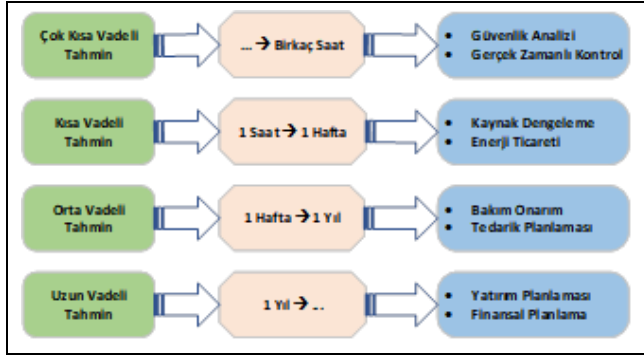
Literatürde tahmin aralığının sınıflandırılması için genel bir kural bulunmamaktadır. Zaman aralığına bağlı olarak, elektrik yük tahmini çok kısa vadeli yük tahmini (VSTLF), kısa vadeli yük tahmini (STLF), orta vadeli yük tahmini (MTLF) ve uzun vadeli yük tahmini (LTLF) olmak üzere dört gruba ayrılabilir [3]-[7]. VSTLF, bir saate kadar ki zaman aralıklarında tahmin yapmak için kullanılır. Genel olarak, güç sistemlerinin gerçek zamanlı kontrolü, güvenlik analizi ve güç ekipmanlarının çalışmasının izlenmesi için kullanılmaktadır [8].

STLF, genellikle bir saat ile bir hafta arası tahmin yapmak için kullanılır. Elektrik fiyatlandırılması, üretim programlarının planlanması ve yük akışlarının yönetilmesi gibi günlük operasyonların planlanması için kullanılmaktadır [9].

MTLF, genellikle bir hafta ile bir yıl arası tahmin yapmak için kullanılır. Güç sistemlerinin işletimi, bakım-onarım çalışmalarının planlanması, risk yönetimi, bilanço hesaplaması için kullanılmaktadır [10].

LTLF, güç sistemlerinin yeni kapasite artışı yatırımlarında bütçeleme kararları için uzun vadeli planlamasında kullanılır. Genel olarak bir yıldan uzun süreli yük tahmini yapmak için kullanılmaktadır [11].

Şekil 1'de tahmin uygulamaları ve zamana bağlı sınıflandırması gösterilmektedir.



Şekil 1. Tahmin uygulamaları ve zamana göre sınıflandırılması.

Figure 1. Forecasting applications and classification by time.

Elektrik tüketim talebi gün içerisinde anlık olarak değişkenlik gösterdiğinden devreye girecek ve devreden çıkacak enerji santrallerinin belirlenmesi için kısa dönem talep tahmin yöntemleri kullanılır. Üretimdeki ve talepteki yük farklılıkları şebeke üzerinde dengesizliğe neden olduğundan piyasa katılımcıları, sistemi dengesizliğe düşürdükleri için dengesizlik maliyetlerine katılmak zorundadır.

14.04.2009 tarih ve 27200 Sayılı Resmî Gazete'de yayımlanan Elektrik Piyasası Dengeleme ve Uzlaştırma Yönetmeliği'nde (DUY) elektrik piyasasında dengesizlik modelleri ve bunlarla ilişkili maliyet hesapları ayrıntılı olarak verilmiştir [12].

Tahmin yöntemleri, istatistiksel ve yapay zekâ yöntemleri olmak üzere iki sınıfa ayrılabilir. İstatistiksel yöntemler, çıktı ile girdi değişkenleri arasındaki ilişkiyi veren açık bir matematiksel model gerektirir. Yapay zekâ yöntemleri ise açık bir model

olmadan girdi ve çıktı değişkenleri arasındaki ilişkiyi öğrenebilir [13].

Tahmin modelinin seçmek ve verilerin işlenerek doğru tahmin yapılması zor bir işlemdir. Seçilen tahmin modeli herhangi bir bölge için uygun olabilirken farklı bir bölgede istenen sonuçları vermeyebilmektedir. Bu farklılıklar, tüketim alışkanlıklarından enerji kaynaklarından ve farklı iklimsel özelliklerden kaynaklanabilmektedir. Bu yüzden her koşul için farklı tahmin yöntemleri kullanılması önerilmektedir [14].

Teknolojik gelişmeler, tüketicilerin enerji tüketiminin akıllı sayaçlar aracılığıyla sürekli olarak izlenmesini sağlamaktadır. Yüksek hassasiyete sahip yük tahmini ile çeşitli alanlarda ve uygulamalarda gelişmiş enerji verimliliği ve daha düşük işletme maliyetleri elde etmek mümkündür.

Pratikte elektriğin depolanmasında sınırlamalar mevcut olduğundan, elektrik üretimi, iletimi ve dağıtım taleple eş zamanlı olarak gerçekleşmek zorundadır. Enerji şebekelerin kesintisiz çalışması için, elektrik enerjisi arzı ve talebi sürekli olarak dinamik bir dengeyi korumalıdır.

Makalenin başlıca katkıları aşağıdaki gibi özetlenebilir:

1. Çalışma, kısa vadeli elektrik tüketim tahmininde kullanılan popüler makine öğrenmesi olan CatBoost modelinin performansı değerlendirilmiştir,
2. Hava durumu verilerinin elektrik tüketim tahminindeki önemini vurgulayan bu çalışma, bu tür verilerin entegrasyonunun tahmin modellerine ne kadar katkı sağladığını gösterir. Bu, veri bilimi alanında bağımsız değişkenlerin seçiminde yol gösterici olabilir,
3. Çalışma, eksik değer doldurma, aykırı değer tespiti ve ölçeklendirme gibi veri ön işleme adımlarının model performansı üzerindeki etkilerini ortaya koyarak, veri hazırlama sürecinin önemini vurgular. Bu, benzer projelerde veri ön işlemenin nasıl yapılması gerektiğine dair rehberlik sağlar,
4. Akademik araştırmalarda ve endüstriyel uygulamalarda, bu çalışmanın sonuçları referans alınarak daha gelişmiş ve optimize tahmin modelleri oluşturulabilir. Bu hem teorik araştırmalara hem de gerçek dünya uygulamalarına yönelik önemli bir bilgi kaynağı oluşturur,
5. Gelecekte yapılacak çalışmalar için çeşitli araştırma alanları ve iyileştirme önerileri sunarak, bu alandaki bilimsel ilerlemeyi teşvik eder. Özellikle, derin öğrenme yöntemleri, model birleştirme teknikleri ve gerçek zamanlı tahmin sistemleri gibi konularda yeni araştırmaların yapılmasına zemin hazırlar.

Bu katkılar hem teorik hem de uygulamalı makine öğrenmesi ve enerji yönetimi alanlarında önemli ilerlemelere yol açabilir. Çalışma, sadece model performanslarını karşılaştırmakla kalmayıp, aynı zamanda veri ön işleme ve bağımsız değişkenlerin seçimi konularında da değerli bilgiler sunar.

Makalenin bundan sonraki içeriği şu şekilde düzenlenmiştir: Bölüm 2, yapay zekâ yöntemlerinin kullanımına ilişkin mevcut literatürün incelenmesine ayrılmıştır. Bölüm 3, kısa vadeli elektrik tüketim tahminini etkileyen faktörlerin ve kullanılan makine öğrenmesi modellerinin kısa bir tanımını sunmaktadır. Bölüm 4'te uygulama detayları verilmekte, karşılaştırmalı sonuçlar sunulmaktadır. Bölüm 5, çalışmayı ve bulgularını özetlemekte ve bazı sonuç değerlendirmeleri sunmaktadır.

2 Literatür araştırması

Yapılan literatür taramasında, elektrik yük tahmini, yapay zekâ yöntemleri ve hava durumunun enerji talebi üzerindeki etkisi gibi konularda yayınlanmış birçok akademik çalışma

incelenmiştir. Çalışma kapsamında elektrik yük tahmini için yapay zekâ tabanlı yöntemlerin kullanıldığı bazı çalışmalar Tablo 1'de özetlenmiştir.

Tablo 1. Literatür özeti.

Table 1. Summay of literature review.

Ref. No	Yıl	Konu / Amaç	Yöntem	Veri Seti	Temel Bulgular
[1]	2023	Alçak gerilim seviyesinde kısa vadeli yük tahmini yapmak için optimize edilmiş derin topluluk öğrenme modelleri geliştirmek	Uyarlanabilir rüzgâr tahrikli optimizasyon (AWDO) algoritması ile optimize edilen Bi-LSTM-AWDO modeli kullanılmıştır. Topluluk öğrenme yöntemleri (torbalama, rastgele alt uzay, güçlendirme) ile modelin doğruluğu artırılmıştır.	Avustralya, Newcastle'daki konut akıllı sayaç ölçümleri	Bi-LSTM-AWDO modeli, bireysel ve toplu yük tahminlerinde diğer derin öğrenme modellerine kıyasla daha yüksek doğruluk sağlamıştır. Bireysel yükler için RMSE değeri 0.121 kW, MAPE değeri %7,55 olarak hesaplanmıştır.
[4]	2022	Ticari binalarda elektrik talebini tahmin etmek ve talep tarafı yönetimi için makine öğrenmesi modellerini kullanmak	LSTM ve Destek Vektör Makineleri (SVM) karşılaştırılmış, özellik seçimi ve veri ön işleme teknikleri uygulanmıştır. Bir saatlik ve bir günlük tahminler için farklı modeller değerlendirilmiştir.	İrlanda'da bulunan University College Dublin'deki bir ticari bina (2013-2018 tarihleri arasındaki 15 dakikalık ölçümler)	LSTM modeli, yeterli veri olduğunda daha yüksek tahmin doğruluğu sağlarken, SVM modeli düşük veri setlerinde daha iyi performans göstermiştir. Özellik seçiminin model doğruluğuna etkisi büyük olmuştur.
[15]	2024	Kısa vadeli yük tahmini için hava durumu ve dış etkenleri dikkate alarak derin öğrenme tabanlı bir model geliştirmek	MoDeNNFE modeli, Mixture of Experts (MoE) yaklaşımını kullanarak DNN tabanlı bir tahmin sistemi sunmaktadır. Model, farklı gün tipleri için özelleştirilmiş derin öğrenme uzmanlarını içermektedir.	Hindistan'daki Mumbai ve Delhi şehirlerinden elde edilen Tata Power ve Hindistan Meteoroloji Departmanı verileri	MoDeNNFE modeli, farklı hava koşulları ve özel günleri dikkate alarak yüksek doğruluklu tahminler üretmiştir. Model, geleneksel yöntemlere göre hata oranlarını önemli ölçüde düşürmüştür.
[16]	2023	Kısa vadeli elektrik yük tahmininde LSTM modelinin hiperparametre optimizasyonu için pekiştirmeli öğrenme (RL) kullanmak	LSTM modeli, pekiştirmeli öğrenme ile hiperparametre optimizasyonuna tabi tutulmuştur. Derin Q Ağı (DQN) kullanılarak en uygun hiperparametreler belirlenmiştir.	Vietnam Elektrik Şirketi'ne ait gerçek yük verileri ve kamuya açık bir veri seti	RL tabanlı hiperparametre optimizasyonu, klasik yöntemlere kıyasla tahmin doğruluğunu artırmıştır. Önerilen model, standart LSTM modeline göre daha düşük hata oranlarına ulaşmıştır.
[17]	2023	Türkiye'deki bazı Fortune 500 şirketlerinin elektrik taleplerini derin öğrenme tekniklerini kullanarak tahmin etmek ve enerji dengesizliğinden kaynaklanan maliyetleri en aza indirmek.	Uzun Kısa Süreli Hafıza (LSTM) ve Çok Katmanlı Algılayıcı (MLP) teknikleri kullanılmıştır. Ayrıca, Loess kullanarak Çoklu Mevsimsel-Trend Ayırıştırması (MSTL) tekniği ilk kez elektrik talebi tahmini probleminde kullanılmıştır.	Türkiye'deki bazı Fortune 500 şirketlerinin elektrik tüketim verileri kullanılmıştır.	MSTL'nin, kütle üretim şeklinde faaliyet gösteren çoğu firma için MLP ve LSTM'den daha iyi performans gösterdiği bulunmuştur. Modelin karmaşıklığının her zaman iyi sonuçları garanti etmediği ve basit yöntemlerin bazen iyi çalışabileceği görülmüştür. Pilot şirkette yaklaşık 378 asgari ücret tutarında maliyet tasarrufu sağlanmıştır.
[18]	2022	Tatil günleri için kısa vadeli yük tahmini (STLF) yapmak ve benzer günlerin yük profillerini kullanarak tahmin doğruluğunu artırmak	Takvim, eğilim, hava durumu ve sayaç arkası PV (BTM PV) kaynaklarını dikkate alan bir tahmin modeli geliştirilmiştir. Genetik algoritma ile hava durumu duyarlılığı optimize edilmiştir.	Güney Kore ulusal tatil günlerine ait saatlik yük verileri ve hava durumu değişkenleri (Korea Power Exchange)	Önerilen yöntem, geleneksel tahmin modellerine kıyasla daha düşük hata oranı sağlamış, özellikle takvim etkisini ve hava durumu değişkenlerini daha doğru bir şekilde değerlendirmiştir.

