T.C.
HALİÇ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE HAKKÂRİ YÜKSEKOVA SELAHADDİN EYYUBİ HAVALİMANI DOĞALGAZ TÜKETİM MİKTARININ TAHMİN EDİLMESİ

MEHMET ERGÜN AZİZOĞLU
MİMARLIK ANABİLİM DALI

YÜKSEK LİSANS TEZİ

DANIŞMAN
Dr. Öğr. Üyesi Dr. Öğr. Üyesi Erdem ÇOBAN

İSTANBUL-2025

T.C.
HALİÇ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE HAKKÂRİ YÜKSEKOVA SELAHADDİN EYYUBİ HAVALİMANI DOĞALGAZ TÜKETİM MİKTARININ TAHMİN EDİLMESİ

MEHMET ERGÜN AZİZOĞLU
MİMARLIK ANABİLİM DALI

YÜKSEK LİSANS TEZİ

DANIŞMAN
Dr. Öğr. Üyesi Erdem ÇOBAN

İSTANBUL-2025

Mehmet Ergün AZİZOĞLU tarafından hazırlanan "MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE HAKKÂRİ YÜKSEKOVA SELAHADDİN EYYUBİ HAVALİMANI DOĞALGAZ TÜKETİM MİKTARININ TAHMİN EDİLMESİ" adlı tez çalışması, aşağıdaki jüri tarafından OY BİRLİĞİ ile Haliç Üniversitesi Lisansüstü Eğitim Enstitüsü Mimarlık Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Dr. Öğr. Üyesi Erdem ÇOBAN

Mimarlık Fakültesi, Haliç Üniversitesi İmza ..……..

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Üye: Dr. Öğr. Üyesi Kemal Ferit ÇETİNTAŞ

Mimarlık Fakültesi, Haliç Üniversitesi İmza ..……..

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Üye: Dr. Öğr. Üyesi Gökhan BALCIOĞLU

İç Mimarlık Fakültesi, İstanbul Gelişim Üniversitesi İmza ..……..

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Tez Savunma Tarihi: 20 / 06 / 2025

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

…….

Ünvanı Adı SOYADI

Lisansüstü Enstitüsü Müdürü

ETİK BEYANI

Haliç Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Mehmet Ergün AZİZOĞLU

Temmuz 2025

ÖZET

MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE HAKKÂRİ YÜKSEKOVA SELAHADDİN EYYUBİ HAVALİMANI DOĞALGAZ TÜKETİM MİKTARININ TAHMİN EDİLMESİ

Haliç Üniversitesi
Lisansüstü Eğitim Enstitüsü
Mimarlık Anabilim Dalı, Yüksek Lisans Tezi
Danışman: Dr. Öğr. Üyesi Erdem ÇOBAN
Temmuz 2025, 80 sayfa

Enerji, modern yaşamın sürdürülebilirliği için temel bir ihtiyaç olup, talep yönetimi ve arz tahmini konuları günümüzde stratejik bir öneme sahiptir. Bu çalışmada, Türkiye'nin doğusunda bulunan Hakkâri Yüksekova Selahaddin Eyyubi Havalimanı'na ait meteorolojik veriler kullanılarak doğalgaz tüketiminin tahmini amaçlanmıştır. Modelleme sürecinde sıcaklık, çiğ noktası, yüzey basıncı, güneş radyasyonu, yatay ve dikey rüzgâr bileşenlerinden oluşan 300 gözlemli çok değişkenli zaman serisi veri seti kullanılmıştır. Makine öğrenmesi algoritmalarından LightGBM, XGBoost, Rastgele Orman (Random Forest-RF), Destek Vektör Regresyonu (Support Vector Regression-SVR) ve Çoklu Doğrusal Regresyon(Multi Lineer Regression-MLR) ile birlikte bir derin öğrenme algoritması olan LSTM (Long Short-Term Memory) uygulanmış ve modellerin başarımları karşılaştırılmıştır. Test verisi üzerindeki MSE değerleri XGBoost için 0.0003, LightGBM için 4.01, LSTM için 4.62, SVR için 5.65, RF için 2.35 ve MLR için 5.80 olarak hesaplanmıştır. R^2 değerlerine göre ise XGBoost 1.0000 ile mükemmel uyum gösterirken, LightGBM 0.9548, LSTM 0.9479, SVR 0.9362, RF 0.9734 ve MLR 0.9345 değerlerine ulaşmıştır. MAE açısından en düşük hata XGBoost (0.01) ve LightGBM (1.58) modellerinde gözlemlenmiştir. SMAPE değerleri ise sırasıyla XGBoost %0.30, LightGBM %29.09, LSTM %31.16, SVR %37.30, RF %18.82 ve MLR %37.30 olarak kaydedilmiştir. Ayrıca, sıcaklık temelli Isıtma Derece Günü (HDD) katsayısı ile doğalgaz tüketimi arasında kurulan doğrusal ilişki değerlendirilmiş ve sıcaklık düşüşlerinin tüketimi anlamlı şekilde artırdığı gözlemlenmiştir. Bulgular, özellikle XGBoost algoritmasının doğalgaz talep tahmininde üstün performans sergilediğini ve enerji yönetimi planlamalarında etkin bir araç olduğunu göstermektedir.

Anahtar Kelimeler: Doğalgaz Tüketimi, LSTM, Makine Öğrenmesi, XGBoost, LightGBM, Isıtma Derece Günü (HDD)

ABSTRACT

PREDICTION OF NATURAL GAS CONSUMPTION AT HAKKÂRI YÜKSEKOVA SELAHADDIN EYYUBI AIRPORT USING MACHINE LEARNING ALGORITHMS

Haliç University
Institute of Graduate Studies
Department of Architecture, Master's Thesis
Supervisor: Asst. Prof. Dr. Erdem ÇOBAN
July 2025, 80 pages

Energy is a fundamental necessity for sustaining modern life, and accurate forecasting of demand and supply has become a strategic priority. In this study, natural gas consumption forecasting was conducted using meteorological data from Hakkâri Yüksekova Selahaddin Eyyubi Airport, located in eastern Turkey. A multivariate time series dataset consisting of 300 observations was prepared, including variables such as temperature, dew point, surface pressure, solar radiation, and horizontal and vertical wind components. Machine learning models such as LightGBM, XGBoost, Random Forest (RF), Support Vector Regression (SVR), and Multiple Linear Regression (MLR), along with a deep learning model Long Short-Term Memory (LSTM), were employed and their performances were compared. On the test set, MSE values were calculated as 0.0003 for XGBoost, 4.01 for LightGBM, 4.62 for LSTM, 5.65 for SVR, 2.35 for RF, and 5.80 for MLR. In terms of R^2 scores, XGBoost achieved a perfect value of 1.0000, followed by RF (0.9734), LightGBM (0.9548), LSTM (0.9479), SVR (0.9362), and MLR (0.9345). Regarding MAE, the lowest values were obtained by XGBoost (0.01) and LightGBM (1.58). The SMAPE metric revealed values of 0.30% for XGBoost, 29.09% for LightGBM, 31.16% for LSTM, 37.30% for both SVR and MLR, and 18.82% for RF. In addition, the Heating Degree Day (HDD) indicator was utilized to model the linear relationship between temperature and gas consumption, revealing a strong positive correlation. The findings underscore the high predictive capability of XGBoost, making it a robust tool for energy demand management and planning.

Keywords: Natural Gas Consumption, LSTM (Long Short-Term Memory), Machine Learning, XGBoost, LightGBM, Heating Degree Day (HDD).

TEŞEKKÜR

Yüksek lisans tez çalışmam süresince bilgi, sabır ve desteğini esirgemeyen, akademik yolculuğumda bana rehberlik eden değerli danışman hocam Sayın Erdem ÇOBAN'a en içten teşekkürlerimi sunarım. Gerek bilimsel yaklaşımı gerekse yönlendirmeleri sayesinde bu çalışmayı şekillendirme fırsatı buldum.

Bu süreçte, manevi desteğini daima yanımda hissettiğim, beni koşulsuz şekilde destekleyen, sabırları ve sevgileriyle her zaman güç veren aileme sonsuz teşekkür ederim. Varlıklarıyla bana güven aşılayan ailem olmasaydı, bu yolculuğu böylesine kararlılıkla sürdüremezdim.

Ayrıca eğitim hayatım boyunca bana katkıda bulunan tüm hocalarıma, değerli arkadaşlarıma ve sürece doğrudan veya dolaylı katkı sağlayan herkese teşekkür ederim.

Bu tez, yalnızca akademik bir çalışma değil; aynı zamanda bana inananların desteğiyle şekillenmiş bir emeğin ürünüdür.

İstanbul, 2025 Mehmet Ergün AZİZOĞLU

İÇİNDEKİLER

Sayfa

ÇİZELGELER DİZİNİ

Sayfa

ŞEKİLLER DİZİNİ

Sayfa

1. GİRİŞ

Enerji, ekonomik kalkınmanın ve toplumsal refahın sürdürülebilirliği açısından vazgeçilmez bir kaynaktır. Küresel ölçekte artan nüfus, kentleşme ve sanayileşme eğilimleri, enerji talebinde ciddi artışlara yol açmakta; buna bağlı olarak enerji kaynaklarının verimli, planlı ve sürdürülebilir biçimde yönetilmesini zorunlu kılmaktadır. Bu bağlamda, enerji tüketim kalıplarının anlaşılması ve geleceğe yönelik tahmin edilmesi, enerji arz-talep dengesinin sağlanması, çevresel etkilerin azaltılması ve stratejik enerji planlamalarının geliştirilmesi açısından kritik rol oynamaktadır.

Fosil yakıtlar arasında yer alan doğalgaz, dünya genelinde ve özellikle Türkiye'de ısınma, elektrik üretimi ve sanayi sektörlerinde yaygın olarak kullanılan birincil enerji kaynaklarından biridir. Temiz yanma özelliği, düşük karbon salımı ve yüksek verimliliği nedeniyle diğer fosil yakıtlara göre daha çevreci bir alternatif olarak değerlendirilen doğalgaz, Türkiye'deki konut tüketiminin yanı sıra kamu hizmeti sağlayan altyapıların da ana enerji girdileri arasında yer almaktadır. Doğalgazın arzının zamanında ve yeterli düzeyde sağlanabilmesi için, tüketim talebinin önceden doğru biçimde tahmin edilmesi büyük önem arz etmektedir.

Bu doğrultuda, enerji talep tahmini literatüründe geleneksel istatistiksel yöntemlerin yanı sıra, son yıllarda yapay zekâ ve makine öğrenmesi tabanlı yöntemlerin yaygınlaştığı görülmektedir. Özellikle büyük veri setleri ve çok değişkenli zaman serisi verilerinin analizi için güçlü yapay öğrenme algoritmaları kullanılarak daha doğru ve hızlı tahminler elde edilebilmektedir. Bu çalışma da benzer bir yaklaşımla, Türkiye'nin doğu bölgesinde yer alan Hakkâri Yüksekova Selahaddin Eyyubi Havalimanı özelinde, doğalgaz tüketiminin makine öğrenimi algoritmaları ile tahmin edilmesini amaçlamaktadır.

Doğalgaz tüketimini etkileyen başlıca faktörlerin başında meteorolojik koşullar gelmektedir. Sıcaklık, çiğ noktası, yüzey basıncı, güneş radyasyonu ve rüzgâr gibi hava durumu parametreleri, ısıtma ihtiyacını doğrudan etkileyerek tüketim miktarlarında dalgalanmalara neden olmaktadır. Bu nedenle, söz konusu meteorolojik

değişkenler doğalgaz talep modellerinde girdi olarak kullanılmaktadır. Ayrıca, sıcaklık temelli ısınma gereksinimini daha somut biçimde yansıtmak amacıyla geliştirilmiş olan Isıtma Derece Günü (HDD) gibi iklimsel göstergeler de bu tür çalışmalarda sıkça başvurulan türetilmiş değişkenlerdendir.

Çalışmada, veri ön işleme, özellik mühendisliği ve algoritma eğitimi süreçleri detaylı bir biçimde ele alınmış, farklı makine öğrenmesi algoritmaları (Multi-Linear Regression, Support Vector Regression, Random Forest, LightGBM, XGBoost, LSTM) performans açısından karşılaştırılmıştır. Modellerin değerlendirilmesinde, MSE (Ortalama Kare Hataları), R^2 (Belirleme Katsayısı), MAE (Ortalama Mutlak Hata), MAPE (Ortalama Mutlak Yüzde Hatası) ve SMAPE (Simetrik Ortalama Mutlak Yüzde Hata) gibi istatistiksel ölçütler kullanılarak tahmin doğruluğu analiz edilmiştir. Bu çerçevede, özellikle LightGBM algoritmasının düşük hata oranları ve yüksek açıklayıcılık düzeyi ile dikkat çektiği gözlemlenmiştir.

Doğalgaz tüketiminin doğru biçimde tahmin edilmesi, sadece enerji planlamasına değil; maliyet yönetimine, karbon emisyonlarının azaltılmasına ve sürdürülebilir kamu hizmetlerinin sürdürülmesine de katkı sağlamaktadır. Ayrıca, bu çalışma kapsamında geliştirilen modelin farklı bölgelerde ve farklı tüketim senaryolarında da genellenebilir bir tahmin yaklaşımı sunduğu düşünülmektedir. Bu yönüyle tez hem yerel enerji yönetimi politikalarına hem de veri tabanlı akıllı sistemlerin geliştirilmesine katkı sunmayı hedeflemektedir.

2. LİTERATÜR ARAŞTIRMASI

Makine öğrenmesi temelli tahmin modelleri, son yıllarda enerji sistemlerinin analizinde ve kestiriminde giderek daha fazla kullanılmaya başlanmıştır. Özellikle meteorolojik değişkenlere bağlı olarak değişkenlik gösteren enerji talebi, bu tür algoritmaların sunduğu esnek ve veri odaklı yaklaşımlar sayesinde daha doğru bir şekilde modellenebilmektedir. Bu bağlamda, farklı enerji sistemlerinde yapılan çok sayıda çalışma, çeşitli makine öğrenimi tekniklerinin tahmin performanslarını karşılaştırmalı olarak analiz etmeyi amaçlamıştır. Yapay sinir ağları, destek vektör makineleri, karar ağaçları ve regresyon modelleri gibi yöntemler, literatürde sıklıkla uygulanan ve başarıyla sonuç veren algoritmalar arasında yer almaktadır. Bu çalışmalar hem tahmin doğruluğu hem de model açıklanabilirliği açısından birbirinden farklı avantajlar sunmakta; böylece enerji verimliliği, sistem optimizasyonu ve stratejik planlama süreçlerine katkı sağlamaktadır.

Sayan (2022) tarafından gerçekleştirilen çalışmada, bir hava ısıtmalı güneş kolektörünün (HGK) deneysel veriler ile hesaplanan enerji verimlilik değerleri, yapay sinir ağı (YSA), karar ağacı (KA), destek vektör makinesi (SVM) ve pace regresyon modelleri kullanılarak tahmin edilmiştir. Söz konusu algoritmalar birbirleriyle karşılaştırılmış ve deneysel verilere en iyi uyum sağlayan modelin karar ağacı olduğu tespit edilmiştir.

Özer (2022), makine öğrenmesi algoritmalarından polinom regresyon ve destek vektör regresyon (SVR) yöntemleri kullanılmıştır. Osmaniye İl Özel İdaresi'ne ait 990 kW kurulu güce sahip Güneş Enerjisi Santrali'nin elektrik üretim tahminine yönelik çalışmada en başarılı tahmin performansı SVR algoritması ile elde edilmiştir. Bu bağlamda, SVR'nin polinom regresyona göre daha uygun bir yöntem olduğu ortaya konmuştur. Ancak, yalnızca bir yıllık veri kullanımı nedeniyle hata oranlarının %10'un üzerinde olduğu görülmüştür.

Korkmaz (2022) tarafından yapılan çalışmada, Bursa ve Çanakkale illerine ait 2015-2019 yılları arasındaki günlük/saatlik güneş ışınımı verileri kullanılarak yapay sinir

ağları ile tahminleme ve makine öğrenmesi algoritmalarıyla sınıflandırma yapılmıştır. Veri seti 1818 adet gözlemden oluşmakta olup, değişkenler arasında toplam yağış, nispi nem, saatlik buharlaşma, güneşlenme süresi ve aktüel basınç yer almaktadır. Güneş ışınımı değerlerinin 0 ile 1 arasında değiştiği göz önünde bulundurularak, sınıflandırma problemi için belirli ölçüm aralıkları sınıf etiketlerine dönüştürülmüştür. Sınıflandırma analizlerinde K-en Yakın Komşu (KNN), Naive Bayes (NB), Tekil Değer Ayrıştırması (SVD) ve Karar Ağaçları (DT) algoritmaları kullanılmıştır. Bursa ili için en yüksek doğruluk oranı karar ağacı algoritması ile elde edilmiştir. Çanakkale ili için ise belirginlik ölçütü doğrultusunda en başarılı sınıflandırma yine karar ağacı algoritması ile sağlanmıştır. Aynı coğrafi bölgede yer almalarına rağmen, iki ilde farklı sonuçlara ulaşılması; enlem, boylam, deniz seviyesinden yükseklik, güneş ışınımının geliş açısı ve toprak türü gibi değişkenlerin etkisini ortaya koymuştur. Ayrıca, bazı saatlerde ölçüm eksikliklerinin olması da tahmin performansını etkilemiştir. Gelecek çalışmalarda bu eksikliklerin giderilmesi ile daha net sonuçlara ulaşılabileceği ifade edilmiştir.

Kaplan (2023) tarafından rüzgâr türbinlerinin farklı rüzgâr hızı koşulları altında üretebilecekleri güçlerin tahmini ve türbin konumlarının enerji üretimi üzerindeki etkileri incelenmiştir. Çalışmalarda Türkiye'de aktif olarak çalışan iki farklı rüzgâr santralinin verileri kullanılmıştır. Tahmin performansı açısından en yüksek başarı %99.49 doğruluk ile karma modelde elde edilmiştir. Bu modeli sırasıyla %99.26, %99.24, %99.17 ve %99.14 doğruluk oranlarıyla gradyan yükseltme, rastgele orman, ekstra ağaç ve k-en yakın komşu algoritmaları takip etmiştir.

Yelgeç (2022) tarafından makine öğrenmesi yöntemleri kullanılarak rüzgâr santrallerinde enerji üretimi tahmini yapılmıştır. Rüzgâr gibi değişken bir yenilenebilir enerji kaynağından elde edilen veriler üzerinden yapılan tahminler için LSTM ve GRU (Gated Recurrent Unit) gibi derin öğrenme modelleri kullanılmıştır. Hiperparametre optimizasyonu Bayes yöntemiyle yapılmış ve bu sayede modellerin tahmin başarısı artırılmıştır. XGBoost algoritması ise düşük hesaplama süresi ve yüksek tahmin başarımı ile dikkat çekmiştir. Çalışmada bir yıllık veri seti eğitim ve test olarak ayrılarak analiz edilmiştir.

Yeşil (2021) tarafından yapılan çalışmada sabit ve hareketli tip güneş kolektörlerinin performansları makine öğrenmesi algoritmaları kullanılarak modellenmiştir. Karar Ağacı ve Yapay Sinir Ağı(YSA) algoritmalarıyla geliştirilen tahmin modellerinin

doğruluğu mutlak hata (MAE), göreceli mutlak hata (RAE), kök ortalama kare hata (RMSE) ve kök göreceli mutlak hata (RRAE) gibi istatistiksel metriklerle değerlendirilmiştir. En düşük RMSE değeri (0.0564), Yapay Sinir Ağı modeli ile elde edilmiştir. Bu bulgular, YSA algoritmasının sabit HGK sisteminin ekserji verimini tahmin etmede oldukça başarılı olduğunu ortaya koymaktadır.

Ayyıldız Koç (2022) tarafından yürütülen çalışmada, güneş panellerinin enerji üretimini tahmin etmek ve üretime etki eden meteorolojik değişkenleri belirlemek amacıyla Lojistik Regresyon, Naive Bayes, Rastgele Orman, AdaBoost, K-En Yakın Komşu, Destek Vektör Regresyonu ve Yapay Sinir Ağları gibi makine öğrenmesi algoritmaları kullanılmıştır. Meteorolojik özniteliklerin enerji üretimi üzerindeki etkileri analiz edilerek öznitelik azaltımı yapılmıştır. 2584 veriden oluşan veri setinin %80'i eğitim, %20'si test amacıyla kullanılmış ve değerlendirme metrikleri olarak MAE, RMSE ve R^2 değerleri kullanılmıştır. En başarılı performans Rastgele Orman algoritması ile elde edilmiştir.

Yeşilyurt (2023) tarafından yapılan çalışmada, Bina Enerji Yönetim Sistemi (BEYS) ile bütünleşmiş şekilde saatlik bina enerji tüketimi tahmin edilmiştir. Rastgele Orman, Gradyan Artırma Ağacı, Destek Vektör Makinesi, Yapay Sinir Ağı ve Derin Sinir Ağı algoritmaları test edilmiş ve en iyi sonuçlar Derin Sinir Ağı ile elde edilmiştir. Derin öğrenme tabanlı yaklaşımların bina enerji tahmini alanında başarılı olduğu gözlemlenmiştir.

Mintemur (2024) çalışmasında, kullanıcıların şarkı beğenilerine dayalı olarak oluşturulan yeni bir veri kümesi üzerinden, öneri sistemi problemi ikili sınıflandırma çerçevesinde ele alınmıştır. Çalışmada LightGBM, Extra Tree ve Random Forest algoritmaları, üç farklı sürü zekâsı optimizasyon yöntemiyle optimize edilerek karşılaştırılmıştır. Elde edilen sonuçlar, LightGBM algoritmasının diğer modellere kıyasla en yüksek sınıflandırma başarımını gösterdiğini ve kullanıcı tercihine dayalı öneri sistemleri için etkili bir yaklaşım sunduğunu ortaya koymuştur. Bu durum, LightGBM'in özellikle büyük ve yapılandırılmış veri kümeleriyle çalışan öneri sistemlerinde yüksek performans sergileyen bir yöntem olduğunu desteklemektedir.

Kibar (2022) tarafından gerçekleştirilen çalışmada, ısıtma, havalandırma ve iklimlendirme (HVAC) sistemine ait çok değişkenli bir zaman serisinde siber saldırıların tespiti için doğrusal regresyon, karar ağaçları, k-en yakın komşu, rastgele

orman ve gradyan artırma gibi makine öğrenmesi algoritmaları karşılaştırılmıştır. Simülasyon verileriyle eğitilen modeller, içinde 16 farklı siber saldırı bulunan test verisiyle değerlendirilmiş ve bazı algoritmaların başarılı sonuçlar verdiği ancak büyük veri işleme yükünün model performansına olumsuz etkileri olduğu tespit edilmiştir.

Günay (2022) tarafından yürütülen çalışmada yapay zekâ ve makine öğrenmesi teknolojilerinin ilerleyerek bir arada kullanılabildiği vurgulanmıştır. Bilgi teknolojilerindeki gelişmelerin mimarlık alanını da dönüştürdüğü, bu süreçte temsil, tasarım ve üretimin değiştiği belirtilmiştir. Makine öğrenmesinin cephe tasarımında, 2D'den 3D model oluşturulmasında, cephe optimizasyonunda, akıllı cephe sistemlerinde ve cephe değerlendirmelerinde kullanıldığı örneklere yer verilmiştir. İncelenen örneklerde, genellikle denetimli ve yarı denetimli makine öğrenmesi yaklaşımlarının tercih edildiği görülmüştür. Yeni yapılarda makine öğrenmesi algoritmalarının örneklerine rastlanırken, eski yapılarda bu tür uygulamalara pek yer verilmemiştir. Makine öğrenmesinin cephelerde kullanılmaya başlanmasıyla birlikte, kullanıcı konforu, enerji verimliliği, iş gücü ve enerji tasarrufu, hızlı üretim, optimum cephe tasarımı gibi olumlu etkiler gözlemlenmiştir. Son yıllarda artan enerji ihtiyacı ile, cephe tasarımında ve kullanımında makine öğrenmesi teknolojilerinin daha sık kullanıldığı örneklerde görülmüştür.

Yıldırım (2023), orman yangınlarının önceden tahmin edilebilmesi amacıyla makine öğrenmesi yöntemleri kullanılarak bir tahmin modeli geliştirilmiştir. Bu amaçla Destek Vektör Makinesi (SVM), Karar Ağacı (DT), Rastgele Orman (RF), Gradyan Artırma (GA), KNN ve NB olmak üzere toplamda 6 farklı sınıflandırma algoritması kullanılmış ve 10 kat çapraz doğrulama ile modeller oluşturulmuştur. Ham veri seti, 49 özellik ve toplamda 1172 örnekten oluşmaktadır. Hedef sınıf değişkeninde, 1012 veri yangının var olduğunu, 160 veri ise yangının olmadığını göstermektedir. Modellerin performanslarını değerlendirmek için, literatürde yaygın olarak kullanılan karışıklık matrisi ile doğruluk, duyarlılık, kesinlik, F1-skoru gibi ölçütler ve ROC eğrisi kullanılmıştır. Modelin doğruluğunu etkileyen gereksiz 5 değişken ve %50'den fazla kayıp değeri olan değişkenler veri setinden çıkarılmış, kalan eksik değerler ise ortalama veri ile tamamlanmıştır. Sonuç olarak, başlangıçta 49 olan özellik sayısı, bu işlemlerle 35'e indirilmiştir. Varyans eşiği ile yapılan özellik seçiminde, 35 özellik 15'e indirgenmiş ve en yüksek doğruluk oranına %97 ile RF algoritması ulaşmıştır. ANOVA yöntemiyle 35 özellik 11'e indirilmiş ve yine en yüksek doğruluk oranı %97

ile RF algoritması elde edilmiştir. Ki-kare testi ile 35 özellik 11'e indirilmiş ve RF, DT ve GA algoritmaları %97 doğruluk oranına ulaşmıştır. PCA yöntemiyle ise 35 özellik 11'e indirilmiş ve RF algoritması doğruluk (%96), AUC (%99) ve diğer performans metriklerinde en yüksek sonuçları vermiştir. Çalışma süreleri açısından ise DT ve NB algoritmalarının, RF'ye göre daha hızlı sonuç verdiği gözlemlenmiştir. Genel olarak, orman yangını tahmininde en yüksek sınıflandırma başarısı sağlayan modelin RF algoritması olduğu, ancak çalışma süresi bakımından NB ve DT algoritmalarının daha hızlı olduğu sonucuna varılmıştır.

Gülşen (2023), erken tanı ve yönlendirmenin ses bozuklukları tedavisinde kritik bir öneme sahip olduğu vurgulanmıştır. Son yıllarda yapılan araştırmalar, ses patolojisi tespit sistemlerinin ses bozukluklarının değerlendirilmesinde etkin bir şekilde kullanılabileceğini ve erken teşhis sağlamada önemli bir rol oynayabileceğini ortaya koymuştur. Bu sistemler, ses patolojisi tespitinde umut verici bir teknoloji olarak kabul edilen makine öğrenimi teknikleriyle geliştirilmiştir. Bu tez çalışmasında, makine öğrenimi modelleri kullanılarak sağlıklı ve patolojik seslerin sınıflandırılması yapılmıştır. Ses örnekleri "Saarbrücken Voice Database" (SVD)'den alınmış ve ses sinyallerinden öznitelik çıkarımı için OpenSMILE algoritması kullanılmıştır. Ayrıca MFCC yöntemi de uygulanmıştır. Modellerin başarımını artırmak için Kendall's Tau öznitelik seçim algoritması kullanılarak sınıflandırma başarısını olumlu etkileyen öznitelikler seçilmiş ve modeller eğitilip test edilmiştir. Sonuç olarak, sağlıklı ve patolojik seslerin sınıflandırılmasında makine öğrenimi modellerinin iyi performans sergilediği ve öznitelik seçim algoritmalarının başarıyı artırdığı gözlemlenmiştir. Bu tezde, patolojik seslerin belirlenmesi için makine öğrenimi teknikleri kullanılmıştır. Çalışmada, patolojik ve sağlıklı ses kayıtlarını içeren bir veri setinden disfonik ses kayıtları seçilmiş, bu kayıtlar sayısal verilere dönüştürülüp öznitelik çıkarımı ve seçimi yapılarak sınıflandırma algoritmalarına sunulmuştur. Başlangıçta SVM ve RF gibi yaygın sınıflandırma algoritmaları tercih edilmiş, ardından daha az yaygın olan NB ve XGB algoritmaları denenmiştir. XGB algoritmasının SVM'den daha iyi sonuçlar vermesi, gelecekte yapılacak çalışmalarda farklı veri tabanları ile genişletildiğinde umut verici sonuçlar elde edilebileceğini göstermektedir.

