T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK BÖLÜMÜ

HAVA KİRLİLİĞİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE TAHMİN EDİLMESİ

Taylan VALUNYA

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Dr. Öğr. Üyesi Coşkun PARİM

Temmuz, 2025

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

HAVA KİRLİLİĞİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE TAHMİN EDİLMESİ

Taylan VALUNYA tarafından hazırlanan tez çalışması 10.07.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Dr. Öğr. Üyesi Coşkun PARİM
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Dr. Öğr. Üyesi Coşkun PARİM, Danışman
Yıldız Teknik Üniversitesi _____

Doç. Dr. Erhan ÇENE, Üye
Yıldız Teknik Üniversitesi _____

Doç. Dr. Seda BAĞDATLI KALKAN, Üye
İstanbul Ticaret Üniversitesi _____

Danışmanım Dr. Öğr. Üyesi Coşkun PARİM sorumluluğunda tarafımca hazırlanan "Hava Kirliliğinin Makine Öğrenme Yöntemleriyle Tahmin Edilmesi" başlıklı tezimde veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Taylan VALUNYA

İmza

TEŞEKKÜR

Bu tez çalışmasının hazırlanmasında bana yol gösteren, değerli bilgi ve deneyimleriyle her zaman destek olan, sabrıyla ve yapıcı eleştirileriyle çalışmamı daha nitelikli hale getiren danışmanım Dr. Öğr. Üyesi Coşkun PARİM'e ve sunumumu sabırla dinleyerek katkı sağlayan çok değerli jüri üyeleri hocalarım Doç. Dr. Erhan ÇENE ve Doç. Dr. Seda BAĞDATLI KALKAN'a en içten teşekkürlerimi sunarım.

Bu süreçte manevi desteklerini hiçbir zaman esirgemeyen, her zaman yanımda olduklarını hissettiren sevgili aileme sonsuz teşekkür ederim. Varlığıyla bana güç veren, sabrıyla ve anlayışıyla en zorlu anlarda bile desteğini esirgemeyen sevgili eşime ise özel olarak teşekkür etmek isterim.

Taylan VALUNYA

İÇİNDEKİLER

SİMGE LİSTESİ vii

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ ix

TABLO LİSTESİ x

ÖZET xi

ABSTRACT xiii

1 GİRİŞ 1

1.1 Literatür Taraması....1
1.2 Hava Kirliliği Hakkında....3
1.3 Hava Kirliliği Değişkenleri....7
1.3.1 Partikül Madde (PM_{10}, $PM_{2,5}$)....8
1.3.2 Azot Dioksit (NO_2)....9
1.3.3 Kükürt Dioksit (SO_2)....10
1.3.4 Karbon Monoksit (CO)....11
1.3.5 Ozon (O_3)....11
1.3.6 Uçucu Organik Bileşikler (UOB)....12

2 VERİ SETİ 13

2.1 Veri Seti....13
2.1.1 Model....14

3 YÖNTEMLER 15

3.1 Makine Öğrenmesi Yöntemleri....16
3.1.1 Lineer Regresyon....17
3.1.2 K-En Yakın Komşu Algoritması....17
3.1.3 Destek Vektör Makine Regresyonu....18
3.1.4 Karar Ağacı Regresyonu....18

3.1.5 Rastgele Orman Regresyonu 19
3.1.6 Ekstra Ağaçlar Regresyonu 19
3.1.7 Bagging Karar Ağacı Regresyonu 19
3.1.8 AdaBoost Regresyonu 20
3.1.9 Gradyan Arttırma Regresyonu 20
3.1.10 XGBoost Regresyonu (Extreme Gradient Boosting) 21
3.1.11 Hafif Gradyan Arttırma Regresyonu 21
3.1.12 Cat Boost Regresyonu 21
3.2 Öznitelik Seçimi 22
3.2.1 Filtre Yöntemleri 22
3.2.1 Sarmalayıcı Yöntemler 22
3.2.2 Gömülü Yöntemler 22
3.3 Hata Metrikleri 23
3.3.1 R-Kare (R-Squared) 23
3.3.2 Ortalama Kare Hata 23
3.3.3 Kök Ortalama Kare Hata 23
3.3.4 Ortalama Mutlak Hata 24
4 BULGULAR **25**
4.1 Değişkenler Arası Korelasyonların İncelenmesi 30
4.2 Özniteliklerin Önem Derecelerinin İncelenmesi 31
4.3 Modellerin Veri Setlerindeki Performansların Karşılaştırılması 32
5 SONUÇ **34**
KAYNAKÇA **36**
TEZDEN ÜRETİLMİŞ YAYINLAR **45**

SİMGE LİSTESİ

PM_{10}	10 µm'den küçük partiküller
y_i	Gerçek değer
n	Gözlem sayısı
°C	Hava sıcaklığı
Km	Kilometre
SO_2	Kükürt dioksit
µm	Mikrometre
µ	Ortalama
O_3	Ozon
o	Rüzgar yönü
Σ	Sigma
$\hat{y}_i$	Tahmin edilen değer

KISALTMA LİSTESİ

HKİ	Hava Kalitesi İndeksi
İBB	İstanbul Büyükşehir Belediyesi
KSİ	Kirlilik Standart İndeksi
MAE	Ortalama Mutlak Hata (Mean Absolute Error)
MGM	Meteoroloji Genel Müdürlüğü
MSE	Ortalama Kare Hatası (Mean Squared Error)
RMSE	Kök Ortalama Kare Hatası (Root Mean Squared Error)
R^2	R Kare
UOB	Uçucu Organik Bileşik

ŞEKİL LİSTESİ

Şekil 1.1 Ulusal hava kalitesi PM_{10} değer grupları (Hava Kalitesi Raporu - Sürekli İzleme Merkezi, 2024)**.....7**

Şekil 2.1 Yalova il sınırları harita görünümü (Harita Genel Müdürlüğü, t.y.)**.....14**

Şekil 4.1 Hava sıcaklığının ortalama değeri**.....25**

Şekil 4.2 PM_{10} ortalama değeri**.....26**

Şekil 4.3 SO_2 ortalama değeri**.....26**

Şekil 4.4 Rüzgar yönü ortalama değeri**.....27**

Şekil 4.5 Rüzgar yönü ve PM_{10} değişkenleri**.....27**

Şekil 4.6 Korelasyon matrisi **.....30**

Şekil 4.7 Öznitelik önem skoru**.....32**

TABLO LİSTESİ

Tablo 1.1 Hava kalitesi renk skalası **5**

Tablo 1.2 Ulusal hava kalitesi indeksi kesme noktaları (Hava Kalitesi Raporu - Sürekli İzleme Merkezi, 2024) **6**

Tablo 1.3 Hava kirliliği değişkenleri **8**

Tablo 4.1 Günlere göre değişken ortalamaları **29**

Tablo 4.2 Yöntemlerin başarı oranlarının karşılaştırılması **33**

ÖZET

Hava Kirliliğinin Makine Öğrenmesi Yöntemleriyle Tahmin Edilmesi

Taylan VALUNYA

İstatistik Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Coşkun PARİM

Hava kirliliği, insan sağlığı, ekosistemler ve genel çevre kalitesi üzerinde geniş kapsamlı etkileri olan, günümüzün en kritik çevresel sorunlarından biridir. Sanayi faaliyetleri, kentleşme, araç emisyonları, enerji üretimi ve tarım gibi insan kaynaklı etkenlerle atmosfere salınan zararlı kimyasal ve fiziksel maddeler, güvenli sınırların üzerinde kirletici yoğunluklarına yol açabilmektedir. Özellikle yoğun nüfuslu veya sanayileşmiş bölgelerde hava kirliliği seviyelerinin önceden tahmin edilmesi, halk sağlığını korumak ve çevresel sürdürülebilirliği sağlamak için büyük önem taşımaktadır.

Bu çalışma, önemli bir hava kirletici olan Partikül Madde (PM_{10}) yoğunluğunu çeşitli makine öğrenmesi teknikleri kullanarak tahmin etmeyi amaçlamaktadır. Çalışmada kullanılan veri seti, 2018 ile 2020 yılları arasında Yalova ilinde bulunan hava kalitesi izleme istasyonundan saatlik olarak elde edilmiştir. PM_{10}'a ek olarak, SO_2, NO, NO_2, NOx ve O_3 gibi diğer hava kirleticiler ile birlikte hava sıcaklığı, rüzgar hızı, rüzgar yönü, bağıl nem ve hava basıncı gibi meteorolojik değişkenler de değerlendirmeye alınmıştır.