Tablo 1. Devamı.

Table 1. Continued.

Ref. No	Yıl	Konu / Amaç	Yöntem	Veri Seti	Temel Bulgular
[19]	2022	Kısa vadeli elektrik talebi tahmini için istatistiksel ve makine öğrenmesi yaklaşımlarını Bayesian optimizasyon algoritması ile geliştirmek	SARIMAX ve NARX modelleri kullanılmış, hiperparametre optimizasyonu için Bayesian optimizasyon algoritması (BOA) uygulanmıştır.	Ontario, Kanada'daki 2013-2019 arası saatlik elektrik tüketim verileri	BOA-NARX modeli, BOA-SARIMAX modeline kıyasla daha düşük hata oranları sağlamıştır (MAPE $\approx$ %3). BOA-NARX modeli, tüm mevsimlerde daha istikrarlı sonuçlar üretmiştir.
[20]	2022	Sanayi ve konut tesislerinde elektrik talep tahmini yapmak ve topluluk makine öğrenmesi modellerini kullanarak doğruluğu artırmak	Eksik verileri tamamlama, aykırı değerleri temizleme ve veri standardizasyonu adımlarını içeren ön işleme uygulanmıştır. Ardından, farklı regresyon modellerinin birleştirilmesiyle bir topluluk model oluşturulmuştur.	İspanya'daki bir sanayi bölgesi, Uruguay'ın toplam elektrik tüketimi ve Montevideo'daki bir dağıtım trafo merkezinden elde edilen saatlik yük verileri	ExtraTreesRegressor modeli, sanayi verilerinde %2.55, toplam tüketim verilerinde %5.17 ve trafo merkezi verilerinde %9.09 MAPE hata oranı ile en iyi tahmin performansını göstermiştir.
[21]	2018	Assam, Hindistan için bölgesel kısa vadeli yük tahmini modeli geliştirmek ve bölgesel iklim koşullarına uyarlamak	Çekirge Optimizasyon Algoritması (GOA) ile optimize edilen Destek Vektör Makineleri (SVM) modeli önerilmiştir. Benzer Gün Yaklaşımı (SDA) kullanılarak geçmiş benzer günler seçilmiştir.	Assam, Hindistan'a ait saatlik elektrik yük verileri ve iklim değişkenleri (sıcaklık, nem)	GOA-SVM modeli, GA-SVM ve PSO-SVM modellerine kıyasla daha düşük hata oranları sağlamıştır. Sıcaklık ve Nem değişkeninin tahmin doğruluğuna önemli katkı sağladığı gösterilmiştir.
[22]	2017	Türkiye'nin elektrik talep tahmininde doğruluğu artırmak için yapay sinir ağları ve uzman sistemleri birleştiren hibrit bir model geliştirmek	EPSİM-NN adı verilen modelde, günlük ortalama saatlik talep miktarı ve 24 saatlik talep şekli için iki ayrı yapay sinir ağı kullanılmıştır. Model, tahmin hatalarını azaltmak için uzman sistem tabanlı bir düzeltme modülü içermektedir.	Türkiye elektrik piyasasına ait saatlik talep verileri, sıcaklık, GSMH ve sanayi üretim endeksi gibi değişkenler	EPSİM-NN modeli, geleneksel yöntemlere göre daha yüksek doğruluk sağlamış, hata oranlarını %2 seviyelerine kadar düşürebilmiştir.
[23]	2024	Kısa vadeli elektrik yük tahmini doğruluğunu artırmak için hibrit CatBoost-XGBoost modellerini kullanmak	CatBoost ve XGBoost algoritmaları Aritmetik Optimizasyon Algoritması (AOA) ile optimize edilerek tahmin performansı iyileştirilmiştir. Duyarlılık analizi ile değişken etkileri incelenmiştir.	2009-2010 saatlik elektrik yük verileri ve sıcaklık verileri	CatBoost-AOA modeli eğitim verisinde en iyi performansı gösterirken, test verisinde XGBoost-AOA modeli daha iyi sonuçlar vermiştir. Sıcaklık ve ay değişkenleri tahmin doğruluğunu en çok etkileyen faktörlerdir.
[24]	2021	Kısa vadeli elektrik yük tahmini (STLF) için makine öğrenmesi modellerini karşılaştırmak ve XGBoost'un performansını incelemek	XGBoost, SVR, Random Forest, KNN ve MLR modelleri kullanılmıştır. Modeller, Panama elektrik şebekesi için 168 saatlik (1 hafta) tahmin yapmak üzere eğitilmiştir.	Panama elektrik şebekesine ait 2015-2020 saatlik elektrik yük verileri, hava durumu değişkenleri ve tatil günleri bilgileri	XGBoost modeli, ANN ve geleneksel yöntemlerden daha iyi performans göstermiştir. XGBoost'un özellik önemi analizi, sıcaklık ve hafta içi değişkenlerinin kritik olduğunu göstermiştir.
[25]	2021	Sanayi müşterileri için kısa vadeli yük tahmini yaparak enerji yönetimini optimize etmek	Varyasyonel Mod Ayrıştırma (VMD) ve Örnek Entropi (SampEn) tabanlı bir ayrıştırma yöntemi (SVMD) geliştirilmiş ve yük verileri trend ve dalgalanma serilerine ayrılmıştır. Trend serisi Lineer Regresyon ile, dalgalanma serileri XGBoost ile modellenmiştir.	Çin ve İrlanda'dan çeşitli sanayi müşterilerine ait akıllı sayaç verileri kullanılmıştır.	SVMD-XGBoost modeli, geleneksel yöntemlere kıyasla tahmin doğruluğunu önemli ölçüde artırmıştır. XGBoost hiperparametreleri Bayes optimizasyonu ile ayarlanarak hata oranları düşürülmüştür.
[26]	2020	Veri odaklı kısa vadeli yük tahmin yöntemlerini karşılaştırarak en uygun model ve özellik seçim yöntemini belirlemek	Zaman serisi, klasik regresyon ve derin öğrenme modelleri karşılaştırılmıştır. Özellik seçimi için Greedy Forward Wrapper Feature Selection (GFW-FS) yöntemi uygulanmıştır.	Çin, Jiangsu Eyaleti'nden saatlik yük verileri (2015-2017) ve hava durumu verileri	MARS, SVR ve MLR modelleri en düşük hata oranlarını sağlamıştır (MAPE < %2.6). Derin öğrenme modelleri klasik regresyon yöntemleri kadar başarılı olmamıştır.

Tablo 1. Devamı.

Table 1. Continued.

Ref. No	Yıl	Konu / Amaç	Yöntem	Veri Seti	Temel Bulgular
[27]	2023	Büyük ölçekli elektrik güç sistemleri için kısa vadeli yük tahmini yapmak ve küresel iklim modellerini (GCM) makine öğrenmesi modellerine entegre etmek	Tek yönlü ve çift yönlü yapay sinir ağları (MLP, LSTM, GRU) kullanılmış, küresel iklim modellerinden elde edilen sıcaklık tahminleri model girdisi olarak değerlendirilmiştir. Diebold-Mariano testi ile farklı modeller karşılaştırılmıştır.	Brezilya'nın Merkezi-Güneydoğu elektrik pazarı (2016-2021 yılları arasındaki saatlik yük ve sıcaklık verileri)	Çift yönlü LSTM ve GRU modelleri en düşük hata oranlarını sağlamış ve geleneksel Brezilya Bağımsız Sistem Operatörü (ISO) modellerinden daha iyi performans göstermiştir (MAPE: %1.18). Küresel iklim modellerinin entegrasyonu tahmin doğruluğunu artırmıştır.

İbrahim ve Hossain alçak gerilim seviyesinde kısa vadeli elektrik yük tahmini için derin topluluk öğrenme modellerine dayalı bir yöntem incelenmiştir. Oluşturulan modeller, bireysel ve toplu yük senaryolarında test edilmiş ve en iyi performans, çift yönlü uzun kısa vadeli bellek (Bi-LSTM) modeli ile elde edilmiştir [1]. Pallonetto ve diğ. ticari binalarda saatlik elektrik talebini tahmin etmek için LSTM ve destek vektör makinesi (SVM) yöntemleri karşılaştırmıştır. Elde edilen sonuçlar, yeterli veri mevcut olduğunda LSTM modelinin daha yüksek doğruluk sunduğunu, ancak veri kısıtlı olduğunda SVM modelinin daha iyi performans gösterdiğini ortaya koymuştur [4].

Kotecha ve diğ. kısa vadeli yük talebi tahminini iyileştirmek için derin öğrenme ve hava durumu verilerini kullanan yeni bir yöntem olan "Mixture of Deep Neural Network Forecasting Experts" (MoDeNNFE) modelini önermiştir. Yapılan çalışmada, hava durumu ve diğer dışsal faktörler dahil edilerek yük talebi tahminlerinin doğruluğu artırılmıştır [15].

Nguyen ve diğ. tarafından uzun kısa süreli bellek (LSTM) modelinin tahmin doğruluğunu artırmak için takviyeli öğrenme (RL) kullanarak hiper-parametre optimizasyonu yapılması önerilmiştir. Ana model, LSTM ve RL'nin birleştirilmesinden oluşmaktadır. RL kullanılarak, LSTM modelinin hiper-parametreleri optimize edilmiş ve tahmin hatalarının azaltılması hedeflenmiştir. Bir gün önceden yapılan tahminlerde, önerilen modelin performansı diğer yöntemlere kıyasla üstün sonuçlar vermiştir [16].

Işık ve diğ. Türkiye'deki bazı Fortune 500 şirketleri için enerji dengesizliği maliyetlerini minimize etmeye yönelik derin öğrenmeye dayalı elektrik talebi tahmin modellerini incelemiştir. Çalışmada, uzun kısa vadeli bellek (LSTM), çok katmanlı algılayıcı (MLP) ve Loess kullanarak çoklu mevsimsel-trend ayrıştırma (MSTL) teknikleri kullanılmıştır. Sonuçlar, MSTL'nin firma düzeyindeki talep tahminlerinde genellikle daha başarılı olduğunu göstermektedir [17].

Son ve ark. tarafından tatil günlerinde gün öncesi kısa vadeli elektrik yük tahmini (STLF) yapmak için benzer günlerin yük profillerinin değiştirilmesine dayalı bir yöntem önerilmiştir. Çalışma, özellikle güneş enerjisi kaynakları gibi dağıtılmış enerji kaynaklarının enerji sistemlerine getirdiği zorluklarla ilgilenmiştir [18].

Sultana ve ark. tarafından kısa vadeli elektrik talebi tahmini için Bayes Optimizasyon Algoritması'na (BOA) dayalı dışsal girdili mevsimsel otoregresif bütünleşik hareketli ortalama (SARIMAX) ve dışsal girdili doğrusal olmayan otoregresif ağlar (NARX) yaklaşımları incelenmiştir. Çalışmada, BOA'nın hiper parametre optimizasyonundaki rolünü değerlendirilmiş ve elektrik tüketimini etkileyen önemli faktörlerin belirlenmesi ve

hiper parametrelerin otomatik olarak optimize edilmesi amaçlanmıştır [19].

Porteiro ve ark. tarafından sanayi ve konut tesislerinde kısa vadeli elektrik talebi tahmini için topluluk makine öğrenmesi modellerinin kullanılması incelenmiştir. Çalışmada, bir günlük yük tahmini modeli, bir saatlik tahminler üzerine kurulan topluluk stratejisiyle geliştirilmiştir. Özellikle seçimi ve veri ön işleme işlemleri yapılarak modellerin doğruluğu artırılmıştır. En iyi sonuçları sunan algoritma optimize edilerek gelecek 24 saate ait tahmin modeli oluşturulmuştur [20].