Lai, Chang, Chen ve Pai (2020) yaptığı çalışmalarda, iklim değişikliği ve küresel ısınmanın etkilerini azaltmak amacıyla yenilenebilir enerjinin kullanımının önem kazandığı vurgulanmış ve yenilenebilir enerjinin tahmin edilebilir bir sistem

olabileceğini göstermek için çeşitli tahmin teknikleri geliştirilmiştir. Çalışmanın ilk aşamasında, yenilenebilir enerji tahminlerinde makine öğrenimi modellerinin kullanımı incelenmiş ve bu konuda bir analiz sunan bir model geliştirilmiştir. İkinci aşamada ise, yenilenebilir enerji tahminleri için makine öğrenimi modellerinde kullanılan veri ön işleme teknikleri, parametre seçim algoritmaları ve tahmin performans ölçümleri gibi farklı prosedürler ele alınmıştır. Son olarak, yenilenebilir enerji kaynaklarının analizi yapılmış ve ortalama mutlak yüzde hata değerleri ile belirleme katsayısı değerleri üzerinden sonuçlar değerlendirilmiştir. Yenilenebilir enerji tahminlerinde makine öğrenimi modellerinin gelecekteki araştırma yönlerinin başında rüzgâr ve güneş enerjisi yer alırken, diğer yenilenebilir enerji türleri, örneğin gelgit enerjisi, biyokütle enerjisi, dalga enerjisi, hidrolik güç ve jeotermal enerji gibi alanlar, gelecekteki çalışmalar için potansiyel fırsatlar sunmaktadır. Ayrıca, yapay zekâ teknikleri ve hibrit modellerin yenilenebilir enerji tahminlerinde umut verici bir yaklaşım olabileceği belirtilmiştir. Diğer bir önemli nokta ise, veri ön işleme yöntemlerinin, makine öğrenimi modellerinin tahmin performansını doğrudan etkilediğidir. Son olarak, parametre seçiminin, yenilenebilir enerji tahminlerinde makine öğrenimi modellerinin performansı üzerinde büyük bir etkisi olduğu sonucuna varılmıştır.

Yao, Lum, Johnston, Mejia-Mendoza, Zhou, Wen, Aspuru-Guzik, Sargent ve Seh (2023) tarafından yürütülen çalışmalarda, enerji araştırmacılarının yenilenebilir enerjinin verimli bir şekilde toplanması, depolanması, dönüştürülmesi ve yönetilmesi için yeni sistemler ve cihazlara olan ihtiyacı ele alınmıştır. Bu eksikliklerin, makine öğrenmesi teknikleri kullanılarak giderilmesi amaçlanmaktadır. Çalışma, enerji araştırmaları için farklı makine öğrenmesi teknikleriyle hızlandırılmış iş akışlarının faydalarını karşılaştırmaya yönelik temel unsurları incelemektedir. Sonuç olarak, enerji hasadı (fotovoltaik), depolama (piller), dönüştürme (elektro kataliz) ve yönetim (akıllı şebekeler) alanlarında makine öğrenmesi uygulamalarının daha fazla fayda sağlayabilecek potansiyel araştırma alanları değerlendirilmiştir. Makine öğrenmesinin, sürdürülebilir enerjinin dağıtımını hızlandırması için, bir sentez prosedürü, karakterizasyon ekipmanı veya kontrol aparatı gibi araçlarla yaygınlaştırılması gerektiği sonucuna ulaşılmıştır.

Liu, Esan, Pan ve An (2021), küresel karbon nötrlüğüne ulaşmak için yeni enerji malzemelerinin geliştirilmesinin önemi ve gerekliliği vurgulanmaktadır. Çalışmada,

makine öğreniminin temelleri hakkında kapsamlı bilgiler verilmiş ve açık kaynaklı veri tabanları, özellik mühendisliği, makine öğrenimi algoritmaları ve bu modellerin analizine dair detaylara yer verilmiştir. Ayrıca, makine öğreniminin başarılı uygulamaları için ipuçları ve gelişmiş enerji malzemelerinin tasarımındaki zorluklar da ele alınmıştır. Başarılı makine öğrenmesi uygulamalarının temel faktörleri arasında veri altyapısının iyileştirilmesi, standartlaştırılması, otomatik kapalı devre optimizasyonu, model görselleştirme, robotlar tarafından yapılan deneysel keşifler, disiplinler arası iş birlikleri ve destekleyici politikalar gibi kalan zorluklar da vurgulanmıştır.

Mahesh (2020), makine öğrenimi algoritmalarının geniş bir uygulama yelpazesi ve gelecekteki beklentiler üzerine bir inceleme yapılmıştır. Makine öğrenimi (ML), bilgisayar sistemlerinin belirli bir görevi açıkça programlanmadan yerine getirmesini sağlayan algoritmalar ve istatistiksel modellerin bilimsel çalışmalarını ifade eder. Bu algoritmalar, veri madenciliği, görüntü işleme, öngörücü analiz gibi çeşitli alanlarda kullanılmaktadır. Makine öğreniminin temel avantajı, bir algoritmanın verilerle nasıl işlem yapacağını öğrendikten sonra, otomatik olarak görevini yerine getirmesidir. İnsanlar, evrimsel süreçlerinde görevleri daha kolay yapabilmek için farklı araçlar geliştirmiştir. Bu araçlar, insan hayatını kolaylaştırarak seyahat, endüstri ve bilgi işlem gibi alanlarda büyük ilerlemelere yol açmıştır. Makine öğrenimi de bu araçlardan biridir ve makinelerin verileri daha verimli kullanabilmesi için uygulanmaktadır. Günümüzde veri kümelerinin bolluğu ile, makine öğrenimine olan talep hızla artmaktadır. Birçok sektör, anlamlı verileri elde edebilmek için makine öğrenimi tekniklerini kullanmaktadır. Makine öğreniminin temel amacı, verilerden öğrenmeyi sağlamaktır. Ayrıca, makinelerin açıkça programlanmadan kendi kendilerine öğrenme süreçleri üzerine birçok çalışma yapılmış ve matematikçilerle programcılar, büyük veri kümeleriyle bu sorunları çözmek için çeşitli yaklaşımlar geliştirmiştir.

Antonopoulos, Robu, Couraud, Kirli, Norbu, Kiprakis, Flynn, Elizondo-Gonzalez ve Wattam (2020), 160'tan fazla makale, 40 şirket ve ticari girişim ve 21 büyük ölçekli projenin sistematik bir incelemesine dayalı olarak Talep Tarafı Tepkisi(DR) uygulamaları için kullanılan Yapay Zekâ(AI) yöntemlerine genel bir perspektiften bakışı ifade etmektedir. Makaleler hem kullanılan AI/MÖ (Makine Öğrenimi) algoritmaları hem de enerji DR'deki uygulama alanı açısından sınıflandırılmıştır. Daha sonra, ticari girişimler (hem yeni kurulan hem de yerleşik şirketler dahil) ve AI

yöntemlerinin enerji DR için kullanıldığı büyük ölçekli inovasyon projeleri sunulmaktadır. Yapay zekâ yaklaşımları, güç sistemlerindeki bu zorlukları ele almak için önemli bir araç olarak tanımlanmıştır. Çalışmada ayrıca araştırma topluluğunun DR sektöründeki AI çözümlerine olan bu artan ilgisinin endüstriyel sektörde de hissedildiğini gösterdiği gözlenmiştir.

Sarker (2021) yaptığı çalışmalarda, Dördüncü Sanayi Devrimi'nin (Endüstri 4.0) getirdiği dijital çağda, Nesnelerin İnterneti (IoT), siber güvenlik verileri, mobil veriler, iş verileri, sosyal medya verileri, sağlık verileri gibi çeşitli veri kaynaklarının mevcut olduğu vurgulanmaktadır. Bu büyük veri setlerini etkili bir şekilde analiz etmek ve akıllı, otomatik uygulamalar geliştirmek için yapay zekâ (AI) ve özellikle makine öğrenimi (ML) bilgisi önemli bir rol oynamaktadır. Çalışmada, denetimli, denetimsiz, yarı denetimli ve takviyeli öğrenme gibi farklı makine öğrenimi algoritmalarının bulunduğu belirtilmektedir. Başarılı bir makine öğrenimi modelinin, kullanılan verilerle birlikte algoritmaların performansına bağlı olduğu ifade edilmiştir. Ayrıca, derin öğrenmenin makine öğrenimi yöntemleri arasında yer alarak, büyük veri setlerini akıllıca analiz etme yeteneğine sahip olduğu da vurgulanmıştır. Makale, bu makine öğrenimi algoritmalarının uygulama zekasını ve yeteneklerini geliştirmek için nasıl kullanılabileceğine dair kapsamlı bir bakış açısı sunmaktadır. Çalışmanın temel katkısı, farklı makine öğrenimi tekniklerinin prensiplerini ve bu tekniklerin siber güvenlik, akıllı şehirler, sağlık, e-ticaret, tarım gibi çeşitli gerçek dünya uygulama alanlarındaki uygulanabilirliğini açıklamaktır.

Stanulov ve Yassine (2023), havacılık sektörü bazlı çalışmada Uzun Kısa Süreli Bellek (LSTM) ağı, Destek Vektör Regresyon Makinesi (SVRM) ve Rastgele Orman (RF) algoritmalarının mimarileri genel bir bakış açısıyla ele alınmış ve karşılaştırılmıştır. Araştırma, bu üç modelin performansını karşılaştırmak için Ortalama Karesel Hata (MSE) ve Kök Ortalama Karesel Hata (RMSE) ölçütleri kullanılarak bir analiz sunmuştur. Kullanılan veri seti, 1 Ocak 2009 ile 31 Aralık 2019 tarihleri arasında Oslo Havalimanı Gardermoen'deki tarifeli uçuşlardan alınan saatlik yolcu sayılarından oluşmakta ve bu veriler saat, gün, ay, yıl gibi tarih ve saat bilgileriyle birlikte hava sıcaklığı ve rüzgâr hızı gibi hava durumu özelliklerini içermektedir. Veri seti toplamda 96185 örneklemden oluşmaktadır. Araştırma sonuçlarına göre, LSTM modeli test veri setinde 0.00445/0.06667 MSE/RMSE ile en yüksek genelleme yeteneğini göstermiştir. SVRM ve RF modellerinin performansı ise sırasıyla 0.00511/0.07147 ve

0.00543/0.07368 MSE/RMSE değerleriyle kaydedilmiştir. Bu performans değerlendirmelerine ek olarak, her bir modelin karmaşıklığı, kararlılığı ve yolcu sayılarındaki saatlik ve günlük dalgalanmaları tahmin etme yetenekleri de dikkate alınmıştır.

Ahmadi (2023), Makine öğreniminin (ML) veri ambarlarına entegrasyonu incelenmiş ve bu süreçte karşılaşılan optimizasyon zorlukları, kullanılan metodolojiler, elde edilen sonuçlar ve gelecekteki eğilimler üzerine odaklanılmıştır. Veri ambarları, genellikle yüksek bakım maliyetleri ve arıza oranları gibi sorunlarla karşı karşıya kalırken, ML ile bu zorlukların üstesinden gelinerek önemli bir dönüşüm sağlanmaktadır. Entegrasyon süreci, sorgu optimizasyonu, dizinleme ve otomatik veri yönetimi gibi yöntemlerle performansın artırılmasına olanak tanımaktadır. Sonuç olarak, ML'nin iş yükü yönetimi, otomatik sorgu optimizasyonu ve uyarlanabilir kaynak tahsisi gibi alanlarda öngörücü analizlerle verimliliği artırdığı ortaya çıkmıştır.

Woodman ve Mangoni (2023) çalışmalarında, makine öğrenimi algoritmalarını kullanarak en uygun tedavi seçeneklerini belirleyen klinik araçların, yönetim organlarından gerekli onayı yavaşça almaya başladığı ve bu araçların sağlık hizmetlerine, dijital tıbbın bir sonraki aşamasında önemli etkiler yaratacağı vurgulanmıştır. Ancak bu araçların uygulanabilirliğinde sadece düzenleyici onay değil, aynı zamanda kullanıcıların güveni ve desteği de önemli bir rol oynamaktadır. Çalışma, makine öğrenimi algoritmalarının çeşitli sınıflandırmalarını sunmakta ve her bir sınıfın hedeflerini, özelliklerini ve özellikle geriatrik tıpta nasıl uygulandığını detaylandırmaktadır. Ayrıca, daha az yorumlanabilirliğe sahip cihazlara duyulan güvenin klinik etkileri ve zorlukları ile açıklanabilir makine öğreniminin geliştirilmesi konusundaki ilerlemelere de odaklanılmaktadır. Bu makale, makine öğrenimindeki çeşitli algoritmaların sağlık hizmetlerinde nasıl kullanıldığını açıklamanın yanı sıra, bu teknolojilerin sürekli olarak geliştirilmesi ve iyileştirilmesi gerektiği, sağlık hizmetlerine daha fazla entegrasyon sağlandıkça hızla ilerlemeler kaydedilebileceği noktasını da vurgulamaktadır.

Musleh, Alotaibi, Alhaidari, Rahman ve Mohammad (2023) tarafından yürütülen çalışmalarda, Nesnelerin İnterneti (IoT) cihazlarının kullanımındaki sürekli artışla birlikte, bu cihazları kötü amaçlı trafikten korumaya yönelik internet güvenliğine artan bir ilgi gösterildiği vurgulanmaktadır. Bu tür tehditleri tespit etmek zor olduğu için, gelişmiş bir saldırı tespit sistemi (IDS) tasarımı önem kazanmıştır. Makine öğrenimi

(ML), IoT gibi farklı alanlarda etkili bir IDS çözümü olarak öne çıkmaktadır. Ancak, ML modellerinin tespit oranı ve doğruluğunu etkileyen en önemli faktör, IoT ortamından uygun özelliklerin çıkarılmasıdır. Bu çalışma, çeşitli ML modelleri ile birlikte farklı özellik çıkarma algoritmalarını değerlendirerek IoT tabanlı ML destekli IDS geliştirmeyi amaçlamaktadır. Araştırmada, VGG-16 ve DenseNet gibi görüntü filtreleri ve transfer öğrenme yöntemleri gibi farklı özellik çıkarıcılar kullanılmıştır. Ayrıca, rastgele orman, K-en yakın komşular, SVM ve farklı yığılmış modeller gibi çeşitli makine öğrenimi algoritmaları da değerlendirilmiştir. IEEE Dataport veri kümesi kullanılarak yapılan değerlendirme sonucunda, VGG-16'nın yığılmış modellerle birlikte %98,3'lük en yüksek doğruluk oranını sağladığı görülmüştür. Bu çalışma, kötü amaçlı trafiği normal trafikten ayırt etme konusunda en yüksek performansa sahip bir model geliştirmeyi hedeflemektedir. Çeşitli özellik çıkarma teknikleri ve makine öğrenimi algoritmaları kullanılarak yapılan deneyler, yüksek performans için özellik çıkarma tekniklerinin ne kadar önemli olduğunu ortaya koymuştur. Ayrıca, bireysel ve yığılmış makine öğrenimi algoritmalarının etkileriyle birlikte, veri bölme oranının modellerin performansı üzerindeki etkisi de araştırılmıştır.

Khan, Qureshi, Daniyal ve Tawiah, (2023), kalp hastalarının doğru tahmin ve karar alma süreçleri için makine öğrenimi algoritmaları kullanılmıştır. Pakistan'daki kalp hastaları için sınıflandırma ve tahmin yapabilmek amacıyla karar ağacı (DT), rastgele orman (RF), lojistik regresyon (LR), Naive Bayes (NB) ve destek vektör makinesi (SVM) gibi farklı makine öğrenimi yöntemleri uygulanmıştır. Tüm algoritmalar için keşifsel analiz ve deneysel çıktılar değerlendirilmiştir. Makine öğrenimi algoritmalarının performansı, her modelin en uygun sonuçları vermesi için farklı koşullar altında test edilmiştir. Sonuçlar, RF algoritmasının kalp hastalıkları için sırasıyla %85,01 tahmin doğruluğu, %92,11 hassasiyet ve %87,73 doğrulukla en yüksek performansı sergilediğini göstermiştir. Ancak, aynı algoritma kalp hastalıkları için sırasıyla %43,48 özgüllük ve %8,70 yanlış sınıflandırma hatalarıyla da en düşük sonuçları almıştır. Bu bulgular, RF algoritmasının kalp sınıflandırması ve tahmini için en uygun yöntem olduğunu ortaya koymaktadır. Önerilen modelin, hastalık sınıflandırması ve tahmini konusunda sağlık sektöründe dünya çapında geniş bir uygulama alanı bulabileceği vurgulanmıştır.

Rashidi Nasab ve Elzarka (2023), Ohio'da bir köprü güvertesinin bozulma süresini doğru şekilde tahmin edebilmek için makine öğrenimi algoritmaları kullanılmıştır. Bu algoritmalar arasında Karar Ağacı (DT), Yapay Sinir Ağları (ANN), K-en Yakın Komşular (k-NN), Lojistik Regresyon (LR) ve Destek Vektör Makineleri (SVR) gibi tek başına çalışan modellerin yanı sıra, Rastgele Orman (RF) ve eXtreme Gradyan Artırma (XGboost) gibi topluluk yöntemleri de yer almaktadır. Bu algoritmalar, köprü güvertesinin durumunu sınıflandırmak için uygulanmıştır. Ohio Ulaştırma Bakanlığı'ndan (ODOT) alınan verilerle yapılan analiz, optimum özellikler kullanıldığında, topluluk tabanlı makine öğrenimi algoritmalarının, tek başına çalışan modellere göre köprü güvertesinin durumunu çok daha doğru bir şekilde tahmin edebildiğini göstermiştir.

Islam, Majumder ve Hussein (2023), Kronik böbrek hastalığı (KBH), erken teşhisini yapabilmek için farklı makine öğrenimi yaklaşımlarının potansiyeli incelenmiştir. Çalışma, 25 değişkenle başlayıp, sonunda bu parametrelerin %30'unu kullanarak en iyi sonuçları veren alt küme üzerinde yoğunlaşmıştır. Denetimli öğrenme ortamında toplamda 12 farklı makine öğrenimi tabanlı sınıflandırıcı test edilmiştir. Bu testlerde en yüksek performans değerleri 0,983 doğruluk, 0,98 kesinlik, 0,98 geri çağırma ve XgBoost sınıflandırıcısı için 0,98 F1 puanı ile elde edilmiştir. Araştırma, makine öğrenimindeki son gelişmelerin ve tahmini modellemenin, böbrek hastalığı gibi sağlık alanlarında tahmin doğruluğunu artırmak için nasıl etkili bir şekilde kullanılabileceğini gösteren yeni çözümler sunduğu sonucuna varmaktadır.

Alkhatib, Sahwan, Alkhatieb ve Schütt (2023), Orman yangınlarının dünya çapında daha sık görülmesi, çevre ve ekonomi üzerinde büyük bir yıkım yarattığı için erken tahmin ve tespit hayati önem taşımaktadır. Orman yangınlarını öngörmek ve tespit etmek için birçok teknoloji ve yöntem geliştirilmiştir. Bu bağlamda, orman yangını olasılığını tahmin etmek ve yangınlardan kaynaklanan zarar riskini değerlendirmek için yapay zekâ teknikleri önemli bir rol oynamaktadır. Orman yangınlarını tanımlamak ve tahmin etmek için kullanılan makine öğrenimi yöntemleri bu araştırmada ele alınmıştır. Bu çalışmanın ana hedefi, orman yangınlarını incelemek için makine öğrenimi tekniklerini kullanan araştırma boşluklarını ve güncel çalışmaları keşfetmektir. Orman özelliklerine göre en uygun makine öğrenimi tekniklerinin seçilmesi, mevcut araştırma sonuçlarının tahmin gücünü artırmaktadır.

Araştırma, orman yangını bilimine makine öğrenimi tekniklerini entegre etmeyi amaçlayan çeşitli çalışmaları incelemiştir.

Md, Kulkarni, Joshua, Vaichole, Mohan ve Iwendi (2023), dünya genelinde karaciğer hastalıkları hızla artmakta ve bu hastalığa sahip pek çok kişi, farkında olmadan hayatını kaybetmektedir. Bu makalede, Hint Karaciğer Hastası Veri Seti'ni (ILPD) kullanarak, karaciğer hastalığını tahmin etmek için topluluk öğrenme ve gelişmiş ön işleme yöntemlerine dayalı yeni bir mimari önerilmektedir. Altı farklı topluluk öğrenme algoritması ILPD'ye uygulanmış ve bu sonuçlar, mevcut çalışmalarda elde edilenlerle karşılaştırılmıştır. Önerilen model, doğruluğu artırmak için veri dengeleme, özellik ölçekleme ve özellik seçimi gibi çeşitli veri ön işleme teknikleri kullanır. Eksik verilerin doldurulması için çok değişkenli tahmin teknikleri uygulanırken, eğik sütunlar üzerinde de log1p dönüşümü, standardizasyon, min-maks ölçekleme, maksimum mutlak ölçekleme ve sağlam ölçekleme gibi işlemler yapılmıştır. Özellik seçimi ise, tek değişkenli seçim, özellik önemi ve korelasyon matrisi gibi yöntemlerle gerçekleştirilmiştir. Bu iyileştirilmiş veriler, Gradient Boosting, XGBoost, Bagging, Random Forest, Extra Tree ve Stacking topluluk öğrenme algoritmaları ile eğitilmiştir. Altı modelin sonuçları birbirleriyle ve diğer mevcut araştırmalarla karşılaştırıldığında, Extra Tree sınıflandırıcı ve Random Forest kullanan önerilen model sırasıyla %91,82 ve %86,06'lık test doğruluğu ile en iyi sonuçları elde ederek diğer yöntemleri geride bırakmıştır. Extra Tree sınıflandırıcı kullanarak geliştirilmiş ön işleme yöntemini içeren modelin, %91,82'lik test doğruluğu ile en yüksek performansı gösterdiği, onu %86,06'lık doğrulukla Random Forest modelinin takip ettiği gözlemlenmiştir.

Elbasi, Zaki, Topcu, Abdelbaki, Zreikat, Cina, Shdefat ve Saker (2023), Bu araştırma, makine öğrenimi algoritmalarını modern tarıma entegre etmenin potansiyel faydalarını incelemektedir. Bu algoritmalar, ekim, sulama ve hasat gibi süreçlerde bilinçli kararlar alarak ürün verimini optimize etmeye ve atıkları azaltmaya yardımcı olmayı hedeflemektedir. Makale, tarımda makine öğreniminin mevcut durumunu tartışmakta, karşılaşılan temel zorlukları ve fırsatları vurgulamakta ve veri analizi algoritmalarının doğruluğunu etkileyen etiket değişikliklerinin deneysel sonuçlarını sunmaktadır. Araştırma bulguları, çiftçilerin, ürün büyümesini etkileyen faktörler hakkında daha bilinçli kararlar alabilmesi için çiftliklerden toplanan geniş verilerin, gerçek zamanlı IoT sensör verileriyle analiz edilmesinin önemini vurgulamaktadır. Bu teknolojilerin bütünleştirilmesi, ürün verimini artırarak atıkları azaltabilir ve modern tarımda devrim

yaratabilir. Çalışmada, tarımda kullanılacak en uygun algoritmaları değerlendirmek amacıyla on beş farklı algoritma göz önünde bulundurulmuş ve yeni bir özellik kombinasyonu şeması ile geliştirilmiş bir algoritma önerilmiştir. Sonuçlar, Bayes Net algoritması ile %99,59, Naive Bayes Sınıflandırıcı ve Hoeffding Ağacı algoritmaları ile ise %99,46 sınıflama doğruluğu elde edilebileceğini göstermektedir. Ayrıca, bu araştırmanın bulguları, çiftçilerin hastalıkları erken tespit etmelerine, ürün verimliliğini artırmalarına ve gıda kıtlığı sırasında fiyatları düşürmelerine yardımcı olabilecek çözümler sunmaktadır. Bu çalışma, makine öğrenimi algoritmalarının ve IoT sensörlerinin tarıma entegrasyonunun önemini vurgulamaktadır. Her sınıflandırma algoritması için doğruluk, hata oranları ve test süreleri gibi etkenler değerlendirilmiş ve etiketlerin değiştirilmesinin veri analizi algoritmalarının doğruluğu üzerindeki etkisi gösterilmiştir. Bulgular, çiftçilerin ürün büyümesini etkileyen faktörler hakkında daha bilinçli kararlar alabilmesi için gerçek zamanlı verilerin ve IoT sensörlerinden elde edilen bilgilerin analizinin önemini ortaya koymaktadır. Çalışmada, farklı makine öğrenimi algoritmaları ile mahsuller genel özelliklerine göre incelenmiş ve doğru tahminlerle değerli sonuçlar elde edilmiştir. Özellikle Bayes Net ve Random Forest algoritmaları, Sıcaklık, Nem, pH ve Yağış gibi özellikler kullanılarak %97,05 ve %97,32 doğruluk oranlarına ulaşmıştır. Bu araştırma, modern tarımda makine öğreniminin potansiyel faydalarını ortaya koymuş ve bu alanda daha fazla araştırma ile mahsul üretiminin optimize edilmesine ve gıda güvenliğinin artırılmasına katkı sağlayacaktır. Gelecekte, farklı coğrafi bölgelerden elde edilen GPS tabanlı IoT ve sensör verileriyle daha fazla mahsul verisi değerlendirilecek ve bu veriler, makine öğrenimi algoritmalarıyla analiz edilerek daha doğru sonuçlar elde edilecektir.

Shin ve Woo (2022), Enerji tüketiminin tahmin edilmesinde, geleneksel ekonometrik ve istatistiksel modeller genellikle kullanılır. Ancak, bu modeller büyük veri analizini gerektiren ve hızla değişen enerji pazarında bazı sınırlamalara sahip olabilir, çünkü karmaşık matematiksel araçlar kullanılarak enerji tüketim kalıplarını ve ilgili değişkenleri analiz etmek gerekir. Mevcut literatürde, Kore'deki enerji tüketimini tahmin etmek için makine öğrenimi algoritmalarını karşılaştıran sınırlı sayıda çalışma bulunmaktadır. Bu boşluğu doldurmak amacıyla, bu makalede üç farklı makine öğrenimi algoritması karşılaştırılmıştır: Rastgele Orman (RF), XGBoost (XGB) ve Uzun Kısa Süreli Bellek (LSTM). Bu algoritmalar, COVID-19 pandemisinin

başlangıcından önceki Dönem 1 ve sonrasındaki Dönem 2 olmak üzere iki farklı dönemde uygulanmıştır. Dönem 1, enerji tüketiminin arttığı bir dönemken, Dönem 2'de enerji tüketiminde bir azalma gözlemlenmiştir. LSTM, Dönem 1'de tahmin doğruluğu açısından en iyi performansı gösterirken, RF ise Dönem 2'de diğer modellere kıyasla daha başarılı olmuştur. Bu sonuçlar, makine öğreniminin enerji tüketimini tahmin etmedeki uygulanabilirliğini vurgulamaktadır. Ayrıca, bulgular, zaman serisinde daha az bilinmeyen düzensizlik olduğunda geleneksel ekonometrik yöntemlerin makine öğreniminden daha iyi performans gösterebileceğini, ancak makine öğreniminin düzensiz ve öngörülemeyen zaman serisi verileriyle daha iyi çalışabileceğini göstermektedir.

Walker, Khan, Katić, Maassen ve Zeiler (2020), Bu çalışmada, 47 ticari binadan elde edilen verilerle, saatlik aralıklarla bireysel bina düzeyinde ve grup düzeyinde elektrik talebini tahmin etmek için çeşitli makine öğrenimi algoritmaları incelenmiştir. Kısa vadeli dinamikleri anlamak için saatlik veri çözünürlüğü kullanmak önem arz etmektedir, ancak çoğu mahalle ölçeğindeki çalışma genellikle yıllık, aylık ya da günlük verilerle sınırlı kalmaktadır. Model eğitimi için iki yıllık veri kullanılmış ve tahminler, eğitilmemiş bir yıllık veri ile yapılmıştır. Güçlendirilmiş ağaç, rastgele orman, SVM doğrusal, ikinci dereceden, kübik, ince Gauss ve yapay sinir ağları (ANN) gibi algoritmalar analiz edilmiştir. Sonuçlar, saatlik tahminlerde en iyi performansı güçlendirilmiş ağaç, rastgele orman ve ANN algoritmalarının verdiğini göstermiştir. Bu çalışma, bina gruplarını dikkate alarak yapılan tahminlerin bireysel tahminlere göre daha fazla değer sunduğunu vurgulamaktadır. Eğitilen modellerden, güçlendirilmiş ağaç ve rastgele ormanın diğer algoritmalara göre daha iyi performans gösterdiği gözlemlenmiştir. Hesaplama gücü ve zamanın yanı sıra hata doğruluğu da dikkate alınarak, bu üç model, elektrik talebi tahmininde tercih edilmiştir. Ayrıca, geleneksel enerji şebekelerinden akıllı şebekelere geçişle birlikte binalardan üretilen veri miktarının artması, verilerin hangi seviyelerde kullanılacağına ve ekonomik bir şekilde nasıl işleneceğine karar verilmesi gerektiğini ortaya koymaktadır. Bu çalışma, saatlik veri çözünürlüğü kullanarak yapılan tahmin sonuçlarının günlük veya haftalık tahminlere göre daha doğru olduğunu göstermektedir.