Veri seti kronolojik olarak %70 eğitim ve %30 test alt kümelerine ayrılmıştır. PM_{10} üzerindeki etkisi en yüksek olan değişkenleri belirlemek için özellik seçimi yöntemleri uygulanmıştır. Birden fazla makine öğrenmesi modeli eğitilerek, R^2, MSE, RMSE ve MAE gibi performans ölçütleri ile değerlendirilmiştir. Sonuç olarak, bu çalışma farklı modellerin başarımlarını karşılaştırarak PM_{10} yoğunluğunu en doğru şekilde tahmin eden yaklaşımı belirlemekte ve veri odaklı hava kalitesi yönetimine katkı sağlamayı hedeflemektedir.

Anahtar Kelimeler: Hava kirliliği, makine öğrenmesi, PM_{10} tahmini, meteorolojik veriler

ABSTRACT

Prediction Of Air Pollution Using Machine Learning Methods

Taylan VALUNYA

Department of Statistics
Master of Science Thesis

Supervisor: Asst. Prof. Dr. Coşkun PARİM

Air pollution is one of the most critical environmental issues of our time, with far-reaching impacts on human health, ecosystems, and overall environmental quality. The release of harmful chemical and physical substances into the atmosphere—originating from industrial activities, urbanization, vehicle emissions, energy production, and agriculture can result in pollutant concentrations that exceed safe limits. Especially in densely populated or industrial regions, predicting air pollution levels in advance is essential for taking timely measures to protect public health and ensure environmental sustainability.

This study focuses on estimating the concentration of Particulate Matter (PM_{10}), a major air pollutant, using various machine learning techniques. The data used in the study consists of hourly observations collected between 2018 and 2020 from a monitoring station located in Yalova, a city of strategic industrial and ecological importance in Türkiye. In addition to PM_{10}, other air pollutants such as SO_2, NO, NO_2, NOx, and O_3 are included, along with meteorological variables such as air temperature, wind speed, wind direction, relative humidity, and air pressure.

The dataset is split chronologically into 70% training and 30% testing subsets. Feature selection methods are applied to determine the most relevant variables influencing PM_{10} concentrations. Multiple machine learning models are trained and evaluated using performance metrics including R^2, MSE, RMSE and MAE. The study ultimately compares these models to identify the most effective approach for accurate PM_{10} prediction, contributing to proactive air quality management strategies.

Keywords: Air pollution, machine learning, PM_{10} prediction, meteorological data

1
GİRİŞ

1.1 Literatür Taraması

Güven (2022), hava kirliliğinin makine öğrenmesi yöntemleriyle tahmin edilmesi isimli tez çalışmasında veriler Çevre, Şehircilik ve İklim Değişikliği Bakanlığı'nın Ulusal Hava Kalitesi İzleme Ağı'ndan elde edilmiş olup, tahminleme modeli kurulurken bağımsız değişken olarak kirletici maddeler, bağıl nem, hava sıcaklığı, rüzgar hızı, ve hava basıncı gibi meteorolojik veriler kullanılmış ve analiz yöntemlerinde ise, k-en yakın komşu, karar ağacı, rastgele orman ve çok katmanlı algılayıcı regresyon modeli uygulanmıştır (Güven, 2022).

Bıyık (2010), hava kirliliğinin yapay zekâ teknikleri ile belirlenmesi isimli tez çalışmasında İstanbul Büyükşehir Belediyesi Çevre Koruma Müdürlüğü'nden elde edilen veriler ile İstanbul ili için SO_2, PM_{10} ve CO kirletici parametreleri için uygun modeller oluşturulmuş ve yapay zekâ yöntemleri ile birlikte bulanık mantık modelinden yararlanılarak hava kirliliğinin tahmini yapılmıştır. Çalışmada ayrıca SO_2, PM_{10}, CO değişkenleri arasındaki korelasyon ilişkisine de bakılmıştır (Bıyık İsmet, 2010).

Kaya (2019), tarafından hazırlanan "Veri Madenciliği Yöntemleri Kullanarak Hava Kirliliği Tahmini" başlıklı tez çalışmasında, veriler bu tezde de kullanılan veri kaynağından, yani Çevre, Şehircilik ve İklim Değişikliği Bakanlığı'nın resmi internet sitesinden elde edilmiştir. Çalışmada, iki katmanlı bütünleşik yapısı, dengesiz verilerden öğrenme konusunda daha başarılı olduğu için iki katmanlı hava kirliliği tahminleme modeli kullanılmış ve PM_{10} değerleri aracılığıyla İstanbul'daki hava kirliliği seviyesi tahmin edilmiştir. Çalışmada Avrupa ve Anadolu yakasındaki toplamda dokuz bölgeye ait Ağustos 2011-Şubat 2018 yılları arasındaki veriler kulanılmış olup CO, NO, NO_2, NO_X, O_3 ve SO_2 kirleticilerinin saatlik yoğunluk

verileri meteoroloji istasyonlarına en yakın kirletici istasyonları seçilerek analiz yapılmıştır (K. Kaya, 2019).

Çelik (2022), makine öğrenmesi yöntemleri ile hava kirliliği tahmini isimli tez çalışmasında kirletici maddeler Çevre Şehircilik Bakanlığı (ÇSB)'ndan, meteorolojik veriler ise Meteoroloji Genel Müdürlüğü (MGM)'nden alınarak Zonguldak ve Manisa iline ait 2016-2021 yılları arasında ölçülen PM_{10}, SO_2, CO, NO_2, NOX, O_3 kirleticileri ile sıcaklık, basınç, yağış, rüzgâr hızı ve nem gibi meteoroloji parametrelerinden oluşan veri seti kullanılmıştır. Çalışma Naive Bayes, Destek Vektör Makineleri, Karar Ağaçları, Rastgele Orman, K-En Yakın Komşu ve Uzun Kısa Süreli Bellek yöntemleriyle modellenmiş ve Ortalama Kare Hatası (MSE), Ortalama Mutlak Hata (MAE), ve Belirlilik Katsayısı (R^2) ile modellerin başarısı çalışmanın sonunda karşılaştırılmıştır (Çelik, 2022).

Gülbağcı (2006), veri madenciliği ile hava kirliliği tahmini yüksek lisans tezi çalışmasında İ.B.B, BBC ve wheather.com'dan alınan verileri yapay sinir ağları yöntemini kullanarak SO_2, PM_{10} ve CO hava kirliliği göstergelerini tahmin eden bir sistemi İstanbul için oluşturarak bu çalışmaya sıcaklık, nem, basınç, rüzgâr yönü, rüzgâr hızı gibi meteorolojik veriler ile hava kirliliği göstergeleri arasındaki ilişkiyi daha da güçlendirerek yorumlamıştır. Ek olarak yapay sinir ağı modeli iyileştirilerek geliştirilmiş, çeşitli deneyler veri kümesi üzerinde yapılmış ve sonrasında çalışmanın sonuçları değerlendirilmiştir (Gülbağcı, 2006).

İçöz (2022), hava kirliliği ölçümü için akıllı sensör geliştirilmesi isimli tez çalışmasında hava izleme istasyonlarının sabit ve veri elde etmenin maliyetli olması sebebiyle optik tabanlı ve daha uygun maliyetli oluşturulabilecek parçacık madde sensörü tasarlanmış, üretilmiş ve ardından test edilmiş sonrasında ise sensör ağı, bulut bilişime veri aktarabilmekte ve uzaktan veri analizine de uygunluk sağlamaktadır (İçöz, 2019).

Türkoğlu (2019), New York John F. Kennedy (JFK) havalimanındaki uçuşların hava kirliliğine etkisini makine öğrenmesi yardımıyla tahmin analizi isimli tez çalışmasında JFK Havalimanı'nda gerçekleşen ticari uçuşlardan kaynaklanan NO_2 emisyonlarını tahmin etmeyi amaçlanmış olup, belirlenen lokasyonda NO_2 emisyonlarını tahminlemek için AERMOD hava dispersiyon modeli ile bir makine öğrenme modelini kullanarak sayısal modellemeyi birleştirmeyi amaçlamış ve

karar ağacı temelli ileri seviye gradyan artırma (XGBoost) yöntemi makine öğrenmesi modeli olarak kullanılmıştır (Türkoğlu, 2019).

Uğurlu (2021) Konya ilinde hava kirliliğine etki eden faktörlerin mekansal analizi isimli çalışmasında Çevre ve Şehircilik Bakanlığına ait veriler kullanılarak Konya ili sınırları içinde farklı şehirlerde hava kirliliğine neden olan SO_2, PM_{10}, CO, NO, NO_2, NO_X, $PM_{2,5}$ ve O_3 kirleticilerinin 2019-2020 yılları için mekânsal analizi yapılmıştır (Uğurlu, 2021).