Barman ve diğ. Hindistan'ın Assam eyaletinde kısa vadeli yük tahmini (STLF) için Çekirge Optimizasyon Algoritması (GOA) ve Destek Vektör Makinelerini (SVM) birleştiren bölgesel hibrit bir model sunulmuştur. Destek Vektör Makinelerini (SVM) birleştiren hibrit bir model üzerinde çalışmıştır. GOA, SVM'nin hiper parametrelerini optimize etmek için kullanılmış ve bu süreçte "benzer gün" yaklaşımı (SDA) ile geçmiş veriler analiz edilmiştir. Araştırmada, sıcaklık ve nemin, Assam gibi tropik bölgelerde yük talebini büyük ölçüde etkilediği görülmüştür [21]. Başoğlu ve Bulut, yapay sinir ağları ve uzman sistemlerin birlikte kullanıldığı, piyasa ve mevsimsel koşulları dikkate alındığı kısa dönemli elektrik talep tahminleri için hibrit bir sistem geliştirilmiştir. Çalışmada 2005-2016 yılları arası geçmiş tüketim değerleri, sıcaklık, gayri-safı milli hâsıla, sanayi üretim endeksi ve günün istisnai özellikleri kullanılmıştır. Gerçekleşen değerler ile tahmin sonuçları karşılaştırılmış ve yüksek doğruluk derecesinde tahminler yapılmıştır [22].

Zhang ve Jánošík kısa vadeli yük tahminini iyileştirmek için iki makine öğrenme modeli olan CatBoost ve XGBoost yaklaşımlarını birleştiren hibrit modellerin kullanımı incelenmiştir. Çalışmada, makine öğrenmesi modelleri daha yüksek tahmin doğruluğu elde etmek için farklı optimizasyon algoritmaları ile desteklenen hibrit modellerin potansiyeli değerlendirilmiştir. Saatlik elektrik yük verilerinin yanı sıra sıcaklık verileri de kullanılarak tahmin doğruluğu artırılmıştır. CatBoost ve XGBoost algoritmalarının performansı farklı veri setlerinde test edilmiş ve sonuçlar, optimizasyon algoritmalarının bu modellerin performansını önemli ölçüde artırdığını göstermiştir. Test verilerinde, XGBoost-Arithmetic Optimization Algorithm (AOA) hibrit modeli, diğer modellere kıyasla üstün performans sergilemiştir. Sonuçlar, sıcaklık verisinin elektrik yük tahminlerinde en etkili değişken olduğunu ve mevsimsel değişimlerin yük tahmininde önemli bir rol oynadığını ortaya koymuştur [23].

Madrid ve Antonio Panama'nın 2019-2020 yılları arasındaki saatlik elektrik yükü ve hava durumu verileri kullanılarak kısa vadeli yük tahmini yapılmıştır. Çoklu doğrusal regresyon (MLR), k-en yakın komşu regresörü (k-NN), destek vektör

regresyonu (SVR), rastgele orman regresörü (RF) ve XGBoost modellerinin karşılaştırılması yapılmıştır. XGBoost ile oluşturulan modellerin diğer algoritmalarla oluşturulan modellere göre üstün performansa sahip olduğu görülmektedir [24].

Zhang ve Song, gün öncesi saatlik bazda yük tahmini için klasik veri madenciliği ve derin öğrenme yaklaşımlarını karşılaştıran bir çalışmayı ele almıştır [23]. Madrid ve ark. tarafından saatlik elektrik yükü ve hava durumu verileri kullanılarak kısa vadeli yük tahmini yapılmıştır. Çoklu doğrusal regresyon (MLR), k-en yakın komşu regresörü (k-NN), destek vektör regresyonu (SVR), rastgele orman regresörü (RF) ve XGBoost modellerinin karşılaştırılması yapılmıştır [24]. Wang ve ark. kısa vadeli elektrik yük tahmini için Uyarlanabilir Varyasyonel Mod Ayrışımı ve Örnek Entropisi (SVMD) ve XGBoost modellerini temel alan bir yöntem önermiştir [25].

Liu ve diğ. klasik veri madenciliği ve derin öğrenme yaklaşımlarını karşılaştıran bir çalışmayı ele almıştır. Çalışmada zaman serisi modelleri, klasik regresyon modelleri ve derin öğrenme modelleri incelenmiştir. Derin öğrenme modellerinin, veri içindeki gizli doğrusal olmayan ilişkileri yakalamada avantaj sağladığı gösterilmiştir. Bununla birlikte, farklı modellerin performansları ve kullanılan özelliklerin etkileri karşılaştırmalı olarak analiz edilmiştir [26].

Morais ve diğ. [27] kısa vadeli elektrik talebi tahmini için yapay sinir ağları (ANN) ve küresel iklim modellerini (GCM) kullanarak büyük ölçekli elektrik güç sistemlerine yönelik bir uygulama sunulmuştur. Çalışmada, uzun kısa vadeli bellek (LSTM) ve kapılı tekrarlayan birim (GRU) gibi yapay sinir ağları yapıları kullanılarak tahminler yapılmış ve bu modellerin performansı, Brezilya Bağımsız Sistem Operatörü'nün (ISO) kullandığı modellerle karşılaştırılmıştır. Sonuçlar, çift yönlü LSTM ve GRU modellerinin daha doğru tahminler sağladığını göstermektedir. Ayrıca, küresel iklim modellerinden elde edilen sıcaklık tahminlerinin ANN modelleriyle birleştirilmesi, doğruluğu önemli ölçüde artırmıştır.

2.1 Kısa dönem elektrik tüketim tahminini etkileyen faktörler

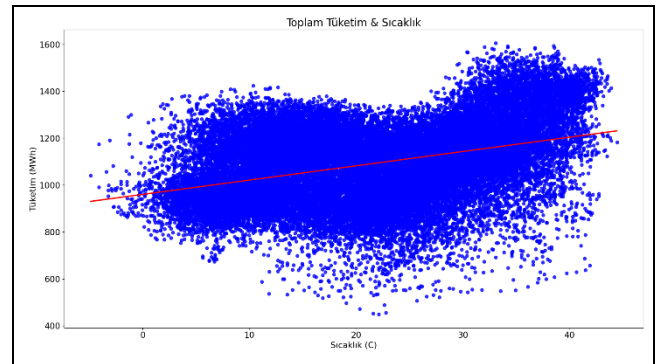
Kısa dönem elektrik tüketim tahmini, genellikle birkaç saat ya da bir gün öncesi gibi kısa bir süre aralığında elektrik talebini tahmin etmeyi içerir. Bu tahminler, enerji şirketlerinin günlük operasyonel kararlarını desteklemek ve elektrik tedarikini planlamak için kullanılır. Kısa dönem tahminler, günlük değişen hava koşulları ve talep dalgalanmaları gibi değişkenleri göz önüne alarak gerçekleştirilir. Elektrik talebinin doğru bir şekilde tahmin edilmesi, enerji şirketlerinin kaynaklarını etkin bir şekilde kullanmasını ve müşteri taleplerini karşılamasını sağlar [25]. Bu sayede, enerji şirketleri optimum düzeyde elektrik üretebilir, gereksiz maliyetleri önler ve sürdürülebilir enerji tedarikini sağlar. Elektrik tüketim tahmini, istatistiksel modeller, veri analitiği ve yapay zekâ teknikleri gibi yöntemler kullanılarak gerçekleştirilir. Bu yöntemler, geçmiş verilere dayanarak gelecekteki elektrik talebini tahmin etmek için kullanılır. Kaliteli tahminler yapabilmek için, doğru ve güncel verilere sahip olmak çok önemlidir. Dahası, günün saatine, haftanın gününe ve hatta tatil dönemlerine göre değişen taleplerin analiz edilmesi gerekmektedir. Bu sayede, enerji şirketleri talebe göre üretimi ve dağıtımını optimize edebilir, olası aksaklıkları önleyebilir ve müşteri memnuniyetini artırabilir. Teknolojik gelişmeler, enerji üretimi ve tüketimi arasındaki dengenin sağlanmasında daha etkili tahmin yöntemlerinin kullanılmasını sağlayarak enerji verimliliğini

artırmaktadır. Bu da çevresel etkileri azaltmak, enerji kaynaklarını daha verimli kullanmak ve sürdürülebilir bir enerji sektörü oluşturmak için büyük bir adımdır. Bu yüzden, doğru tahminler yapabilmek için sürekli olarak yeni yöntemler ve teknolojilerin geliştirilmesi önemlidir [3].

Çalışmada kullanılan veriler arasında geçmiş tüketim verileri, zaman faktörleri ve hava durumu verileri bulunmaktadır.

2.1.1 Hava durumu verileri

Hava durumu verileri, elektrik dağıtım yükünün tahmin edilmesinde kullanılan temel faktörlerden biridir. Yük tahmini için çeşitli hava değişkenleri dikkate alınabilir. Sıcaklık ve nem en sık kullanılan yük tahmin edicileridir [28]. Hava durumunun etkisi, en çok evsel ve tarımsal kullanım tüketicilerinde görülür bununla beraber endüstriyel tüketicilerin de yük profilini değiştirebilir. Sıcaklık ve elektrik tüketimi arasındaki güçlü ve doğrusal olmayan ilişki yapılan çalışmalar ile gösterilmiştir [29]. Nem, rüzgâr hızı ve rüzgâr yönünün hissedilen sıcaklık üzerinde etkisi vardır. Ülkemiz için kuzeyden esen rüzgarlar hissedilen sıcaklığı düşürürken güneyden esen rüzgarlar hissedilen sıcaklığı artırmaktadır. Yüksek nem hissedilen sıcaklığın değerini artırır. Rüzgârlı bir yaz gününde daha az soğutma cihazı kullanılacağından elektrik tüketimi daha düşük olacaktır. Rüzgârlı bir kış gününde ise daha fazla soğutma cihazı kullanılacağından elektrik tüketimi daha yüksek olacaktır. Yağışın sıcaklığı düşürücü etkisi olduğundan elektrik tüketimi üzerinde olumlu veya olumsuz etki yaratabilir [30]. Şekil 2'de tüketim ve sıcaklık arasındaki dağılım grafiği gösterilmiştir. Grafik, yüksek sıcaklık ve elektrik tüketimi arasında daha büyük bir korelasyon göstermektedir.



Şekil 2. Tüketim-Sıcaklık ilişkisi.

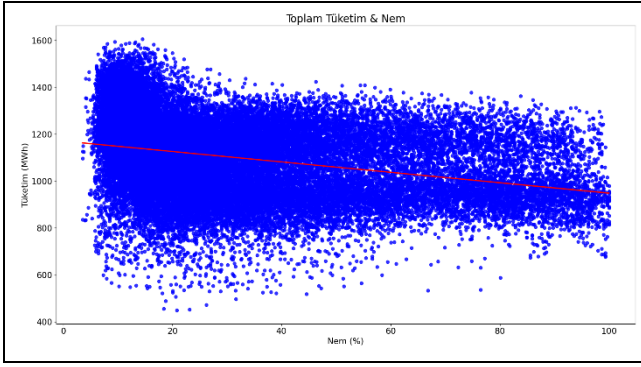
Figure 2. Consumption-Temperature relationship.

2.1.2 Geçmiş tüketim verileri

Geçmiş tüketim verileri, elektrik tüketimini anlamak için temel bir bilgi kaynağıdır. Bu veriler, geçmiş dönemlerdeki elektrik tüketim desenlerini ve trendlerini analiz etmek için kullanılmış, elektrik talebindeki değişiklikleri anlamak için geçmiş sıcaklıklara bağlı olarak bu veriler detaylı bir şekilde incelenmiştir.

Şekil 3'te tüketim ve nem arasındaki dağılım grafiği gösterilmiştir. Şekilden de görüleceği üzere tüketim ve nem arasında negatif korelasyon ilişkisi vardır.

Bu analizler sayesinde, enerji yönetimini geliştirmek ve enerji tasarrufu sağlamak mümkün olmuştur. Ayrıca, elektrik tüketiminde tahminlerde bulunmak ve enerji talebinin gelecekte nasıl değişeceğini öngörmek için kullanılan bir yöntemdir.