Olu-Ajayi, Alaka, Sulaimon, Sunmola ve Ajayi (2022), enerji verimsiz binaların inşaatını azaltmak amacıyla erken tasarım aşamasında potansiyel bina enerji tüketimini tahmin etmek için makine öğrenimi yöntemlerinin uygunluğunu araştıran

çalışma yapmıştır. Bunun için makine öğrenmesi yöntemlerinden, Yapay Sinir Ağı (ANN), Gradyan Arttırma (GB), Derin Sinir Ağı (DNN), Rastgele Orman (RF), Yığınlama, K En Yakın Komşu (KNN) gibi çeşitli makine öğrenme tekniklerinin kullanmıştır. , Konut binalarından oluşan geniş bir veri kümesi kullanarak yıllık bina enerji tüketimini tahmin etmek için Destek Vektör Makinesi (SVM), Karar Ağacı (DT) ve Doğrusal Regresyon (LR) kullanılmıştır. Büyük veri setinde iyi sonuçlar ürettiği bilinen ANN tabanlı model galip gelmiş ve GB tabanlı model enerji tahmini alanında fazla ilgi görmeyen üçüncü en iyi tahmin modeli olarak ortaya çıkmıştır. DNN, ANN, GB, RF ve SVM tabanlı modeller KNN, DT, istifleme ve LR gibi doğrudan karşılaştırılabilmektedir. DNN daha iyi MAE, MSE, RMSE ve R2 ile ANN ve GB'den biraz daha iyi performans göstermektedir.

Eşidir (2025) çalışmasında, Türkiye İstatistik Kurumu'nun (TÜİK) 2022 yılına ait Yetişkin Eğitimi Araştırması mikro veri setini kullanarak bireylerin örgün eğitime katılım durumlarını ve bu durumu etkileyen faktörleri incelemiştir. Toplamda 14 değişken ve 24.462 gözlem içeren veri seti üzerinde çeşitli makine öğrenimi algoritmaları uygulanmış ve özellikle LightGBM algoritmasının %94 doğruluk oranı, 0,92 AUC skoru ve 0,75 F1 skoru ile üstün performans sergilediği belirlenmiştir. Sayısal değişkenlerin ölçeklendirilmesinde Standard Scaler yöntemi tercih edilmiştir. Model başarımı; doğruluk, duyarlılık, kesinlik, F1 skoru ve AUC gibi performans metrikleriyle değerlendirilmiştir. Çalışmanın sonuçlarına göre "yaş", "eğitim seviyesi" ve "medeni durum" değişkenleri, bireylerin örgün eğitime katılımında belirleyici faktörler olarak öne çıkmıştır. Elde edilen bulgular, LightGBM algoritmasının eğitim verilerinin analizinde yüksek doğruluk ve dengeli performans ile etkili bir araç olduğunu ve veri temelli eğitim politikalarının oluşturulmasında kullanılabileceğini göstermektedir.

Kılıç, Alp, Akyüz, Abut, Bozdemir, Çetinkaya, Çelik, Demir, Göktaş, Güneş, Koç ve Yılmaz (2022) tarafından yürütülen çalışmalarda, e-ticarette sıklıkla karşılaşılan ve ciddi operasyonel maliyetlere yol açan anlaşmazlık içeren iadelerin çözüm süreci ele alınmıştır. Bu doğrultuda, söz konusu süreci otomatikleştirmek amacıyla makine öğrenimi tabanlı modeller geliştirilmiş; Logistic Regression, CatBoost ve LightGBM algoritmalarının performansları karşılaştırılmıştır. Değerlendirme sonuçları, LightGBM modelinin en yüksek doğruluk, duyarlılık ve AUC(Area Under the Curve) skorlarına ulaştığını ve en başarılı performansı sergilediğini ortaya koymuştur. Elde

edilen bulgular doğrultusunda, LightGBM destekli otomatik karar destek sisteminin, insan müdahalesini azaltarak e-ticaret iadeleri sürecinde verimliliği önemli ölçüde artırma potansiyeline sahip olduğu sonucuna varılmıştır.

Kaynar, Taştan ve Demirkoparan (2012) tarafından yürütülen çalışmada, günlük doğalgaz tüketim verileri üzerinde gerçekleştirilen tahmin analizlerinde, ARIMA ve Çok Katmanlı Algılayıcı (MLP) modellerinin performansları karşılaştırılmıştır. Elde edilen sonuçlar, her iki modelin genel tüketim eğilimlerini başarıyla yakaladığını göstermekle birlikte, MLP modelinin hem Ortalama Kare Hata (MSE) hem de Ortalama Mutlak Yüzde Hata (MAPE) kriterlerinde daha düşük hata oranlarıyla daha yüksek doğruluk sağladığını ortaya koymuştur. Özellikle MLP modelinin MSE değeri 1807,82 ve MAPE oranı %4,42 olarak hesaplanmış; bu değerler ARIMA modelinin 2310,30'luk MSE ve %5,07'lik MAPE değerlerinden daha iyi performans sergilediğini göstermiştir. Bu bulgular, yapay sinir ağları tabanlı modellerin, doğrusal varsayımlara dayanan klasik zaman serisi modellerine kıyasla doğalgaz tüketim tahmininde daha etkin ve güvenilir sonuçlar verdiğini desteklemektedir.

Bu çalışmada, 2002–2014 dönemine ait 156 aylık veri kullanılarak Türkiye'de doğalgaz tüketimi tahmini yapılmıştır. Modelde gayri safi yurtiçi hasıla (GSYH), sıcaklık ve doğalgaza erişim oranı bağımsız değişkenler olarak kullanılmış ve yapay sinir ağları ile bulanık mantık prensiplerini birleştiren Adaptive Neuro-Fuzzy Inference System (ANFIS) yöntemi uygulanmıştır. Veri seti, eğitim, doğrulama ve test olmak üzere ayrılarak modelin genelleme performansı optimize edilmiştir. Modelin performansı, test verisi üzerinde hesaplanan 290,85 RMSE ve %5,77 MAPE değerleri ile değerlendirilmiş; bu değerler ANFIS modelinin doğalgaz tüketimi tahmininde yüksek doğruluk ve güvenilirlik sağladığını ortaya koymuştur. Ayrıca, modelde kullanılan bulanık çıkarım sistemi sayesinde hem doğrusal hem de doğrusal olmayan veri yapıları etkin şekilde işlenmiş ve karmaşık ilişkiler başarılı şekilde modellenmiştir.

3. MATERYAL VE YÖNTEM

37° 33′ 0″ Kuzey enlemi ve 44° 14′ 16″ Doğu boylamı koordinatlarında konumlanan bölge, Türkiye'nin güneydoğusunda yer alan Hakkâri ilinin Yüksekova ilçesi sınırları içerisinde bulunmaktadır. Deniz seviyesinden yaklaşık 1.950 metre yükseklikte konumlanan Selahaddin Eyyubi Havalimanı hem yüksek rakımı hem de karasal iklim özellikleri ile öne çıkmakta; uzun ve sert kış mevsimi ile kısa, serin yaz aylarının yaşandığı bir iklim tipine sahiptir. Meteorolojik veriler ışığında, bölge yıllık ortalama sıcaklık bakımından 7–9 °C aralığında seyretmekte olup, yıllık ortalama yağış miktarı 700–900 mm civarındadır; kış aylarında sıcaklık genellikle sıfırın altına düşmekte ve yoğun kar örtüsü görülmektedir. Ayrıca bölgenin dağlık topoğrafyası ve sınır hattına yakın stratejik konumu, iklimsel verilerin sadece meteorolojik değil, aynı zamanda sosyoekonomik, enerji ve çevresel planlama açısından da değerlendirilmesini gerekli kılmaktadır.

3.1. Veri Seti

ERA5, Avrupa Orta Vadeli Hava Tahminleri Merkezi (ECMWF) tarafından geliştirilen ve Copernicus İklim Değişikliği Servisi (C3S) aracılığıyla sunulan bir yeniden analiz (reanalysis) veri kümesidir. Yeniden analiz yöntemi, tarihsel atmosferik koşulların hem gözlemsel veriler hem de fiziksel atmosfer modelleri yardımıyla yeniden yapılandırılmasını amaçlar. Bu yaklaşım sayesinde farklı kaynaklardan elde edilen veriler, tutarlı ve sürekli bir formatta birleştirilerek eksiksiz veri alanları oluşturulmaktadır. ERA5 veri seti, 1950 yılından günümüze kadar olan dönemi kapsamakta olup saatlik zaman çözünürlüğü ve yaklaşık 31 km (0.25° x 0.25°) uzamsal çözünürlüğe sahiptir. Veri seti; 2 metre hava sıcaklığı, yüzey basıncı, rüzgâr bileşenleri, toplam yağış, radyasyon ve toprak nemi gibi pek çok temel meteorolojik değişkeni içermektedir. Bu çok değişkenli ve yüksek çözünürlüklü yapı sayesinde ERA5, iklim değişikliği analizi, hidrolojik modelleme, çevresel etki değerlendirmesi ve enerji talep tahminleri gibi çok sayıda akademik ve uygulamalı çalışmada yaygın olarak kullanılmaktadır. Özellikle enerji tüketimine yönelik analizlerde, ERA5'in

sunduğu 2 metre hava sıcaklığı verisi kullanılarak Isıtma Derece Günleri (Heating Degree Days - HDD) hesaplanmakta ve bu değişken doğalgaz tüketim tahmin modellerinde girdi olarak değerlendirilmektedir. ERA5 verisine erişim, Copernicus Climate Data Store (CDS) platformu üzerinden sağlanmakta olup hem manuel indirme yoluyla hem de Python tabanlı CDS API aracılığıyla otomatik veri çekimi gerçekleştirilebilmektedir. Bunun yanı sıra, Google Earth Engine gibi bulut tabanlı analiz platformlarında da ERA5 verisinin çeşitli alt türevleri (örneğin ERA5-Land) kullanılabilmektedir.

3.2. Makine Öğrenmesi

Makine öğrenmesi, yapay zekânın bir alt dalı olup, bilgisayar sistemlerinin açıkça programlanmadan verilerden öğrenmesini ve bu öğrenme süreci sonucunda tahmin ya da kararlar verebilmesini sağlar. Bu yöntem, sistemlerin geçmiş verilerden edindiği örüntüler aracılığıyla yeni durumlara uyum sağlamasına imkân tanır.

Makine öğrenmesi üç temel kategoriye ayrılır: denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme.

- **Denetimli öğrenme** (supervised learning), girdi ve buna karşılık gelen çıktı etiketlerinin bulunduğu verilerle çalışır. Bu yöntem genellikle sınıflandırma ve regresyon problemlerinde kullanılır.
- **Denetimsiz öğrenme** (unsupervised learning), sadece girdi verilerini kullanır ve veri içerisindeki yapıları ya da kümeleri ortaya çıkarmayı amaçlar.
- **Pekiştirmeli öğrenme** (reinforcement learning) ise bir ajanın, ortamla etkileşim içinde ödül-maliyet sistemine göre öğrenme süreci yürüttüğü bir yaklaşımdır.

Makine öğrenmesi günümüzde finans, sağlık, enerji, güvenlik, pazarlama, meteoroloji ve daha birçok alanda yaygın olarak kullanılmaktadır. Bu çalışmada da makine öğrenmesi algoritmalarının veri analizi ve tahmin performansları karşılaştırmalı olarak değerlendirilmiştir.

3.2.1. Makine Öğrenmesi Tanımı ve Sınıflandırması

Geçmişten günümüze ulaşan veriler ve yapılan araştırmalar sonucu ulaşılan sonuçlarla ortaya çıkan yapay zekâ sistemli makine öğrenmesi uygulamaları, makinelerin herhangi bir komut veya talimat verilmeden insanlar gibi düşünebilmesi ve karar verebilmelerini sağlayan programlar olarak tanımlanmıştır. Makine öğrenmesi uygulamaları makinelerden yapılması istenebilecek görevleri yerine getirebilmeleri için hiçbir komut ve talimat girilmeden öngörülebilir verilere dayanarak tutarlı ve gerçeğe en yakın olan sonuçları ortaya çıkartan bir sistem olarak kullanılmaktadır. Bu çalışma prensibi, kendisine aktarılan ve eğitim modeli verisi olarak adlandırılan veri depolarıyla desteklenen matematiksel bir yöntemle işletilmektedir. 1980'li yıllara kullanılmaya başlanılan makine öğrenmesi tanımı, optik karakterleri tanıyarak başlayan bu teknolojinin aktif olarak ilk kullanıldığı alan ise 1990'lı yıllarda istenmeyen e-postaların tespit edilerek ayrılması işlemini sağladığı bir filtreleme sistemi olduğu bilinmektedir. Buna benzer uygulamalarda her geçen gün artarak devam eden bir model olarak makine öğrenmesi uygulamalarının kullanımının artış gösterdiği saptanmıştır.

3.2.2. Makine Öğrenmesi Tarihsel Süreci

İnsan gibi düşünen ve hareket eden makine ve robotların var olması hayali 17. yüzyıl tarihine dayandırılmış olsa da tarih öncesi dönem olan 2500 yıl öncesine bakıldığında Yunan mitolojisi incelendiğinde o dönemki yaşayan insanlar tarafından yapay insan oluşturma fikrinin var olduğu görülmektedir. Charles Babbage, 19. yüzyılın başlarında insanlar gibi aklıyla davranabilen bir makine yapabilme fikrini deneyler yaparak ortaya çıkarmayı amaçlamıştır. 1819 yılında, bir bilim insanı olan arkadaşı John Herscel ile buhar motoruyla çalışabilecek bir hesap makinesini geliştirmek üzere çeşitli çalışmalar yürütmüşlerdir. "The Difference Engine" ismiyle anılmaya başlanacak olan bu makine, farkların kullanılarak hesap yapılabilen herhangi matematiksel bir fonksiyona ait tabloların hesaplanması ve yazılması amaçlarıyla kullanılması düşünülmüştür. Makine öğrenmesi algoritmaları eldeki verilerin niteliği ve analiz edilmesi sonucu elde edilen sonuçların irdelenebilmesi için programlanmış bir veya birden fazla algoritma içerisinden en az bir tanesinin seçilmesini gerekli kılmaktadır.

3.2.3. Makine Öğrenmesi Algoritmaları

3.2.3.1. Multi-Lineer Regression

Multilineer regresyon (Çoklu doğrusal regresyon), bir bağımlı değişkenin (y) birden fazla bağımsız değişken (x [illegible] **doğrusal ilişki** içinde olduğunu varsayan istatistiksel bir modeldir. Basit doğrusal regresyonun genelleştirilmiş halidir.

- **Çoklu Doğrusal Regresyon Denklemleri**
  - **Temel Çoklu Doğrusal Regresyon Denklemi (Genel Gösterim)**

Çoklu doğrusal regresyon (Multiple Linear Regression, MLR), bir bağımlı değişken ile birden fazla bağımsız değişken arasındaki doğrusal ilişkiyi modellemek amacıyla kullanılan temel istatistiksel yöntemlerden biridir. Modelin genel formu aşağıdaki 3.1'deki denklemle ifade edilir:

$$Y = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \cdots + \beta_\rho\chi_\rho + \varepsilon \quad (3.1)$$

Bu denklemdeki terimler şu şekilde tanımlanır: Y: Modelin bağımlı değişkeni olup, açıklanmaya veya tahmin edilmeye çalışılan değeri temsil eder. $X^1, X^2, \ldots, X$[illegible]: Modelde yer alan bağımsız değişkenlerdir ve Y üzerinde açıklayıcı etkisi olduğu varsayılır. β^0, sabit terim (intercept) olup, tüm bağımsız değişkenlerin sıfır olduğu durumda Y 'nin aldığı değeri gösterir. $\beta^1, \beta^2, \ldots, \beta$[illegible]: Her bir bağımsız değişkene karşılık gelen regresyon katsayılarıdır. Bu katsayılar, ilgili değişkenin bağımlı değişken üzerindeki doğrusal etkisinin büyüklüğünü ve yönünü ifade eder. ε: Modelin açıklayamadığı kısmı temsil eden hata terimidir. Gözlemlenen değerler ile model tarafından tahmin edilen değerler arasındaki farkı içerir. Bu yapı doğrusal varsayımlar altında çalışır ve model parametreleri genellikle En Küçük Kareler Yöntemi (OLS – Ordinary Least Squares) ile tahmin edilir. Çoklu doğrusal regresyon modeli, bağımlı değişkenin birden çok faktörden nasıl etkilendiğini analiz etmek ve geleceğe yönelik tahminlerde bulunmak için yaygın olarak kullanılmaktadır.

  - **Tahmin Edilen Regresyon Denklemi**

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1\chi_1 + \hat{\beta}_2\chi_2 + \cdots \hat{\beta}_\rho\chi_\rho \quad (3.2)$$

3.2'deki denklemde; $\hat{Y}$, tahmin edilen değeri; $\hat{\beta}_i$ ise tahmini regresyon katsayılarını ifade etmektedir.

- **Matris Formunda Çoklu Doğrusal Regresyon Denklemi**

Çoklu doğrusal regresyon modeli, birden fazla bağımsız değişken ile bir bağımlı değişken arasındaki ilişkiyi analiz etmek amacıyla kullanılır. Bu modelin daha kompakt ve genel bir gösterimi matris cebiri kullanılarak ifade edilebilir. Matris formundaki temel regresyon denklemi 3.3 sayılı denklemdeki gibidir:

$$Y = X\beta + \varepsilon \tag{3.3}$$

Burada: Y:$n \times 1$boyutunda bağımlı değişken vektörüdür, her satır bir gözleme karşılık gelir. X:$n \times p$ boyutundaki bağımsız değişkenler tasarım matrisidir. (gözlem sayısı n, değişken sayısı p). İlk sütun genellikle sabit terimi (intercept) temsil eden 1'lerden oluşur. β:$p \times 1$ boyutunda regresyon katsayıları vektörüdür. ε: $n \times 1$ boyutundaki hata terimi (residual) vektörüdür.

Model parametrelerinin tahmini, genellikle En Küçük Kareler Yöntemi (Ordinary Least Squares, OLS) ile yapılır. Bu yöntemde, tahmin edilen katsayı vektörü $\widehat{\beta}$, aşağıdaki kapalı formülle elde edilir:

$$\widehat{\beta} = \left(X^{\intercal}X\right)^{-1}X^{\intercal}Y \tag{3.4}$$

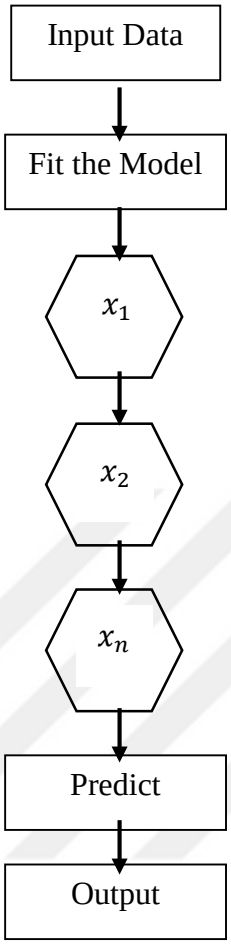
Bu ifade, hata kareleri toplamını minimize eden katsayı vektörünü verir. Burada $X^{\intercal}$ matrisin transpozunu, $\left(X^{\intercal}X\right)^{-1}$ ise tersini temsil etmektedir. OLS çözümü, $X^{\intercal}X$ matrisinin terslenebilir olması koşuluyla geçerlidir. Bu matris temsili, özellikle çok boyutlu veri kümeleriyle çalışırken hem teorik analiz hem de sayısal hesaplama açısından büyük kolaylık sağlar(Denklem 3.4).

Çoklu doğrusal regresyon modelinde temel amaç, modelin tahmin ettiği değerler ($\hat{y}_i$) ile gerçek gözlem değerleri (y_i) arasındaki farkın mümkün olduğunca küçük olmasıdır. Bu farkların kareleri alınarak toplam hata hesaplanır ve bu toplam hata, regresyon katsayılarını optimize etmek için kullanılır.

Bu bağlamda kullanılan en yaygın yöntemlerden biri **En Küçük Kareler Yöntemi**dir (Ordinary Least Squares, OLS). OLS, tüm gözlemler için tahmin hatalarının karelerinin toplamını minimize eden katsayıları bulmayı hedefler. Matematiksel olarak bu amaç 3.5'teki denklemle ifade edilir:

$$\min_{\beta} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 \tag{3.5}$$

Burada: y_i: i gözlem için gerçek (bağımlı) değer. $\hat{y}_i$: i gözlem için modelin tahmin ettiği değer.



3.1. Multilineer regresyon akış diyagramı

Çoklu Doğrusal Regresyon (MLR), bağımlı ve bağımsız değişkenler arasındaki ilişkiyi açıklayan bir makine öğrenimi algoritmasıdır(Zhao, Caiafa, Mandic, Chao, Nagasaka, Fujii ve Cichocki, 2012).

Çoklu doğrusal regresyonda, birden fazla bağımsız değişken (x_1, x_2, …, x_n) ile bir bağımlı değişken "y" arasındaki ilişki elde edilir. Her bir bağımsız değişken ile bağımlı değişken arasındaki ilişki, denklem 3.6 yardımıyla belirlenebilir.

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n \tag{3.6}$$

Bu şekilde, doğrusal 'b' katsayısı yerine 'n' doğrusal regresyon katsayısı elde edilir. Hem basit hem de çoklu regresyon yöntemlerinde bu işlevi sağlamak için kullanılan yöntem, en küçük kareler yöntemidir. Ayrıca, MLR, en küçük mutlak sapmalar regresyonu gibi diğer yöntemlere veya lasso (L1-norm cezası) ve ridge regresyonu

(L2-norm cezası) gibi en küçük kareler maliyet fonksiyonlarına da uyarlanabilir. (Sykes, 1993), Uyanık ve Güler, 2013), Yan ve Su, 2009).

□3.2.3.2. Support Vector Regression

Support Vector Regression (SVR), Support Vector Machine (SVM) algoritmasının regresyon (sürekli değer tahmini) için uyarlanmış halidir. Amaç, verilerle en fazla sayıda noktayı kapsayan bir marj aralığında tahmin yapan bir fonksiyon bulmaktır. SVR'nin Temel Fikri; veri noktalarının birçoğunu belirli bir hata toleransı (ϵ\epsilonϵ) içinde tutan düz bir çizgi veya non-lineer bir yüzey bulmaya çalışır.

- **Support Vector Regression (SVR) Denklemleri**
  - **Lineer SVR Temel Tahmin Denklemi:**

Destek Vektör Regresyonu (Support Vector Regression, SVR), regresyon problemlerinde doğrusal veya doğrusal olmayan ilişkileri modellemek için kullanılan güçlü bir makine öğrenmesi yöntemidir. Lineer SVR modeli, veriler ile hedef değişken arasında doğrusal bir ilişki olduğunu varsayar. Bu modelin tahmin fonksiyonu aşağıdaki denklem 3.7 ile ifade edilir:

$$f(x) = \langle w, x \rangle + b = w \quad x + b \tag{3.7}$$

Burada: $f(x)$: Girdi x için modelin tahmin ettiği çıktı değeridir. w: Ağırlık vektörü, doğrusal regresyon düzleminin yönünü belirler. x: Girdi örneğini temsil eden özellik vektörüdür. $\langle w, x \rangle$: w ve x vektörleri arasındaki iç çarpımı temsil eder. b: Bias (sapma) terimi, modelin tahmin düzlemini dikey olarak kaydırmasını sağlar. Lineer SVR, bu tahmin fonksiyonu ile, tüm verilerin $f(x)$ tahminlerine olan uzaklığının belirli bir tolerans sınırı (epsilon-insensitive zone) içinde kalmasını amaçlar. Hedef, yalnızca bu sınırı aşan tahmin hatalarına ceza vererek hem modelin genellenebilirliğini artırmak hem de daha kararlı bir tahmin gerçekleştirmektir.

  - **SVR Amaç Fonksiyonu (ε-tüplü regresyon):**

Destek Vektör Regresyonu (Support Vector Regression, SVR), regresyon problemlerinde hem doğruluk hem de modelin genellenebilirliğini dikkate alan bir yaklaşımdır. SVR'nin temelinde, verilerin tahmin edilen değerlerden belirli bir tolerans (ε) sınırı içinde kalması hedeflenir. Bu sınırın dışına çıkan örnekler için ceza uygulanır. ε -tüplü regresyon olarak bilinen bu yaklaşımda amaç fonksiyonu şu şekilde formüle edilir:

$$\min_{w,\xi,\xi^*} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \quad (3.8)$$

Burada w: Ağırlık vektörüdür ve modelin tahmin düzlemini tanımlar. $\| w \|^2$: Modelin **karmaşıklığını** temsil eder; daha düz bir regresyon doğrusunu tercih ederek aşırı öğrenmeyi önler. $\xi_i,\ \xi_i^*$: Her bir gözlem için pozitif **hata toleransı (slack variables)** terimleridir; gözlem değerinin ε sınırını aştığı durumlarda kullanılır. C: **Ceza parametresi**, modelin hata toleransı ile genelleme kabiliyeti arasında denge kurar. Büyük C değerleri, hatalara daha az tolerans gösterilmesini sağlar. n: Gözlem sayısıdır(Denklem 3.8).

Bu optimizasyon problemi, iki temel hedefi dengelemeye çalışır:

1. Regresyon fonksiyonunun düz ve basit kalmasını sağlamak (yani $\| w \|^2$ terimini minimize etmek),
2. ε sınırının dışına çıkan tahmin hatalarını ceza terimleri $\xi_i + \xi_i^*$ aracılığıyla kontrol altına almak.

Amaç fonksiyonuna karşılık gelen kısıtlar (constraints) ile bu yapı, SVR'nin optimizasyon problemini tamamlar. Böylece model, belirli bir esneklik içinde yüksek doğruluk ve düşük genelleme hatası hedefler.

Destek Vektör Regresyonu (Support Vector Regression – SVR) algoritmasında, amaç fonksiyonunun minimize edilmesi süreci belirli **kısıt koşulları** altında gerçekleştirilir. Bu kısıtlar, tahmin edilen değerlerin gerçek değerlere belirli bir tolerans (ε) dahilinde yakın olmasını sağlamaya yöneliktir. SVR'nin **ε-tüplü regresyon** yaklaşımında kullanılan kısıt denklemleri 3.9 denklemindeki gibidir:

$$\begin{gathered} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_{i \leq} \varepsilon + \xi_i \\ \xi_i,\ \xi_i^* \geq 0 \end{gathered} \quad (3.9)$$

Burada: w: Modelin **ağırlık vektörüdür** ve regresyon düzleminin eğimini belirler. x_i:i gözleme ait bağımsız değişken vektörüdür. b: **Bias terimi**, regresyon doğrusunun konumunu belirler. y_i: Gerçek (gözlenen) bağımlı değişken değeridir. ε: Tolerans (ε -tüpü) parametresidir; modelin tahminleriyle gerçek değerler arasında kabul edilebilir farkı belirler. $\xi_i,\ \xi_i^*$: Pozitif **gevşetme değişkenleri (slack variables)** olup, ε

sınırının dışına çıkan sapmaları temsil eder. $\langle w, x_i \rangle = w \top xi \langle w, x_i \rangle = w \top xi$: Ağırlık ve girdi vektörlerinin iç çarpımıdır.

Bu kısıtlar şunu ifade eder:

- İlk eşitsizlik, model tahmininin gerçek değerden **en fazla** $\varepsilon + \xi_i \varepsilon + \xi_i$ kadar düşük olmasına izin verir.
- İkinci eşitsizlik, model tahmininin gerçek değerden **en fazla** $\varepsilon + \xi_i^* \varepsilon + \xi_i^*$ kadar yüksek olmasına izin verir.
- Üçüncü ve dördüncü koşullar ise, gevşetme değişkenlerinin negatif olmamasını garanti eder.

Bu yapı sayesinde, SVR modeli hem **doğrusal regresyon düzlemini** optimize eder hem de gerçek değerlerle tahminler arasındaki farkı belirli bir hata toleransı içinde kontrol altında tutar.

- **Çekirdekli SVR için Tahmin Denklemi (Dual Form):**

Destek Vektör Regresyonu (SVR), doğrusal olmayan ilişkilere sahip veri setlerinde daha yüksek esneklik ve tahmin gücü sağlayabilmek için çekirdek (kernel) yöntemleri ile genişletilebilir. Bu durumda, orijinal veri uzayı daha yüksek boyutlu bir özellik uzayına dönüştürülür ve regresyon işlemi bu uzayda gerçekleştirilir. Bu tür SVR modelleri, dual formüzerinden çözülür ve tahmin fonksiyonu denklem 3.10'daki şekilde tanımlanır:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \, \mathrm{K}(x_i, x) + b \qquad (3.10)$$

Burada: $f(x)$: Girdi x için modelin tahmin ettiği çıktıdır. α_i, α_i^*: Her bir eğitim örneğine karşılık gelen **Lagrange çarpanlarıdır**. Bu çarpanlar, SVR optimizasyonunun dual formunun çözümünde ortaya çıkar. $\mathrm{K}(x_i, x)$:x_i ile x arasındaki **çekirdek fonksiyonudur** ve giriş uzayında hesaplanan benzerliği temsil eder. b: **Bias (sapma)** terimidir ve modelin regresyon düzlemini düşey eksende kaydırmasına olanak tanır. n: Eğitim örneği sayısıdır.

Çekirdek fonksiyonu $\mathrm{K}(x_i, x)$, verilerin yüksek boyutlu uzayda iç çarpımını hesaplamayı sağlar; böylece doğrusal olmayan desenler de doğrusal SVR çerçevesinde modellenebilir. Yaygın çekirdek fonksiyonları arasında şunlar bulunur:

- Lineer çekirdek: $(x_i, x) = x_i \top x$

- RBF (Gaussian) çekirdeği: K (x_i, x)=exp$(-\gamma \parallel x_i - x \parallel 2)$
- Polinom çekirdeği: K $(x_i, x) = (x_i \top x + c)^d$
- Bu yapı sayesinde, yalnızca destek vektörleri (yani $\alpha_i - \alpha_i{}^* \neq 0$ olan örnekler) model tahminine katkıda bulunur. Bu da SVR'nin hem hesaplama verimliliğini artırır hem de modelin genelleme kapasitesini güçlendirir.

o Matematiksel Tanım

Verilen bir regresyon problemi için, gözlemler $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$şeklinde tanımlanmış bir veri seti üzerinden, amaçlanan; giriş uzayındaki x vektörlerini, sürekli bir hedef değişken y ile ilişkilendiren bir tahmin fonksiyonu $f(x)$ elde etmektir.