1.2 Hava Kirliliği Hakkında

Hava, yeryüzünde yaşayan her canlı için hayati derecede önemli olarak atmosferde yer alan gazlardan oluşan bir gaz kütlesidir. Havada yaklaşık olarak %78 oranında azot, %21 oksijen, %1 karbondioksit ve su buharı, argon, neon, helyum, metan, kripton, hidrojen, ozon ve ksenon elementlerinden bulunmaktadır (Akbulut Zencirci, 2017).

Dünya nüfusunun hızla artması ile endüstri gelişimi, artan enerji kullanımı, insan sağlığı gibi kaçınılmaz sonuçlar ekolojik dengeyi tehdit edecek boyutlara ulaşmıştır. İnsan ve nüfusun artmasıyla yoğunlukla yapay kaynaklar dolayısıyla hava kirliliği oluşuyor olsa da orman yangınları, yanardağ faaaliyetleri ve çöl tozları gibi doğal kaynaklar da hava kirliliği oluşumunda önemli rol oynamaktadır.

Tüm ekosistemi etkisi altına alan hava kirliliğini incelemek ve bunu üzerine oluşabilecek sonuçları belirlemek üzere Hava Kalitesi İndeksi (HKİ) tanımlama sistemi oluşturulmuştur. Bu tanımlamaya göre "İyi"den "Tehlikeli"ye kadar 6 ayrı kategoride derecelendirilmiştir (Aksoy, 2023).

Hava Kalitesi İndeksi belirli hava kirleticilerinin (PM_{10}, CO, SO_2, NO_2, O_3 vb.) konsantrasyonunu ölçerek hava kalitesini sayısal birer değere dönüştürür. Bu çalışma kapsamında kullanılan hava kalitesi değerlendirmesi Tablo 1.1'de açıkça gösterilmektedir. Bu tablo, Hava Kalitesi İndeksi (HKİ) esas alınarak sınıflandırılmıştır. HKİ, atmosferdeki kirletici maddelerin insan sağlığı üzerindeki olası etkilerini göstermek amacıyla geliştirilen sayısal bir ölçüttür. Tabloya göre, HKİ değeri 0–50 arasında olduğunda hava kalitesi "İyi" olarak kabul edilir ve bu aralıkta halk sağlığı açısından herhangi bir risk bulunmamaktadır. 51–100 aralığı "Orta" olarak sınıflandırılır; bu seviyede hava kalitesi genel anlamda kabul

edilebilir olmakla birlikte, özellikle astım gibi solunum sorunları olan hassas bireyler için sınırlı düzeyde risk oluşturabilir.

HKİ değeri 101–150 olduğunda hava kalitesi “Hassas” kategorisine girer ve bu durum özellikle çocuklar, yaşlılar ve solunum rahatsızlığı olan bireyler gibi duyarlı gruplar için sağlık riski oluşturabilir. 151–200 aralığındaki “Sağlıksız” düzeyinde ise hem hassas gruplarda hem de genel nüfusta sağlık etkileri gözlemlenebilir. HKİ’nin 201–300 arasında olduğu “Kötü” kategorisi, hava kalitesinin halk sağlığı açısından belirgin bir tehdit oluşturduğunu ve özellikle açık hava aktivitelerinin kısıtlanması gerektiğini gösterir. Son olarak, 301–500 aralığı “Tehlikeli” olarak tanımlanmakta ve bu durumda tüm bireylerin etkilenme olasılığı yüksek olduğundan ciddi önlemlerin alınması gerekmektedir.

Her bir indeks aralığına karşılık gelen renkler daha kolay anlatılabilmesi açısından sırasıyla yeşil, sarı, turuncu, kırmızı, mor ve koyu kahverengi olmak üzere hava kalitesi hakkında hızlı bir şekilde bilgilendirme yapmayı amaçlamaktadır.

Tablo 1.2). Bu kirleticilerin değişikliklerinin kaydını alan ölçüm cihazlarına ek, formüllerle birlikte HKİ değerlerine dönüşmektedir.

Amerika Birleşik Devletleri Çevre Koruma Ajansı (U.S. Environmental Protection Agency (USEPA)) tarafından benimsenen ilk yapay indekslerden biri, Ott ve Hunt tarafından geliştirilen Kirlilik Standart İndeksi (KSİ) olduğu söylenebilir (Ott & Hunt, 1976). 1999'da EPA, Hava Kalitesi İndeksi'ni benimsemiştir. Bu yeni indeks, yer seviyesinde ozon ve ince partikül gibi iki yeni alt endeks içermektedir (EPA, 2023).

EPA’nın 1999 yılında yaptığı çalışmada HKİ, belirli değerler arasında belirli değerlere karşılık gelecek şekilde ifade edilmektedir ve bu seviyede insan sağlığında oluşacak riskleri Tablo 1.1’de göstermiştir.

Tablo 1.1 Hava kalitesi renk skalası

İndeks	HKİ	Hava Kalitesi Açıklaması	Renk
İyi	0 – 50	Risk barındırmamaktadır.	Yeşil
Orta	51 – 100	Bu hava kalitesi kabul edilebilir ancak hassas insanlar için risk barındırabilecek seviyededir.	Sarı
Hassas	101 – 150	Hassas gruplar için risk oluşturabilir ancak genel halkın etkilenme olasılığı düşüktür	Turuncu
Sağlıksız	151 – 200	Sağlıkla ilgili etkileri olabilir.	Kırmızı
Kötü	201 – 300	Sağlık etkileri riski herkes için artar.	Mor
Tehlikeli	301 – 500	Herkesin etkilenme olasılığı daha yüksektir.	Koyu Kahverengi

Bu çalışma kapsamında kullanılan hava kalitesi değerlendirmesi Tablo 1.1'de açıkça gösterilmektedir. Bu tablo, Hava Kalitesi İndeksi (HKİ) esas alınarak sınıflandırılmıştır. HKİ, atmosferdeki kirletici maddelerin insan sağlığı üzerindeki olası etkilerini göstermek amacıyla geliştirilen sayısal bir ölçüttür. Tabloya göre, HKİ değeri 0–50 arasında olduğunda hava kalitesi "İyi" olarak kabul edilir ve bu aralıkta halk sağlığı açısından herhangi bir risk bulunmamaktadır. 51–100 aralığı "Orta" olarak sınıflandırılır; bu seviyede hava kalitesi genel anlamda kabul edilebilir olmakla birlikte, özellikle astım gibi solunum sorunları olan hassas bireyler için sınırlı düzeyde risk oluşturabilir.

HKİ değeri 101–150 olduğunda hava kalitesi "Hassas" kategorisine girer ve bu durum özellikle çocuklar, yaşlılar ve solunum rahatsızlığı olan bireyler gibi duyarlı gruplar için sağlık riski oluşturabilir. 151–200 aralığındaki "Sağlıksız" düzeyinde ise hem hassas gruplarda hem de genel nüfusta sağlık etkileri gözlemlenebilir. HKİ'nin 201–300 arasında olduğu "Kötü" kategorisi, hava kalitesinin halk sağlığı açısından belirgin bir tehdit oluşturduğunu ve özellikle açık hava aktivitelerinin kısıtlanması gerektiğini gösterir. Son olarak, 301–500 aralığı "Tehlikeli" olarak tanımlanmakta ve bu durumda tüm bireylerin etkilenme olasılığı yüksek olduğundan ciddi önlemlerin alınması gerekmektedir.

Her bir indeks aralığına karşılık gelen renkler daha kolay anlatılabilmesi açısından sırasıyla yeşil, sarı, turuncu, kırmızı, mor ve koyu kahverengi olmak üzere hava kalitesi hakkında hızlı bir şekilde bilgilendirme yapmayı amaçlamaktadır.