Şekil 3. Tüketim-Nem ilişkisi.

Figure 3. Consumption-Humidity relationship.

Bu şekilde, enerji arzı ve talebi arasındaki dengeyi sağlamak, enerji kaynaklarının sürdürülebilir kullanımını sağlamak amacıyla çalışmalar yapılmaktadır. Geçmiş tüketim verilerinin analizi, enerji sektörü için stratejik planlamayı desteklemekte ve daha verimli bir enerji yönetim sistemine katkıda bulunmaktadır. Tüm bu çalışmalar, enerji verimliliğini artırmak, sürdürülebilir bir enerji geleceği oluşturmak ve çevresel etkileri azaltmak için oldukça önemlidir. Bu nedenle, geçmiş tüketim verilerini analiz etmek ve bu verilerden elde edilen bilgileri kullanmak, enerji sektöründe başarılı sonuçlar elde etmek için kritik bir konudur.

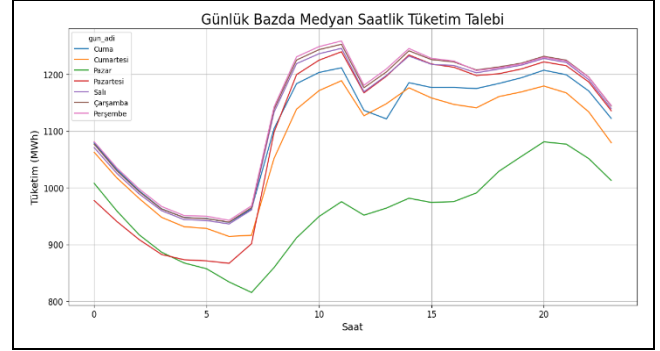
2.1.3 Zaman faktörü

Zaman faktörü, veri setlerinde saatlik, günlük örüntüler gibi örüntülerin bulunması nedeniyle STLF için en önemli faktördür [11].

Zaman faktörleri yılın zamanını, haftanın gününü ve günün saatini içerir. Hafta içi, hafta sonu ve özel (resmî ve dini bayram günleri) günler arasında önemli tüketim farklılıkları bulunmaktadır. Hafta içi farklı günlerindeki elektrik tüketimi de farklı olabilmektedir [28]. Dolayısıyla takvim günleri, tahmin sürecinde dikkate alınması gereken önemli bir özelliktir. Farklı özel günlerin bölgeye ve haftanın gününe bağlı olarak gün içindeki elektrik tüketimi üzerinde farklı etkileri olduğunu unutmamak gerekir [31]. Örneğin, hafta içi ve hafta sonu elektrik tüketim desenleri arasında belirgin farklılıklar olabilir ve bu farklılıkların kesin olarak modellere yansıtılması gerekmektedir. Buna ek olarak, mevsimsel değişiklikler ve tatil günleri gibi faktörler de elektrik tüketimini etkileyebilir ve bu da modellerin daha da geliştirilmesini gerektirebilir. Tüm bu verilere dayanarak, elektrik yükünün gelecekteki talebini öngörebilmek için daha karmaşık bir analiz yapılması ve doğru tahminlerde bulunulması gerekmektedir. Bu nedenle, gün özellikleri ve zaman belirleyicileri, elektrik talebi analizi için vazgeçilmez bir rol oynamaktadır.

Tatil günleri, elektrik talebini etkileyebilecek önemli bir faktördür. Bu bağımsız değişken, özellikle tatil günlerindeki elektrik tüketimi ve talebindeki değişiklikleri analiz etmek için kullanılmıştır. Çünkü tatillerde genellikle endüstriyel tüketim azalırken, konut tüketimi artabilmektedir. Dolayısıyla, tatil günlerinin elektrik yük tahmininde dikkate alınması gerekmektedir. Tatil günlerinin elektrik yükü üzerindeki etkisi birçok farklı faktöre bağlı olabilir. Bunlar arasında tatil günlerinde insanların daha fazla evde olması, aile toplantıları ve kutlamalar nedeniyle artan ev tipi elektrikli cihazların kullanımı gibi etkenler yer almaktadır. Bunun yanı sıra, tatil bölgelerindeki otel ve tatil köylerinde de elektrik tüketimi

artabilir. Turistlerin sayısındaki artış, havuzlar, klima ve diğer tesisat sistemlerinin daha yoğun bir şekilde kullanılmasına neden olabilir. Bu sayede enerji şirketleri ve elektrik sağlayıcıları, tatil dönemlerindeki artan talebe göre kaynaklarını ve üretim kapasitelerini ayarlayabilirler. Şekil 4'te bağımlı değişken (Tüketim) ile iki bağımsız değişken (günün saati ve haftanın günü) arasındaki gösterilmektedir.



Şekil 4. Haftanın günleri ve saatlik bazda tüketim verilerinin istatistiksel özellikleri.

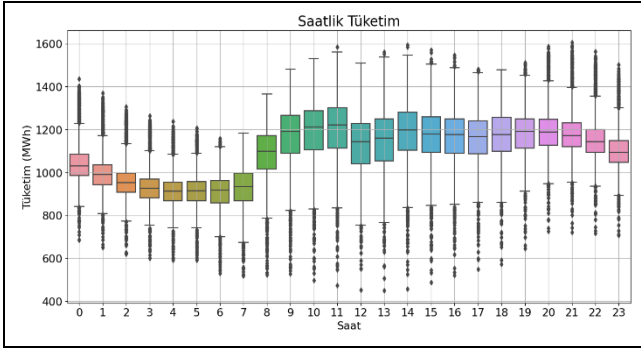
Figure 4. Statistical characteristics of consumption data on days of the week and hourly basis.

Şekil 5, günün saatine göre tüketim değişimini, Şekil 6, farklı aylardaki tüketim değişimi ve Şekil 7, hafta içi ve hafta sonu tüketimindeki değişimi göstermektedir. Bu kutu grafikleri, tüketim verilerinin yayılımını ve çarpıklığını göstermektedir. Yaz ayları olan Temmuz ve Ağustos'un en yüksek tüketim değerlerine sahip olduğu, Nisan ve Ekim'in ise daha düşük tüketim değerlerine sahip olduğu görülmektedir. Tüketim değerlerinin saat 0.00'dan sabah saat 6:00'ya kadar düşük olduğu ve azaldığı, saat 6:00'den itibaren saat 11:00'e kadar yükseldiği, saat 12:00'da bir miktar azaldığı ve sonrasında tekrar yükseldiği ve saat 14:00'den sonra gün sonuna kadar tekrar azaldığı görülmektedir. Pazar günü tüketim değerlerinin diğer günlerin tüketimlerinden çok daha düşük olduğu görülmektedir.

2.2 CatBoost, LightGBM ve XGBoost makine öğrenme modelleri

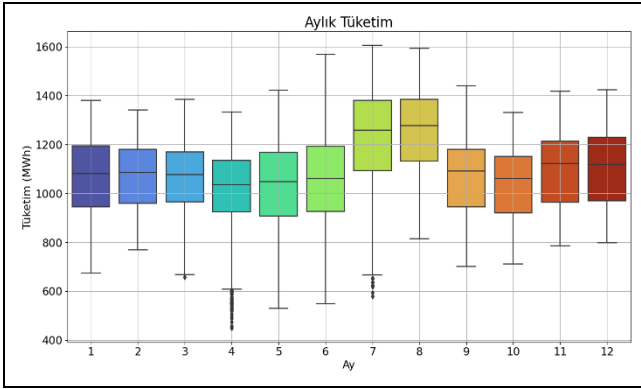
CatBoost, LightGBM ve XGBoost gibi güçlü makine öğrenmesi modelleri, veri bilimi ve makine öğrenimi alanında yaygın bir şekilde kullanılır. Bu modeller, yüksek doğruluk ve güvenilirlik sağlama yetenekleri sayesinde, elektrik yük tahmini alanında mükemmel sonuçlar elde etmek için kullanılmaktadır. Özellikle büyük ölçekli elektrik dağıtım şebekelerinde bu makine öğrenmesi modelleri, talebe dayalı optimizasyonda ve enerji kullanımının verimli bir şekilde yönetilmesinde önemli bir rol oynamaktadır.

CatBoost, XGBoost ve LightGBM, birkaç ortak özelliği paylaşan popüler gradyan artırma algoritmalarıdır [32]. Hepsî ağaç tabanlı öğrenme algoritmalarıyla gradyan artırma kullanır; büyük veri kümeleriyle verimli bir şekilde çalışır, düzenleme teknikleri içerir, ayarlanabilir hiper parametreler sunar ve yerleşik çapraz doğrulamayı destekler. Ayrıca, bu kütüphaneler hesaplama süresini azaltmak için eğitim aşamasında paralelleştirmeden yararlanır. Benzerliklerine rağmen, her çerçevenin kendine özgü güçlü yönleri vardır ve bu da onları farklı kullanım durumları ve veri türleri için uygun hale getirir. Bu çalışmada CatBoost modeli kullanılmıştır.



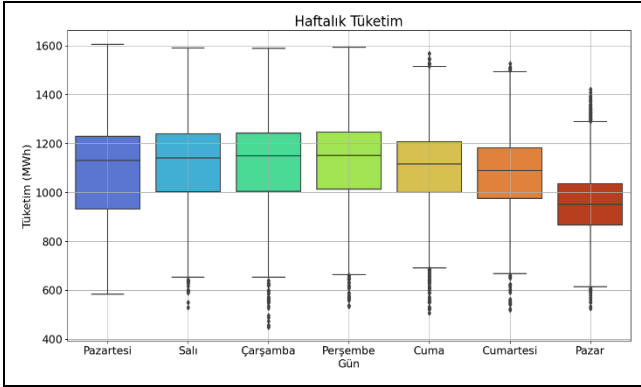
Şekil 5. Saatlik bazda tüketim değişimi.

Figure 5. Hourly consumption change.



Şekil 6. Aylık bazda tüketim değişimi.

Figure 6. Monthly consumption change.



Şekil 7. Günlük bazda tüketim değişimi.

Figure 7. Change in consumption on a daily basis.

“Category” ve “Boosting” kelimelerinin birleşiminden oluşan CatBoost [33], 2017 yılında Yandex şirketi tarafından geliştirilmiştir. Eksik değerleri işleyebilen, gradyan artırma tabanlı açık kaynak kodlu bir makine öğrenmesi algoritmasıdır. Ön tanımlı parametrelerle bile iyi sonuçlar sağladığından veri hazırlığı aşamalarını kısaltmaktadır. Öğrenme hızı yüksektir. Hem sayısal hem de sayısal olmayan veriler ile çalışabilmektedir.

CatBoost'un arkasındaki ana fikir, kategorik verileri işlemek için özel olarak tasarlanmış gradyan artırma modelleri oluşturmaktır. CatBoost'un temel özelliklerinden biri, gerçek dünya veri kümelerinde yaygın bir sorun olan verilerdeki eksik değerleri otomatik olarak işleyebilmesidir. Bunu, her eksik değer için ayrı bir ağaç oluşturarak yapar, bu da modelin

verilerde eksik değerler olduğunda bile daha doğru tahminler yapmasını sağlamaktadır [34].

LightGBM (Light Gradient Boosting Machine) [35], Microsoft tarafından 2017 yılında geliştirilmiştir. Eğitimi hızlandırmak için bellek kullanımını ve hesaplama süresini azaltan histogram tabanlı algoritmalar kullanır. Ağaçları seviye bazında büyüten diğer gradyan artırma algoritmalarının aksine, LightGBM ağaçları yaprak bazında büyütür. Büyük veri kümelerini verimli bir şekilde işleyebilir.

XGBoost (eXtreme Gradient Boosting) [36], 2016 yılında Tianqi Chen ve Carlos Guestrin tarafından geliştirilmiştir. Gradyan Artırma algoritmasının hız ve verimlilik için optimize edilmiş yüksek performanslı biçimidir. Aşırı uyumun önlenmesine yardımcı olan yerleşik L1 (Lasso) ve L2 (Ridge) düzenleme tekniklerine sahiptir. Eksik değerleri işleyebilir.