Destek Vektör Regresyonu (SVR) algoritması bu amaçla, denklem 3.11'de tanımlanan doğrusal bir fonksiyonu öğrenmeyi hedefler:

$$f(x) = \omega^T x + b \tag{3.11}$$

Burada: ω: Ağırlık vektörüdür ve modelin yönünü belirler. b: Bias (sapma) terimidir ve modelin sabit bileşenini temsil eder. SVR'nin temel varsayımı, bu fonksiyonun tüm tahminlerinin, gerçek gözlem değerlerinden en fazla ε kadar sapmasına izin verilmesidir. Bu varsayım şu şekilde ifade edilir:

$$|y_i - f(x_i)| \leq \epsilon \tag{3.12}$$

Denklem 3.12, modelin yalnızca belirli bir **tolerans bölgesi (ε-tüpü)** içinde yer alan tahmin hatalarını "önemsiz" olarak değerlendirdiğini ve bu bölgenin dışına çıkan hatalar için ceza uygulanacağını gösterir. Ancak, tüm verilerin bu tolerans içinde kalması çoğu zaman mümkün değildir. Bu nedenle, SVR modeli tolerans sınırını aşan durumları temsil etmek üzere **gevşetme değişkenleri (slack variables)** olan ξ_i ve $\xi_i{}^*$ terimlerini tanımlar.

Bu terimler, sırasıyla modelin $y_i - f(x_i) > \varepsilon y_i - f(x_i) > \varepsilon$ ve $f(x_i) - y_i > \varepsilon$ $f(x_i) - y_i > \varepsilon$ koşullarını ihlal ettiği durumlarda ortaya çıkan sapmaları ifade eder. Böylece SVR, hem doğrusal tahmin fonksiyonunu hem de tolerans ihlallerine karşılık gelen ceza bileşenlerini birlikte optimize ederek daha esnek ve genellenebilir bir regresyon modeli oluşturur. Ancak bu her zaman mümkün olmayacağı için, toleransı aşan durumlar için **ceza terimleri (slack variables)** tanımlanır.

- **SVR ve Diğer Regresyonlara Göre Farkları**

Amaç, hataları **tam minimize etmek değil**, onları ϵ\epsilonϵ-tüp içinde tutmak.

Marj aralığı (tüp) tanımı, modelin genelleştirme kabiliyetini artırır.

Lineer olmayan durumlar için **kernel trick** (örneğin RBF kernel) kullanılabilir.

- **SVR Avantajları**

- Karmaşık ve lineer olmayan ilişkileri modelleyebilir.
- Aykırı değerlere karşı daha dayanıklıdır (ayarlara bağlı olarak).
- Yüksek boyutlu veriyle iyi çalışır.

Vapnik (1995) tarafından önerilen destek vektör regresyonu, istatistiksel öğrenme teorisindeki yeniliklere dayanan ve hem sınıflandırma hem de regresyon problemleri için kullanılan güçlü bir makine öğrenme yöntemidir. Destek vektör regresyonu (SVR), genelleme hatasının üst sınırını minimize etmeye dayalı risk minimizasyonu ilkesiyle çalışır. Bu özellik, SVR'yi aşırı uyum (overfitting) sorunlarına karşı dirençli hale getirir ve farklı zaman serisi tahmin problemlerini çözerken yüksek bir genelleme performansı elde etmesini sağlar (Fan, Chen ve Lee, 2008). Bu yöntem, deneysel risk minimizasyonu ile hataları en aza indirmeye çalışan yapay sinir ağlarının (ANN) yerel minimumlarda takılma gibi dezavantajlarını ortadan kaldırır. Destek vektör regresyonu (SVR) ilkesi, verileri daha yüksek boyutlu bir özellik uzayına dönüştürmeye dayanır. SVR'nin öğrenme kapasitesi, özellik uzayının boyutundan bağımsızdır, bu nedenle yüksek performans sergiler(Oğcu, Demirel ve Saim, 2012).

SVM, özellikle küçük örneklem büyüklüğüne sahip problemleri çözmek için uygundur (Chen, Lu, Yang ve Li, (2004). SVM, 1974'te Vapnik ve Chervonenkis tarafından geliştirilen istatistiksel öğrenme teorisi çerçevesine dayanır. SVR, giriş verilerini x'i, doğrusal olmayan bir dönüşüm Φ ile daha yüksek boyutlu bir özellik uzayına haritar ve ardından bu özellik uzayında bir doğrusal regresyon problemi elde edilip çözülür (Gao, Bompard, Napoli ve Cheng, 2007).

Zouzou ve Çıtakoğlu (2021) yaptığı çalışmalarda; Destek vektör makinesi (SVM) algoritması, başlangıçta bir sınıflandırıcı olarak geliştirilmiştir. Bu algoritma, farklı değişkenleri ayıran en uygun hiper düzlemi bulmaya çalışır. Ancak, çoğu durumda, değişkenler orijinal boyutlu uzayda bir hiper düzlemle ayrılamaz. Çekirdek fonksiyonları, verileri daha yüksek boyutlu bir uzaya dönüştürerek, değişkenlerin bir

hiper düzlemle ayrılabileceği yeni bir uzay sağlar. Bu çalışmada, radyal baz fonksiyon çekirdeği SVM çekirdeği olarak kullanılmıştır. Aynı algoritma ile, sınıflandırma hiper düzlemi, marjın içine düşen örnek sayısını maksimuma çıkaracak şekilde bir regresyon hiper düzlemi olarak da kullanılabilir.

3.2.3.3. Rastgele Orman (Random Forest) Algoritması

Random Forest, makine öğrenmesinde kullanılan bir topluluk yöntemidir (ensemble method). Temel olarak birçok karar ağacından oluşur ve her bir ağaç bağımsız olarak eğitilir. Bu yöntemin temel amacı, birden fazla modelin çıktısını birleştirerek daha doğru ve kararlı sonuçlar elde etmektir.

- **Regresyon için Random Forest Denklemi**

Random Forest regresyon problemlerinde, her bir karar ağacı bir tahminde bulunur ve nihai sonuç bu tahminlerin ortalaması alınarak elde edilir:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t\,(x) \tag{3.13}$$

Modelin tahmin sürecine ilişkin olarak, 3.13 denkleminde $\hat{y}$, girdi verisi x için modelin elde ettiği tahmin değerini ifade eder. Rastgele Orman (Random Forest) algoritmasında T, modelde yer alan toplam karar ağacı sayısını temsil etmektedir. Her bir karar ağacının çıktısı ise $h_t(x)$ ile gösterilir; burada t , ilgili karar ağacının indeksidir. Sonuç olarak, modelin nihai tahmini, tüm ağaçlardan alınan bireysel tahminlerin birleştirilmesiyle elde edilir.

- **Sınıflandırma için Random Forest Denklemi**

Sınıflandırma problemlerinde Random Forest, her ağacın oy kullanması prensibine dayanır. Her bir ağaç, verilen girdiye karşılık bir sınıf tahmin eder. En fazla oy alan sınıf, nihai tahmin olarak belirlenir:

$$\hat{y} = mode\,(\{h_t(x)\})_{t=1}^{T}) \tag{3.14}$$

Sınıflandırma problemlerinde, Rastgele Orman (Random Forest) algoritması her bir karar ağacının bağımsız olarak yaptığı sınıf tahminlerini dikkate alır. Bu bağlamda, $h_t(x)$, t indeksli karar ağacının girdi verisi x için gerçekleştirdiği sınıf tahminini temsil eder. Nihai tahmin ise tüm ağaçlardan gelen sınıf tahminleri arasında en sık tekrar eden değerin (mode) seçilmesiyle belirlenir; bu yöntem, çoğunluk oyu prensibine dayanmaktadır (Denklem 3.14).

- **Sınıflandırma İçin Olasılık Tahmini Denklemi (Soft Voting)**

Makine öğrenmesi kapsamında sınıflandırma problemlerinde yaygın olarak kullanılan topluluk (ensemble) yöntemlerinden biri olan Rastgele Orman (Random Forest) algoritması, birçok karar ağacından oluşan bir modeldir. Bu algoritmada, her bir karar ağacı, gözlem verisi için bağımsız bir sınıf tahmininde bulunur. Bu tahminlerin toplu biçimde değerlendirilmesi, sınıf kararının belirlenmesinde kritik rol oynar. Soft voting (yumuşak oylama) yaklaşımı, nihai sınıf kararını doğrudan çoğunluk oyuna göre belirlemek yerine, her bir sınıf için olasılık değerleri üzerinden karar verme esasına dayanır. Bu yöntemde, modelin herhangi bir gözlem x için belirli bir sınıfa ($y = c$) ait olma olasılığı, tüm karar ağaçlarının bu sınıfa yönelik tahminlerinin ortalaması alınarak hesaplanır. Söz konusu olasılık tahmini şu şekilde matematiksel olarak ifade edilir:

$$p(y = c|x) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}(h_t(x) = c) \qquad (3.15)$$

3.15 sayılı denklemde, $p(y = c|x)$, gözlem x'in sınıf c'ye ait olma olasılığını ifade etmektedir. T, topluluk içerisindeki toplam karar ağacı sayısını göstermektedir. $h_t(x)$, t karar ağacının gözlem x için yaptığı sınıf tahminidir. $\mathbb{I}$, girdi koşulunun sağlanması durumunda 1, aksi hâlde 0 döndüren gösterge (indicator) fonksiyonudur.

Soft voting yöntemi, her sınıf için elde edilen bu olasılık değerlerini karşılaştırarak, en yüksek olasılığa sahip sınıfı nihai çıktı olarak belirler. Bu yaklaşım, özellikle sınıflar arasında dengesizlik bulunan veri kümelerinde, karar ağaçlarının tekil çoğunluğuna dayanan hard voting (sert oylama) yöntemine kıyasla daha istikrarlı ve genelleştirilebilir sınıflandırma sonuçları üretme potansiyeline sahiptir.

- **Yöntemin Temel Özellikleri**

Rastgele Orman (Random Forest) algoritması, topluluk öğrenmesi (ensemble learning) yaklaşımına dayanan ve hem sınıflandırma hem de regresyon problemlerinde yüksek başarı sağlayan bir makine öğrenmesi yöntemidir. Bu yöntemde her bir karar ağacı, eğitim verisinin rastgele seçilmiş bir alt kümesi ile eğitilir. Bu işlem **bootstrap örnekleme** (bootstrap sampling) adı verilir ve her ağacın farklı bir veri alt kümesine dayanarak öğrenmesini sağlar.

Modelin rastgeleliğini artıran bir diğer unsur ise, her bölme (split) işlemi sırasında karar verilecek özelliğin, tüm öznitelik kümesi yerine bu kümeden rastgele seçilmiş

bir alt küme içerisinden belirlenmesidir. Bu mekanizma, bireysel ağaçlar arasındaki korelasyonu azaltarak modelin çeşitliliğini artırır.

Modeldeki bu çift yönlü rastgelelik — hem veri örneklemesinde hem de özellik seçiminde — aşırı öğrenme (overfitting) riskini azaltmakta ve modelin daha iyi genellenebilirlik performansı göstermesine katkıda bulunmaktadır. Tüm bu özellikleri sayesinde, Random Forest algoritması çeşitli veri yapılarında istikrarlı, esnek ve yüksek doğrulukla çalışan güçlü bir yöntem olarak literatürde geniş uygulama alanı bulmaktadır.

- **Random Forest Akış Diyagramı (Adım Adım)**

- **Veri Girişi (Dataset)**

Tüm eğitim verileri sisteme girer.

- **Bootstrap Örnekleme**

Eğitim setinden rastgele (tekrar seçmeli) örnekler alınarak *n* tane farklı alt veri kümesi oluşturulur.

- **Karar Ağaçlarının Eğitimi (n adet)**

Her bir alt veri kümesiyle bir karar ağacı eğitilir. Her düğümde rastgele *k* özellik seçilir. Bu özellikler içinden en iyi bölünme seçilir.

- **Tüm Ağaçların Eğitimi Tamamlanır**

Her ağaç kendi başına eğitilmiş olur, ağaçlar birbirinden bağımsızdır.

- **Tahmin Aşaması (Test Verisi Girilir)**

Her ağaç test verisine ayrı ayrı tahmin üretir.

- **Sonuçların Birleştirilmesi**

Sınıflandırma: Ağaçların oyu ile çoğunluk oyu alınır.
Regresyon: Tüm ağaçların tahminlerinin ortalaması alınır.

- **Nihai Tahmin Çıktısı**

Rastgele Orman Algoritması; Farklı karar ağaçlarının birleşimine dayanan Rastgele Orman Algoritması Breiman (2001) tarafından geliştirilmiştir. Oluşturulan her bir ağaç için bağımsız sınıflandırma yapılması amaçlanmıştır. Yapılan sınıflandırılmalar neticesinde oylama yapılarak en yüksek değeri alan sınıf sonuç olarak kabul

edilmektedir (Ercire & Ünsal, 2021: 910). Rastgele orman regresyonu, birçok regresyon ağacını bir araya getirerek çalışan bir topluluk öğrenme yöntemidir. Breiman (2001) Bu yöntemde kullanılan her bir regresyon ağacı, kökten başlayıp yaprak düğümlerine kadar ilerleyen, belirli kurallar veya koşulların sırasıyla uygulandığı hiyerarşik bir yapıya sahiptir (Rodriguez-Galiano, Mendes, Garcia-Soldado, Chica-Olmo ve Ribeiro, 2014).

Rastgele Orman algoritmasında, her ağaç bağımsız şekilde tahmin üretir ve bu tahminler bir araya getirilerek genellikle ortalama alınarak sonuç belirlenir (Breiman, 2001).

Rastgele Orman (Random Forest) algoritması hem sınıflandırma hem de regresyon problemlerinde kullanılabilen, çok sayıda karar ağacının bir araya gelmesiyle çalışan bir topluluk öğrenme yöntemidir. Sınıflandırma işlemlerinde en sık görülen sınıf sonucu, regresyonda ise tahminlerin ortalaması nihai sonuç olarak kabul edilir. Bu çoklu yapı sayesinde, tek bir karar ağacında oluşabilecek hatalar azaltılarak daha güvenilir ve dengeli sonuçlar elde edilir.

Algoritmanın çalışma prensibi şu adımlardan oluşur:

1. Her karar ağacı için eğitim verisinden **bootstrap** yöntemiyle (yeniden örnekleme yapılarak) rastgele alt kümeler oluşturulur.
2. Bu alt kümeler üzerinde her bir karar ağacı ayrı ayrı eğitilir.
3. Ağaçlar oluşturulurken, aşırı öğrenmeyi önlemek amacıyla tüm özellikler yerine sadece rastgele seçilen bir alt kümesi değerlendirmeye alınır.
4. Sınıflandırma işlemlerinde, veri bölünmelerinin kalitesini ölçmek için genellikle **Gini safsızlık indeksi** veya **bilgi kazancı** gibi kriterler kullanılır. Bu sayede oluşturulan düğümlerin veriyi ne kadar iyi temsil ettiği değerlendirilir.
5. Eğitilen tüm ağaçların çıktıları birleştirilerek, ortalama veya çoğunluk kararı ile nihai tahmin üretilir.
6. Sonuç olarak, Random Forest algoritması, farklı veri alt kümeleri ve özellikler kullanılarak oluşturulan birçok karar ağacını birleştirerek genel tahmin başarısını artıran, esnek ve güçlü bir yöntemdir. (Svetnik vd., 2003; Biau ve Scornet, 2016). Rastgele Orman, yüksek boyutlu verilerle etkili şekilde çalışabilmesi ve sağlam yapısıyla öne çıkar; bu sayede tek bir karar ağacına

kıyasla aşırı uyum (overfitting) gibi problemlere karşı daha dirençlidir. Aynı zamanda, birden fazla karar ağacından oluşan yapısı sayesinde, modelin belirsizlik düzeyi doğal olarak ölçülebilir ve bu durum tahminlerin güvenilirliğini artırır. (Dutschmann ve Baumann, 2021; Li, 2023).

3.2.3.4. LightGBM

LightGBM, özellikle büyük veri setlerinde yüksek işlem hızı, düşük bellek kullanımı ve modelleme verimliliği ile öne çıkan güçlü bir gradyan artırma yöntemidir. Bu algoritma, verileri histogram tabanlı ayrıştırma (histogram-based splitting) yöntemiyle işleyerek, bölgesel sıcaklık dinamikleri gibi karmaşık ilişkileri etkili biçimde modelleyebilmektedir. LightGBM'in başarısı; model doğruluğu, hesaplama hızı, tutarlılık ve kullanım kolaylığı gibi çeşitli performans kriterleri açısından değerlendirilmektedir. Büyük ölçekli veri setleriyle çalışırken sunduğu işlem avantajları, makine öğrenimi uygulamalarında geniş bir kullanım alanı sağlamaktadır. Bu özellikleri sayesinde LightGBM, model geliştirme süreçlerinde hem esneklik hem de yüksek tahmin gücü sunan etkili bir algoritma olarak tercih edilmektedir (Ke, Meng, Finley, Wang, Chen, Ma ve Liu, 2017).

LightGBM (Light Gradient Boosting Machine), karar ağaçları üzerinden çalışan ve gradyan artırma (Gradient Boosting) mantığını kullanan bir makine öğrenimi algoritmasıdır. Temelinde GBDT (Gradient Boosting Decision Tree) algoritması yer alır. Aşağıda LightGBM'in matematiksel temeli sadeleştirilmiş biçimde açıklanmıştır.

Temel GBDT (Gradient Boosting Decision Tree) Formülü

Gradyan Artırma Karar Ağaçları (Gradient Boosted Decision Trees – GBDT), aşamalı öğrenme yaklaşımını benimseyen güçlü bir topluluk (ensemble) yöntemidir. Model, her bir iterasyonda mevcut tahminleri iyileştirmek amacıyla yeni bir zayıf öğrenici (weak learner) ekleyerek tahminlerini günceller. Bu süreç 3.16'daki denklemde matematiksel olarak ifade edilir.

$$\hat{y}^{(t)} = \hat{y}^{(t-1)} + f_t(x) \tag{3.16}$$

B$urada, \hat{y}^{(t)}$: t'inci iterasyondaki tahmin, $f_t(x)$: t'inci iterasyondaki öğrenilen karar ağacı (weak learner), Gradyan Yaklaşımını temsil etmektedir. Her yeni öğrenici, önceki modelin hatalarını telafi edecek şekilde öğrenilmekte ve bu süreçte belirli bir kayıp fonksiyonu (loss function) esas alınmaktadır. Temel amaç, bu kayıp

fonksiyonunu minimize edecek şekilde$f_t(x)$fonksiyonunun (yani karar ağacının) yapılandırılmasıdır. Bu yaklaşım, modelin hem doğruluğunu artırmakta hem de karmaşık verilerle başa çıkmasını sağlamaktadır.

- **Gradyan Artırmalı Karar Ağaçlarında (GBDT) İkinci Dereceden Optimizasyon ve Split Kararı**

Her iterasyonda, Gradyan Artırmalı Karar Ağaçları (GBDT) algoritması, modelin hatalarını azaltmak için kayıp fonksiyonunun gradyanına (birinci türev) ve Hessian'ına (ikinci türev) başvurarak yeni bir karar ağacı oluşturur. Her gözlem için bu türevler 3.17'deki denklemle tanımlanır.

$$g_i = \frac{\partial \mathcal{L}\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}}, h_i = \frac{\partial^2 \mathcal{L}\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)^2}} \quad (3.17)$$

Burada, g_i:i. gözlem için kayıp fonksiyonunun gradyanını (birinci türev), h_i:i. gözlem için Hessian değerini (ikinci türev), $L(\cdot)$ ise kullanılan kayıp fonksiyonu ifade etmektedir. (örneğin, kare hatalar toplamı - MSE veya log-loss).

- **Bilgi Kazancı (Gain) Fonksiyonu ile Split Kararı**

GBDT tabanlı algoritmalarda, özellikle LightGBM ve XGBoost gibi modern varyantlarda, her ağaç bölünmesi (split) kararında en uygun ayrımı belirlemek için bilgi kazancı (Gain) hesaplanır. Bilgi kazancı formülü denklem 3.18'deki gibidir:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \quad (3.18)$$

Burada; G_L ve G_R sol ve sağ bölme için gradyanların toplamını ifade ederken, H_L ve H_R ise sol ve sağ bölme için Hessian'ların toplamını göstermektedir. λ : L2 regularizasyon parametresi iken, γ ise bölme başına sabit ceza parametresidir. (Split penalizasyonu) Bu formül, her bölmenin model üzerindeki katkısını ölçer ve maksimum kazanç sağlayan split'in seçilmesini sağlar. Böylece modelin genel hata oranı daha hızlı şekilde düşürülür.

- **Yaprak Bazlı (Leaf-wise) Büyüme**

LightGBM, geleneksel GBDT algoritmalarında kullanılan level-wise (seviye bazlı) genişletme stratejisi yerine, leaf-wise(yaprak bazlı) genişletmeyi tercih etmektedir. Bu yöntem, her iterasyonda mevcut yapraklar arasında en fazla kayıp azaltımı sağlayan yaprağı belirleyerek yalnızca bu yaprağı genişletir. Bu sayede daha derin ve optimize

ağaçlar oluşturulur. Her ne kadar bu yöntem aşırı öğrenme riskini artırabilse de yüksek veri hacmine sahip durumlarda daha iyi genel performans ve doğruluk sağlamaktadır.

3.2.3.5. XGBoost Algoritması

XGBoost (Extreme Gradient Boosting), karar ağaçlarına dayalı, denetimli (supervised) bir makine öğrenmesi algoritmasıdır. Boosting tekniğini kullanarak tahmin doğruluğunu artırır ve genellikle hem sınıflandırma hem de regresyon problemlerinde en güçlü algoritmalardan biridir.

- **XGBoost'un Temel Özellikleri**
- **Boosting Mantığı:**

Zayıf öğrenicileri (genellikle karar ağaçları) ardışık şekilde eğitir. Her yeni ağaç, önceki ağaçların hatalarını azaltmaya çalışır.

- **Gradienet Tabanlı:**

Hataları azaltmak için **gradyan inişi** (gradient descent) mantığını kullanır. Bu nedenle adı "Gradient Boosting"dir.

- **Düzenleme (Regularization):**

Aşırı öğrenmeyi (overfitting) önlemek için L1 ve L2 ceza terimleri içerir. Bu, onu klasik Gradient Boosting'den ayıran en güçlü yanlardan biridir.

- **Ağaç Yapısı:**

Karar ağaçları *derin* ve *karmaşık* olabilir, bu sayede karmaşık veri ilişkilerini yakalayabilir.

- **Paralel Hesaplama ve Hız:**

Split aramalarında paralel işlem yapabilir. Bellek ve zaman açısından çok verimli çalışır. Bu nedenle Kaggle yarışmalarında çok tercih edilir.

XGBoost (Extreme Gradient Boosting), karar ağaçları temelli güçlü bir boosting algoritmasıdır. Her bir model, önceki modellerin hatalarını düzeltmek amacıyla ardışık şekilde inşa edilir. Aşağıda XGBoost algoritmasının temel formülasyonu ve bileşenleri açıklanmıştır:

- **Tahmin Fonksiyonu**

XGBoost, toplam K adet zayıf modelin (karar ağaçlarının) çıktılarının toplamını alarak tahmin yapar. 3.19'daki denklemle ifade edilir.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\chi_i), f_k \in \mathcal{F} \quad (3.19)$$

Makine öğrenmesi kapsamında topluluk modelleri ile yapılan tahmin süreçlerinde, her bir gözleme ilişkin model çıktıları belirli bileşenler aracılığıyla tanımlanabilir. Bu bağlamda, $\hat{y}_i$ gözlem için modelin ürettiği tahmin değerini temsil etmektedir. Topluluğu oluşturan karar ağaçlarının sayısı K ile gösterilir ve bu ağaçların her biri, zayıf öğrenici (weak learner) niteliğinde olan f_k ile ifade edilir. Burada f_k, k karar ağacına karşılık gelir.

Tüm bu karar ağaçları, belirli yapısal ve işlevsel sınırlamalara sahip bir uzayda tanımlanır; bu uzaya $\mathcal{F}$ denir. $\mathcal{F}$, örneğin maksimum derinliği sınırlandırılmış karar ağaçlarından oluşan bir fonksiyon kümesini ifade eder. Bu yapı, modelin kapasitesini kontrol altında tutarak aşırı öğrenmeyi engellemeye ve daha güçlü genelleme yeteneklerine sahip tahminler üretmeye olanak tanır.

- **Amaç Fonksiyonu (Objective Function)**

Modelin öğrenmesi sırasında optimize edilen toplam kayıp fonksiyonu denklem 3.20'deki şekildedir:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega(f_k) \quad (3.20)$$

Topluluk tabanlı öğrenme yöntemlerinde, modelin eğitimi sırasında hem tahmin hatalarının en aza indirilmesi hem de modelin aşırı karmaşık hale gelmesinin önlenmesi amacıyla belirli matematiksel bileşenler kullanılır. Bu süreçte, l, modelin tahmin performansını ölçen **kayıp fonksiyonu** olarak tanımlanır. Kayıp fonksiyonu, model tahminleri ile gerçek değerler arasındaki farkı nicel olarak değerlendirir; örneğin regresyon problemlerinde genellikle Ortalama Kare Hata (MSE), sınıflandırma problemlerinde ise log-loss kullanılmaktadır.

Modelin yalnızca doğruluğuna değil, aynı zamanda basitliğine de önem verilmesi için bir düzenleme terimi olan $\Omega(f_k)$ kullanılır. Bu terim, öğrenicinin karmaşıklığını cezalandırarak aşırı öğrenme (overfitting) riskini azaltmayı hedefler. $\Omega(f_k)$, genellikle ağaçların yapısal özelliklerine (örneğin, yaprak sayısı veya dallanma derinliği) bağlı olarak tanımlanır.

Ayrıca, iteratif öğrenme süreçlerinde her yineleme (iteration) sonucunda elde edilen tahminler $\hat{y}_i^{(t)}$ ile gösterilir; burada t, ilgili iterasyonun sırasını ifade eder. Bu gösterim, özellikle boosting gibi ardışık modelleme yaklaşımlarında, her adımda modelin nasıl geliştiğini izlemek açısından önemlidir.

- **Model Karmaşıklığı Ceza Terimi (Regularization)**

Ağaç tabanlı topluluk modellerinde, özellikle gradyan artırmalı karar ağaçları (Gradient Boosted Decision Trees, GBDT) gibi yöntemlerde, modelin yapısal karmaşıklığı ve öğrenme süreci belirli parametrelerle denetlenir. Bu bağlamda, T, bir karar ağacında bulunan toplam yaprak (leaf) sayısını ifade eder. Yapraklar, nihai tahminin belirlendiği ağaç son düğümleri olup, her biri bir skor değeri taşır. Bu skorlar, ω_j^2 ile gösterilir ve j yaprağa karşılık gelen model çıktısını belirtir.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{3.21}$$

Modelin aşırı uyum sağlamasını (overfitting) engellemek ve daha sade yapılarla güçlü genelleme performansı elde edebilmek amacıyla çeşitli düzenleme (regularization) parametreleri kullanılır. Bu parametrelerden γ, her ek yaprak için maliyet cezası uygulayarak gereksiz yaprakların oluşumunu sınırlar; böylece modelin yapısal karmaşıklığı kontrol altına alınır. Diğer bir düzenleme parametresi olan λ ise yaprak skorlarına uygulanan L2 normu üzerinden ceza ekleyerek modelin ağırlıklarının büyüklüğünü düzenler. Bu iki hiperparametre birlikte, modelin hem yapısal hem de parametrik basitliğini teşvik ederek genellenebilirliğini artırır (Denklem 3.21).

- **Dereceden Taylor Yaklaşımı ile Optimizasyon**

Gradyan artırmalı karar ağaçları (Gradient Boosting Decision Trees – GBDT) ve özellikle XGBoost algoritması, her iterasyonda modele yeni bir ağaç ekleyerek toplam hatayı minimize etmeyi amaçlar. Bu süreçte, modelin hedef fonksiyonu doğrudan optimize edilmek yerine, ikinci dereceden bir Taylor genellemesi ile yaklaşık olarak ifade edilir. Bu yaklaşım, optimizasyon sürecinin daha hesaplanabilir ve kararlı olmasını sağlar.