Tablo 1.2 Ulusal hava kalitesi indeksi kesme noktaları (Hava Kalitesi Raporu - Sürekli İzleme Merkezi, 2024)

İndeks	HKİ	SO2 [µg/m³]	NO2 [µg/m³]	CO [µg/m³]	O3 [µg/m³]	PM10 [µg/m³]
		1 Sa. Ort.	1 Sa. Ort.	8 Sa. Ort.	8 Sa. Ort.	24 Sa. Ort.
İyi	0 – 50	0-100	0-100	0-5500	0-120	0-50
Orta	51 – 100	101-250	101-200	5501-10000	121-160	51-100
Hassas	101 – 150	251-500	201-500	10001-16000	161-180	101-260
Sağlıksız	151 – 200	501-850	501-1000	16001-24000	181-240	261-400
Kötü	201 – 300	851-1100	1001-2000	24001-32000	241-700	401-520
Tehlikeli	301 – 500	>1101	>2001	>32001	>701	>521

Hava Kalitesi İndeksi (HKİ) aralıklarına göre çeşitli hava kirleticilerin (SO_2, NO_2, CO, O_3 ve PM_{10}) konsantrasyon eşikleri Tablo 2.1'de verilmektedir. Her kirletici için, ilgili saatlik veya günlük ortalama sınır değerleri HKİ kategorilerine karşılık gelecek şekilde düzenlenmiştir. "İyi" kategorisinde tüm kirleticilerin konsantrasyonları insan sağlığı açısından güvenli sınırlar içerisindedir. HKİ değeri yükseldikçe kirletici yoğunlukları da artmakta, bu durum özellikle hassas gruplar başta olmak üzere halk sağlığı açısından giderek artan riskler oluşturmaktadır. "Orta" ve "Hassas" kategorileri, belirli gruplar için etkiler yaratabilecekken; "Sağlıksız", "Kötü" ve "Tehlikeli" seviyeleri, tüm bireyler için sağlık riski oluşturan durumlara işaret etmektedir. Kirletici sınır değerlerinin saatlik ya da sekiz/dokuz saatlik ortalamalarla sunulması, hava kalitesinin kısa vadeli değişimlerini ve bu değişimlerin sağlık üzerindeki potansiyel etkilerini değerlendirmeye olanak tanımaktadır.



Şekil 1.1 Ulusal hava kalitesi PM_{10} değer grupları (Hava Kalitesi Raporu - Sürekli İzleme Merkezi, 2024).

Şekilde 1.1'de PM_{10} konsantrasyon değerlerine göre hava kalitesi sınıflandırması görsel olarak sunulmuştur. Her bir daire, farklı bir hava kalitesi seviyesini ve bu seviyeye karşılık gelen PM_{10} aralığını temsil etmektedir. Renk skalası yeşilden koyu kahverengiye doğru ilerleyerek hava kalitesinin kötüleştiğini ve sağlık riskinin arttığını göstermektedir. Bu görsel, PM_{10} değerlerine göre hava kirliliğinin halk sağlığı üzerindeki etkilerini sade ve anlaşılır biçimde yansıtmaktadır.

1.3 Hava Kirliliği Değişkenleri

Hava kirliliğini ölçmek için birçok farklı parametre kullanılabilir. Bu parametrelerden akademik çalışmalarda kullanılanlardan en yaygın olanları; partikül maddeler (PM_{10}, ve $PM_{2,5}$), Azot Dioksit (NO_2), Kükürt dioksit (SO_2), Karbon Monoksit (CO), Ozon (O_3) ve uçucu organik bileşikler (UOB'ler, ingilizce ise VOC'ler) olarak adlandırılır.

Bu değişkenlere ek olarak bazı akademik çalışmalarda hava sıcaklığı, bağıl nem, rüzgar yönü ve rüzgar şiddeti gibi değişkenlerin de eklendiği gözlemlenmiştir.

Tablo 1.3 Hava kirliliği değişkenleri

Hava Kirleticileri	Ana Kaynakları	Sağlığa Etkisi
Partikül madde (PM)	Endüstriyel kaynaklar,taşıt emisyonları,yakıt kullanımı,tarım ve ikincil kimyasal reaksiyonlar	Kanser ,kalp problemleri,solunum yolu hastalıkları,bebek ölüm oranlarında artış
Azot oksitler (NO_X)	Taşıt emisyonları ve yakıt kullanımı ve yüksek sıcaklıkta yakma prosesleri	Göz ve solunum yolu hastalıkları,asit yağmurları
Kükürt dioksit (SO_2)	Taşıt emisyonları ve yakıt kullanımı	Solunum yolu hastalıkları,asit yağmurları
Karbon monoksit (CO)	Taşıt emisyonları	Kandaki hemoglobin ile birleşerek oksijen taşıma kapasitesinde azalma ve ölüm
Ozon (O_3)	Trafikten kaynaklanan azot oksitler ve uçucu organik bileşiklerin (UOB) güneş ışığıyla değişimi	Solunum sistemi problemleri,göz ve burunda irritasyon,astım,vücut direncinde azalma
Uçucu Organik Bileşikler (UOB)	Motorlu taşıtlar yakıt kullanımı,temizlik, yapı ve kozmetik ürünler,doğal kaynaklar	Solunum yolu hastalıkları,nörolojik etkiler,organlarda hasar ve yetmezlikler

Tablo 1.3'te hava kirliliğini gösteren değişkenlerin isimleri, nereden kaynaklandığını ve sağlığa etkilerinin neler olduğu açıkça belirtilmiştir.

1.3.1 Partikül Madde (PM_{10}, $PM_{2,5}$)

Partikül maddeler, maddelerin organik ve inorganik karşımı ile meydana gelebilen ve havada asılı kalabilen maddelerdir. Katı ve sıvı halde doğada bulunabilirler. Aerodinamik çapı 10 μm'den daha küçük partiküller PM_{10} olarak adlandırılırken, aerodinamik çapı 2,5 μm'den daha küçük partiküller $PM_{2,5}$ olarak adlandırılır.

1.3.1.1 Partikül maddelerin (PM_{10} ve $PM_{2,5}$) insan sağlığına etkileri

Partikül büyüklüklerinin insan sağlığına doğrudan ilişkisi mevcuttur. Özellikle Aerodinamik çapı 10 μm'den daha küçük partiküller alt solunum yollarına kadar ulaşabildiğinden çok daha ciddi sağlık sorunlarına sebep olabilmektedir. Bu durum doğrudan akciğer ve kalp hastalığına sebebiyet verebilir. Ayrıca, kalp ritim düzensizliği, akciğer fonksiyonunda azalma, nefes almada güçlük ve dolaylı birçok hastalığın temel risk grubu oluşturan çocuklar ve yaşlıları yoğunlukla etkileyebilir.

Partikül maddeye dayalı yapılan analizler, hava kirliliğinin dünya çapında yaklaşık olarak nefes borusu, bronş ve akciğer kanserinin %5'ini, yetersiz oksijene bağlı ölümlerin %2'sini ve solunum enfeksiyonları ölümlerinin yaklaşık olarak %1'ini oluşturduğunu tahmin etmektedir. Bu, yaklaşık olarak 0.8 milyon (%1.4) ölüm ve 7.9 milyon (%0.8) sakatlığa ayarlanmış yaşam yılına (ingilizcede disability adjusted life year daly olarak ifade edilmektedir) karşılık gelmektedir. Bu yük çoğunlukla gelişmekte olan ülkelerde meydana gelmektedir. Partikül maddeye

bağlı sebepten yaşam yıllarının azalma durumu %42'si “Batı Pasifik Bölgesi’nde” ve %19'u “Güneydoğu Asya Bölgesi’nde” meydana gelmektedir. Alt bölgeler içinde, toplam yükün en yüksek oranları Batı Pasifik Bölgesi, ve Avrupa bölgelerinin alt bölümlerinde meydana gelmektedir. Burada hava kirliliği hastalık yükünün %0.6-1.4'ünü oluşturmaktadır. Bu tahminler, hava kirliliğinin sadece ölümler üzerindeki etkisini dikkate alır ve epidemiolojik veritabanındaki sınırlamalar nedeniyle ölüm oranını içermez. Eğer hava kirliliği hem hastalık sıklığını hem de ölümleri aynı oranda artırıyorsa, hastalık oranı daha fazla olacaktır (WHO, 2002).

Farklı bir şekilde yine hava kirliliğinin Avrupa’da bir örneğine dikkat çekilecek olursa; 2020'de Avrupa Birliği'nde kentsel nüfusun %96'sı, Dünya Sağlık Örgütü tarafından belirlenen sağlık tabanlı kılavuz seviyesinin üzerinde ince partikül madde seviyelerine maruz kalmıştır (EEA, 2022).

1.3.1.2 Partikül maddelerin (PM_{10} ve $PM_{2,5}$) çevresel etkileri

Özellikle büyük şehirlerde $PM_{2,5}$ partikül maddesinin çevrede görsel olarak yansıması yoğun pus ve sis olarak gözlemlenmektedir. Partiküller rüzgârın da etkisi ile, uzun mesafeler boyunca taşınabilir ve daha sonra yere veya suya çökebilir. Bu çökelmenin etkisi ile birlikte bölgede; toprak, sulak alanlar, ormanlardaki besin dengesi değişir, asitleşir, bu da başta tüm ekosistem olmak üzere, tarım ürünlerini etkiler dolaylı olarak tüketimi ve insan sağlığını etkiler (EPA, 2023).