2.3 Performans değerlendirmede kullanılan metrikler

Modellerin performansını değerlendirmek için literatürde birçok doğruluk ölçütü kullanılmaktadır. En yaygın olarak kullanılan doğruluk ölçütleri aşağıdaki denklemlerde gösterilmiştir. Burada y_t gerçek tüketim değerini, f_t tahmini tüketim değerini ve n toplam tahmin sayısını göstermektedir. Daha iyi tahmin performansı, daha düşük MAE (Ortalama Mutlak Hata), MSE (Ortalama Karesel Hata) ve MAPE (Ortalama Mutlak Yüzdesel Hata) değerleri ile gösterilir. MAE, tahmin edilen ve gerçek değerler arasındaki ortalama mutlak farktır. MSE, tahmin değerleri ile gerçek değerler arasındaki karesel hataların toplamını karşılaştırır.

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - f_t| \quad (1)$$

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - f_t)^2 \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - f_t}{y_t} \right| \times 100 \quad (3)$$

Lewis [37] tarafından yapılan çalışmada, MAPE değerinin diğer ölçümlere göre daha etkin sonuçlar elde edildiği ve belirli değerler arasındaki MAPE değerlerinin değişimi açıklamakta yeterli olduğunu belirtilmiştir. Bu çalışmada MAPE değerinin %10'dan küçük olduğu durumda modelin yüksek doğruluğa sahip olduğu, MAPE değerinin %10 ile %20 aralığında olduğunda ise orta derecede doğruluğa sahip olduğu, %20 ile %50 aralığındaki MAPE değerinde ise düşük derecede doğruluğa sahip olduğu ancak kabul edilebilir, %50'nin üzerinde olduğunda ise “tahmin değeri yoktur” şeklinde tanımlanmıştır.

2.4 Ölçeklendirme

Ölçeklendirme, farklı ölçeklerdeki öznitelikler için standart bir karşılaştırma tabanı sağlar. Çok büyük ya da çok küçük değer aralıklarındaki öznitelikler öğrenme sürecinde hatalara yol açabilir. Ölçeklendirme bu aralıkları daraltır. Ölçeklendirme ile oluşturulan modelin performansı artar, aykırı değerlerin etkisi azalır ve verilerin aynı ölçekte olması sağlanır. Bu nedenle ölçeklendirme, tutarlı ve başarılı modelleme için çok önemli bir veri ön işleme adıdır.

Makine öğrenmesinde yaygın olarak kullanılan başlıca ölçeklendirme (scaling) yöntemleri şunlardır:

Standartlaştırma: Öznitelikleri ortalama 0 ve standart sapma 1 olacak şekilde ölçeklendirir.

$$x' = \frac{x_i - x_{mean}}{\sigma} \quad (4)$$

Min-Max Ölçeklendirme: Doğrusal dönüşüm kullanarak tüm değerleri 0-1 aralığına sığdırır.

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (5)$$

Robust Ölçeklendirme: Ortalama ve standart sapma yerine ortanca ve çeyreklik aralığı kullanır. Aykırı değerlere karşı daha dayanıklıdır.

$$Q_1 = \left(\frac{n-1}{4}\right)^{inci} \text{ terim} \quad (6)$$

$$Q_3 = 3 * \left(\frac{n-1}{4}\right)^{inci} \text{ terim} \quad (7)$$

$$IQR = Q_3 - Q_1 \quad (8)$$

$$x' = \frac{x_i - x_{median}}{IQR} \quad (9)$$

x' : Ölçeklendirilmiş Veri

x_i : Girdi Değeri

x_{max} : Girdi Verilerindeki En Büyük Değer

x_{min} : Girdi Verilerindeki En Küçük Değer

σ : Standart Sapma

x_{median} : Ortanca Değer

Uygun ölçeklendirme yönteminin seçilmesi verinin dağılımına ve model türüne bağlıdır. Ölçeklendirme model performansını önemli ölçüde etkileyebilir.

2.5 Korelasyon analizi

Korelasyon analizi, iki veya daha fazla değişken arasındaki ilişkinin gücünü ve yönünü ölçmek için kullanılan istatistiksel bir yöntemdir [38], [39]. Temel amacı, değişkenler arasındaki doğrusal ilişkinin varlığını ve derecesini belirlemektir.

Korelasyon analizi, iki değişken arasındaki ilişkiyi -1 ile +1 arasında değişen bir korelasyon katsayısı ile ifade eder:

- (+1): Mükemmel pozitif korelasyon (Bir değişken arttıkça diğeri de artar),
- (-1): Mükemmel negatif korelasyon (Bir değişken arttıkça diğeri azalır),
- (0): Hiç korelasyon yoktur (Değişkenler birbirinden bağımsızdır).

Korelasyon analizinin temel kullanım alanları:

1. Değişkenler arası ilişkileri keşfetmek,
2. Bağımsız değişkenleri seçmek (model kurulumunda),
3. Çoklu doğrusal bağlantı sorununu belirlemek,
4. Veri ön işlemede öznitelik seçimi yapmak.

Makine öğrenmesinde, korelasyon analizleri özellikle öznitelik seçimi ve modelin anlamlılığının değerlendirilmesi için kullanılır. Yüksek korelasyona sahip öznitelikler modelden

çıkarılabilir veya modele dahil edilebilir. Tablo 2’de korelasyon katsayı ilişkisi gösterilmiştir.

Tablo 2. Korelasyon katsayısı ilişkisi.

Table 2. Correlation coefficient relationship.

Korelasyon	Negatif	Pozitif
Çok zayıf	-0.25-0,00	0.00-0.25
Zayıf	-0.49- -0,26	0.26-0.49
Orta	-0.69 - -0,50	0.50-0.69
Yüksek	-0.89 - -0,70	0.70-0.89
Çok Yüksek	-0.99 - -0,90	0.90-0.99
Mükemmel	-1.00	1.00

2.6 Döngüsel özellik kodlama

Döngüsel özellik kodlama (Cyclical Feature Encoding), periyodik özellikler (örneğin saat, gün, ay gibi) için kullanılan bir kodlama tekniğidir. Bu özellikler döngüsel bir yapıya sahiptir, yani değerler belli bir noktadan sonra tekrar etmeye başlar.

Geleneksel tek değerli kodlama yöntemlerinde, periyodik özelliklerin döngüsel doğası göz ardı edilir. Bunun yerine döngüsel özellikler farklı kategorik değerler olarak kodlanır. Ancak bu yaklaşım, komşu kategoriler arasındaki ilişkiyi göz ardı eder.

Döngüsel özellik kodlama tekniği ise özelliğin döngüsel yapısını koruyan ve komşu değerlerin ilişkisini yansıtan bir kodlama sağlar. Her döngüsel özellik için sinüs ve kosinüs bileşenleri olmak üzere iki yeni özellik tanımlanır. Orijinal periyodik değer, denklem (10) ve (11)’te gösterilen sinüs ve kosinüs fonksiyonlarına göre kodlanır. Böylece komşu değerler birbirine yakın kodlanır ve özelliğin döngüsel yapısı korunmuş olur.

$$x_{sin} = \sin\left(\frac{2 * \pi * x}{x_{max}}\right) \quad (10)$$

$$x_{cos} = \cos\left(\frac{2 * \pi * x}{x_{max}}\right) \quad (11)$$

Döngüsel özellik kodlama, özellikle zaman serisi problemlerinde ve takvim verilerini içeren regresyon/sınıflandırma görevlerinde faydalı olabilir. Model bu sayede periyodik yapıları daha iyi yakalar.

3 Analiz ve tartışma

Etkili ve doğru bir tahmin modeli tasarlamak için sistemin özelliklerini iyi anlamak ve tüketim davranışlarını etkileyecek çeşitli faktörleri göz önünde bulundurmak gerekir. Kısa vadeli tüketim tahmini için geçmiş tüketim verileri, zaman faktörü ve hava durumu gibi faktörler önemli olurken orta ve uzun vadeli tahminler için geçmiş yük ve hava durumu, farklı kategorilerdeki müşteri sayısı, ekonomik ve demografik veriler gibi faktörlerin dikkate alınması gerekir [40],[28],[30].

Tahmin modellerini geliştirmek için Python programlama dili kullanılmıştır. Açık kaynak kodlu ve tahmin uygulamaları için çok popüler bir programlama dili olan Python, model geliştirmek için çok sayıda kütüphaneye sahiptir ve bu kütüphaneler sürekli geliştirilmektedir.

Geliştirme ortamı olarak Jupyter Notebook kullanılmıştır. Jupyter Notebook başlangıçta Python ve R dilinde yazılmış veri bilimi uygulamaları için geliştirilmiş, canlı kod, denklemler, görselleştirmeler ve anlatı metni içeren belgeler oluşturmak,

yürütmek ve paylaşmak için tasarlanmış açık kaynak etkileşimli bir geliştirici ortamıdır [41]. Bu nedenle veri bilimi, makine öğrenmesi ve derin öğrenme konularında çok kullanılmaktadır. Bu bölümde yapılan çalışma detaylandırılmıştır. Osmangazi Elektrik Dağıtım A.Ş. (OEDAŞ) bölgesi için 01.01.2021-31.12.2023 tarihleri arasındaki saatlik elektrik tüketim verileri EPIAŞ Şeffaflık Platformundan alınmıştır [42]. Aynı tarihler için hava durumu verileri açık kaynaklı ücretsiz bir hava durumu API'si sunan Open-Meteo sitesinden alınmıştır [43]. Bunun ardından veri ön işleme için çalışmalara başlanmıştır. Tüketim verisi bağımlı değişken, diğer alanlar ise bağımsız değişken olarak tanımlanmıştır. Çalışmada kullanılan veriler; takvim verileri, tüketim verileri ve meteorolojik veriler olarak sınıflandırılmıştır ve Tablo 3'te gösterilmiştir. Veri setinde 01.01.2021-31.12.2023 tarihleri arasında 1095*24=26280 saat bulunmaktadır. Kullanılan verilerde ilk olarak eksik değer kontrolü yapılmıştır. Tüketim ve hava durumu verilerinde eksik değer bulunmamaktadır. Veri kontrollerinin peşinden aykırı değer kontrolü yapılmıştır. Bu kontrollerde denklem (6),(7),(8)'deki uç değerler yöntemi tercih edilmiştir. Birinci çeyrek değer ve üçüncü çeyrek değer elde edildikten sonra bunların farkı hesaplanmıştır. Daha sonra alt ve üst değerler denklem (12) ve (13)'e göre hesaplanmıştır.

$$Alt\ Değer = Q_1 - 1.5 * IQR \quad (12)$$

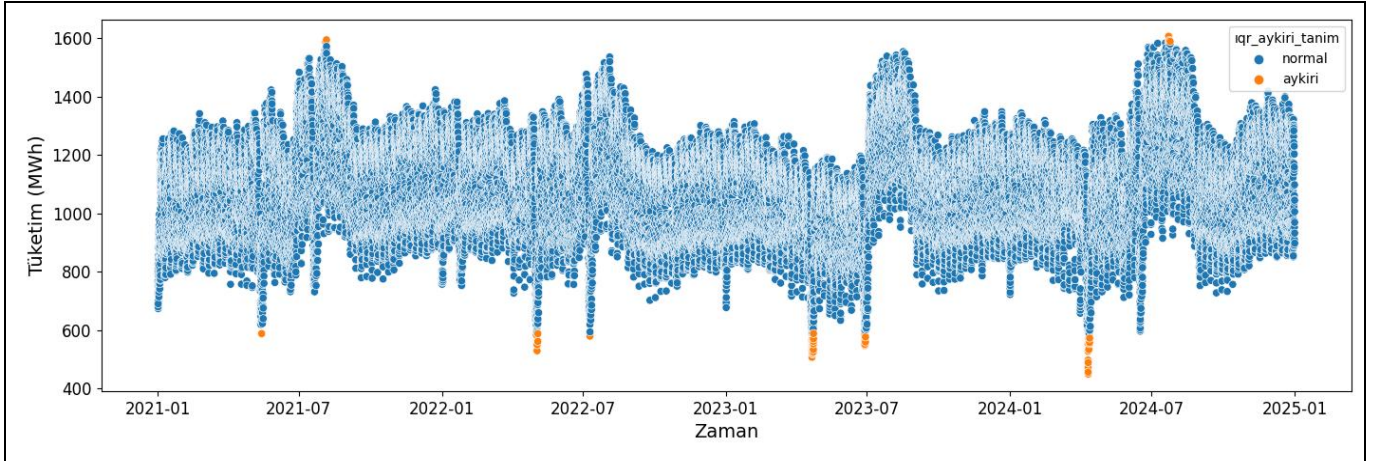
$$Üst\ Değer = Q_3 + 1.5 * IQR \quad (13)$$

Alt değer ve üst değer arasında kalmayan tüketim değerleri ve bu değerlere ait bağımsız değişkenler veri setinden çıkarılmıştır. Veri setindeki aykırı değerler Şekil 8'de gösterilmiştir. Korelasyon analizi için Pearson korelasyon yöntemi kullanılmıştır. Şekil 9 korelasyon matrisini göstermektedir. Tüketim ile arasındaki korelasyon katsayı çok zayıf olan yüzey basıncı, rüzgâr hızı, rüzgâr yönü, toprak nemi ve dağınık radyasyon verisi veri setinden çıkarılmıştır. Veri setindeki döngüsel özellik gösteren bağımsız değişkenlerin (saat, gün, hafta, ay) denklem (10) ve (11) yardımıyla iki boyuta dönüşümleri gerçekleştirilmiştir. Bu işlemlerin ardından ölçeklendirme işlemine geçilmiştir. Ölçeklendirme yöntemi olarak Robust ölçeklendirme kullanılmıştır. Veri ön işleme aşamaları bu işlemlerle birlikte sonlanmıştır. Bu işlemlerin yapılmasının ardından toplam 35 adet bağımsız değişkene sahip bir veri seti oluşmuştur. Bu veri seti eğitim ve test olmak üzere iki kısma bölünmüştür. Veri setinin %80'i eğitim, %20'si test için kullanılmıştır.