Her iterasyonda, toplam kayıp fonksiyonu $\mathcal{L}^{(t)}$, denklem 3.22'deki şekilde ikinci dereceden açılım ile yaklaşık olarak ifade edilir:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[g_i, f_t(x_i) + \frac{1}{2} h_i {f_t}^2(x_i) \right] + \Omega(f_t) \tag{3.22}$$

Bu denklemde yer alan terimler şu şekilde açıklanabilir:

- g_i: Örnek için mevcut modelin kayıp fonksiyonuna göre hesaplanan **birinci türev** (gradyan) değeridir. Gradyan, tahmin ile gerçek değer arasındaki sapmanın yönünü ifade eder.
- h_i: Aynı örnek için kayıp fonksiyonunun **ikinci türevidir** (Hessian). Bu terim, kayıp fonksiyonunun eğim değişimini ve dolayısıyla gradyanın ne kadar hızlı değiştiğini temsil eder.
- $f_t\ (x_i)$: İterasyonda eklenen **yeni karar ağacının**, $i.$ gözlem üzerindeki tahminidir. Bu ağaç, önceki modelin yapamadığı doğrulamaları hedefleyerek hatayı düzeltmeyi amaçlar.
- $\Omega(f_t)$: Modelin karmaşıklığını kontrol altına alan **düzenleme (regularization)** terimidir. Bu terim, ağaç yapısındaki yaprak sayısı ve yaprak skorları gibi faktörlere bağlı olarak hesaplanır.
- **En İyi Bölünmenin Skoru (Gain)**

Karar ağaçlarının oluşturulmasında, her düğümde hangi özelliğe ve eşik değerine göre bölünme yapılacağına karar verilirken bilgi kazancı (Gain) temel kriter olarak kullanılır. Bu süreç, modelin hatasını azaltacak en uygun bölünmenin seçilmesini sağlar. Özellikle XGBoost gibi ikinci türev tabanlı algoritmalarda, bilgi kazancı yalnızca hata azaltımıyla değil aynı zamanda model karmaşıklığıyla da ilişkilendirilerek hesaplanır.

Herhangi bir bölünmenin skoru denklem 3.23'teki formül ile hesaplanır:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \tag{3.23}$$

Burada:

- G_L ve G_R: Sırasıyla sol ve sağ alt düğümler için toplam **gradyan** (birinci türev) değerleridir.
- H_L ve H_R: Sırasıyla sol ve sağ alt düğümler için toplam **Hessian** (ikinci türev) değerleridir.
- λ: Ağaç yapısındaki yaprak skorlarını düzenlemek için kullanılan **L2 regülarizasyon** parametresidir.

- γ: Her yeni yaprak eklenmesine karşılık gelen **karmaşıklık ceza terimidir.**

Bu formülde, sol ve sağ alt düğümlerin ayrı ayrı bilgi kazançları ile birleşik yapının kazancı karşılaştırılır. Eğer elde edilen Gain değeri pozitifse, söz konusu bölünme modelin toplam kaybını azaltmakta ve bu nedenle uygulanmaya değer kabul edilmektedir. Tersi durumda, bölünme yapılmayarak mevcut yapı korunur.

Bu yöntem sayesinde, yalnızca istatistiksel olarak anlamlı ve genellenebilir bölünmelerin yapılması sağlanır; bu da modelin aşırı öğrenme (overfitting) riskini azaltarak performansını artırır.

- **Özet**

  • XGBoost, zayıf modelleri ardışık olarak toplar.

  • Her yeni model, kalan hataları minimize etmeye çalışır.

  • İkinci türev bilgisiyle optimize edilir, bu da öğrenmeyi hızlandırır.

  • Düzenleme (regularization) ile aşırı öğrenme azaltılır ve genel performans artırılır.

- **XGBoost'un Avantajları**

  • Çok hızlı ve verimli

  • Aşırı öğrenmeye karşı dirençli

  • Eksik verilerle iyi çalışır

  • Özellik önem derecesi çıkarımı yapılabilir

  - **XGBoost Akış Diyagramı**

  ⟶Başla

  ⟶Eğitim verisini ve parametreleri yükle

  ⟶İlk tahminleri başlat(genellikle sabit bir değer, örn. Ortalama)

  ⟶Döngü: t = 1 to T(ağaç sayısı kadar)

  ⟶Her gözlem için:

  - Gradient $g_i = \partial L(y_i, \hat{y}_i)$
  - Hessian $h_i \partial^2 L(y_i, \hat{y}_i)$

  ⟶Yeni ağaç $f_t(x)$ $oluştur$:

  - Bölümleri gradient & Hessian toplamlarına göre seç

- Gain (kazanç) fonksiyonunu kullan

⟶Ağaç yaprak değerlerini hesapla (Output weights)

⟶Modeli güncelle: $\hat{y}_i \leftarrow \hat{y}_i + \eta\ x\ f_t(x_i)$

⟶Tahmin : $\hat{y} = \sum_{t=1} T\ f_t(x)$

⟶Bitiş

XGBoost, gradyan artırma temelli bir topluluk öğrenme tekniğidir; bu yöntem, hataları azaltmak amacıyla zayıf tahmin modellerini ardışık olarak bir araya getirerek genel tahmin başarımını artırmayı hedefler (Chen ve Guestrin, 2016). XGBoost, yüksek ölçeklenebilirliği ve esnek yapısı sayesinde, gradyan artırma tabanlı kütüphaneler arasında öne çıkan ve yaygın olarak tercih edilen bir yöntem haline gelmiştir (Ding, Jin, Wang, Chen, Li ve Xie, 2021). Boosting yöntemleri arasında yer alan bu algoritma, zayıf öğrenicileri ardışık olarak geliştirerek daha güçlü ve etkili modeller oluşturmayı hedefler (Abdulazeez, 2024) ve Birden çok karar ağacını bir araya getirerek güçlü ve dayanıklı bir sınıflandırma modeli meydana getirir (Wang, Wang, Chen, Jin ve Che, 2020; Liang, Luo, Zhao ve Wu, 2020).

XGBoost, Chen ve Guestrin'in (2016) geliştirdiği, topluluk tabanlı ağaç modellerine dayanan güçlü bir makine öğrenme algoritmasıdır. Friedman'ın (2001) gradyan artırma algoritmasından faydalanarak geliştirilmiştir (Zhou, Qiu, Khandelwal, Zhu ve Zhang, 2021). XGBoost, tek başına kullanılan yöntemlerden daha iyi tahmin performansı sağlayan ve karar ağaçlarını verimli bir şekilde birleştirerek oluşturulan bir kolektif modeldir. Bu model, karmaşıklığı azaltmak, aşırı öğrenmeyi engellemek ve öğrenme sürecini hızlandırmak için amaç fonksiyonunda normalizasyon tekniklerini kullanır (Jabeur, Mefteh-Wali ve Viviani, 2021).

Gradyan Artırma (Gradient Boosting) algoritması, birden fazla zayıf tahminciyi birleştirerek güçlü bir tahminci oluşturmayı hedefleyen bir makine öğrenimi yöntemidir (Yeşilyurt ve Dalkılıç, 2021). Gradyan Artırma, genellikle karar ağaçları gibi basit modelleri kullanarak zayıf öğrenicileri birleştirir ve bu modelleri sırayla eğitir.

3.1. Gradyen artırma algoritması gradyen artırma algoritması

1. Veri setini yükle

2. Hedef değişkeni belirle

3. Başlangıç tahmini yap (örneğin, ortalama değeri kullan)

4. Başlangıç artıkları hesapla: Hedeflenen- Başlangıç tahmini

5. Döngü başlat

   - a. Zayıf öğrenici oluştur - Veri setini ve artıkları kullanarak bir karar ağacı eğit
   - b. Yeni tahmin yap: Önceki tahmin + Zayıf öğrenicinin tahmini
   - c. Yeni artıkları hesapla: Hedeflenen- Yeni tahmin
   - d. Artıkları güncelle
   - e. Döngüden çıkma kriterini kontrol et (örneğin, belirli bir iterasyon sayısı veya artıkların belirli bir eşiği)

6. Döngüyü sonlandır

İlk adımda, basit bir tahminci olarak veri setinin ortalaması kullanılır. Bu başlangıç tahmini ile gerçek değerler arasındaki farklar, "artıklar" olarak adlandırılır. Bu artıkları tahmin etmek için bir dizi zayıf öğrenici eğitilir ve genellikle karar ağaçları tercih edilir. Her ağaç, önceki ağaçların tahminleriyle gerçek değerler arasındaki farkları azaltmaya odaklanır. Son olarak, tüm zayıf öğrenicilerin tahminleri ağırlıklı bir şekilde birleştirilir ve nihai tahmin yapılır. Bu yöntem, hedef değişkeninin karmaşık ve doğrusal olmayan ilişkilerini modellemeye yardımcı olur ve hem sınıflandırma hem de regresyon problemlerinde yaygın olarak kullanılır. XGBoost, Gradyen Artırma algoritmasını kullanarak, ağaç tabanlı modellerin performansını artırmak için optimize edilmiş ve paralel çalışabilen bir makine öğrenmesi modelidir (Yeşilyurt ve Dalkılıç, 2021).

3.2.3.6. LSTM

LSTM (Long Short-Term Memory), derin öğrenme ve özellikle zaman serisi verileriyle çalışırken kullanılan bir tür yapay sinir ağıdır. LSTM, RNN (Recurrent Neural Networks) modelinin bir çeşididir ve özellikle uzun vadeli bağımlılıkları öğrenmede oldukça etkilidir. Geleneksel RNN'ler, önceki adımların bilgilerini hatırlama konusunda sınırlıdır çünkü öğrenme sürecinde uzun vadeli bağımlılıkları tutma yetenekleri yoktur. LSTM, bu sorunu aşmak için özel bir yapıya sahiptir.

LSTM'nin temel özellikleri şunlardır:

Hücre Durumu (Cell State): LSTM, hücre durumu adı verilen bir mekanizma ile verilerin uzun süreli hatırlanmasına olanak tanır. Hücre durumu, ağın "hafızası" gibi çalışır ve zaman içinde korunur.

- **LSTM Denklemleri**
    - **Giriş Kapısı (Input Gate),**

Yeni bilgiye ne kadar yer açılacağını ve hangi bilginin hücre durumuna eklenmesi gerektiğini kontrol eder.

$$i_t = \sigma(W_i x_t + \upsilon_i h_{t-1} + b_i) \tag{3.24}$$

x_t: Mevcut giriş vektörüdür. Zaman adımı t'de modele sunulan giriş verisini temsil eder. Bu veri hem kapı fonksiyonlarının hem de hücre durumu güncellemelerinin temel girdisidir. h_{t-1}: Önceki zaman adımına ait gizli durumdur. LSTM hücresinin zaman adımı $t-1$'de ürettiği çıktıdır ve hem kapı mekanizmaları hem de hücre durumu güncellemelerinde geçmiş bilginin aktarımını sağlar. W_i, υ_i: Ağırlık matrisleridir. W_i, giriş vektörü x_t ile çarpılan ağırlıkları; υ_i ise önceki gizli durum h_{t-1} ile çarpılan ağırlıkları ifade eder. Her kapı (giriş, unutma, çıkış) için ayrı ağırlık matrisleri tanımlanır. b_i: Bias (sapma) terimidir. Giriş kapısı için kullanılan bu sabit değer, lineer dönüşümlerden sonra modelin daha esnek öğrenmesini sağlar. σ: Sigmoid aktivasyon fonksiyonudur. Bu fonksiyon, giriş değerlerini 0 ile 1 arasında sıkıştırarak kapıların hangi bilginin ne ölçüde geçmesine izin vereceğini belirler. Böylece ağ, bilgi akışını yumuşak biçimde kontrol edebilir(Denklem 3.24).

Bu bileşenler, LSTM ağlarında zaman boyunca hem kısa vadeli hem de uzun vadeli bağımlılıkların öğrenilmesini mümkün kılar.

- **Unutma Kapısı (Forget Gate):**

LSTM (Long Short-Term Memory) ağlarının temel avantajlarından biri, zaman boyunca bilgiyi seçici bir şekilde hatırlayabilme veya unutabilme yeteneğidir. Bu işlev, hücre içinde yer alan unutma kapısı (forget gate) aracılığıyla gerçekleştirilir. Unutma kapısının görevi, bir önceki hücre durumunda yer alan bilgilerin ne kadarının korunacağını ve ne kadarının silineceğini belirlemektir.

Unutma kapısının matematiksel ifadesi denklem 3.25'teki gibidir:

$$f_t = \sigma\left(W_f x_t + \upsilon_f h_{t-1} + b_f\right) \tag{3.25}$$

Burada: f_t: Zaman adımı t'deki **unutma kapısı çıktısıdır**. Bu çıktı, [0, 1] aralığında değer alır ve önceki hücre durumu bilgisine ne ölçüde izin verileceğini belirler. x_t: Mevcut giriş vektörüdür. h_{t-1}: Önceki zaman adımındaki gizli durumdur. W_f, υ_f: Sırasıyla giriş ve gizli durum için tanımlanmış **unutma kapısı ağırlık matrisleridir.** b_f: Unutma kapısına ait **bias (sapma)** terimidir. $\sigma(\cdot)$: Sigmoid aktivasyon fonksiyonu olup, çıktıyı 0 ile 1 arasına sıkıştırarak bilgi akışının derecesini belirler.

- **Hücre Adayı (Candidate Cell State)**

LSTM ağlarında, her zaman adımında hücre durumunun güncellenebilmesi için yeni bilgi kaynaklı bir aday hücre durumu oluşturulur. Bu bileşen, mevcut giriş bilgisi ve önceki gizli durum kullanılarak hesaplanır ve hücreye ne kadar yeni bilginin ekleneceği, giriş kapısı tarafından belirlenir.

Aday hücre durumu (candidate cell state), 3.26 sayılı denklemle tanımlanır:

$$\hat{c}_t = \tan h\left((W_c x_t + \upsilon_c h_{t-1} + b_c)\right) \quad (3.26)$$

Burada:

- $\hat{c}_t$: Zaman adımı t'de hesaplanan **aday hücre durumu**dur.
- x_t: Mevcut zaman adımındaki giriş vektörüdür.
- h_{t-1}: Önceki zaman adımındaki gizli durumdur.
- W_c, υ_c: Aday hücre durumu için giriş ve gizli durumla ilişkili **ağırlık matrisleridir.**
- b_c: Aday hücre durumu için bias (sapma) terimidir.
- $\tan h\,(\cdot)$: **Hiperbolik tanjant aktivasyon fonksiyonu**dur; çıktıyı $[-1,1]$ aralığına sıkıştırarak hem pozitif hem de negatif bilgi aktarımına izin verir.

- **Hücre Durumu Güncellemesi**

LSTM (Long Short-Term Memory) ağlarının temel işlevlerinden biri, zaman boyunca bilgiyi taşıyabilme ve düzenli olarak güncelleyebilme yeteneğidir. Bu işlem, hücre durumu olarak adlandırılan uzun vadeli belleğin güncellenmesiyle gerçekleştirilir. Hücre durumu hem önceki bilgileri korumak hem de yeni bilgileri entegre etmek için giriş ve unutma kapıları ile birlikte kontrol edilir. Hücre durumu güncelleme denklemi 3.27'deki denklemle ifade edilir:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (3.27)$$

Burada:

- c_t: Zaman adımı t'deki güncellenmiş hücre durumu.
- c_{t-1}: Önceki zaman adımındaki hücre durumu.
- f_t: Unutma kapısı çıktısı; c_{t-1}'in ne kadarının korunacağına karar verir.
- i_t: Giriş kapısı çıktısı; yeni bilginin ne ölçüde hücreye ekleneceğini belirler.
- $\tilde{c}_t$: Aday hücre durumu; güncel girdi ve geçmiş gizli durumdan türetilmiş yeni bilgi adayıdır.
- $\odot$: Elemanlara göre çarpım (Hadamard çarpımı) işlemini ifade eder.

Bu denklem iki temel işlevi birleştirir:

1. Unutma mekanizması: $f_t \odot c_{t-1}$, eski bilgilerin seçici biçimde silinmesini sağlar.
2. Bilgi güncellemesi: $i_t \odot \tilde{c}_t$, yeni bilgilerin kontrollü şekilde hücreye eklenmesini sağlar.

Bu yapı, LSTM'nin geçmişte öğrenilen bilgileri gerektiği kadar süreyle saklamasına ve yalnızca önemli bilgileri entegre ederek zaman bağımlı problemler (örneğin zaman serisi tahmini, dil modelleme) için yüksek performanslı çözümler üretmesine olanak tanır.

- **Çıkış Kapısı (Output Gate):**

LSTM (Long Short-Term Memory) hücresinin son aşamasında, modelin o zaman adımında dışarıya ne tür bir bilgi aktaracağı çıkış kapısı (output gate) tarafından belirlenir. Çıkış kapısı, güncellenmiş hücre durumuna dayanarak gizli durumu (hidden state) üretir; bu çıktı, hem sonraki zaman adımında giriş olarak kullanılacak hem de modelin nihai çıktısı olarak görev yapacaktır.

Çıkış kapısının matematiksel tanımı denklem 3.28'deki şekildedir:

$$o_t = \sigma(W_o x_t + v_o h_{t-1} + b_o) \quad (3.28)$$

Burada:

- o_t: Zaman adımı t'deki **çıkış kapısı çıktısıdır**.

- x_t: Mevcut giriş vektörü.
- h_{t-1}: Önceki zaman adımındaki gizli durum.
- W_o, v_o: Çıkış kapısına ait **ağırlık matrisleridir.**
- b_o: Çıkış kapısına ait **bias (sapma)** terimidir.
- $\sigma(\cdot)$: Sigmoid aktivasyon fonksiyonudur; çıkış değerlerini [0, 1] aralığına sıkıştırır.

Bu kapıdan elde edilen o_t değeri, hücre durumunun aktivasyonlu hâliyle birlikte kullanılarak gizli durum (yani çıktı) aşağıdaki şekilde hesaplanır:

- **Gizli Durum (Hidden State)**

$$h_t = o_t \odot \tan h\,(c_t) \tag{3.29}$$

3.29 sayılı denklemde:

- c_t: Güncellenmiş hücre durumu.
- h_t: Zaman adımı t'deki **gizli durum**, yani ağın çıktısıdır.
- $\odot$: Eleman bazlı çarpımı (Hadamard çarpımı) gösterir.

Bu yapı sayesinde LSTM, hem geçmişten gelen uzun vadeli bilgileri koruyabilir hem de mevcut bağlamda uygun çıktılar üretebilir. Böylece hem zaman serisi tahminlerinde hem de dil modellemede başarılı sonuçlar elde edilir.

Uzun Vadeli Bağımlılıklar: LSTM, geçmişteki bilgileri daha uzun süre tutarak, zaman serisi veya dil modelleme gibi uygulamalarda etkili sonuçlar elde edilmesini sağlar. Özetle, LSTM'ler özellikle dil işleme, zaman serisi analizi ve video işleme gibi alanlarda yaygın olarak kullanılır, çünkü bu tür verilerde uzun vadeli ilişkiler genellikle çok önemlidir.

LSTM, derin öğrenme alanında kullanılan bir tekrarlayan sinir ağı (Recurrent Neural Network-RNN) modelidir. RNN mimarilerinin en temel özelliği, diğer modellere göre hatırlama yeteneğine sahip olmalarıdır. Bu ağlar, verilen girdiler arasındaki ilişkileri keşfeder ve bu ilişkileri hatırlayarak bir sonraki adımda kullanır. LSTM'nin içindeki yenilemeli yapısı ve sonucu bir sonraki girdiye aktarma özelliği sayesinde, tekrarlayan sinir ağlarında hatırlama işlemi gerçekleştirilir (Zhang, Zheng, Cui, Zong ve Li, 2019).

Literatürde, zaman serisi analizlerinde derin öğrenme modellerinden sıklıkla geri dönüşümlü yapay sinir ağı (RNN) kullanıldığı belirtilmektedir (Chen, Yeo, Lau ve

Lee, 2018; Saud ve Shakya, 2020). Ancak, RNN modelinin başlangıçtaki giriş bilgilerini unutması ve model parametrelerinin kontrolsüz şekilde güncellenmesi gibi sorunlar nedeniyle uzun dönem bağımlılıklarını öğrenmekte zorluklar yaşadığı ifade edilmektedir (Kong, Dong, Jia, Hill, Xu ve Zhang, 2019). Bu sorunları çözmek için LSTM modeli geliştirilmiştir. Hochreiter ve Schmidhuber (1997) LSTM mimarisini hafıza kapasitesi sayesinde uzun zaman serileri üzerinde kullanım için önermiştir. LSTM, zaman serisi verilerini ve sıralı bilgileri modellemek amacıyla tasarlanmış bir yapay sinir ağı türüdür. LSTM mimarisi, uzun vadeli bağımlılıkları daha iyi öğrenebilmek için özel hücre yapıları ve kapı mekanizmalarına sahiptir.

Derin sinir ağlarının eğitimi, her katmanın girdilerinin ve önceki katmanın parametrelerinin eğitim sürecinde değişmesi nedeniyle karmaşık bir süreçtir (Rehman, Malik, Raza ve Ali, 2019), (Kervancı ve Akay, 2023).

LSTM, kaybolan gradyan sorununu çözmek ve uzun vadeli bağımlılıkları daha etkili bir şekilde öğrenmek için özel olarak tasarlanmış bir RNN modelidir. LSTM mimarisi, giriş, unutma ve çıkış kapıları gibi bir dizi kapıdan oluşan bellek hücrelerinden meydana gelir. Bu kapılar, ağ içindeki bilgi akışını kontrol ederek her zaman adımında bilgilerin seçici bir şekilde hatırlanmasını veya unutulmasını sağlar. Bu geçit mekanizması, LSTM'lerin uzun diziler boyunca önemli bilgileri yakalayarak saklamasını mümkün kılar, böylece uzun menzilli bağımlılıkları modellemek için uygun hale gelir. Giriş kapısı, bellek hücresine eklenmesi gereken yeni bilgileri düzenlerken, unutma kapısı, geçmiş bilgilerin ne kadarının saklanacağına karar verir. Çıkış kapısı ise, bellek hücresinden bir sonraki zaman adımına aktarılacak bilgi miktarını belirler (Koca ve Kılıç, 2023).

3.2.3.7. Isıtma Derece Günü (Heating Degree Day - HDD)

Isıtma Derece Günü (Heating Degree Day - HDD), bir bölgedeki binaların ısıtma gereksinimini değerlendirmek amacıyla kullanılan önemli bir iklim göstergesidir. Bu gösterge, dış ortam sıcaklığının belirli bir referans sıcaklığın (genellikle 18 °C) altına düştüğü günlerde, binaların iç ortam konfor sıcaklığını koruyabilmek için gereken ısıtma enerjisini dolaylı olarak temsil eder.

HDD, günlük ortalama sıcaklık ile referans sıcaklık arasındaki farkın pozitif olduğu durumlar için hesaplanır. Aşağıdaki Denklem 3.30, HDD'nin matematiksel ifadesini göstermektedir:

$$\mathrm{HDD}_{gün} = \max(0, \mathrm{T}_{base} - \mathrm{T}_{ort}) \quad (3.30)$$

Burada:

- T_{base}: Referans sıcaklık (genellikle 18 °C olarak alınır),
- T_{ort}: Günlük ortalama dış hava sıcaklığıdır.

Eğer günlük ortalama sıcaklık referans sıcaklıktan yüksekse, yani ısınma ihtiyacı yoksa, HDD değeri sıfır olur. Ancak sıcaklık bu eşik değerin altına düştüğünde, HDD pozitif bir değer alır ve bu değer arttıkça ısıtma enerjisi ihtiyacı da artmaktadır.

Bu gösterge, enerji tüketimi modellemeleri, bina enerji verimliliği analizleri ve ısıl konfor planlamaları gibi birçok mühendislik ve çevresel değerlendirme çalışmasında temel bir parametre olarak kullanılmaktadır.

- Doğalgaz Tüketim Modellerinde HDD Kullanımı

Isıtma Derece Günü (HDD) göstergesi, özellikle konut ve endüstriyel sektörlerde doğalgaz tüketimini modellemek için yaygın olarak kullanılmaktadır. HDD'nin artması, hava sıcaklığının konfor seviyesinin altına düştüğünü ve dolayısıyla ısınma ihtiyacının arttığını göstermektedir. Bu bağlamda, HDD ile doğalgaz tüketimi arasında çoğu zaman **doğrusal bir ilişki** olduğu varsayılır.

$$Doğalgaz\ Tüketimi = \alpha + \beta \cdot HDD + \varepsilon \quad (3.31)$$

3.31 sayılı denklemde; α: Sabit terim olup, ısıtma ihtiyacının olmadığı koşullarda bile oluşan bazal doğalgaz tüketimini (örneğin pişirme veya sıcak su ihtiyacı gibi) temsil eder. β: HDD başına düşen doğalgaz tüketim katsayısıdır; ısıtma ihtiyacındaki her bir birim artışın doğalgaz tüketimine etkisini gösterir. ε: Modelin açıklayamadığı rastgele hata terimidir.

Bu model yaklaşımı, ısıl konforun korunması için gerekli olan enerji talebinin tahmininde ve enerji yönetim sistemlerinin optimizasyonunda önemli bir rol oynamaktadır. Ayrıca, mevsimsel sıcaklık değişimlerinin enerji tüketimine etkisini istatistiksel olarak değerlendirmek açısından da etkili bir yöntemdir.

5.3 Yapıya Özgü HDD Tabanlı Doğalgaz Tüketim Modeli

Bu çalışmada, doğalgaz tüketim tahminlerinin doğruluğunu artırmak amacıyla, ısıtma derece günü (Heating Degree Days, HDD) göstergesine dayalı doğrusal bir regresyon

modeli geliştirilmiştir. Modelde, HDD göstergesi ile aylık toplam doğalgaz tüketimi arasındaki ilişki aşağıdaki formülle ifade edilmiştir:

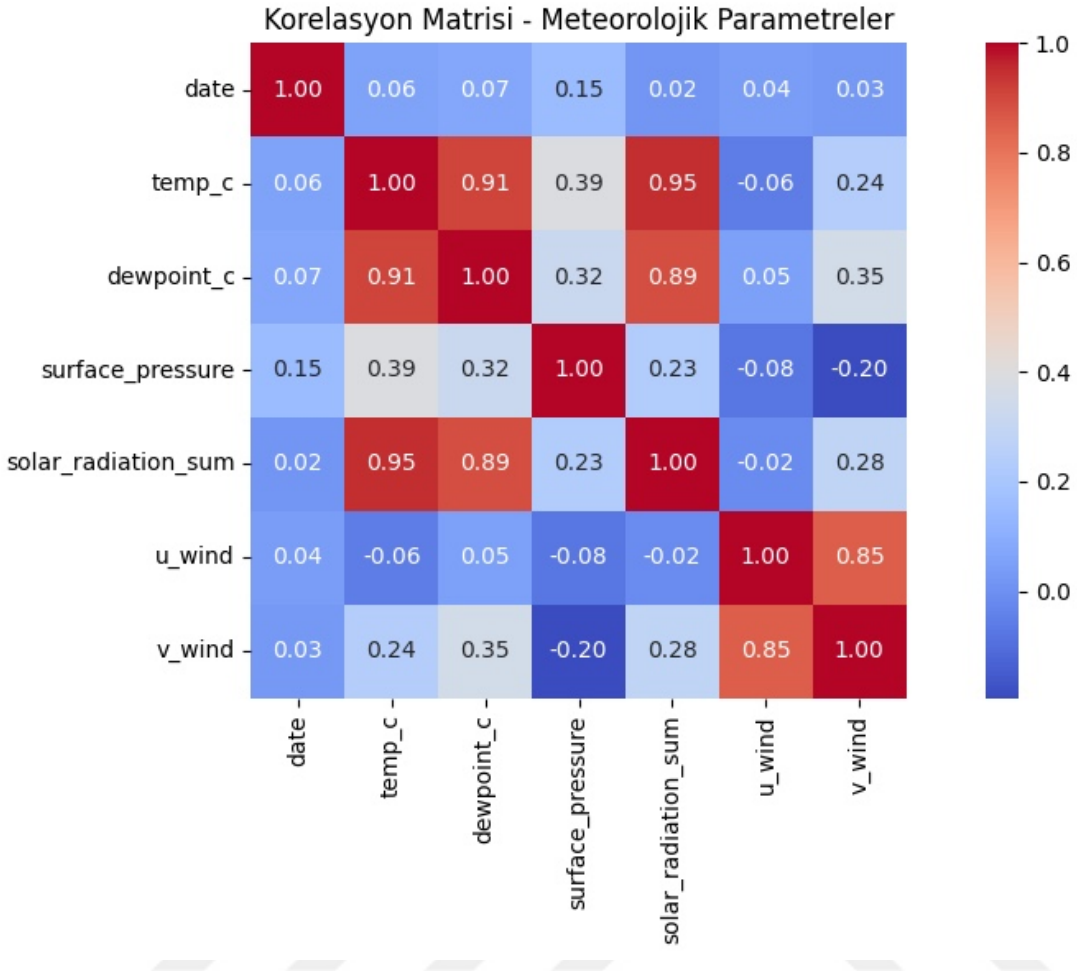
$$\hat{y} = \alpha + \beta \cdot HDD + \varepsilon \tag{3.31}$$

Burada $\hat{y}$ tahmini doğalgaz tüketimini, α sabit tüketim bileşenini (baz yük), β ise her bir HDD birimi için doğalgaz tüketimindeki artışı temsil etmektedir. ε ise modelin açıklayamadığı varyansı, yani hata terimini ifade eder.