Kentsel hava kirliliği büyük ölçüde artarak, ulaşım, enerji üretimi ve diğer insan faaliyetleri için fosil yakıtların yanması sonucunda ortaya çıkmaktadır. Yanma süreçleri, hem dizel ısıtma parçacıkları ve kurşun gibi birincil emisyonları hem de kükürt içeren yakıtların yanmasından oluşan ozon ve sülfat parçacıkları gibi atmosferik dönüşüm ürünlerini içeren karmaşık bir kirletici karışımı üretir (WHO, 2002).

1.3.2 Azot Dioksit (NO_2)

Nitrik oksit (NO) ve azot dioksit (NO_2) olmak üzere iki temel bileşenden oluşan azot oksitleri (NOx), özellikle motorlu taşıtlar ve güç santralleri gibi yüksek sıcaklıklı yanma süreçlerinde oluşur (Atkinson, 2000).

Azot oksitler, orman yangınları, topraktaki organik madde ayrışması, mikrobiyal işlemler, yıldırım gibi doğal süreçlerle atmosfere salınabilir.

1.3.2.1 Azot Dioksit (NO_2) insan sağlığına etkileri

Azot oksitler, kandaki hemoglobin ile etkileşime girerek kompleksler oluşturabilmektedirler. Bununla birlikte, akciğerdeki nem ile bir araya geldiğinde nitrik asitin oluşumu ile doğrudan solunum sistemi hastalıklarına sebebiyet verebilmektedir.

1.3.2.2 Azot Dioksit (NO_2) çevresel etkileri

En belirgin çevresel etkisi, atmosferik parçacıkların oluşumuna yol açmasıdır. Hava ve su kirliliği ile birlikte asit yağmurlarının oluşumuna direk etki sağlarken, küresel ısınma sürecine ciddi katkısı bulunmaktadır (E. Kaya, 2007).

Dolayısıyla, motorlu araçların yakıtları ve güç santrallerinin kullanımları dolayısıyla bu bileşiklerin salınımının ve konsantrasyonunun azaltılması kritik öneme sahiptir.

1.3.3 Kükürt Dioksit (SO_2)

Kükürt dioksit (SO_2); renksiz, keskin kokulu, yanıcı olmayan, zehirli bir gazdır. Bu gazın suda çözünmesi dolayısıyla, alt solunum yollarına ulaşmadan büyük oranda solunum bölgesinde elimine edildiği bilinmektedir.

Kükürt, genellikle katı yakıtların bileşiminde organik ve inorganik formlarda bulunmaktadır. İnorganik kükürt, toprakta sülfat şeklinde bulunurken, organik kükürt ise yakıtlardaki karbonun yanması sonucu oluşan SO_2 olarak atmosfere salınmaktadır. Hava kirliliği emisyonlarının en yaygın olanı, kükürt oksitleridir ve bunun sonucunda kükürtdioksit (SO_2) atmosfere salınır (İlkılıç & Behçet, 2006). Ayrıca, yoğunlukla çeşitli maden cevherlerinde, petrol rafinerilerinde, metal üretim alanlarında gerçekleşen işlemler sonucu atmosfere salınmaktadır. Ek olarak orman yangınları ve volkanik faaliyetler gibi doğal süreçlerde de salınabilmektedir. Kükürt dioksitin suda çözünürlüğü sebebiyle atmosferdeki nem ile çözünerek, güneş ışığı ve bazı kimyasalların da etkisiyle sülfürik asite sebep olur, bunun da asit yağmurları oluşumuna önemli bir katkı sağlar, dolaylı olarak insan sağlığı ve çevreye önemli zararlara yol açmaktadır (Sağlık Bakanlığı Temel Sağlık Hizmetleri Genel Müdürlüğü, 2010).

Kükürt dioksit ve atmosferde bulunan diğer kükürt bileşikleri, yüksek oranlarda solunum yoluyla alındığında, kükürt dioksitin büyük bir kısmı üst solunum yollarından emilir. Bu durum, bronşit gibi solunum yolu hastalıklarına neden olabilir. İnsanlar, SO_2 gazını soluduklarında, akciğerlerde bulunan nemle birleşerek H_2SO_3 ve daha sonra H_2SO_4 asidine dönüşür. Bu asit, solunum yollarını tahriş eder ve solunum yolu ile ilgili hastalıklara yol açabilir. SO_2 gazının yoğunluğu ve maruziyet süresi, solunum organlarına verdiği zararın şiddetini belirler ve hatta ölümcül olabilir (İlkılıç & Behçet, 2006).

1.3.4 Karbon Monoksit (CO)

Karbon monoksit (CO); renksiz, kokusuz ve oldukça zehirli bir gazdır. Benzin, doğal gaz ve odun gibi karbon temelli yakıtların eksik yanması sonucunda oluşur (Grahame & Schlesinger, 2009).

Ülkemizde karbon monoksit zehirlenmelerinin en fazla kış aylarında gerçekleştği görülmektedir, bunun sebebi ise kışın ısınmada kullanılan yakıtlardan çıkan gazlardır. Mevsim karşılaştırması yapıldığında ise; en fazla kış (%59.5) ikinci olarak ilkbahar (%22.3) başvuru olurken aylara göre bakıldığında en çok Ocak (%22.9), Aralık (%18.42) ve Şubat (%18.1) aylarında görüldüğü saptanmıştır (GİNİŞ ve ark., 2014).

1.3.5 Ozon (O_3)

Ozon (O_3) moleküler oksijenin (O_2) üç atomlu bir şeklidir. Keskin kokulu, zehirli, soluk mavi, kararsız bir gazdır. Ozon, özellikle stratosferde dünya yüzeyinden 19 ile 30 km uzaklıktadır (Maleki ve ark., 2019).

Azot oksitleri (NOx) ve uçucu organik bileşikler (İngilizce volatile organic compound, kısaca VOC'ler), atmosferde güneş ışınlarının etkisiyle ozon oluşumuna yol açabilir. Geçmişte, bu maddelerin salınımını kontrol altına almak için hava kalitesi düzenlemeleri yapılmadan önce, ozon seviyeleri çok yüksek seyretmekteydi ve bu da gözlerde ve burunda tahriş, solunum hastalıkları gibi acil sağlık problemleri ve akciğer fonksiyonu bozulması gibi ciddi etkilere neden olmaktaydı. Günümüzde ise ozon seviyelerindeki günlük dalgalanmaların, astım vakaları ve kaza dışı ölümlerle ilişkili olduğu tespit edilmiştir. Yapılan yeni araştırmalar, ozon seviyelerinin mevcut düzenleyici sınırların altında kalsa bile hem kısa hem de uzun vadeli maruz kalmanın solunum ve kalp hastalıklarından

kaynaklanan ölümleri artırabileceğini göstermektedir. Ozon ile bu sağlık sorunları arasındaki bağlantıyı açıklamak için, ozonun güçlü bir oksidan olduğunu ve hücrelere, hava yollarındaki sıvılara zarar vererek bağışıklık sistemini ve iltihaplanma yanıtlarını tetiklediğini anlamak gerekir. Bu özellikleri, ozonun sağlığımız üzerinde yarattığı olumsuz etkilerin temelini oluşturur (Zhang ve ark., 2019).

1.3.6 Uçucu Organik Bileşikler (UOB)

Yapısında en az bir karbon ve hidrojen atomu bulunduran kimyasal bileşikler organik bileşikler olarak adlandırılırlar. Bu organik bileşikler, uçucu organik bileşikler, yarı uçucu organik bileşikler ve uçucu olmayan organik bileşikler olmak üzere üç ana grupta ayrılmaktadırlar (Anderson ve ark., 2017).

2
VERİ SETİ

2.1 Veri Seti

Hava kirliliği tahmin çalışması için Yalova ilinde yer alan istasyondan alınan veriler kullanılmıştır. Çalışmada yer alan kirletici maddeler ve meteorolojik göstergelere ait veriler, T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı resmi internet sitesinden temin edilmiştir. Veri seti, 01.01.2018 - 30.09.2020 yılları arasında 2 yıllık PM_{10}, SO_2, CO, NO_2, NO_X, ve O_3 parametlerine ilişkin günlük ortalama değerleri (µg/m3) ve rüzgâr yönü (o), rüzgâr hızı (km/h), hava sıcaklığı (°C), basınç ve bağıl nem değerlerini kapsamaktadır. Özellikle yıl anlamında 2018-2020 arasının kullanılması, bu tarihten sonra günlük meteorolojik verilerin sağlanamamasından kaynaklıdır. Veri setinde bulunan eksik gözlemler saatlik ortalama değerler kullanılarak atanmış ve eksik gözlemler bu şekilde giderilmiştir.