Tablo 3. Verilerin sınıflandırılması.

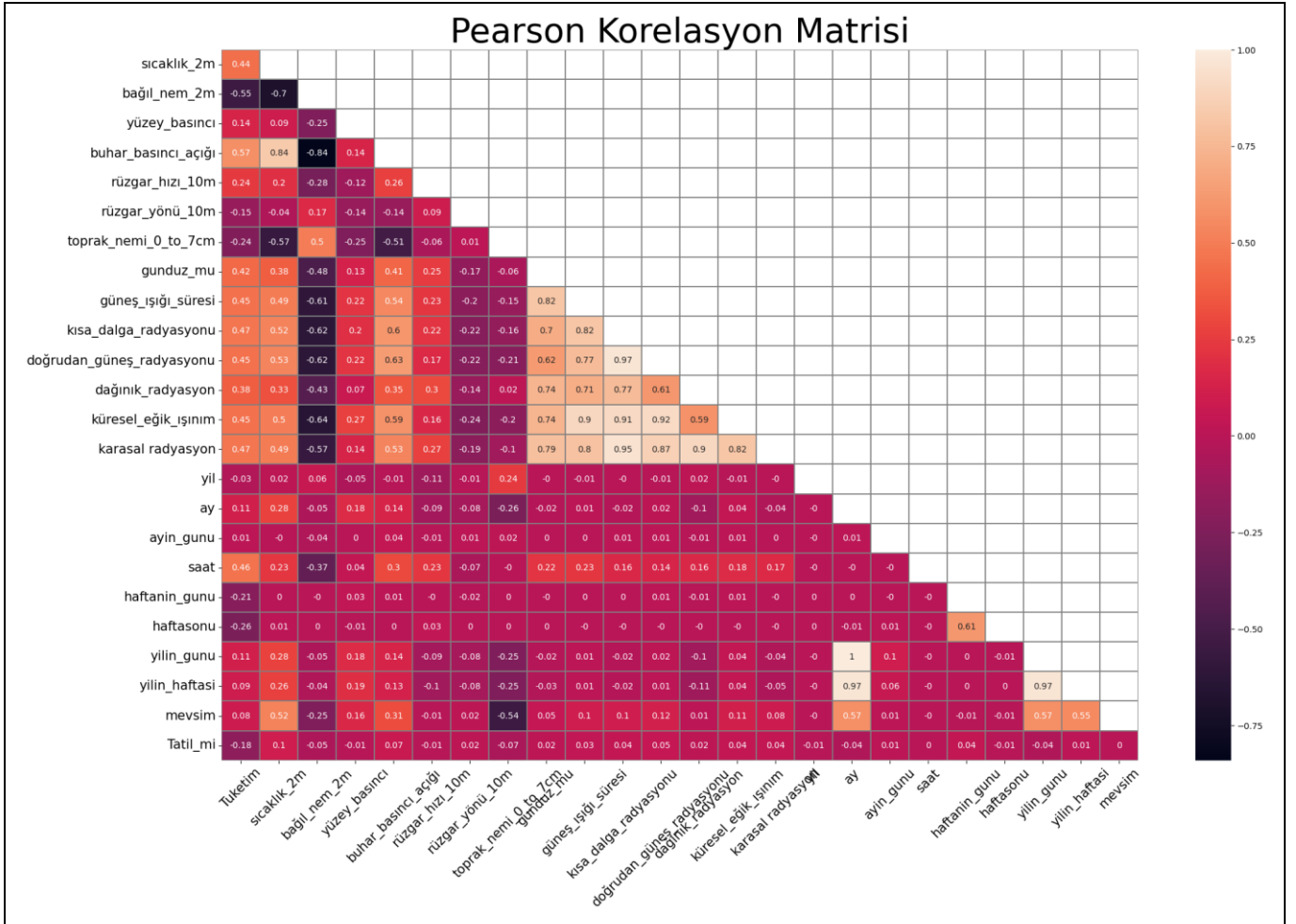
Table 3. Classification of data.

Veri Türü	Değişken	Açıklama
Takvim Verileri	Yıl	2021-2022-2023
	Mevsim	1-Kış, 2-İlkbahar, 3-Yaz, 4-SonBahar
	Ay	1-12
	Hafta	1-52
	Gün	0-6
	Saat	0-23
	Gün Türü	0-Hafta içi, 1-Hafta sonu
	Resmî Tatil	0-Hayır, 1-Evet
	Dini Bayram	0-Hayır, 1-Evet
Tüketim	Geçmiş Tüketim Verisi	t-1, t-2, t-3, t-7
Hava Durumu	Sıcaklık (2 m)	Yerden 2 m yükseklikte hava sıcaklığı (°C)
	Bağıl Nem (2 m)	Yerden 2 metre yükseklikte bağıl nem (%)
	Yüzey Basıncı	Ortalama deniz seviyesine (msl) indirgenmiş atmosferik hava basıncı (hPa)
	Kısa Dalga Radyasyonu	Önceki saatin ortalaması olarak kısa dalga güneş ışıması (W/m ²)
	Doğrudan Güneş Radyasyonu	Normal düzlemde (güneşe dik) bir önceki saatin ortalaması olarak doğrudan güneş ışıması (W/m ²)
	Dağınık radyasyon	Önceki saatin ortalaması olarak yayılan güneş radyasyonu (W/m ²)
	Küresel Eğik Işınım	Önceki saatin ortalaması olarak eğik bir camdan alınan toplam radyasyon (W/m ²)
	Güneş ışıği süresi	WMO tanımına göre, 120 W/m ² 'yi aşan doğrudan normalleştirilmiş ışıınım ile hesaplanan, önceki saatte saat başına düşen güneş ışıği saniye sayısı (Saniye)
	Rüzgâr hızı (10 m)	Yerden 10 yükseklikte rüzgâr hızı. 10 m.'deki rüzgâr hızı standart seviyededir (km/sa.)
	Rüzgâr yönü (10 m)	Yerden 10 metre yükseklikte rüzgâr yönü (°)
	Toprak nemi (0 -7 cm)	0-7 cm derinlikte hacimsel karışım oranı olarak ortalama toprak su içeriği (m ³ /m ³)



Şekil 8. Veri setindeki aykırı değerler.

Figure 8. Outliers in the data set.

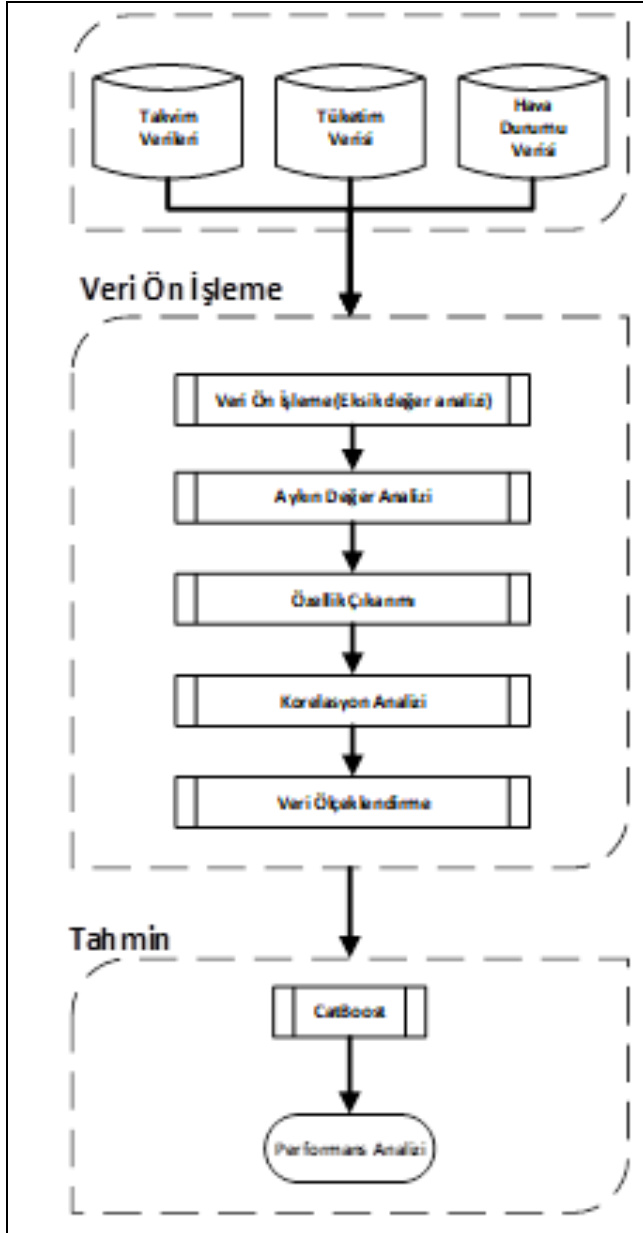


Şekil 9. Korelasyon ilişkisi.

Figure 9. Correlation relationship.

Modellerin eğitim ve test işlemleri sırasında zaman serisi verileri için özel olarak tasarlanan, scikit-learn kütüphanesindeki TimeSeriesSplit çapraz doğrulama yöntemi kullanılmıştır. Veri 5 parçaya bölünerek (n_splits=5) eğitim ve test aşamaları tamamlanmıştır. Şekil 10'da yapılan çalışmalara

ait iş akışı adımları gösterilmiştir. Çalışmada kullanılan modellerin hava durumu verileri dahil edilmeden ve hava durumu verileri dahil edilerek ayrı ayrı eğitim ve performans ölçümleri gerçekleştirilmiştir.



Şekil 10. İş Akışı adımları.

Figure 10. Workflow steps.

Hava durumu verileri dahil edilmeden yapılan tahminlerde, geçmiş yük verileri, gün özellikleri ve tatil günleri gibi bağımsız değişkenler kullanılarak makine öğrenmesi modelleri oluşturulmuş ve modellerin hata oranları kaydedilmiştir. Daha sonra hava durumu verileri, veri setine eklenerek makine öğrenmesi modelleri tekrar oluşturulmuş ve modelleri hata oranları kaydedilmiştir. Kullanılan modellere ait hiper-parametre değerleri Tablo 4'te gösterilmiştir.

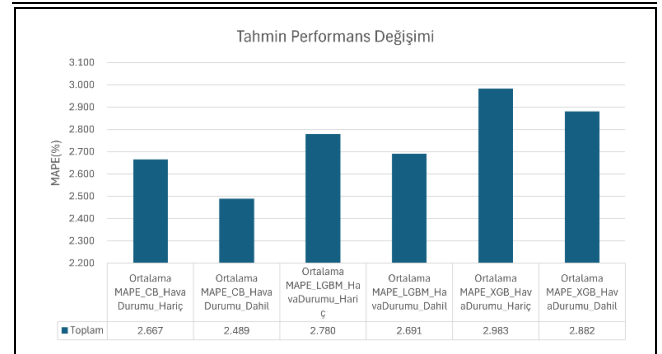
Hava durumu verileri dahil edilmediğinde; Catboost için ortalama MAPE değeri %2.667, LightGBM için ortalama MAPE değeri, %2.780, XGBoost için ortalama MAPE değeri, %2.983'tür. Hava durumu verileri dahil edildiğinde; Catboost için ortalama MAPE değeri, %2.489, LightGBM için ortalama MAPE değeri, %2.691, XGBoost için ortalama MAPE değeri, %2.882'dir. Sonuçlar, hava durumu verilerinin dahil edildiği durumda model performanslarının önemli ölçüde arttığını göstermiştir. Hava durumu verilerinin veri setine eklenmesiyle

birlikte MAPE değerlerinde LightGBM modelinde %3.213, XGBoost modelinde %3.404 ve CatBoost modelinde %6.671 performans iyileşmesi olduğu tespit edilmiştir. Şekil 11'de Modellerin MAPE performans değişimleri gösterilmiştir.