Bu modelin temel varsayımı, ortam sıcaklığı konfor seviyesinin (bu çalışmada 18°C) altına düştükçe doğalgaz tüketiminin doğrusal olarak artacağı yönündedir. Ancak, doğrusal ilişkinin katsayıları olan α ve β, yalnızca iklimsel değişkenlere değil; aynı zamanda yapının fiziksel özelliklerine, yalıtım kalitesine, ısıtma sisteminin verimliliğine, kullanım alışkanlıklarına ve veri örneklemesinin sınırlılığına bağlı olarak değişkenlik göstermektedir. Bu nedenle model parametreleri, kullanılan yapıya özgü gerçek sıcaklık ve doğalgaz tüketimi verileri üzerinden kalibre edilmiştir.

Yapıya özgü bu parametreleme sayesinde, yapıdan yapıya değişen ısı kayıpları, konstrüktif özellikler, enerji verimliliği düzeyi gibi etkenler modelin içinde dolaylı olarak temsil edilmiştir. Literatürde ideal koşullarda kullanılan sabit parametrelerin, küçük örneklem boyutu veya sınırlı veri setleriyle uygulandığında tahmin performansını düşürdüğü bilinmektedir. Bu bağlamda çalışma, sahaya özgü sıcaklık-tüketim ilişkilerini HDD üzerinden yeniden yapılandırarak daha gerçekçi bir tüketim modeli sunmayı hedeflemiştir.

Ek olarak, bu modelin yalnızca gerçek sıcaklık verileriyle değil, aynı zamanda çalışmada geliştirilen makine öğrenmesi temelli sıcaklık tahminleriyle de uygulanması sağlanmıştır. Böylece, aynı yapıya ait α ve β değerleri korunarak; farklı sıcaklık kaynaklarıyla (gerçek vs. tahmin) oluşturulan HDD girdilerinin modele etkisi karşılaştırmalı olarak analiz edilmiştir. Bu yaklaşım, tahmin edilen sıcaklık değerlerinin doğruluğunun yalnızca sıcaklık özelinde değil, nihai enerji tüketimi tahmini üzerindeki etkisi açısından da değerlendirilmesine olanak tanımıştır.



3.2. Meteorolojik parametreler arası korelasyon matrisi

Bu çalışmada doğalgaz tüketimi ile ilişkili olabilecek meteorolojik değişkenler arasındaki doğrusal ilişkiler, korelasyon matrisi üzerinden analiz edilmiştir. Şekil 3.2'de sunulan matris, parametre çiftleri arasındaki Pearson korelasyon katsayılarını renk skalası eşliğinde göstermektedir. 1.00 değeri tam pozitif korelasyonu, –1.00 değeri ise tam negatif korelasyonu ifade etmektedir.

Sıcaklık (temp_c) ile çiğ noktası sıcaklığı (dewpoint_c) arasında oldukça yüksek ve pozitif bir korelasyon gözlemlenmiştir ($\mathbf{R^2}$ = 0.91). Bu durum, sıcaklığın arttığı günlerde nem oranının da arttığını, yani çiğ noktasının yükseldiğini göstermektedir. Benzer şekilde, **sıcaklık ile güneş radyasyonu toplamı (solar_radiation_sum)** arasındaki korelasyon da oldukça yüksektir ($\mathbf{R^2}$ = 0.95). Bu bulgu, güneşli günlerde sıcaklık artışının belirgin olduğunu ve bu iki değişkenin birbiriyle kuvvetli bir şekilde ilişkili olduğunu ortaya koymaktadır.

Ayrıca, **çiy noktası ile güneş radyasyonu ($\mathbf{R^2}$ = 0.89)** arasındaki pozitif ilişki, nemli günlerin genellikle güneşli günlerle örtüştüğünü düşündürmektedir. Rüzgâr bileşenleri

olan **yatay rüzgâr hızı**(u_wind**)** ile **dikey rüzgâr hızı(**v_wind**)** arasında yüksek pozitif korelasyon (**R^2** = 0.85) bulunmuş olup, bu da farklı yönlerdeki rüzgâr bileşenlerinin sıklıkla birlikte değiştiğini göstermektedir.

Buna karşın, **sıcaklık ile rüzgâr bileşenleri** (temp_c – u_wind: **R^2** = –0.06; temp_c – v_wind **R^2** = 0.24) ve **yüzey basıncı ile rüzgâr bileşenleri** (surface_pressure – v_wind: **R^2** = –0.20) arasındaki korelasyonların düşük olması, bu değişkenlerin birbirlerinden bağımsız hareket etme eğiliminde olduğunu göstermektedir.

Elde edilen bu bulgular, doğalgaz tüketiminin modellenmesinde kullanılacak değişkenlerin seçiminde rehberlik edici olacaktır. Özellikle yüksek korelasyon gösteren değişkenlerin birlikte kullanılmaları, çoklu doğrusal regresyon gibi modellerde **çoklu bağlantı (multicollinearity)** problemlerine yol açabileceği için dikkatle değerlendirilmelidir.

3.2.3.8. Değerlendirme Metrikleri

Ortalama Kare Hataları (MSE - Mean Squared Error)

Gerçek (gözlenen) değerler ile model tarafından tahmin edilen değerler arasındaki farkların karelerinin ortalamasıdır. MSE (Ortalama Kare Hatası), hata büyüklüğünü dikkate alır ve büyük hataları daha fazla cezalandırır. Bu yönüyle modelin büyük hatalardan ne derece kaçındığını vurgular. Küçük hataları da göz ardı etmez. Hataların karesi alındığı için aykırı (uç) değerlere karşı daha duyarlıdır ve bu nedenle hassas bir ölçüt olarak kabul edilir. MSE, denklem 3.32'de gösterildiği şekilde hesaplanır. Burada; N, toplam veri sayısını, y_j gözlenen (gerçek) değeri, $\hat{y}_j$ ise model tarafından tahmin edilen değeri ifade eder.

$$\mathrm{MSE} = \frac{1}{N}\sum_{j=i}^{N}\left(y_j - \hat{y}_j\right)^2 \tag{3.32}$$

- **Ortalama Mutlak Hata (MAE - Mean Absolute Error)**

Ortalama Mutlak Hata (MAE), tahmin edilen değerler ile gerçek gözlem değerleri arasındaki farkların mutlak değerlerinin ortalamasını alarak hesaplanan bir regresyon performans ölçütüdür. MAE, her bir tahminin ortalama ne kadar hata içerdiğini açık ve yorumlanabilir bir şekilde sunar. Hata birimleri, tahmin edilen değişken ile aynı olduğundan, modelin tahmin başarısı doğrudan yorumlanabilir.

MAE, büyük hataları daha az şiddetle cezalandırdığı için RMSE gibi metriklere göre aşırı uç değerlere (outliers) karşı daha az hassastır. Bu özelliği sayesinde, verilerinde yüksek varyans olmayan durumlarda daha güvenilir bir değerlendirme aracı olarak öne çıkar.

MAE, aşağıdaki Denklem 3.33 ile hesaplanır:

$$MAE: \frac{1}{N}\sum_{j=1}^{N}\left|y_j - \hat{y}_j\right| \tag{3.33}$$

Burada; y_j: Gerçek (gözlenen) değeri, $\hat{y}_j$: Model tarafından tahmin edilen değeri, N: Toplam gözlem sayısını temsil eder. MAE değeri sıfıra yaklaştıkça modelin doğruluğu artmakta, yüksek MAE değerleri ise modelin tahmin hatalarının daha büyük olduğunu göstermektedir.

- **Belirleme Katsayısı (R^2 or Coefficient of Determination)**

Belirleme Katsayısı (R^2), bağımlı değişkendeki varyansın ne kadarının modeldeki bağımsız değişkenler tarafından açıklandığını ölçer. Başka bir deyişle, çıktıda gözlemlenen toplam değişimin ne kadarının modelin tahminleri tarafından yakalandığını yansıtır. R^2 değeri 1'e yaklaştıkça, modelin açıklayıcılık gücü artar. R^2'nin 0.0 olması modelin hiçbir açıklama gücü olmadığını, 1.0 olması ise tahminlerin tamamen doğru yapıldığını, yani mükemmel bir öngörü sağlandığını gösterir.

Belirleme Katsayısı (R^2), aşağıdaki Denklem 3.34 ile hesaplanır:

$$R^2 = 1 - \frac{\sum_{j=i}^{N}(y_j - \hat{y}_j)^2}{\sum_{j=i}^{N}(Y_j - \bar{y}_j)^2} \tag{3.34}$$

- **Ortalama Mutlak Yüzde Hatası (MAPE - Mean Absolute Percentage Error)**

Ortalama Mutlak Yüzde Hatası (MAPE), özellikle iş tahminleri ve zaman serisi analizleri gibi alanlarda regresyona dayalı tahminlerin doğruluğunu değerlendirmek için yaygın olarak kullanılan bir ölçüttür. Gerçek ve tahmin edilen değerler arasındaki mutlak yüzde farklarının ortalamasını hesaplayarak modelin performansına yorumlanabilir ve birime bağlı olmayan bir değerlendirme sağlar. Hataları yüzde cinsinden ifade ettiği için, farklı ölçeklere veya birimlere sahip veri setleri arasında anlamlı karşılaştırmalar yapılmasına olanak tanır.

MAPE, aşağıdaki Denklem 3.35 ile hesaplanır:

$$MAPE = \frac{1}{N}\sum_{j=1}^{N}\left[\frac{y_j - \hat{y}_j}{y_j}\right] * 100 \quad (3.35)$$

- **Simetrik Ortalama Mutlak Yüzde Hata (SMAPE – Symmetric Mean Absolute Percentage Error)**

Simetrik Ortalama Mutlak Yüzde Hata (SMAPE), regresyon tabanlı tahmin modellerinin doğruluğunu ölçmek için kullanılan bir metriktir ve özellikle geleneksel MAPE ölçütünün bazı dezavantajlarını gidermek amacıyla geliştirilmiştir. MAPE, sıfıra yakın gerçek değerlerde çok yüksek hata yüzdeleri üretebilirken; SMAPE bu durumu simetrik hale getirerek hem gerçek hem de tahmin edilen değerleri paydada dikkate alır. Böylece daha dengeli ve istikrarlı bir hata ölçümü sağlar.

SMAPE, hataları yüzde olarak ifade ettiği için farklı ölçek veya birimlere sahip veri kümeleri arasında karşılaştırma yapılmasına da olanak tanır. Ayrıca, tahmin edilen değerin gerçek değerden ne kadar saptığını pozitif bir ölçüyle ve 0–100 aralığında gösterdiği için yorumlanması kolaydır.

SMAPE, aşağıdaki Denklem 3.36 ile hesaplanır:

$$SMAPE: \frac{100}{N}\sum_{j=1}^{N}\frac{\left|y_j - \hat{y}_j\right|}{\frac{\left(|y_j| + |\hat{y}_j|\right)}{2}} \quad (3.36)$$

Burada: y_j; gerçek (gözlenen) değerini, $\hat{y}_j$; model tarafından tahmin edilen değerini, N; toplam veri sayısını ifade eder. SMAPE hem düşük hem de yüksek değerli tahminlerde tutarlı bir değerlendirme sunar ve %0 ile %100 arasında değişen bir performans metriSMAPE değeri sıfıra yaklaştıkça modelin başarısı artar.

3.2. Verilere ait istatistiksel bilgiler tablosu

	Sıcaklık	Çiğ noktası	Yüzey basıncı	Güneş radyasyonu	Yatay Rüzgâr	Dikey Rüzgâr
Veri Sayısı	300	300	300	3,00E+02	300	300
Ortalama	6,38	-2,1	78063,29	4,03E+10	0,09	0,09
Standart	9,74	5,58	219,36	2,11E+08	0,24	0,33
Minimum	-12,22	-17,37	77436,78	9,39E+07	-0,7	-0,8
25%	-2,63	-6,51	77928,67	1,96E+08	-0,06	-0,11
50%	6,58	-0,97	78066,3	3,81E+08	0,09	0,11
75%	15,83	2,71	78210,21	6,31E+08	0,26	0,33
Maksimum	22,14	7,46	78695,01	7,40E+08	0,65	0,86
Çarpıklık/Eğrilik	0,05	-0,52	0,000497	1,38E-01	-0,02	-0,19
Basıklık	-1,4	-0,86	-0,11	-1,49E+00	-0,11	-0,33

Bu çalışmada kullanılan veri setine ait temel tanımlayıcı istatistikler Çizelge 3.1'de sunulmuştur. Sıcaklık tahmini amaçlı olarak değerlendirilen bu veri seti; sıcaklık, çiğ noktası, yüzey basıncı, güneş radyasyonu, yatay rüzgâr ve dikey rüzgâr olmak üzere toplam altı meteorolojik değişken içermektedir. Her bir değişken için minimum, maksimum, ortalama, standart sapma, çeyrek değerler (25%, 50%, 75%) ile çarpıklık (skewness) ve basıklık (kurtosis) değerleri hesaplanarak, dağılım özellikleri detaylı biçimde analiz edilmiştir.

Sıcaklık değişkeni, ortalama 6,38 °C değerinde olup, -12,22 ile 22,14 °C arasında değişmektedir. Yüksek standart sapma değeri (9,74), sıcaklık değerlerinin oldukça değişken bir yapıya sahip olduğunu göstermektedir. Çarpıklık değeri 0,05 ile dağılımın yaklaşık olarak simetrik olduğunu, basıklık değeri -1,40 ise dağılımın normalden daha basık, yani uç değer içermeyen bir profile sahip olduğunu ortaya koymaktadır.

Çiğ noktası ortalaması -2,10 °C olup genellikle negatif değerler göstermiştir. -17,37 ile 7,46 °C arasında değişim göstermekte, çarpıklık (-0,52) ve basıklık (-0,86) değerleri ise dağılımın sola çarpık ve basık olduğunu işaret etmektedir. Bu durum, çiğ noktası değerlerinin büyük ölçüde düşük seviyelerde yoğunlaştığını göstermektedir.

Yüzey basıncı, ortalama 78.063 Pa civarında seyretmiş ve diğer değişkenlere kıyasla oldukça düşük standart sapma (219,36) değeri ile daha sabit bir dağılım sergilemiştir. Çarpıklık (yaklaşık 0,0005) ve basıklık (-0,11) değerlerinin sıfıra yakın olması, yüzey basıncı değişkeninin normal dağılıma oldukça yakın bir yapı gösterdiğini ortaya koymaktadır.

Güneş radyasyonu, veri setindeki en büyük ölçekli değişkenlerden biri olarak dikkat çekmektedir. Ortalama 4,03×10 ^{8} W/m^{2} olan bu değiş
maksimum 7,40×10 ^{8} W/m^{2} değerleri arasın
(2,11×10 ^{8}) güneş radyasyonu miktarındaki önemli de
Çarpıklık değeri (0,138) sağa hafif çarpık bir dağılımı, basıklık değeri (-1,49) ise uç değerlerin daha seyrek olduğunu ve dağılımın basık olduğunu göstermektedir. Yatay ve dikey rüzgâr bileşenleri, sırasıyla ortalama 0,09 m/s ile oldukça düşük değerlere sahiptir. Standart sapma değerleri de sırasıyla 0,24 ve 0,33 olarak düşük düzeydedir. Bu durum, rüzgâr bileşenlerinin veri setinde sınırlı bir değişkenlik gösterdiğini ortaya koymaktadır. Yatay rüzgâr neredeyse simetrik bir dağılım sergilerken (çarpıklık = -0,02), dikey rüzgârda hafif bir negatif çarpıklık (-0,19) gözlenmiştir. Genel olarak

değerlendirildiğinde, Çizelge 3.1'deki tanımlayıcı istatistikler, model kurulum sürecinde uygulanacak veri ön işleme adımlarına ilişkin önemli bilgiler sunmaktadır. Özellikle sıcaklık ve güneş radyasyonu gibi değişkenlerin geniş dağılım aralıkları, bu özniteliklerin modele katkı potansiyelinin yüksek olduğunu göstermektedir. Ayrıca, bazı değişkenlerin basık veya çarpık dağılımları, dönüşüm (örneğin log, Box-Cox) veya yeniden ölçeklendirme gibi ek işlemlerin gerekebileceğini düşündürmektedir. Bu nedenle, veri seti üzerinde gerçekleştirilecek modelleme adımlarında, bu istatistiksel farklılıklar dikkate alınarak uygun normalizasyon ve seçici öznitelik mühendisliği stratejileri uygulanmıştır.

4. BULGULAR

Bu çalışmada, Hakkâri Yüksekova Selahaddin Eyyubi Havalimanına ait enerji tüketimi verileri ele alınmıştır. Farklı makine öğrenmesi algoritmalarının doğalgaz tüketimi tahminindeki başarı düzeyleri, eğitim ve test veri kümeleri üzerinde çeşitli hata metrikleri (MSE, R^2, MAE, MAPE ve SMAPE) kullanılarak karşılaştırmalı biçimde değerlendirilmiştir. Elde edilen bulgular, modellerin hem öğrenme kapasitesi hem de genelleme yetenekleri açısından detaylı biçimde analiz edilmesini mümkün kılmıştır.

4.1. Değerlendirme metrikleri tablosu

	Train					**Test**				
	MSE	R^2	MAE	MAPE	SMAPE	MSE	R^2	MAE	MAPE	SMAPE
MLR	4.2148	0.9562	1.5650	0.5039	30.18%	5.8085	0.9345	1.9022	0.4843	37.30%
SVR	3.8811	0.9596	1.58	0.44%	29.83%	5.6553	0.9362	1.92	0.45%	37.30%
RF	0.5266	0.9945	1.11	0.12	11.11%	2.3519	0.9734	1.11	0.193	18.82%
LightGBM	0.7661	0.9920	0.60	0.17	14.15%	4.0123	0.9548	1.58	0.37	29.09%
XGBoost	0.0411	0.9996	0.15	0.05	4.86%	0.0003	1.0000	0.01	0.00	0.30%
LSTM	3.7662	0.9608	1.43	0.36%	23.73%	4.6239	0.9479	1.66	0.36%	31.16%

İlk olarak, XGBoost algoritması genel performans açısından açık ara en başarılı model olmuştur. Eğitim setinde 0.0411 MSE, 0.9996 R^2 ve yalnızca %4.86 SMAPE değeri ile neredeyse mükemmel bir uyum sergileyen XGBoost, test verisinde de bu başarıyı korumuş; MSE değeri 0.0003 ve R^2 skoru 1.0000 ile sıfıra yakın hata ve tam doğruluk sunmuştur. Diğer tüm modellerle kıyaslandığında hem eğitim hem de test aşamasında en düşük mutlak hata (MAE = 0.01) ve yüzde hata oranı (MAPE = 0.00) bu modele aittir. Bu sonuçlar, XGBoost'un yüksek veri adaptasyonu, öğrenme kabiliyeti ve aşırı öğrenmeden kaçınma başarısıyla öne çıktığını göstermektedir.

LightGBM, XGBoost'un ardından en başarılı ikinci model olarak öne çıkmaktadır. Eğitim setinde R^2 = 0.9920 ve test setinde R^2 = 0.9548 değerleriyle yüksek açıklayıcılığa sahip olan bu modelin test verisindeki MAE değeri 1.58 iken, SMAPE

değeri %29.09'dur. Her ne kadar LightGBM, XGBoost kadar düşük hata oranlarına ulaşamasa da, özellikle modelin hızı ve büyük veri setleriyle çalışma verimliliği düşünüldüğünde ciddi avantajlar sunmaktadır. Bununla birlikte, SMAPE değerinin test setinde %30'a yaklaşması, uç değerlerin etkisine karşı daha hassas olduğunu göstermektedir. Bu yönüyle LightGBM, doğruluk açısından XGBoost'un gerisinde kalmakla birlikte, bazı durumlarda hesaplama maliyetleri açısından tercih edilebilir bir seçenek olmaktadır.

Random Forest (RF) modeli ise eğitim setinde $R^2 = 0.9945$ ile yüksek bir öğrenme performansı sergilemiş ve MAE = 1.11 ile istikrarlı tahminler üretmiştir. Test setindeki performansı da güçlüdür ($R^2 = 0.9734$; SMAPE = %18.82). Özellikle LightGBM ile kıyaslandığında, RF'nin daha düşük SMAPE ve benzer MAE değerleri sunması, bu algoritmanın uç değer etkisine karşı daha dirençli olduğunu göstermektedir. Bununla birlikte, XGBoost'un ulaştığı yüksek doğruluk ve düşük hata seviyelerine erişememiştir. Bu bağlamda RF, doğruluk ve genelleme arasında dengeli bir yapı sunarak orta düzeyde veri karmaşıklığı için ideal bir model olma niteliği taşır.

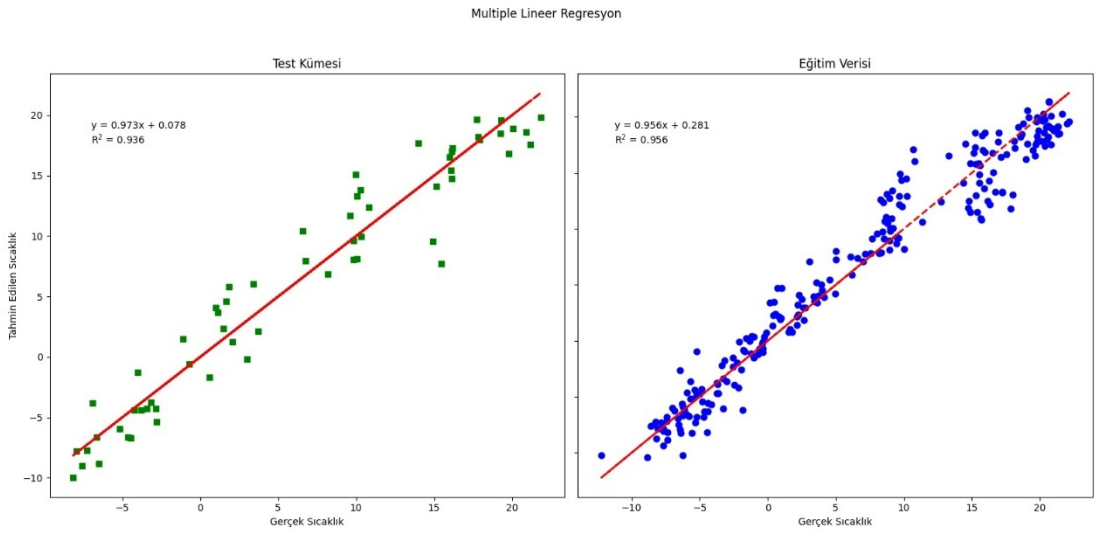
LSTM modeli, özellikle zaman serisi verilerinde yapay sinir ağlarının potansiyelini göstermesi açısından önemli olmakla birlikte, bu çalışmada beklentilerin altında bir performans sergilemiştir. Eğitim aşamasında 3.7662 MSE ve %23.73 SMAPE değerleri ile yüksek hata oranları kaydedilirken, test setinde bu değerler sırasıyla 4.6239 ve %31.16 olarak gerçekleşmiştir. Diğer modellerle kıyaslandığında LSTM hem mutlak hem de yüzde hata oranları açısından daha zayıf bir tahmin yeteneği sunmuş ve yüksek varyans göstermiştir. Bu durum, LSTM'nin daha fazla hiperparametre ayarlaması ve daha geniş veri seti gereksinimi olduğunu ortaya koymaktadır. Ayrıca, LSTM'nin XGBoost'a göre yaklaşık 5 kat daha fazla MSE üretmesi, klasik gradyan artırma yöntemlerinin kısa vadeli tahminlerde daha başarılı olduğunu düşündürmektedir.

Destek Vektör Regresyonu (SVR) ise doğrusal olmayan ilişkileri modelleme potansiyeline sahip olmasına rağmen hem eğitim hem de test setinde yüksek hata değerleri üretmiştir. Eğitim setinde $R^2 = 0.9596$ ve test setinde $R^2 = 0.9362$ gibi görece düşük açıklama değerleriyle birlikte %29.83 ve %37.30 SMAPE değerleriyle istikrarsız bir performans sergilemiştir. RF ve LightGBM gibi ağaç tabanlı modellerle kıyaslandığında, SVR'nin değişkenliğe daha duyarlı olduğu ve genelleme yeteneğinin sınırlı kaldığı gözlemlenmiştir.

Son olarak, Çoklu Doğrusal Regresyon (MLR) modeli, en basit yapılandırma olmasına rağmen en düşük performansı sunmuştur. Eğitim ve test setlerinde sırasıyla %30.18 ve %37.30 SMAPE değerleri ile en yüksek yüzde hata oranlarına ulaşan MLR, diğer modellerin oldukça gerisinde kalmıştır. Bu durum, doğrusal modellerin doğalgaz tüketimi gibi mevsimsel ve çok değişkenli problemlerde yetersiz kaldığını açık biçimde ortaya koymaktadır. XGBoost ve LightGBM gibi modeller, doğrusal olmayan ilişkileri daha başarılı şekilde modellediğinden daha düşük hata seviyelerine ulaşabilmektedir.

Genel kıyaslama yapıldığında, XGBoost; doğruluk, genelleme ve hata minimizasyonu açısından mutlak üstünlük sağlamaktadır. LightGBM ve Random Forest, istikrarlı performanslarıyla alternatif çözümler sunarken; LSTM gibi derin öğrenme yaklaşımları, daha ileri düzey yapılandırmalara ihtiyaç duymaktadır. SVR ve MLR gibi modeller ise bu bağlamda yetersiz kalmış, karmaşık enerji tüketimi desenlerini modellemede başarısız olmuştur.

Bu çalışmada kullanılan makine öğrenmesi algoritmalarından elde edilen sonuçlar incelendiğinde grafiklerde gösterilen sıcaklık değerleri yükseldikçe regresyon doğrusundan uzaklaşan sonuçlara ulaşılmıştır. Sıcaklık değerleri arttıkça dalgalanmaların yaşandığı gözlenmiştir.



4.1. Multi lineer regresyon modeli ile gerçek ve tahmin edilen sıcaklık değerlerinin karşılaştırılması

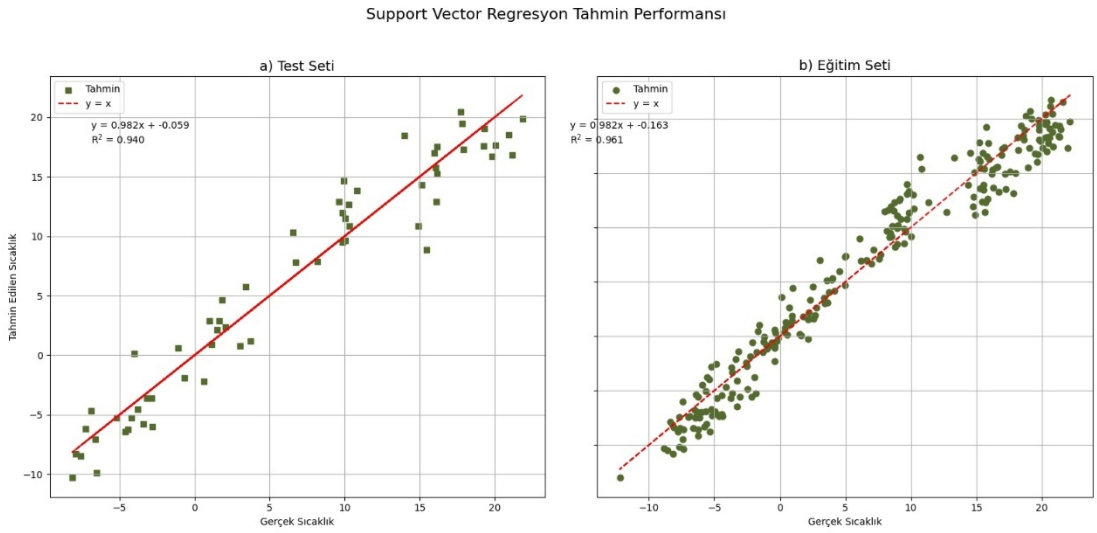
Şekil 4.1'de çoklu doğrusal regresyon (Multiple Linear Regression, MLR) modeli kullanılarak gerçekleştirilen sıcaklık tahminine ait test ve eğitim kümeleri karşılaştırmalı olarak gösterilmiştir. Bu grafiklerde yatay eksende gözlemlenen

(gerçek) sıcaklık değerleri, düşey eksende ise model tarafından tahmin edilen sıcaklık değerleri yer almaktadır. Her iki grafik de modelin tahmin başarımını görselleştirmek amacıyla oluşturulmuştur. Regresyon doğruları, tahminlerin genel eğilimini temsil etmekte olup, kırmızıçizgiyle gösterilmiştir.

Test kümesi için elde edilen regresyon doğrusu denklemi y = 0.973x + 0.078, determinasyon katsayısı (R^2) ise 0.936 olarak hesaplanmıştır. Bu durum, modelin daha önce görmediği veriler üzerinde yüksek doğrulukta tahmin yapabildiğini ve genelleme yeteneğinin güçlü olduğunu göstermektedir. Ayrıca test verisine ait ortalama kare hata (MSE) 5.8085, ortalama mutlak hata (MAE) 1.9022, ortalama mutlak yüzeysel hata (MAPE) 0.4843, ve simetrik ortalama mutlak yüzeysel hata (SMAPE) %37.30 olarak belirlenmiştir. Bu değerler, modelin test verisinde düşük hata ile çalıştığını göstermektedir. Eğitim verisine ait regresyon doğrusu ise y = 0.956x + 0.281 şeklindedir ve R^2 değeri 0.9562 olarak elde edilmiştir. Modelin eğitim verisindeki ortalama kare hatası (MSE) 4.2148, MAE değeri 1.5650, MAPE değeri 0.5039 ve SMAPE değeri %30.18 olarak hesaplanmıştır. Bu performans ölçütleri, modelin öğrenme sürecinde veriye güçlü şekilde uyum sağladığını göstermektedir.