Şekil 2.1 Yalova ilinin coğrafi konumunu göstermektedir. Ayrıca; Yalova, 1995 yılında il statüsü kazanmıştır. Türkiye'nin kuzeybatısında, Marmara Bölgesi'nin güneydoğu kısmında konumlanan il; Armutlu Yarımadası'nın kuzey kıyısı ile Samanlı Dağları'nın kuzey etekleri arasına kurulmuştur. Yalova, 28° 45' ile 29° 35' doğu boylamları ve 40° 28' ile 40° 45' kuzey enlemleri arasında yer alır. Karasal komşuları Bursa ve Kocaeli illeri olmakla birlikte, deniz komşusu olarak da İstanbul bulunmaktadır. Bu yanıyla, Türkiye'nin önemli coğrafi bölgelerinden birinde yer almaktadır.



Şekil 2.1 Yalova il sınırları harita görünümü (Harita Genel Müdürlüğü, t.y.)

2.1.1 Model

Yukarıdaki aşamalarda da belirttiğimiz üzere modelimizin kurgusu aşağıdaki şekilde olacak ve bağımsız değişkenlerimizle birlikte değerini tahmin etmek için modelimizi aşağıdaki şekilde oluşturuyoruz.

$PM_{10} \sim NOx + SO_2 + O_3 +$ Bağıl Nem+ Rüzgar Yönü + Hava sıcaklığı + Rüzgar Hızı

3
YÖNTEMLER

Verilerin analiz edilmesi için farklı “Makine Öğrenmesi Yöntemleri” kullanılmış olup, “Makine Öğrenmesi”, mevcut verilere dayanarak sonuçlar çıkarmak için istatistiksel ve matematiksel yöntemler kullanan ve bu sonuçları algoritmalar ve modelleme kullanarak geleceğe dair tahminler yapmak için kullanan bir yapay zekâ alanıdır (Çağlayan Akay, 2018).

Makine öğrenmesi; (Machine Learning kısaca ML) insan zekasına benzeyen algoritmalar üreten bilgisayar biliminin bir alt dalıdır. Bu algoritmalar istenilen ürünü üretmek için insan müdahalesi olmadan yapılandırılmış veriden beslenebilir (Andriy Burkov, 2019).

Makine öğrenmesi modellerinin geliştirilmesinde kullanılan algoritmaların seçimi kullanılan versi setine bağlı olup bazı algoritmalarda doğruluk çok yüksek çıkabildiği gibi bazı algoritmalarda ise tahmin doğruluğu düşük çıkabilir. Aynı verisetlerinde dahi farklı algoritmalar kendi yapıları gereği farklı doğruluk oranları ile tahminleme yapabilir dolayısıyla analiz edenler birden fazla makine öğrenmesi yöntemi kullanarak bu algoritmaların tahminleme sonuçlarını karşılaştırabilir ve bunun sonucunda tahmin doğruluğu en yüksek algoritmayı seçerek sonuca daha yüksek değerlerle ulaşmayı amaçlamaktadır (Hasan ve ark., 2024).

Makine öğrenmesi yöntemini oluşturan algoritmalar 1959 yılında bilgisayar biliminin bir alt grubu olarak doğmuştur. Bu tür algoritmalar, tahminlerde bulunmak ve verileri kullanarak kararlar almak amacıyla örnek modeller üreterek çalışır ve makinelerin insanların öğrenme sistemini kurgulayarak çalışabileceğinin mümkün olduğu düşüncesine dayanır (Zilyas & Yılmaz, 2023).

Bu yöntem ilk kez 1959 senesinde “Arthur Lee Samuel” tarafından IBM için geliştirilen dama oyununu tasarlanırken bu terimi ilk kez kullanılmıştır (Mockenhaupt & Schlagenhauf, 2024).

3.1 Makine Öğrenmesi Yöntemleri

Makine öğrenmesi yöntemleri, genel olarak birçok kaynakta ikiye ayrılmıştır, bazı kaynaklarda birden fazla yöntemin bulunması sonradan diğer yöntemler de eklendiği içindir.

Makine öğrenmesi denetimli ve denetimsiz olarak ikiye ayrılır. Yarı denetimli ve pekiştirmeli öğrenme gibi yöntemler sonradan eklenmiştir (Koyuncugil & Özgülbaş, 2009).

Denetimli öğrenmede, bir dizi girdi değişkeninden yararlanılarak çıktı değerini tahmin etmek amaçlanır (Hastie ve ark., 2009).

Her bir gözlem için, yani her bir tahminleyici ölçümü x_i (i = 1, ..., n) ile ilişkili bir yanıt ölçümü y_i bulunur. Amaç, yanıtı tahminlerle (gözlemlerden elde edilen verilere) bağlayan bir model oluşturmak ve bu modeli kullanarak gelecekteki gözlemler için doğru tahminler yapmak (tahmin) veya yanıt ile tahminlerle arasındaki ilişkiyi daha iyi anlamak olacaktır (James ve ark., 2021).

Eğitim verilerinin hem girdi vektörlerini hem de bunlara karşılık gelen hedef vektörlerini içerdiği problemlere denetimli öğrenme problemleri denir. Her bir giriş vektörünün belirli bir sayıda ayrık kategoriye atanmasının amaçlandığı, örneğin rakam tanıma gibi durumlar, sınıflandırma problemleri olarak adlandırılır. Eğer istenen çıktı bir veya birden fazla sürekli değişkenden oluşuyorsa, bu görev regresyon olarak adlandırılır. Örnek olarak girdi faktörlerinden bazılarının sıcaklık ve basınçla birlikte tepkime maddelerinin konsantrasyonu olacağı bir kimyasal üretim süreci için verimi tahmin etmek olabilir (Bishop, 2006).

Denetimli öğrenmenin aksine, her giriş için istenen çıktı belirtilmez. Bunun yerine, görev yoğunluk tahmini olarak formalize edilir. Denetimsiz öğrenme, insan ve hayvan öğrenmesiyle daha tipik bir şekilde karşılaştırılabilir. Ayrıca, denetimli öğrenmeye göre daha geniş bir uygulama alanına sahiptir çünkü verileri manuel olarak etiketlemek için bir insan uzmanına ihtiyaç duyulmaz. Etiketli veriler yalnızca edinmesi pahalı olmakla kalmaz, aynı zamanda nispeten az bilgi içerir; kesinlikle karmaşık modellerin parametrelerini güvenilir bir şekilde tahmin etmek için yeterli değildir (Murphy, 2012).

Denetimsiz öğrenmeye özellikle etiketli veri elde etmenin zor, sınırlı olduğu ya da bulunmadığı durumlarda başvurulmaktadır.

Yarı denetimli öğrenme ise makine öğrenmesi yöntemleri arasında denetimli öğrenme ve denetimsiz öğrenme arasındaki bir yaklaşımdır. Etiketlenmiş veriler kısıtlı bulunduğu ve etiketlenmemiş veriler zengin bulunduğu durumlarda bu yöntem uygulanabilir (van Engelen & Hoos, 2020).

Pekiştirmeli öğrenme, bir problem ve bu problemlere uygun çözüm yöntemleri ve bu çözümleri inceleyen bir alan olarak tanımlanabilir. Pekiştirmeli öğrenme yöntemleri, bir durumu eyleme dönüştürmeyi ve sayısal bir çözüm elde etmeyi öğrenmeyi içerir. Bu tür problemler, eylemlerin çıktılarının sonraki durumları etkileyebileceği kapalı bir döngüye sahiptir. Uygulayıcı, hangi eylemleri yapması gerektiğini doğrudan öğrenmez, bunun yerine denemeler yaparak hangi eylemlerin daha yüksek çıktılar sağladığını keşfeder. Pekiştirmeli öğrenmenin en önemli özellikleri; kapalı döngü olması, doğrudan talimat olmaksızın eylemleri keşfetmesi ve eylemlerin sonuçlarının zaman içinde etkili olmasıdır (Sutton & Barto, 2014). Bu tez çalışmada aşağıda verilen tüm makine algoritması yöntemleri kullanılmasa da, bilgi vermek amacıyla en çok kullanılan yöntemlerden bahsedilmiştir.