Tablo 4. Modellerin hiper-parametre değerleri.

Table 4. Hyper-parameterization of the models.

Hiper-parametre	CatBoost	LightGBM	XGBoost
iterations/n_estimators	1000	100	100
learning_rate	0.06	0.1	0.3
max_depth	6	-1	6
num_leaves	64	31	-
min_child_samples	-	20	-
min_child_weight	-	0.001	1
subsample	0.8	1	1
subsample_freq	-	0	-
colsample_bytree	-	1	1
reg_lambda	3	0	1
loss_function/objective	RMSE	regression	reg:squarederror
threading	thread_count=-1	n_jobs=-1	n_jobs=None
verbosity	False	-1	0

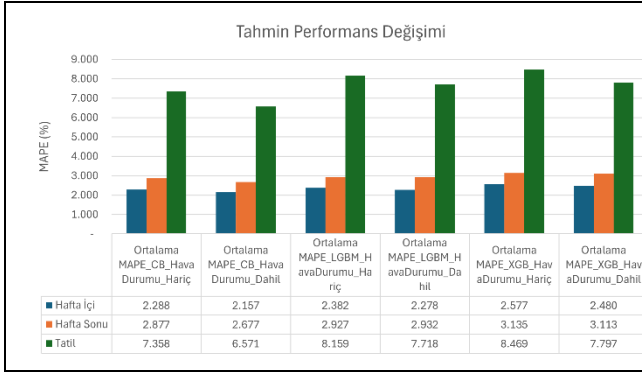


Şekil 11. MAPE tahmin performans değişimi.

Figure 11. MAPE prediction performance change.

Bu durum, hava durumu verilerinin elektrik tüketim tahmini üzerindeki etkilerinin ve yapılan tahminlerin doğruluğu üzerindeki önemli bir rol oynadığını vurgulamaktadır. Hava durumu verilerinin kullanılması, modelin daha iyi bir genelleme yapabilmesine olanak sağlamış ve tahminlerin daha doğru olmasını sağlamıştır. Ayrıca, hava durumu verilerinin dahil edilmesiyle birlikte modelin öğrenme yeteneği ve performansı da artmıştır. Bu da hava durumu verilerinin büyük bir değer taşıdığını ve modellerin daha güvenilir sonuçlar üretebilmesini sağladığını göstermektedir. Özellikle elektrik yük tahmini gibi hava koşullarına bağlı olarak değişkenlik gösteren tahminlerde, hava durumu verilerinin dahil edilmesi büyük bir önem arz etmektedir. Tüm bu faktörler göz önüne alındığında, hava durumu verilerinin model oluştururken dikkate alınması gereken kritik bir faktör olduğu söylenebilir.

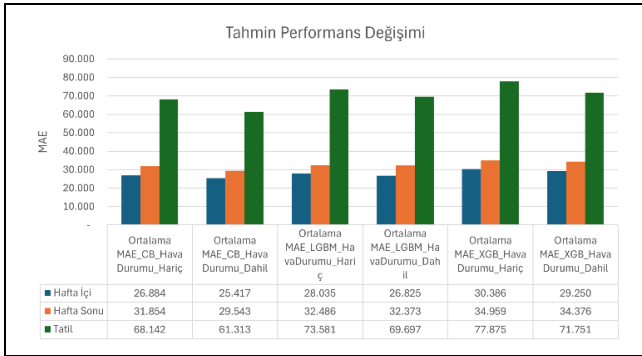
Şekil 12'de farklı gün özelliklerine (hafta içi, hafta sonu, tatil) göre hava durumu verilerinin dahil edildiği ve edilmediği durumlara göre MAPE değişimleri gösterilmiştir.



Şekil 12. Farklı Gün Özelliklerine Göre MAPE değişimleri.

Figure 12. MAPE changes according to different day characteristics.

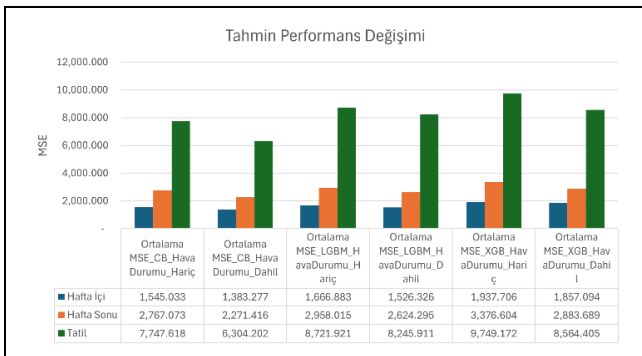
Şekil 13'te farklı gün özelliklerine (hafta içi, hafta sonu, tatil) göre hava durumu verilerin dahil edildiği ve edilmediği durumlara göre MAE değişimleri gösterilmiştir.



Şekil 13. Farklı Gün Özelliklerine Göre MAE değişimleri.

Figure 13. MAE changes according to different Day Characteristics.

Şekil 14'te farklı gün özelliklerine (hafta içi, hafta sonu, tatil) göre hava durumu verilerin dahil edildiği ve edilmediği durumlara göre MSE değişimleri gösterilmiştir.



Şekil 14. Farklı Gün Özelliklerine Göre MSE değişimleri.

Figure 14. MSE changes according to different day characteristics.

4 Sonuçlar

Bu çalışmada, hava durumu verilerinin makine öğrenmesi yöntemiyle yapılan kısa vadeli elektrik yük tahmini üzerine etkisi araştırılmıştır. Hava durumu verilerinin dahil edildiği ve edilmediği veri setleri kullanılarak modeller eğitilmiş ve test

edilmiştir. Veri setleri, eksik değerlerin kontrolü, aykırı değer tespiti ve ölçeklendirme gibi ön işlemlerden geçirilmiştir. Her iki durum için de eğitim ve test veri setleri ayrılarak modeller bu veri setleri üzerinde değerlendirilmiştir.

Çalışma sonuçları, hava durumu verilerinin elektrik yük tahmini performansı üzerinde belirgin bir etkiye sahip olduğunu göstermektedir. Tüm modellerin hava durumu verilerinin dahil edilmediği (önceki) ve edildiği (sonraki) ortalama saatlik MAPE değerleri için Wilcoxon işaretli sıralar testi uygulanmıştır. Tablo 5'te modellere ait Wilcoxon testi p-değerleri gösterilmiştir. Test sonuçları önceki ve sonraki değerler arasında istatistiksel olarak anlamlı bir fark olduğunu göstermektedir.

Tablo 5. Wilcoxon işaretli sıralar test sonuçları.

Table 5. Wilcoxon signed-rank test results.

Model Adı	p-değeri
CatBoost	0.000018
LightGBM	0.000021
XGBoost	0.000102

Bununla birlikte, makalede incelenen CatBoost, LightGBM ve XGBoost gibi farklı makine öğrenmesi yöntemleri arasında performans açısından belirgin farklılıklar olduğu gözlemlenmiştir. Bu farklılıklar, modellerin veri analizi ve öğrenme süreçlerinde farklı stratejiler izlediğini göstermektedir. CatBoost modeli hava durumu verilerinin dahil edildiği ve edilmediği her iki durumda da diğer modellere kıyasla daha üstün performans göstermiştir. CatBoost, daha düşük hata oranı ve daha yüksek doğruluk ile elektrik tüketim tahminlerinde en iyi sonuçları vermiştir.

Çalışma sonuçları, hava durumu verilerinin elektrik yük tahmini performansı üzerinde belirgin bir etkiye sahip olduğunu göstermektedir. Ayrıca, makine öğrenmesi modellerinin hava durumu verileriyle beslenmesinin tahmin doğruluğunu artırdığı tespit edilmiştir. Elektrik tüketimi ve sıcaklık arasında pozitif bir korelasyon olduğu, sıcaklık arttıkça elektrik tüketiminin artma eğilimi olduğu Şekil 2'de gösterilmiştir. Elektrik tüketimi ve nem arasında negatif bir korelasyon olduğu, nem arttıkça tüketimin azalma eğiliminde olduğu Şekil 3'te gösterilmiştir. Hava durumu verilerinin veri setine eklenmesiyle birlikte LightGBM modelinde %3.213, XGBoost modelinde %3.404 ve CatBoost modelinde %6.671 MAPE değerlerinde performans iyileşmesi olduğu gözlemlenmiştir.

Bu sonuçlar, enerji sektöründe planlama ve operasyonel süreçlerin daha güvenilir bir şekilde yürütülmesine katkı sağlayabilir ve elektrik talebinin yönetimi ve güç üretim planlaması gibi kritik alanlarda daha kesin ve etkili kararlar almak için kullanılabilir. Hava durumu verilerinin doğru bir şekilde analiz edilmesi, enerji şirketlerinin talep tahminlerini iyileştirmelerine olanak tanır ve böylece kaynakların daha verimli bir şekilde kullanılmasını sağlar.

Hava durumu tahminlerinin güvenilirliği arttıkça, enerji tedarikçileri arasında yapılan anlaşmalar ve enerji dağıtım stratejileri daha etkili bir şekilde yönetilebilir. Yüksek doğruluk seviyelerine dayalı bir tahmin sistemi, enerji şirketlerine talep dalgalanmalarına hızlı bir şekilde tepki verebilme yeteneği sağlar. Ayrıca, hava durumu verilerini kullanarak enerji tüketim desenlerinin analizi kolaylaşır ve enerji şirketlerine enerji kaynaklarının kullanımını optimize etme ve enerji verimliliğini artırma konusunda önemli bilgiler sağlar.

Ancak, bu alanda daha ileri araştırmalar ve iyileştirmeler yapılabilir. Gelecekte yapılabilecek çalışmalar aşağıdaki gibi özetlenebilir:

- Daha geniş ve farklı bölgelerden elde edilen veri setleri kullanılarak, farklı hava durumu özelliklerinin elektrik yük tahmini performansı üzerindeki etkilerini incelenebilir,
- Elektrik tüketimini etkileyebilecek başka değişkenler eklenerek modellerin performansı artırılabilir. Örneğin, sosyal etkinlikler, ekonomik göstergeler ve enerji fiyatları gibi faktörler tüketim üzerinde önemli etkilere sahip olabilir,
- Hiper-parametre optimizasyonu ve model düzenleme teknikleri kullanılarak modellerin performansları daha da iyileştirilebilir. Bu bağlamda, Bayes optimizasyonu, genetik algoritmalar ve grid search gibi yöntemler kullanılabilir,
- LSTM, GRU gibi zaman serisi tahmininde güçlü olan derin öğrenme modelleri ile mevcut modeller karşılaştırılabilir. Derin öğrenme modellerinin kompleks yapıları, büyük veri setlerinde daha iyi performans gösterebilir,
- Farklı modellerin bir araya getirilerek oluşturulan topluluk yöntemleri kullanılabilir. Bu teknik, bireysel modellerin zayıf yönlerini dengeleyerek genel performansı artırabilir,
- Geliştirilen modellerin gerçek zamanlı tahmin sistemlerine entegre edilmesi ve gerçek zamanlı veri akışı ile sürekli güncellenmesi sağlanabilir. Böylece tahminlerin doğruluğunu ve güvenilirliğini artırabilir.

Sonuç olarak, bu çalışma hava durumu verilerinin elektrik yük tahmini modellerine dahil edilmesinin performansı artırdığını ortaya koymakta olup, gelecekteki çalışmalar için çeşitli araştırma ve geliştirme alanları önermektedir. Bu öneriler doğrultusunda yapılacak çalışmalar, elektrik tüketim tahminlerinin doğruluğunu ve güvenilirliğini daha da artırabilir.