Eğitim ve test kümelerindeki hata metriklerinin birbirine yakın olması, modelin aşırı öğrenme (overfitting) sorunu yaşamadığını ve genellenebilirliğinin başarılı olduğunu göstermektedir. R^2 değerlerinin yüksek olması, bağımsız değişkenlerin (örneğin çiy noktası, rüzgâr hızı, güneş radyasyonu gibi meteorolojik faktörler) hedef değişken olan sıcaklığı açıklamada güçlü katkı sağladığını göstermektedir.

Sonuç olarak, çoklu doğrusal regresyon modeli hem eğitim hem de test verisi üzerinde yüksek doğruluk ve düşük hata ile başarılı bir tahmin performansı sergilemiştir. Bu bulgular, Şekil 4.1'de görsel olarak da desteklenmektedir.

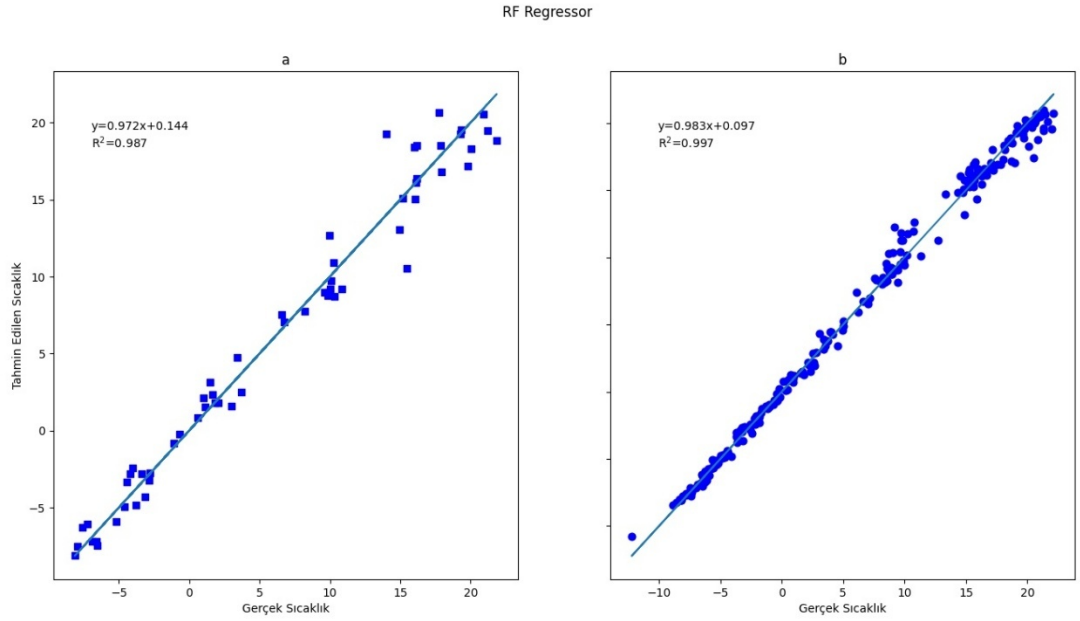


4.2. SVR Modeli ile gerçek ve tahmin edilen sıcaklık değerlerinin karşılaştırılması

Şekil 4.2'de SVR algoritması ile yapılan sıcaklık tahminlerinin performansı sunulmuştur. Şeklin sol tarafında (a) test veri setine, sağ tarafında (b) ise eğitim veri setine ait tahmin sonuçları yer almaktadır. Grafiklerde yatay eksende gerçek sıcaklık değerleri, düşey eksende ise tahmin edilen sıcaklık değerleri gösterilmekte olup, ideal durumu temsil eden $y = x$ doğrusu kırmızı kesikli çizgiyle belirtilmiştir. Eğitim verisine ait regresyon doğrusu y=0.982x+(−0.163) şeklindedir ve bu veri setindeki determinasyon katsayısı (R^2) 0.961 olarak elde edilmiştir. Eğitim setine ilişkin ortalama kare hata (MSE) 3.8382, ortalama mutlak hata (MAE) 1.56, ortalama mutlak yüzde hata (MAPE) %0.44 ve simetrik ortalama mutlak yüzde hata (SMAPE) %29.60'tır. Bu değerler, modelin eğitim verisine oldukça iyi uyum sağladığını göstermektedir. Şekil 4.2 (b)'de tahmin edilen değerlerin $y = x$ doğrusuna yakın konumlandığı ve dağılımın yoğunlaştığı gözlenmektedir.

Test verisi üzerinden elde edilen regresyon doğrusu y=0.982x+(−0.059) olup, determinasyon katsayısı (R^2) 0.940 olarak hesaplanmıştır. Bu durum, modelin daha önce görmediği veriler üzerinde de güçlü bir genelleme yeteneğine sahip olduğunu göstermektedir. Test verisi için hesaplanan MSE 5.4889, MAE 1.90, MAPE %0.44 ve SMAPE %36.33 olarak belirlenmiştir. Test setindeki SMAPE değerinin eğitim setine göre daha yüksek olması, özellikle tahmin edilen değerlerin göreli uzaklıklarının arttığını ve bu durumun bazı uç noktalarda etkili olduğunu göstermektedir. Şekil 4.2(a)'da bazı sapmalar görülmekle birlikte genel dağılım, regresyon doğrusuna yakın bir şekilde yoğunlaşmıştır.

Genel olarak SVR modeli, doğrusal olmayan ilişkileri modelleme kapasitesi ve kenar vektörlerine (support vectors) dayalı karar yapısıyla sıcaklık tahmininde başarılı bir performans ortaya koymuştur. Hem eğitim hem de test veri setlerinde yüksek R^2 değerleri ve kabul edilebilir hata metrikleri ile istikrarlı bir sonuç elde edilmiştir. Ancak, diğer bazı algoritmalara kıyasla SMAPE değerinin daha yüksek olması, modelin bazı değer aralıklarında daha düşük isabetle tahminde bulunduğuna işaret etmektedir.



4.3. RF Modeli ile gerçek ve tahmin edilen sıcaklık değerlerinin karşılaştırılması

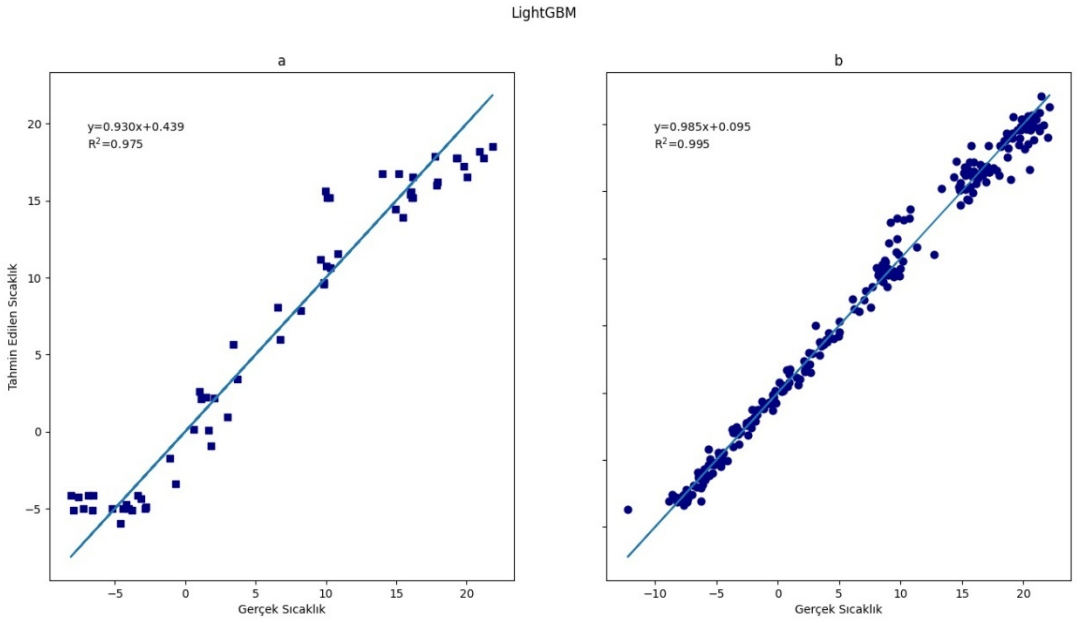
Şekil 4.3'te RF regresyon modeli ile tahmin edilen sıcaklık değerleri ile gözlemlenen gerçek sıcaklık değerlerinin karşılaştırılması sunulmaktadır. Şeklin sol tarafında (a) test verisine, sağ tarafında (b) ise eğitim verisine ait dağılım grafikleri yer almaktadır. Her iki grafikte de yatay eksende gerçek sıcaklık değerleri, düşey eksende ise model tarafından tahmin edilen sıcaklık değerleri yer almaktadır. Grafiklerde gösterilen regresyon doğruları, modelin genel tahmin eğilimini yansıtmaktadır.

Eğitim verisi üzerinde elde edilen regresyon doğrusu $y = 0.983x + 0.097$ şeklindedir ve determinasyon katsayısı (R^2) 0.997 olarak hesaplanmıştır. Bu sonuç, modelin eğitim verisine son derece yüksek uyum sağladığını göstermektedir. Eğitim verisi için ortalama kare hata (MSE) 0.5266, ortalama mutlak hata (MAE) 1.1104, ortalama mutlak yüzde hata (MAPE) 0.1272 ve simetrik ortalama mutlak yüzde hata (SMAPE) %11.11 olarak hesaplanmıştır. Bu değerler, modelin tahmin hatalarının düşük

seviyede olduğunu ve yüksek isabetle çalıştığını göstermektedir. Şekil 4.3(b)'deki dağılım incelendiğinde, tahmin edilen değerlerin büyük ölçüde regresyon doğrusuna yakın olduğu ve sapmaların minimum düzeyde kaldığı görülmektedir.

Test verisi üzerinde elde edilen regresyon doğrusu ise y = 0.972x + 0.144 şeklindedir ve R^2 değeri 0.987 olarak bulunmuştur. Bu sonuç, modelin daha önce görmediği veri üzerinde de oldukça başarılı bir tahmin performansı gösterdiğini ortaya koymaktadır. Test verisi için hesaplanan ortalama kare hata (MSE) 2.3520, MAE değeri 1.1104, MAPE değeri 0.1930 ve SMAPE değeri %18.82'dir. Test ve eğitim verisi üzerindeki benzer hata düzeyleri, modelin aşırı öğrenme (overfitting) eğiliminde olmadığını ve genellenebilirliğinin güçlü olduğunu göstermektedir.

Sonuç olarak, Random Forest regresyon modeli hem eğitim hem de test veri kümelerinde yüksek doğrulukla tahminler üretmiş, düşük hata oranları ve yüksek R^2 katsayıları ile dikkat çekmiştir. Modelin doğrusal olmayan ilişkileri başarıyla öğrenebilmesi, sıcaklık tahmini gibi kompleks meteorolojik problemlerde etkili bir araç olduğunu göstermektedir. Bu bulgular Şekil 4.3 ile de görsel olarak desteklenmektedir.

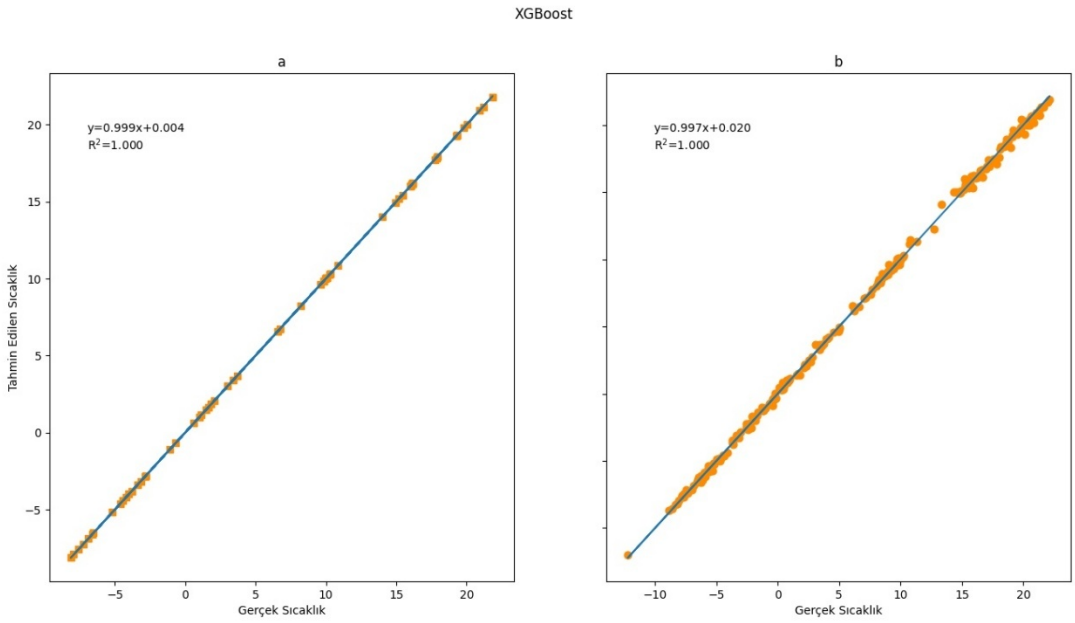


4.4. LightGBM modeli ile gerçek ve tahmin edilen sıcaklık değerlerinin karşılaştırılması

Şekil 4.4'te LightGBM algoritması kullanılarak gerçekleştirilen sıcaklık tahminine ait model performans grafikleri yer almaktadır. Şeklin sol tarafında (a) test verisi, sağ tarafında (b) ise eğitim verisine ilişkin tahmin sonuçları sunulmuştur. Grafiklerde

yatay eksende gözlenen (gerçek) sıcaklık değerleri, düşey eksende ise tahmin edilen sıcaklık değerleri yer almakta olup, mavi çizgiler regresyon doğrusunu temsil etmektedir.

Eğitim verisi üzerinde elde edilen regresyon denklemi y=0.985x+0.095 olup, determinasyon katsayısı (R^2) 0.995'tir. Modelin eğitim setindeki hata metrikleri şu şekildedir: ortalama kare hata (MSE) 0.9854, ortalama mutlak hata (MAE) 0.69, ortalama mutlak yüzde hata (MAPE) %0.17, simetrik ortalama mutlak yüzde hata (SMAPE) ise %14.33 olarak hesaplanmıştır. Bu değerler modelin eğitim verisi üzerinde yüksek doğrulukla çalıştığını göstermektedir. Şekil 4.4(b)'de yer alan dağılım, tahminlerin büyük ölçüde $y = x$ doğrusu üzerinde yer aldığını ve sapmaların düşük olduğunu ortaya koymaktadır. Test verisi için elde edilen regresyon doğrusu ise y=0.930x+0.439 şeklindedir. Test verisine ait R^2 değeri 0.975 olarak bulunmuştur. Bu, modelin daha önce görmediği veriler üzerinde de yüksek açıklayıcılığa sahip olduğunu göstermektedir. Test setine ait ortalama kare hata (MSE) 4.4700, ortalama mutlak hata (MAE) 1.68, MAPE %0.37 ve SMAPE %33.38 olarak hesaplanmıştır. Test verisindeki hata düzeyleri eğitim verisine kıyasla biraz daha yüksek olsa da, tahminlerin genel olarak başarılı olduğu söylenebilir. Şekil 4.4(a)'da tahmin edilen değerlerin $y = x$ doğrusuna yakın konumlandığı, ancak özellikle yüksek sıcaklık değerlerinde hafif sapmaların oluştuğu gözlemlenmektedir. Genel olarak LightGBM modeli, düşük hesaplama süresi, yüksek doğruluk ve geniş veri setlerine olan adaptasyon kabiliyeti sayesinde sıcaklık tahmininde etkili bir performans sergilemiştir. Model hem eğitim hem de test veri setlerinde kabul edilebilir hata seviyeleri ve yüksek R^2 katsayıları ile başarılı sonuçlar üretmiştir. Ancak SMAPE değerlerindeki artış, özellikle test setinde göreli hatanın arttığını ve bazı noktalarda tahmin gücünün düşebileceğini göstermektedir. Bu durum, modelin performansını iyileştirmek amacıyla hiperparametre optimizasyonu veya melez modellerin değerlendirilmesi gibi ileri yöntemlerin kullanılabileceğine işaret etmektedir.

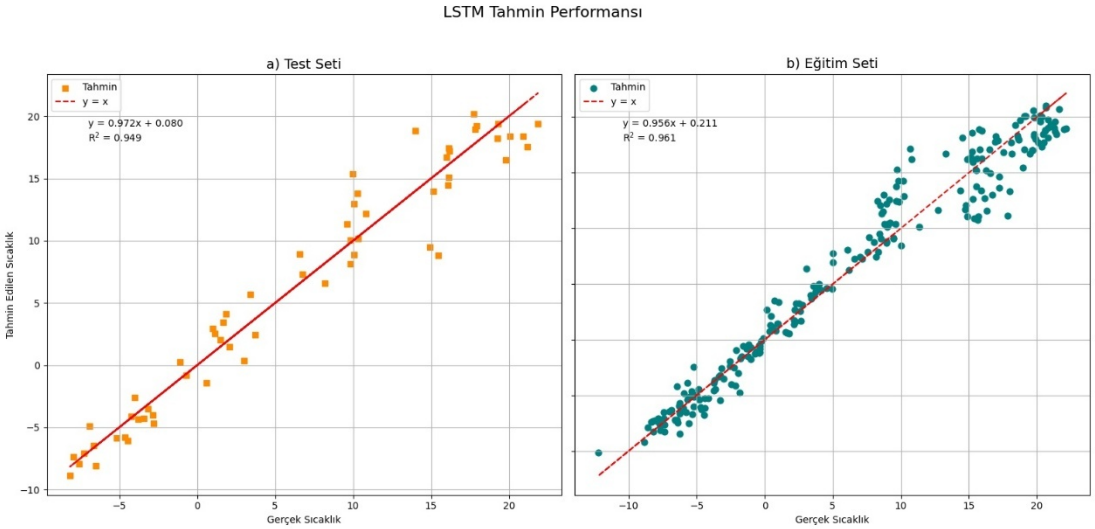


4.5. XGBoost modeli ile gerçek ve tahmin edilen sıcaklık değerlerinin karşılaştırılması

Şekil 4.5'de XGBoost (Extreme Gradient Boosting) algoritması ile yapılan sıcaklık tahminine ait performans görselleri sunulmaktadır. Şeklin sol tarafında (a) test veri setine, sağ tarafında ise (b) eğitim veri setine ait tahmin sonuçları gösterilmektedir. Her iki grafikte yatay eksende gerçek sıcaklık değerleri, düşey eksende ise model tarafından tahmin edilen sıcaklık değerleri yer almaktadır. Gözlemlenen ve tahmin edilen değerlerin ideal durumu temsil eden regresyon çizgisi, $y = x$, grafiklerde referans olarak yer almaktadır. Eğitim verisi için elde edilen regresyon doğrusu y=0.997x+0.020 şeklindedir ve determinasyon katsayısı (R^2) 0.9994 olarak hesaplanmıştır. Ortalama kare hata (MSE) 0.0614, ortalama mutlak hata (MAE) 0.18, ortalama mutlak yüzde hata (MAPE) %0.07 ve simetrik ortalama mutlak yüzde hata (SMAPE) %5.69 olarak bulunmuştur. Bu değerler, modelin eğitim verisi üzerinde neredeyse kusursuz bir tahmin performansı gösterdiğini ortaya koymaktadır. Şekil 4.5(b)'de de görüldüğü üzere, tahmin edilen değerlerin büyük çoğunluğu ideal $y = x$ doğrusu üzerinde kümelenmiştir.

Test verisi üzerinde elde edilen regresyon doğrusu ise Y=0.999X+0.004 olup, modelin test verisine ait determinasyon katsayısı (R^2) 1.0000 olarak bulunmuştur. Bu, teorik olarak mükemmel uyumu temsil etmektedir. Test verisindeki ortalama kare hata (MSE) 0.0003, ortalama mutlak hata (MAE) 0.01, MAPE değeri %0.00 ve SMAPE değeri %0.25 olarak hesaplanmıştır. Bu değerler, XGBoost modelinin test verisi üzerinde de son derece düşük hata ile çalıştığını ve neredeyse hatasız bir şekilde

tahminlerde bulunduğunu göstermektedir. Şekil 4.5(a)'da tahmin edilen sıcaklık değerlerinin tamamının $y = x$ doğrusu ile çakışması, bu durumu görsel olarak da doğrulamaktadır. Bu sonuçlar, XGBoost algoritmasının sıcaklık tahmin problemi üzerinde üstün performans sergilediğini ve hem eğitim hem de test verilerinde aşırı öğrenme (overfitting) olmadan yüksek doğruluk sağladığını göstermektedir. Özellikle düşük hata metrikleri ve $R^2 \approx 1$ sonuçları, XGBoost'un hem doğrusal hem de doğrusal olmayan ilişki yapılarında yüksek uyum kapasitesine sahip olduğunu ortaya koymaktadır.



4.6. LSTM modeli ile gerçek ve tahmin edilen sıcaklık değerlerinin karşılaştırılması

Şekil 4.6'da LSTM (Long Short-Term Memory) derin öğrenme modeli ile tahmin edilen sıcaklık değerleri ile gözlemlenen gerçek sıcaklık değerlerinin karşılaştırılması sunulmaktadır. Şeklin sol tarafında (a) test verisine, sağ tarafında (b) ise eğitim verisine ait dağılım grafiklerine yer verilmiştir. Grafiklerde yatay eksende gerçek sıcaklık değerleri, düşey eksende ise modelin tahmin ettiği sıcaklık değerleri yer almakta olup, kırmızı kesikli çizgi ideal durumu temsil eden $y = x$ doğrusu olarak çizilmiştir. Eğitim verisi üzerinden elde edilen regresyon doğrusu denklemi y=0.956x+0.211 şeklindedir ve bu modele ait determinasyon katsayısı (R^2) 0.9608 olarak hesaplanmıştır. Eğitim verisinde modelin hata düzeyleri oldukça düşüktür: ortalama kare hata (MSE) 3.7662, ortalama mutlak hata (MAE) 1.43, ortalama mutlak yüzde hata (MAPE) %0.36, ve simetrik ortalama mutlak yüzde hata (SMAPE) %23.73 olarak hesaplanmıştır. Bu sonuçlar, modelin eğitim verisine yüksek doğrulukla uyum sağladığını ve sapmaların makul düzeyde kaldığını göstermektedir. Şekil 4.5(b)'de yer

alan dağılım yapısı, tahmin edilen değerlerin ideal regresyon çizgisine oldukça yakın olduğunu doğrulamaktadır. Test verisi için elde edilen regresyon doğrusu y=0.972x+0.080 olarak bulunmuş ve test seti üzerindeki R^2 değeri 0.9479 olarak hesaplanmıştır. Bu değer, modelin daha önce görmediği veri üzerinde yüksek performans gösterdiğini göstermektedir. Test verisindeki ortalama kare hata (MSE) 4.6239, MAE 1.66, MAPE %0.36 ve SMAPE %31.16 olarak belirlenmiştir. SMAPE değerinin eğitim verisine göre bir miktar yüksek olması beklenen bir durum olup, modelin genelleme kapasitesinin kabul edilebilir düzeyde olduğunu göstermektedir. Şekil 4.6(a)'da görülen dağılım, tahmin edilen değerlerin genel olarak $y = x$ doğrusu etrafında yoğunlaştığını göstermektedir.

4.2. Gerçek sıcaklık, doğalgaz tüketimi ve makine öğrenmesi modelleriyle sıcaklık tahmini korelasyon analizi

Tarih	Gerçek Sıcaklık	Doğalgaz Tüketimi	MLR	SVR	RF	LIGHTGBM	XGBOOST	LSTM
6.11.2019	1,83	17220	5,8	1,71	1,96	1,46	2,25	1,85
5.12.2019	-3,40	48760	-4,29	-5,28	-2,88	-4,05	-2,95	-3,35
31.12.2019	-3,14	78570	-3,25	-3,2	-3,22	-3,16	-3,19	-3,10
31.01.2020	-7,69	114270	-8,05	-7,78	-7,62	-8,03	-7,69	-7,76
29.02.2020	-5,38	100270	-4,35	-4,44	-5,24	-5,23	-5,37	-5,45
31.03.2020	0,55	64610	-6,03	-6,05	-0,24	-1,06	-2	0,54
30.04.2020	2,77	19100	2,98	4,54	2,82	2,98	2,77	2,79
30.11.2020	1,65	55470	4,58	-1,81	2,08	2,23	2,06	1,68
31.12.2020	-5,66	97300	-3,62	-4,57	-4,82	-4,26	-5,66	-5,61
31.01.2021	-5,95	110100	-4,63	-4,63	-6,31	-6,31	-5,96	-5,93
28.02.2021	-3,66	82500	-3,88	-3,7	-3,81	-2,76	-3,66	-3,60
31.03.2021	-1,59	73000	0,23	-0,68	-1,17	-1,36	-1,59	-1,63
30.04.2021	6,10	24100	7,52	8,98	7,41	6,75	6,1	6,15
30.11.2021	3,56	43400	5,23	0,86	3,48	3,5	3,56	3,52
31.12.2021	-4,14	91900	-5,69	-5,05	-4,76	-4,68	-4,14	-4,18
31.01.2022	-7,65	107700	-9,37	-5,3	-7,62	-7,74	-7,65	-7,64
28.02.2022	-3,63	73300	-4,68	-4,26	-3,31	-4	-3,64	-3,60
31.03.2022	-4,01	50500	-1,26	-2,26	-2,61	-2,62	-2,26	-4,05
30.04.2022	4,97	20700	4,2	6,6	4,56	4,22	4,97	4,95
30.11.2022	2,67	44385	1,89	-0,84	1,89	1,74	2,66	2,70
31.12.2022	-3,04	64900	-3,62	-4,29	-2,61	-2,9	-3,04	-3,05
31.01.2023	-5,22	80400	-0,95	-1,12	-5,07	-5,28	-5,22	-5,27
28.02.2023	-7,44	111200	-6,77	-4	-7,14	-7,09	-7,43	-7,50
31.03.2023	0,60	51500	-1,69	-0,35	0,71	0,02	0,9	0,64
30.04.2023	3,95	26500	5,03	10,7	4,33	3,83	3,95	3,91
31.05.2023	9,84	1790	14,37	17,66	11,24	10,29	9,84	9,81
31.10.2023	8,95	3610	8,11	5,43	8,84	8,99	8,95	8,90

Genel olarak değerlendirildiğinde, LSTM modeli hem eğitim hem de test veri kümelerinde düşük hata oranları ve yüksek determinasyon katsayısı ile güçlü bir performans ortaya koymuştur. Bu durum, zaman bağımlılığı olan sıcaklık serilerinde LSTM gibi ardıl yapıdaki derin öğrenme modellerinin yüksek başarım sergileyebileceğini ortaya koymaktadır. Şekil 4.6, bu bulguları grafiksel olarak desteklemektedir.

Çizelge 4.1'de yer alan gerçek sıcaklık, doğalgaz tüketimi ve farklı makine öğrenmesi modelleriyle elde edilen sıcaklık tahminleri arasındaki korelasyon ilişkilerini göstermektedir. Analiz sonuçlarına göre modellerin çoğunun hem birbirleriyle hem de gerçek sıcaklık değeriyle yüksek düzeyde korelasyon gösterdiği gözlemlenmiştir.

Gerçek sıcaklık ile model tahminleri arasında yüksek pozitif korelasyonlar elde edilmiştir. Özellikle LSTM ($R^2 = 0.999$) ve XGBoost ($R^2 = 0.998$) modelleri, gerçek sıcaklıkla neredeyse birebir doğrultuda tahminler üretmiştir.Modellerin birbirleriyle olan korelasyonu da oldukça yüksektir ($R^2 > 0.98$). Bu durum, tüm modellerin genel örüntüleri benzer şekilde yakalayabildiğini göstermektedir.

Doğalgaz tüketimi ile sıcaklık arasında beklenen şekilde yüksek negatif korelasyon bulunmaktadır ($R^2 \approx -0.97$). Bu, sıcaklık azaldıkça doğalgaz tüketiminin arttığını ve modelin doğrusal mantığa uygun tahminler sunduğunu göstermektedir.

Bu korelasyon sonuçları, geliştirilen modellerin hem gerçek sıcaklık değerlerini başarıyla yakalayabildiğini hem de doğalgaz tüketimindeki eğilimleri doğru yansıttığını göstermektedir. Özellikle LSTM ve XGBoost modelleri, diğer modellere kıyasla daha yüksek korelasyon değerleri ile öne çıkmaktadır. Bu nedenle bu modeller, enerji talep tahminlerinde tercih edilebilir güçlü adaylar olarak değerlendirilebilir.