3.1.1 Lineer Regresyon

Lineer regresyon, bir bağımlı değişken ile bir veya birden fazla bağımsız değişken arasındaki ilişkiyi incelemek ve bu ilişki doğrultusunda bağımlı değişkeni tahmin etmek amacıyla kullanılan bir analiz yöntemidir. Bu yöntem, değişkenler arasındaki ilişkiyi doğrusal (lineer) bir denklemle açıklar. Amaç, bağımsız değişken veya değişkenler yardımıyla bağımlı değişkenin en uygun şekilde modellenmesidir ve analiz edilebilmesidir (Memnun, 2013).

3.1.2 K-En Yakın Komşu Algoritması

K-en yakın komşu İngilizce'de kNN (K-Nearest Neighbors) olarak anılır. Sınıflandırma, regresyon görevleri için güçlü ve sezgisel bir veri madenciliği modelidir. Yeni nesneleri eğitim verilerindeki bilinen nesnelere olan benzerliklerine göre sınıflandırır. Parametrik modellerin aksine, kNN parametrik olmayan bir yöntemdir ve veri dağılımları hakkında varsayımlara dayanmaz. kNN algoritması, yeni gözlem ile mevcut veri noktaları arasındaki mesafeleri hesaplar. En küçük mesafelere sahip k, en yakın komşu seçilir ve bunların çoğunluk sınıfı

veya ortalama değeri sınıflandırma veya regresyon için kullanılır. kNN'nin avantajı, basit olması ve büyük verilerde verimli şekilde işleyebilme yeteneğidir. Dezavantajlarından biri, her yeni gözlemin sınıflandırılması gerektiğinde tüm eğitim verilerini taraması gerektiğinden, büyük veri setleri için zaman alıcı olabilmesidir. kNN, farklı hedef değişkenler ve dağılımlar için esnek ve sağlam olmasına rağmen, büyük değerlere sahip değişkenlerden kaynaklanan yanlılığı önlemek için bağımsız değişkenlerin standartlaştırılmasını gerektirir. Ayrıca, yüksek boyutlu uzaylarda performansın artan seyrekliğe bağlı olarak düşmesi anlamına gelen "boyutlanma laneti"nden muzdariptir. Sınırlamalarına rağmen, kNN, doğrusal olmayan ilişkilerle ve veri hakkında katı varsayımların bulunmadığı durumlarla başa çıkmada değerli bir araç olmaya devam etmektedir. Dikkatli veri ön işleme ve k değerinin optimize edilmesi, çeşitli uygulamalarda performansını arttırabilir (Acito, 2023).

3.1.3 Destek Vektör Makine Regresyonu

Destek vektör makineleri (Support Vector Regression kısaca SVM) gerçek değerli fonksiyonların tahmininde etkili bir yöntem olarak kabul edilmektedir. Yüksek ve düşük yanlış tahminleri eşit şekilde cezalandıran simetrik bir kayıp fonksiyonu kullanarak eğitilmektedir. Tahmin edilen fonksiyonun etrafında minimum yarıçaplı esnek bir tüp simetrik olarak oluşturulmakta ve belirli bir eşik değeri epsilon'dan küçük olan hata mutlak değerleri hem tahminin üstünde hem de altında göz ardı edilmektedir. Bu şekilde, tüpün dışındaki noktalar cezalandırılırken, tüp içinde kalan noktalar, fonksiyonun üstünde veya altında olsun, ceza almamaktadır. SVR'nin başlıca avantajlarından biri, hesaplama karmaşıklığının giriş uzayının boyutuna bağlı olmamasıdır. Ayrıca, yüksek tahmin doğruluğuyla en iyi genelleme yeteneğine sahip olduğu da görülmektedir (Awad & Khanna, 2015).

3.1.4 Karar Ağacı Regresyonu

Regresyon ağacı yöntemi bağımlı ve bağımsız değişkenler arasındaki ilişkiyi saptayarak veri setini homojen alt kümelere bölebilen bir analiz yöntemidir (Kayri & Boysan, 2008).

Bu yöntemde bağımlı ve bağımsız değişkenler arasında ilişkinin incelenmesinin yanısıra bağımsız değişkenlerin birbirleri arasındaki ilişki de incelenmektedir.

Regresyon ağacı yönteminin tercih edilen yöntemlerden biri olmasının sebebi daha geniş veri setlerine uygulanabilir olması, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tespit etmesi ve görsel açıdan da daha kolay anlaşılabilir olması denilebilir.

3.1.5 Rastgele Orman Regresyonu

Rastgele orman, temelinde regresyon ağaçları ve sınıflandırma kullanarak esnek yapıda kullanımı kolay ancak verimli makine öğrenmesi yöntemlerinden biridir. Hem sınıflandırma hem de regresyon problemlerinde kullanılabilmesi sebebiyle tercih edilen yöntemlerden birisidir. Büyük veri setlerinde yararlı olan ancak saklı yapıdaki bilgileri ayırt edebilmek için tercih edilen yöntemlerden biri olmuştur. Bu algoritma çalıştırıldığında eğitim seti (T), algoritma özellikleri (F) ve ormandaki ağaç sayısı (K) algoritma tarafından okunur ancak ön yükleme işlemi gerçekleştirilmeden önce (P) boş küme şeklindedir. Daha sonrasında veri setine göre fonksiyon alt küme oluşturmaya ve olasılıkları üretmeye başlar ve olasılıklar da olasılık kümelerine aktarılarak süreç tamamlanır. Sonrasında alt kümeler arasında en küçük olan (f) bulunur, en küçük alt küme en iyi özelliğe göre bölünür ve algoritmanın sonucu döndürülür (Kumral ve ark., 2022).

3.1.6 Ekstra Ağaçlar Regresyonu

Ekstra ağaçlar (Extra-Trees, Extremely Randomized Trees) algoritması, klasik yukarıdan aşağıya yaklaşımına göre budanmamış karar ya da regresyon ağaçlarından oluşan bir topluluk oluşturur. Bu algoritmanın diğer ağaç tabanlı topluluk yöntemlerinden iki temel farkı vardır: Birincisi, düğümleri ayırırken kesim noktalarını tamamen rastgele seçmesidir; ikincisi ise, ağaçları büyütmek için tüm öğrenme örneklemini (bootstrap örneklemesi yerine) kullanmasıdır (Geurts ve ark., 2006).

3.1.7 Bagging Karar Ağacı Regresyonu

Bagging Karar Ağacı **(**Bootstrap Aggregating) yöntemi, makine öğrenmesinde sınıflandırma ve regresyon problemlerinde model varyansını azaltmak için kullanılan bütünleşik bir yöntemdir. Bu yaklaşım, eğitim verilerinin farklı önyüklemeli örnekleri üzerinde eğitilerek aynı modelin birden fazla versiyonunu oluşturur. Bu modellerin çıktıları daha sonra birleştirilir ve sınıflandırma problemleri için çoğunluk oyu, regresyon problemleri için ise ortalama oy alınır.

Bagging yönteminin temel amacı, özellikle büyük varyansa sahip olan karar ağaçları gibi modellerin genelleme hatasını azaltmaktır. Her model, orijinal veri kümesinden tekrarlı örnekleme yoluyla elde edilen farklı bir alt küme üzerinde eğitilir. Bu şekilde modeller arasında farklılıklar ortaya çıkmakta ve bu farklılıklar toplandığında daha kararlı ve hata payı daha düşük bir sonuç ortaya çıkmaktadır (Breiman, 1996).

3.1.8 AdaBoost Regresyonu

AdaBoost yöntemi ingilizce "adaptive boosting" olarak geçer, zayıf öğrenicileri (örneğin karar ağaçları gibi basit modelleri) bir araya getirerek güçlü bir sınıflandırma modeli oluşturan bir güçlendirme (boosting) yöntemidir. İlk olarak Yoav Freund ve Robert Schapire tarafından ortaya atılmıştır. AdaBoost'un önelikli amacı her adımda hatalı sınıflandırılan örneklere daha fazla ağırlık vererek bir sonraki zayıf modelin bu örneklere daha fazla odaklanmasını sağlamaktır. Başlangıçta tüm örnekler eşit ağırlıktadır. İlk model eğitildikten sonra, yanlış sınıflandırılan örneklerin ağırlıkları artırılır ve doğru sınıflandırılan örneklerin ağırlıkları azaltılır. Bu süreç hata belirli bir seviyeye düşene kadar devam eder. Son olarak her zayıf modelin çıktısı, kendi doğruluğuna göre ağırlıklandırılır ve ağırlıklı oylama ile nihai karar verilir. AdaBoost genellikle aşırı uyuma karşı oldukça dayanıklıdır ve düşük bias'a sahip modeller üretir. Adaboost doğru verilerde iyi sonuçlar yakalarken random forest büyük veri setlerinde daha iyi sonuçlar yakalar (Freund & Schapire, 1997).