5 Conclusions

In this study, the impact of weather and different machine learning methods on short-term electricity load forecasting is investigated. Models are trained and tested using datasets with and without the inclusion of weather data. The datasets were preprocessed by checking for missing values, outlier detection and scaling. For both cases, training and test datasets were separated and the models were evaluated on these datasets.

The results show that the models that include weather data perform better than the other models. Wilcoxon signed-rank test was applied for the average hourly MAPE values of all models without (before) and with (after) the inclusion of weather data. Table 5 shows the Wilcoxon test p-values of the models.

Table 5. Wilcoxon signed-rank test results.

Model Name	p-value
CatBoost	0.000018
LightGBM	0.000021
XGBoost	0.000102

The test results show that there is a statistically significant difference between the before and after values. However, there are significant differences in performance between the

different machine learning methods analyzed in this paper, such as CatBoost, LightGBM and XGBoost. These differences indicate that the models follow different strategies in data analysis and learning processes. The CatBoost model outperformed the other models both with and without the inclusion of weather data. CatBoost gave the best results in electricity consumption forecasts with lower error rate and higher accuracy.

The study results show that weather data has a significant impact on electricity load forecasting performance. It is also found that feeding machine learning models with weather data improves the forecast accuracy. Figure 2 shows that there is a positive correlation between electricity consumption and temperature, and electricity consumption tends to increase as temperature increases. Figure 3 shows that there is a negative correlation between electricity consumption and humidity, and consumption tends to decrease as humidity increases. With the addition of weather data to the dataset, it is observed that there is a performance improvement with MAPE values of 3.213% for the LightGBM model, 3.404% for the XGBoost model and 6.671% for the CatBoost model.

These results can contribute to more reliable planning and operational processes in the energy sector and can be used to make more precise and effective decisions in critical areas such as electricity demand management and power generation planning. Accurate analysis of weather data allows energy companies to improve their demand forecasts, thus enabling more efficient use of resources.

As the reliability of weather forecasts increases, agreements between energy suppliers and energy distribution strategies can be managed more effectively. A forecasting system based on high levels of accuracy provides energy companies with the ability to react quickly to demand fluctuations. Furthermore, analysis of energy consumption patterns using weather data becomes easier and provides energy companies with important information to optimize the use of energy resources and improve energy efficiency.

However, further research and improvements can be made in this area. Future work can be summarized as follows:

- Using larger data sets from different regions, the effects of different weather characteristics on electrical load forecasting performance can be studied. The performance of the models can be enhanced by incorporating additional variables that may influence electricity consumption. For instance, factors such as social events, economic indicators, and energy prices can have significant impacts on consumption.
- The performance of the models can be further improved by using hyperparameter optimization and model regularization techniques. In this context, methods such as Bayesian optimization, genetic algorithms and grid search can be used.
- Deep learning models such as LSTM, GRU which are powerful in time series forecasting can be compared with existing models. The complexity of deep learning models may perform better on large data sets.
- Ensemble methods that combine different models can be used. This technique can improve overall performance by compensating for the weaknesses of individual models.

- The developed models can be integrated into real-time forecasting systems and continuously updated with real-time data flow. Thus, it can increase the accuracy and reliability of forecasts.

In conclusion, this study demonstrates that incorporating weather data into electricity load forecasting models improves performance and suggests several research and development areas for future work. Future studies in line with these recommendations can further improve the accuracy and reliability of electricity consumption forecasts resources.

6 Yazar katkı beyanı

Gerçekleştirilen çalışmada Murat Ünlü fikrin oluşması, tasarımın yapılması, literatür taraması, makale yazılması başlıklarında ve elde edilen sonuçların değerlendirilmesi; Habil Ergen literatür taraması, elde edilen sonuçların değerlendirilmesi, sonuçların incelenmesi ve makale yazılması başlıklarında katkı sunmuşlardır.

7 Etik kurul onayı ve çıkar çatışması beyanı

"Hazırlanan makalede etik kurul izni alınmasına gerek yoktur".
"Hazırlanan makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır".

8 Kaynaklar

- [1] Ibrahim IA, Hossain MJ. "Short-term multivariate time series load data forecasting at low-voltage level using optimised deep-ensemble learning-based models". *Energy Conversion and Management*, 296, 117663, 2023.
- [2] Hafiz F, Queiroz ARD, Husain I. "Solar generation, storage, and electric vehicles in power grids: Challenges and solutions with coordinated control at the residential level". *IEEE Electrification Magazine*, 6(4), 83-90, 2018.
- [3] Pallonetto F, Jin C, Mangina E. "Forecast electricity demand in commercial building with machine learning models to enable demand response programs". *Energy and AI*, 7, 100121, 2022.
- [4] Mehigan L, Deane JP, Gallachóir BPÓ, Bertsch V. "A review of the role of distributed generation (DG) in future electricity systems". *Energy*, 163, 822-836, 2018.
- [5] Hong T, Fan S. "Probabilistic electric load forecasting: A tutorial review". *International Journal of Forecast*, 32(3), 914-938, 2016.
- [6] Şekerçi H. "Load demand forecast of organized industrial zone and imbalance cost analysis". *Innovations in Intelligent Systems and Applications Conference*, Ankara, Turkey, 16-18 October 2024.
- [7] Chandrasekaran R, Paramasivan SK. "Advances in deep learning techniques for short-term energy load forecasting applications: a review," *Archives of Computational Methods in Engineering*, 32, 663-692, 2024.
- [8] Ekonomou L, Christodoulou CA, Mladenov V. "A short-term load forecasting method using artificial neural networks and wavelet analysis". *International Journal of Power Systems*, 1, 64-68, 2016.
- [9] Tsalikidis N, Mystakidis A, Tjortjis C, Koukaras P, Ioannidis D. "Energy load forecasting: one-step ahead hybrid model utilizing ensembling". *Computing*, 106, 241-273, 2024.
- [10] Kuo PH, Huang CJ. "A high precision artificial neural networks model for short-Term energy load forecasting". *Energies*, 11(1), 213, 2018.
- [11] Farsi B, Amayri M, Bouguila N, Eicker U. "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach". *IEEE Access*, 9, 31191-31212, 2021.
- [12] Enerji Piyasası Düzenleme Kurumu (EPDK), "Elektrik Piyasası Dengeleme ve Uzlaştırma Yönetmeliği". <https://www.epdk.gov.tr/Detay/Icerik/3-0-159-3/yonetmelikler> (01.02.2024).
- [13] Iwafune Y, Yagita Y, Ikegami T, Ogimoto K. "Short-term forecasting of residential building load for distributed energy management". *IEEE International Energy Conference (ENERGYCON)*, Cavtat, Croatia, 13-16 May 2014.
- [14] Ekinci F. "YSA ve ANFIS tekniklerine dayalı enerji tüketim tahmin yöntemlerinin karşılaştırılması". *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 7(3), 1029-1044, 2019.
- [15] Kotecha R, Ukarande S, Hosalikar K, Maniar P, Maru S, Pallikuth D, Biswas T, Rane V. "Short-Term load demand forecasting based on weather and influencing factors using deep neural network experts for sustainable development goal 7". *SN Computer Science*, 5(2), 253, 2024.
- [16] Nguyen NA, Dang TD, Verdú E, Solanki EK. "Short-term forecasting electricity load by long short-term memory and reinforcement learning for optimization of hyper-parameters". *Evolutionary Intelligence*, 16(5), 1729-1746, 2023.
- [17] Işık G, Ögüt H, Mutlu M. "Deep learning based electricity demand forecasting to minimize the cost of energy imbalance: A real case application with some fortune 500 companies in Türkiye". *Engineering Applications of Artificial Intelligence*, 118, 105664, 2023.
- [18] Son J, Cha J, Kim H, Wi YM. "Day-Ahead short-term load forecasting for holidays based on modification of similar days' load profiles". *IEEE Access*, 10, 17864-17880, 2022.
- [19] Sultana N, Hossain SMZ, Almuhammadi S, Düşteğör D. "Bayesian optimization algorithm-based statistical and machine learning approaches for forecasting short-term electricity demand". *Energies*, 15(9), 3425, 2022.
- [20] Porteiro R, Callejo LH, Nesmachnow S. "Electricity demand forecasting in industrial and residential facilities using ensemble machine learning". *Universidad de Antioquia*, 102, 9-25, 2022.
- [21] Barman M, Choudhury NBD, Sutradhar S. "A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India". *Energy*, 145, 710-720, 2018.
- [22] Başoğlu B, Bulut M. "Kısa dönem elektrik talep tahminleri için yapay sinir ağları ve uzman sistemler tabanlı hibrit sistem geliştirilmesi". *Journal of the Faculty of Engineering and Architecture of Gazi University*, 32(2), 575-583, 2017.
- [23] Zhang L, Jánošík D. "Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches". *Expert Systems with Applications*, 241, 122686, 2024.
- [24] Madrid EA, Antonio N. "Short-term electricity load forecasting with machine learning". *Information*, 12(2), 1-21, 2021.

- [25] Wang Y, Sun S, Chen X, Zeng X, Kong Y, Chen Y, Wang T. "Short-term load forecasting based on SVM and XGBoost". *International Journal of Electrical Power & Energy Systems*, 129, 106830, 2021.
- [26] Zhang XL, Song Z. "A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning". *Renewable and Sustainable Energy Reviews*, 119, 109632, 2020.
- [27] Morais LBS, Aquila G, Faria VAD, Lima LMM, Lima JWM, Queiroz AR. "Short-term load forecasting using neural networks and global climate models: An application to a large-scale electrical power system". *Applied Energy*, 348, 121439, 2023.
- [28] Feinberg EA, Genethliou D. *Load Forecasting*. Editors: Chow JH, Wu FF, Momoh J. Applied Mathematics for Restructured Electric Power Systems, 269-285, New York, USA, Springer Press, 2005.
- [29] Sobhani M, Hong T, Marti C. "Temperature anomaly detection for electric load forecasting". *International Journal of Forecasting*, 36(2), 324-333, 2020.
- [30] Fahad MU, Arbab N. "Factor affecting short term load forecasting". *Journal of Clean Energy Technologies*, 2(4), 305-309, 2014.
- [31] Erişen E, Iyigun C, Tanrısever F. "Short-term electricity load forecasting with special days: an analysis on parametric and non-parametric methods". *Annals of Operations Research*, 2017. <https://doi.org/10.1016/j.soidyn.2019.106004>.
- [32] Friedman JH. "Greedy function approximation: a gradient boosting machine". *Annals of Statistics*, 29(5), 1189-1232, 2001.
- [33] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. "CatBoost: unbiased boosting with categorical features". *32nd International Conference on Neural Information Processing Systems*, Montréal, Canada, 3-8 December 2018.
- [34] Aydın ZE, Erdem Bİ, Çiçek ZİE. "Prediction bike-sharing demand with gradient boosting methods". *Pamukkale University Journal of Engineering Sciences*, 29(8), 824-832, 2023.
- [35] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. "LightGBM: A highly efficient gradient boosting decision tree". *Advances in Neural Information Processing Systems Conference (NIPS)*, Long Beach CA, USA, 4-9 December 2017.
- [36] Chen T, Guestrin C. "XGBoost: A scalable tree boosting system". in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Association for Computing Machinery. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, 13-17 August, 2016.
- [37] Lewis CD. "Industrial and Business Forecasting Methods: a practical guide to exponential smoothing and curve fitting". *Journal of Forecasting*, 2(2), 194-196, 1982.
- [38] Weisstein EW. "Correlation". <https://mathworld.wolfram.com/Correlation.html> (08.02.2024).
- [39] Rodgers JL, Nicewander WA. "Thirteen Ways to Look at the Correlation Coefficient". *The American Statistician*, 42, 186-195, 1988.
- [40] Emhamed AA, J. Shrivastava J. "Electrical load distribution forecasting utilizing support vector model (SVM)". *Materials Today: Proceedings*, 47(1), 41-46, 2021.
- [41] Microsoft. "microsoft.com/tr-tr". <https://learn.microsoft.com/tr-tr/azure/cosmos-db/notebooks-overview> (25.02.2024).
- [42] EPİAŞ. "Seffaflik". <https://seffaflik.epias.com.tr/home> (20.01.2024).
- [43] OpenMeteo, "OpenMeteo". <https://open-meteo.com/en/docs/historical-weather-api> (15.01.2024).