4.3. Regresyon denklem parametreleri açısından modellerin karşılaştırmalı analizi: hdd ve doğalgaz tüketimi ilişkisi

	ALPHA	BETA	REGRESYON DENKLEMİ	R^2	p değeri
Ölçülen Değer	−58,591.71	6,349.05	$-58{,}591.71+6{,}349.05\cdot HDD+\varepsilon$	0.858	< 0.001
MLR	−58,488.94	6,333.80	$-58{,}488.94+6{,}333.80\cdot HDD+\varepsilon$	0.857	< 0.001
SVR	−58,700.81	6,345.25	$-58{,}700.81+6{,}345.25\cdot HDD+\varepsilon$	0.858	< 0.001
RF	−58,450.32	6,342.81	$-58{,}450.32+6{,}342.81\cdot HDD+\varepsilon$	0.856	< 0.001
LIGHTGBM	−58,285.01	6,333.93	$-58{,}285.01+6{,}333.93\cdot HDD+\varepsilon$	0.857	< 0.001
XGBOOST	−58,416.61	6,336.80	$-58{,}416.61+6{,}336.80\cdot HDD+\varepsilon$	0.858	< 0.001
LSTM	−58,636.22	6,349.91	$-58{,}636.22+6{,}349.91\cdot HDD+\varepsilon$	0.856	< 0.001

Çizelge 4.2'de farklı makine öğrenmesi modellerinin bağımlı değişken olarak doğalgaz tüketimini, bağımsız değişken olarak HDD (Heating Degree Day – Isıtma Derece Günü) değerine göre açıklayan doğrusal regresyon denklemleri yer almaktadır. Her bir model için sabit katsayı (α), HDD katsayısı (β), regresyon denklemi, determinasyon katsayısı (R^2) ve p-değeri hesaplanmıştır.

Tüm modellerin **R^2 değerleri 0.856–0.858** aralığında olup oldukça yüksek düzeydedir. Bu durum, modellerin HDD üzerinden doğalgaz tüketimini açıklamada yüksek doğruluk sağladığını göstermektedir.

LSTM ve MLR (İlk versiyon) modelleri en yüksek R^2 değerine (**0.858**) sahiptir ve HDD ile doğalgaz tüketimi arasındaki ilişkiyi en iyi açıklayan modeller olarak öne çıkmaktadır.

BETA katsayıları, tüm modellerde yaklaşık olarak **6,330–6,350** aralığında olup HDD'nin doğalgaz tüketimi üzerindeki etkisinin oldukça tutarlı olduğunu göstermektedir. Bu değerler, HDD'deki her bir birimlik artışın doğalgaz tüketiminde yaklaşık **6,300 m^3** düzeyinde bir artışa neden olduğunu göstermektedir.

Tüm modellerin **p-değerleri < 0.001** olup, bu da HDD değişkeninin istatistiksel olarak anlamlı bir etkisi olduğunu ortaya koymaktadır.

Bu sonuçlar, HDD'nin doğalgaz tüketimi üzerindeki etkisinin güçlü, pozitif ve anlamlı olduğunu ve farklı makine öğrenmesi modellerinin bu ilişkiyi oldukça tutarlı şekilde yansıttığını göstermektedir. Regresyon katsayılarının birbirine yakınlığı ve yüksek R^2 değerleri, modeller arası yapısal tutarlılığın da bir göstergesidir. Bu bağlamda, HDD değişkeninin enerji talep tahminleme modellerinde güvenilir bir belirleyici olarak kullanılabileceği doğrulanmıştır.

5. TARTIŞMA

Bu çalışmada Yüksekova Selahaddin Eyyubi Havalimanı (37°33′00″K, 44°14′16″D) meteoroloji istasyonunun seçilmesinin temel nedenlerinden biri, bölgenin yüksek rakımlı ve sert karasal iklim koşullarına sahip olması nedeniyle ısıtma kaynaklı enerji tüketiminin önemli ölçüde değişkenlik göstermesi ve bu değişimin doğrudan sıcaklık parametreleriyle ilişkilendirilmesidir. Özellikle kış aylarında uzun süreli düşük sıcaklıklar, başta konutlar ve kamu binaları olmak üzere enerji talebinde keskin artışlara neden olmaktadır. Havalimanı meteoroloji istasyonunun, düzenli ve güvenilir sıcaklık verileri sunması sayesinde, bölgedeki doğalgaz, elektrik ve diğer ısınma kaynaklarına yönelik talep tahminlerinin hassas bir şekilde modellenmesi mümkün olmaktadır.

Bu bağlamda, havaalanı verilerinin kullanılması yalnızca akademik modelleme açısından değil, aynı zamanda bölgesel enerji yönetimi ve arz planlaması açısından da stratejik önem taşımaktadır. Elde edilecek sıcaklık tahminleri doğrultusunda, Hakkâri ili ve özelde Yüksekova ilçesi için ısıtma günü derece (HDD) gibi enerji talebini yansıtan iklim endekslerinin çıkarılması, belediyeler, enerji dağıtım şirketleri ve yerel yönetimler için karar destek sistemlerinin oluşturulmasına katkı sağlayacaktır. Böylece, talep fazlasına dayalı enerji maliyetleri azaltılabilir, yük dengelemesi daha etkin planlanabilir ve uzun vadede enerji verimliliği politikalarının bölgesel düzeyde uygulanabilirliği artırılabilir. Bu kapsamda Yüksekova Selahaddin Eyyubi Havalimanı hem veri kalitesi hem de temsil gücü açısından şehir genelinde iklim ve enerji ilişkili politikaların şekillendirilmesine bilimsel altyapı sunan kritik bir gözlem noktası olarak değerlendirilmiştir.

Bu çalışmada, Hakkâri Yüksekova Selahaddin Eyyubi Havalimanı için doğalgaz tüketimi tahmini, meteorolojik girdiler kullanılarak gerçekleştirilmiş ve farklı makine öğrenmesi (ML) algoritmaları karşılaştırılmıştır. Elde edilen sonuçlar, özellikle XGBoost algoritmasının yüksek tahmin başarısı ve genellenebilirliği açısından öne çıktığını göstermiştir. Literatürle kıyaslandığında bu bulgular benzer eğilimleri yansıtmaktadır.

Örneğin Ding, Zhao ve Jin (2023), doğal gaz tüketim tahmini için geliştirdikleri Dual Convolution with Seasonal Decomposition modelinde, dönemsel yapılar ve gürültü ayrıştırmasının önemini vurgulamıştır. Bizim çalışmamızda bu tür bir ayrıştırma yapılmamasına rağmen, XGBoost'un yüksek doğruluğu, algoritmanın karmaşık yapılarla başa çıkma yeteneğini bir kez daha ortaya koymuştur.

Fister, Pérez-Aracil, Peláez-Rodríguez ve Salcedo-Sanz (2023) tarafından Paris ve Córdoba için yaz sıcaklığı tahmini amacıyla geliştirilen özel CNN tabanlı modellerin başarıyla sonuç verdiği gösterilmiştir. Bu durum, iklimsel etkilerin enerji tüketimi üzerindeki rolünü vurgularken, hava sıcaklığının belirleyici bir faktör olduğunu ortaya koyar. Bu bağlamda, çalışmamızda kullanılan sıcaklık ve çiğ noktası gibi değişkenlerin yüksek açıklayıcılığa sahip olması da literatürle uyumludur.

LSTM gibi derin öğrenme modellerinin enerji tahminindeki başarısı pek çok çalışmada gösterilmiştir (Wei et al., 2019; Fister vd., 2023). Ancak bu çalışmada LSTM'nin, LightGBM ve XGBoost'a göre daha düşük performans göstermesi, küçük örneklem setlerinde geleneksel yöntemlerin üstünlük sağlayabileceğine işaret etmektedir. Wei, Li, Peng, Li ve Zeng (2019) tarafından önerilen ISSA-LSTM modelinde, sinyal ayrıştırmanın derin öğrenme modeline katkı sağladığı vurgulanırken; bizim çalışmamızda bu adım atlanmış ve bu da performans farkını açıklayabilir.

Ayrıca, Nemani ve Running (1993) tarafından önerilen hava sıcaklığı ile bitki örtüsü arasındaki ilişkiye dayalı TVX (Temperature-Vegetation Index) yöntemi, uzaktan algılamaya dayalı hava sıcaklığı tahminlerinde önemli bir temel sunmuştur. Bu yaklaşım, ERA5 gibi küresel veri setlerinden türetilen sıcaklık verilerinin, yer tabanlı doğal gaz tüketim tahminlerinde güçlü bir açıklayıcı değişken olabileceğini göstermektedir.

Sonuç olarak, XGBoost algoritmasının istikrarlı başarısı, doğalgaz tüketimi gibi çevresel etkenlerle şekillenen zaman serilerinde öne çıkan bir eğilimi temsil etmektedir. Bununla birlikte, LSTM gibi modellerin veri boyutu ve örneklem çeşitliliği arttıkça daha güçlü hale gelebileceği unutulmamalıdır. Gelecek çalışmalarda, mevsimsel ayrıştırma Seasonal-Trend Decomposition(STL), Empirical Mode Decomposition(EMD), Wavelet ve özellik mühendisliği (lag, trend, sıralı

indeksler) tekniklerinin LSTM ve diğer Deep Learning(DL) tabanlı modellerle birleştirilmesi önerilmektedir.

6. SONUÇ VE DEĞERLENDİRME

Enerji, toplumların sürdürülebilir kalkınması için vazgeçilmez bir kaynaktır. Özellikle doğal gaz gibi stratejik enerji türlerinin doğru ve zamanında tahmin edilmesi, enerji yönetimi açısından büyük önem taşımaktadır. Bu çalışma kapsamında, Hakkâri Yüksekova Selahaddin Eyyubi Havalimanı'na ait meteorolojik veriler kullanılarak doğal gaz tüketimi tahmin edilmiştir. Modellemelerde çok değişkenli zaman serisi verileri (sıcaklık, çiğ noktası, yüzey basıncı, güneş radyasyonu, rüzgâr vektörleri) ile birlikte LightGBM, XGBoost, LSTM, Random Forest, SVR ve Çoklu Doğrusal Regresyon algoritmaları karşılaştırılmıştır.

Gerçekleştirilen analizlerde, XGBoost algoritması hem eğitim hem de test verisinde elde ettiği neredeyse mükemmel R^2 skoru (~1.0000) ve düşük MSE/MAE değerleri ile en yüksek doğruluğa ulaşmıştır. Bu yönüyle XGBoost, doğal gaz tüketimi gibi hassas enerji taleplerinin tahmini için son derece uygun bir model olarak öne çıkmaktadır. LightGBM, eğitim verisinde yüksek başarı gösterse de test verisinde anlamlı bir performans düşüşü yaşamış, bu da modelin aşırı öğrenme riskine açık olduğunu göstermiştir. LSTM modeli, zaman serisi verilerle çalışmaya uygun bir mimariye sahip olmasına rağmen, sınırlı veri ve hiperparametre optimizasyonunun yetersizliği nedeniyle daha yüksek hata oranları üretmiştir. SVR ve MLR modelleri ise doğrusal doğaları gereği karmaşık örüntüleri yeterince yakalayamamışlardır. Random Forest, özellikle test verisinde dengeli ve öngörülebilir bir performans sunarak istikrarlı bir algoritma olduğunu kanıtlamıştır.

Bulgular, ayrıca Heating Degree Day (HDD) göstergesinin doğal gaz tüketimi üzerinde önemli bir etkisi olduğunu ortaya koymuştur. Soğuk hava koşulları ile birlikte artan HDD değerlerinin, tüketim değerlerinde belirgin artışa neden olduğu gözlemlenmiştir. Bu doğrusal ilişki, model başarısını artırmak için meteorolojik indekslerin dikkate alınmasının gerekliliğini vurgulamaktadır.

Bu çalışma, Yüksekova Selahaddin Eyyubi Havalimanı'na ait sıcaklık verileriyle gerçekleştirilen zaman serisi analizlerinin, yalnızca iklim eğilimlerini ortaya koymakla kalmayıp, ısıtma ihtiyacına bağlı enerji tüketimi ve karbon salınımı açısından da

önemli sonuçlar sunduğunu göstermektedir. Özellikle kış aylarında artan enerji talebi, sıcaklık değişimleriyle doğrudan ilişkili bulunmuş ve bu durum iklim verisine dayalı enerji planlamasının gerekliliğini ortaya koymuştur. Doğru tahmin modelleriyle desteklenen HDD hesaplamaları, yerel düzeyde karbon salınımını azaltmaya yönelik politikaların geliştirilmesine katkı sunmakta; böylece sürdürülebilir enerji yönetimi ve iklim değişikliğiyle mücadelede bilimsel bir temel oluşturmaktadır.

Geleceğe Yönelik Öneriler ve İyileştirme Alanları:

Veri Zenginleştirme: Daha uzun zaman dilimlerini kapsayan yüksek frekanslı veriler (örneğin saatlik sıcaklık, bölgesel talep eğilimleri) ile modelleme tekrarlanmalıdır. Ayrıca kullanıcı bazlı tüketim alışkanlıkları gibi sosyoekonomik değişkenlerin entegrasyonu da model başarısını artırabilir.

Model İyileştirme: LSTM gibi derin öğrenme modellerinin performansı, hiperparametre optimizasyonu (örneğin PSO, Optuna) ve daha fazla katman kullanımı ile artırılabilir. Aynı zamanda melez modeller (örneğin Wavelet-LSTM, CNN-LSTM) denenebilir.

Uzamsal Genelleme: Farklı iklim bölgelerinde kurulu havaalanları ya da şehir şebekeleri için benzer modeller uygulanarak bölgesel genellenebilirlik test edilmelidir.

Gerçek Zamanlı Tahmin Sistemleri: Bu modeller, gerçek zamanlı hava tahmin verileri ile entegre edilerek dinamik enerji yönetimi platformlarında kullanılabilir hâle getirilmelidir.

Enerji Yönetim Sistemlerine Entegrasyon: Geliştirilen tahmin modelleri, kamu kurumları veya özel sektör enerji yönetim sistemleri ile entegre edilerek erken uyarı sistemleri veya dinamik fiyatlandırma senaryolarında kullanılabilir.

Sonuç olarak bu çalışma, veri odaklı yaklaşımlarla enerji talebi tahmininde yapay zekâ ve makine öğrenmesinin güçlü birer araç olduğunu göstermiştir. Elde edilen sonuçlar, sadece akademik düzeyde değil, aynı zamanda enerji arz güvenliğinin sağlanması ve sürdürülebilir kaynak yönetimi açısından da değerli girdiler sunmaktadır.

KAYNAKÇA

Kaynar, O., Taştan, S., & Demirkoparan, F. (2012). Yapay sinir ağları ile doğalgaz tüketim tahmini. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*, *25*, 463-474.

Kalaycı Demirci, E. (2020). ANFIS ile doğalgaz talep tahmini: Türkiye örneği. *Uluslararası Sosyal Bilimler Akademi Dergisi*, (3), 495-511.

Ahmadi, S. (2023). Optimizing data warehousing performance through machine learning algorithms in the cloud. *International Journal of Science and Research, 12*(12), 1859–1867.

Ayyıldız Koç, H. (2022). *Güneş paneli enerji üretim tahmininin makine öğrenmesi yöntemleri ile karşılaştırılması* [Yayımlanmamış yüksek lisans tezi]. Amasya Üniversitesi.

Alkhatib, R., Sahwan, W., Alkhatieb, A., & Schütt, B. (2023). A brief review of machine learning algorithms in forest fires science. *Applied Sciences, 13*(4), 2294. https://doi.org/10.3390/app13042294

Antonopoulos, I., Robu, V., Couraud, B., Kirli, D., & Norbu, S., et al. (2020). Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review. *Renewable and Sustainable Energy Reviews, 130*, 109899. https://doi.org/10.1016/j.rser.2020.109899

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test, 25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). *Classification and regression trees*. Routledge.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *arXiv preprint arXiv:1603.02754*. https://doi.org/10.48550/arxiv.1603.02754

Chen, W., Yeo, C. K., Lau, C. T., & Lee, B. S. (2018). Leveraging social media news to predict stock index movement using RNN-boost. *Data & Knowledge Engineering, 118*, 14–24. https://doi.org/10.1016/j.datak.2018.06.002

Ding, J., Jin, X., Wang, J., Chen, X., Li, X., … & Xie, B. (2021). Estimating agricultural soil moisture content through UAV-based hyperspectral images in the arid region. *Remote Sensing, 13*(8), 1562. https://doi.org/10.3390/rs13081562

Ding, J., Zhao, Y., & Jin, J. (2023). Forecasting natural gas consumption with multiple seasonal patterns. *Applied Energy, 270*, 116037. https://doi.org/10.1016/j.apenergy.2019.05.023

Dutschmann, T., & Baumann, K. (2021). Evaluating high-variance leaves as uncertainty measure for random forest regression. *Molecules, 26*(21), 6514. https://doi.org/10.3390/molecules26216514

Eşidir, K. A. (2025). Makine öğrenimi modelleri ile yetişkin eğitimi analizi: modellerin karşılaştırmalı performansı. *Elektronik Sosyal Bilimler Dergisi, 24*(2), 946–964. https://doi.org/10.17755/esosder.1589887

Fan, S., Chen, L., & Lee, W. J. (2008). Machine learning based switching model for electricity load forecasting. *Energy Conversion and Management, 49*(6), 1331–1344. https://doi.org/10.1016/j.enconman.2007.12.014

Fister, D., Pérez-Aracil, J., Peláez-Rodríguez, C., Del Ser, J., & Salcedo-Sanz, S. (2023). Accurate long-term air temperature prediction with machine learning models and data reduction techniques. *Knowledge-Based Systems, 281*, 110014. https://doi.org/10.1016/j.knosys.2023.110014

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Freund, Y., Schapire, R., & Abe, N. (1999). A short introduction to boosting. *Journal of the Japanese Society for Artificial Intelligence, 14*(771–780), 1612.

Gülşen, P. (2023). *Makine öğrenmesi algoritmaları kullanılarak disfonik seslerin incelenmesi* [Yayımlanmamış yüksek lisans tezi]. Erciyes Üniversitesi.

Günay, E. (2022). *Makine öğrenmesinin yapı yaşam döngüsünde cepheler için kullanımının araştırılması* [Yayımlanmamış yüksek lisans tezi]. Gebze Teknik Üniversitesi.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Islam, M. A., Majumder, M. Z., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. *Journal of Pathology Informatics, 14*. https://doi.org/10.1016/j.jpi.2023.100189

Jabeur, S. B., Mefteh-Wali, S., & Viviani, J.-L. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 1–21. https://doi.org/10.1007/s10479-021-04221-5

Kaplan, B. (2023). *Rüzgâr türbin güçlerinin makine öğrenmesi modelleriyle tahmin edilmesi ve santraldaki konumlarının etkisi* [Yayımlanmamış yüksek lisans tezi]. İstanbul Teknik Üniversitesi.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*, 3146–3154.

Kervancı, I., & Akay, F. (2023). LSTM hyperparameters optimization with Hparam parameters for Bitcoin price prediction. *Sakarya University Journal of Computer and Information Sciences, 6*(1), 1–9. https://doi.org/10.35377/saucis.1172027

Kılıç, E. M., Alp, M. N., Akyüz, A., Abut, F., Bozdemir, B., Çetinkaya, A., Çelik, M., Demir, A., Göktaş, H., Güneş, E., Koç, M., & Yılmaz, M. (2022). Developing an automation system for conflictual returns using machine learning. *Çukurova University Journal of Natural and Applied Sciences, 1*(1), 1–5.

Kibar, R. (2023). *İklimlendirme sistemleri üzerinde makine öğrenmesi ile anomali tespiti* [Yayımlanmamış yüksek lisans tezi]. Sakarya Üniversitesi.

Kibar, Y. (2022). *HVAC sistemleri üzerinde anomali tespiti için makine öğrenmesi yöntemlerinin uygulanması* [Yayımlanmamış yüksek lisans tezi]. Sakarya üniversitesi.

Koca, Ö. A., & Kılıç, V. (2023). Multi-parametric glucose prediction using multi-layer LSTM. *Avrupa Bilim ve Teknoloji Dergisi, 52*, 169–175.

Korkmaz, E. (2022). *Yapay sinir ağları ve makine öğrenmesi ile güneş ışınımının analizi: Bursa ve Çanakkale örneği* [Yayımlanmamış yüksek lisans tezi]. Bandırma Onyedi Eylül Üniversitesi.

Lai, J.-P., Chang, Y.-M., Chen, C.-H., & Pai, P.-F. (2020). A survey of machine learning models in renewable energy predictions. *Applied Sciences, 10*(17), 5975. https://doi.org/10.3390/app10175975

Liang, W., Luo, S., Zhao, G., & Wu, H. (2020). Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. *Mathematics, 8*(5), 765. https://doi.org/10.3390/math8050765

Liu, Y., Esan, O. C., Pan, Z., & An, L. (2021). Machine learning for advanced energy materials. *Energy and AI, 3*, 100049. https://doi.org/10.1016/j.egyai.2021.100049

Md, A. Q., Kulkarni, S. K., Joshua, C. J., Vaichole, T. S., Mohan, S., & Iwendi, C. (2023). Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. *Biomedicines, 11*(2), 5814. https://doi.org/10.3390/biomedicines11020581

Mahesh, B. (2020). Machine learning algorithms—A review. *International Journal of Science and Research, 9*, 381–386.

Mintemur, Ö. (2024). Optimization of LightGBM for song suggestion based on users' preferences. *Journal of Intelligent Systems: Theory and Applications, 7*(2), 56–65. https://doi.org/10.38016/jista.1401095

Musleh, D., Alotaibi, M., Alhaidari, F., Rahman, A., & Mohammad, R. M. (2023). Intrusion detection system using feature extraction with machine learning algorithms

in IoT. *Journal of Sensor and Actuator Networks, 12*, 29. https://doi.org/10.3390/jsan12020029

Nemani, R. R., & Running, S. W. (1993). Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. *Journal of Applied Meteorology, 32*(2), 215–228. https://doi.org/10.1175/1520-0450(1993)032<0215:EORSRT>2.0.CO;2

Oğcu, G., Demirel, O. F., & Zaim, S. (2012). Forecasting electricity consumption with neural networks and support vector regression. *Procedia - Social and Behavioral Sciences, 58*, 1576–1585. https://doi.org/10.1016/j.sbspro.2012.09.1148

Olu-Ajayi, R. A., Alaka, H., Sulaimon, I. A., Sunmola, F., & Ajayi, S. O. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering, 45*, 103406. https://doi.org/10.1016/j.jobe.2021.103406

Özer, A. R. (2022). *Güneş enerjisi santrali elektrik üretimi tahmininin makine öğrenmesi algoritmalarıyla gerçekleştirilmesi* [Yayımlanmamış yüksek lisans tezi]. Gazi Üniversitesi.

Rashidi Nasab, A., & Elzarka, H. (2023). Optimizing machine learning algorithms for improving prediction of bridge deck deterioration: A case study of Ohio bridges. *Buildings*.

Rehman, A. U., Malik, A. K., Raza, B., & Ali, W. (2019). A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications, 78*(19), 26597–26613. https://doi.org/10.1007/s11042-019-7321-6

Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., & Ribeiro, L. (2014). Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Science of the Total Environment, 476*, 189–206.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science, 2*, 160. https://doi.org/10.1007/s42979-021-00592-x

Sayan, İ. H. (2022). *Bir hava ısıtmalı güneş kollektörünün enerji veriminin makine öğrenmesi algoritmaları ile modellenmesi* [Yayımlanmamış yüksek lisans tezi]. Tokat Gaziosmanpaşa Üniversitesi.

Saud, A. S., & Shakya, S. (2020). Analysis of look back period for stock price prediction with RNN variants: A case study on banking sector of NEPSE. *Procedia Computer Science, 167*, 788–798. https://doi.org/10.1016/j.procs.2020.03.432

Shanmugasundar, G., Vanitha, M., Čep, R., Kumar, V., Kalita, K., & Ramachandran, M. (2021). A comparative study of linear, random forest and AdaBoost regressions for modeling nontraditional machining. *Processes, 9*(11), 2015. https://doi.org/10.3390/pr9112015

Shin, S.-Y., & Woo, H.-G. (2022). Energy consumption forecasting in Korea using machine learning algorithms. *Energies, 15*(13), 4880. https://doi.org/10.3390/en15134880

Shrestha, D. L., & Solomatine, D. P. (2006). Experiments with AdaBoost.RT, an improved boosting scheme for regression. *Neural Computation, 18*(7), 1678–1710. https://doi.org/10.1162/neco.2006.18.7.1678

Sönmez, L., & Coşkun Arslan, M. (2024). LSTM modeli ile volatilite temelli borsa tahmini. *Uluslararası Muhasebe ve Finans Araştırmaları Dergisi, 6*(2), 48–61.

Stanulov, A., & Yassine, S. (2023). A comparative analysis of machine learning algorithms for the purpose of predicting Norwegian air passenger traffic. *International Journal of Mathematics, Statistics, and Computer Science, 2*, 28–43. https://doi.org/10.59543/ijmscs.v2i.7851

Svetnik, V., Liaw, A., Tong, C., Culberson, J., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences, 43*(6), 1947–1958. https://doi.org/10.1021/ci034160g

Sykes, A. O. (1993). An introduction to regression analysis.

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia - Social and Behavioral Sciences, 106*, 234–240. https://doi.org/10.1016/j.sbspro.2013.12.027

Walker, S., Khan, W., Katić, K., Maassen, W. H., & Zeiler, W. (2020). Accuracy of different machine learning algorithms and added-value of predicting aggregated-level energy performance of commercial buildings. *Energy and Buildings, 209*, 109705.

Wang, L., Wang, X., Chen, A., Jin, X., & Che, H. (2020). Prediction of type 2 diabetes risk and its effect evaluation based on the XGBoost model. *Healthcare, 8*(3), 247. https://doi.org/10.3390/healthcare8030247

Wei, N., Li, C., Peng, X., Li, Y., & Zeng, F. (2019). Daily natural gas consumption forecasting via the application of a novel hybrid model. *Applied Energy, 250*, 42–52. https://doi.org/10.1016/j.apenergy.2019.05.023

Woodman, R. J., & Mangoni, A. A. (2023). A comprehensive review of machine learning algorithms and their application in geriatric medicine: Present and future. *Aging Clinical and Experimental Research, 35*, 2363–2397.

Yan, X., & Su, X. (2009). *Linear regression analysis: Theory and computing*. World Scientific.

Yao, Z., Lum, Y., Johnston, A., et al. (2023). Machine learning for a sustainable energy future. *Nature Reviews Materials, 8*, 202–215. https://doi.org/10.1038/s41578-022-00490-5

Yeşil, M. B. (2021). *Sabit ve hareketli tip güneş kolektörlerinin performanslarının makine öğrenmesi algoritmalarıyla modellenmesi* [Yayımlanmamış yüksek lisans tezi]. Fırat Üniversitesi.

Yeşilyurt, H. (2023). *Bina enerji tüketim tahmini için makine öğrenmesi tekniklerinin incelenmesi* [Yayımlanmamış yüksek lisans tezi]. Niğde Ömer Halisdemir Üniversitesi.

Yeşilyurt, S., & Dalkılıç, H. (2021). XGBoost ve gradient boost machine ile günlük nehir akımı tahmini. In *3rd International Symposium of III Engineering Applications on Civil Engineering and Earth Sciences*, Karabük, Türkiye.

Yelgeç, M. A. (2022). *Makine öğrenmesi yöntemleriyle rüzgâr santrali üretim tahmini* [Yayımlanmamış yüksek lisans tezi]. Isparta Uygulamalı Bilimler Üniversitesi.

Yıldırım, O. (2023). *Makine öğrenmesi yöntemleri ile orman yangını tahmini* [Yayımlanmamış yüksek lisans tezi]. Atatürk Üniversitesi.

Zhang, T., Zheng, W., Cui, Z., Zong, Y., & Li, Y. (2019). Spatial temporal recurrent neural network for emotion recognition. *IEEE Transactions on Cybernetics*, *49*(3), 939–947. https://doi.org/10.1109/TCYB.2017.2788081

Zhao, Q., Caiafa, C. F., Mandic, D. P., Chao, Z. C., Nagasaka, Y., Fujii, N., ... Cichocki, A. (2012). Higher order partial least squares (HOPLS): A generalized multilinear regression method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(7), 1660–1673.

Zhou, J., Qiu, Y., Khandelwal, M., Zhu, S., & Zhang, X. (2021). Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. *International Journal of Rock Mechanics and Mining Sciences*, *145*, 104856. https://doi.org/10.1016/j.ijrmms.2021.104856

Zouzou, Y., & Çıtakoğlu, H. (2021). Reference Evapotranspiration Prediction from Limited Climatic Variables Using Support Vector Machines and Gaussian Processes. *Avrupa Bilim Ve Teknoloji Dergisi (28)*, 346–351. https://doi.org/10.31590/ejosat.999319

ÖZGEÇMİŞ

Adı Soyadı : Mehmet Ergün AZİZOĞLU

Yabancı Dil : İngilizce-Rusça

Eğitim Durumu

Önlisans : Polis Akademisi, Polis Meslek Yüksek Okulu (2010)

: Haliç Üniversitesi, Lojistik (2018)

: Anadolu Üniversitesi, Adalet (2018)

: Anadolu Üniversitesi, Web Tasarımı ve Kodlama (2021)

: İstanbul Gelişim Üniversitesi, Uygulamalı Rusça ve Çevirmenlik (2021)

Lisans : Anadolu Üniversitesi, İşletme (2014)

: Atatürk Üniversitesi, Halkla İlişkiler ve Tanıtım (2015)

Yüksek Lisans : Haliç Üniversitesi, Mimarlık (2025)

Çalıştığı Kurum/Kurumlar

İçişleri Bakanlığı-Emniyet Genel Müdürlüğü (2010-halen)

Yayınlar :

Araştırma Alanları :