3.1.9 Gradyan Arttırma Regresyonu

Gradyan artırma (Gradient Boosting), tahmin hatalarını aşamalı olarak azaltmayı hedefleyen güçlü bir güçlendirme (boosting) yöntemidir. Bu yöntem, zayıf öğrenicileri genellikle karar ağaçlarını ardışık olarak eğitir. Her yeni model, önceki modelin hatalarını gradyan inişi (gradient descent) mantığıyla düzeltmeye çalışır. Yani model, her iterasyonda kayıp fonksiyonunun negatif gradyanına karşılık gelen artıkları (residuals) tahmin eder ve bu hatalara göre yeni bir zayıf öğrenici (örneğin küçük bir karar ağacı) eklenir. Bu şekilde, model adım adım daha iyi hale getirilir. Gradyan artırma özellikle düşük biaslı ve yüksek doğruluklu tahminler üretmesiyle bilinir, ancak dikkatli ayarlanmazsa aşırı uyum (overfitting) riski taşıyabilir (Hastie ve ark., 2009).

3.1.10 XGBoost Regresyonu (Extreme Gradient Boosting)

İleri seviye gradyan arttırma (XGBoost), gradyan artırma yönteminin (Gradient Boosting) güçlü bir iyileştirilmiş versiyonudur. XGBoost ("Extreme Gradient Boosting"), özellikle büyük veri setlerinde ve yüksek boyutlu özelliklerde son derece hızlı ve verimli sonuçlar sağlayan bir topluluk öğrenme (ensemble learning) algoritmasıdır.

XGBoost, gradyan artırma yönteminin geliştirilmiş bir versiyonudur ve özellikle regülerleşme (L1, L2) teknikleriyle aşırı öğrenmeyi engellemeye odaklanır. Hafif gradyan arttırma (LightGBM) ise büyük veri setlerinde daha hızlı ve bellek verimli bir model oluşturmak için histogram tabanlı öğrenme ve yaprak bazlı (leaf wise) büyüme stratejisi kullanırken, CatBoost ise kategorik değişkenleri doğrudan işleyebilme yeteneği ile öne çıkar ve sıralı hedef kodlama teknikleri ile önyargısız öğrenme sağlar. Bu üç algoritma da paralel hesaplama ve GPU desteği sunarak büyük veri ile daha verimli çalışırken, her biri farklı avantajlar sunar: XGBoost regülerleşme ve hız ile, LightGBM bellek verimliliği ve hız ile, CatBoost ise kategorik verileri doğal işleyebilme yeteneği ile farklılaşır (Chen & Guestrin, 2016).

3.1.11 Hafif Gradyan Arttırma Regresyonu

Hafif gradyan arttırma yöntemi ingilizce Light Gradient Boosting Machine (LGBM) Microsoft tarafından geliştirilmiş, gradient boosting tabanlı, hızlı ve yüksek verimli bir topluluk öğrenme (ensemble learning) algoritmasıdır. Temelinde besleyici zayıf öğrenicilerin (genellikle karar ağaçları) ardışık olarak eğitilmesi vardır. Ancak geleneksel boosting algoritmalarından farklı olarak LightGBM, özellikle büyük veri setlerinde ve yüksek boyutlu özelliklerde hem hız hem de bellek verimliliği açısından öne çıkar. LightGBM yapraksal büyüme yani düşey büyüme yöntemini izlerken, diğer gradyan artırma yöntemleri daha çok seviye bazlı (level-wise) büyüme stratejisini kullanırlar (Ke ve ark., 2017).

3.1.12 Cat Boost Regresyonu

CatBoost, Yandex tarafından geliştirilen ve özellikle kategorik veri içeren veri setleri için optimize edilmiş bir gradyan arttırma algoritmasıdır. Adı da içeriğine uygun olarak "Categorical + Boosting" ingilizce kelimelerinden türetilmiştir. CatBoost'un temel farkı, kategorik değişkenleri önceden dönüştürmeden doğrudan

kullanabilmesidir. Bu sayede bilgi kaybı olmadan daha doğru modeller kurulabilir. Ayrıca sıralı hedef kodlama ve rastgele kullanılan tekniklerle aşırı öğrenme riskini azaltır. Ayrıca, CatBoost eğitim sırasında simetrik ağaç yapısı (oblivious trees) kullanır. Bu yapı, her seviye için aynı koşullu özelliği değerlendirerek eğitim ve tahmin sürecini hızlandırır ve modeli daha stabil hale getirir (Prokhorenkova ve ark., 2017).

3.2 Öznitelik Seçimi

Öznitelik seçimi bir veri setinde yer alan tüm değişkenler arasından en anlamlı olanların belirlenmesi ve yalnızca bu değişkenlerin modele dahil edilmesidir. Bu süreç; modelin sonuçtaki başarı oranını artırmak, analiz maliyetini düşürmek aşırı öğrenmeyi azaltmak gibi süreçlere destek olur (Breiman, 2001).

Öznitelik seçimi tekniikleri ise temelde 3 ana gruba ayrılır bunlar; filtre yöntemleri, sarmalayıcı yöntemler, gömülü yöntemlerdir (Guyon & De, 2003).

Bu tez çalışmasında rastgele orman algoritması öznitelik seçimi kullanılmıştır.

3.2.1 Filtre Yöntemleri

Filtre yöntemleri, öznitelikleri herhangi bir modele bağlı kalmaksızın istatistiksel ölçütler aracılığıyla değerlendirir. Her bir öznitelik puanlanır, bu puanlara göre sıralanır ve belirlenen bir eşik değeri kullanılarak seçilir (Liu & Motoda, 1998).

Genel olarak Pearson ve Ki-kare yöntemleri en çok tercih edilen yöntemlerdir.

3.2.1 Sarmalayıcı Yöntemler

Sarmalayıcı yöntemlerde (wrapper methods) ise özniteliklerin farklı alt kümeleri seçilerek, her alt kümenin modele olan etkisi incelenir ve modelin başarısı göz önünde bulundurularak en uygun öznitelik alt kümesi belirlenir. Modelin doğruluk veya hata oranı gibi performansını gösteren oranlar hangi özniteliklerin dikkate alınacağını gösterir. Sarmalayıcı yöntemler genellikle arama algoritmaları (ileri seçim, geri eleme, genetik algoritmalar vb.) kullanarak en iyi öznitelik kombinasyonunu bulmaya çalışır (Guyon, I., & Elisseeff, A.,2003)

3.2.2 Gömülü Yöntemler

Gömülü (embedded) yöntemler, filtre ve sarmalayıcı yöntemlerin avantajlarını birleştirir yani filtre yöntemlerindeki hız ve gömülü modellerin doğrudan model

performansını optimize eder. Gömülü yöntemler, genellikle modelin doğrudan parçası olduğu için daha verimli ve genelleştirilebilir öznitelik seçimi sağlar. Özniteliklerin model üzerindeki etkisini doğrudan yansıttığı için ve aşırı öğrenmeyi önlemek için regülasyon teknikleriyle birlikte kullanılabildiğinden tercih edilir (Kuhn & Johnson, 2019).

3.3 Hata Metrikleri

3.3.1 R-Kare (R-Squared)

R-Kare (R-Squared), değeri modelin ne kadar iyi açıklanabildiğini gösteren değerdir yani gerçek değerle modelin tahminleri arasındaki bağı ölçer. 0 ile 1 arasında değer alır ve ne kadar 1'e yaklaşırsa modeli o kadar iyi açıklar (Botchkarev, 2018).

Formülü ise aşağıdaki gibidir;

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{3.1}$$

3.3.2 Ortalama Kare Hata

Ortalama kare hata (MSE – Mean Squared Error) veya hata karelerinin ortalaması ise modelin tahmin edilen değerleri ile gerçek değerleri arasındaki farkların karelerinin ortalamasıdır (Power, 2011).

Formülü ise aşağıdaki gibidir;

$$\text{MSE} = \frac{1}{n}\sum(y_i - \hat{y}_i)^2 \tag{3.2}$$

3.3.3 Kök Ortalama Kare Hata

Ortalama kare hatanın (RMSE-Root Mean Square Error) kareköküdür uzun bir söyleyişle tahmin edilen değerler ile gerçek değerler arasındaki farkların karelerinin ortalamasının kareköküdür. Modelin tahminlerinin ne kadar sapma gösterdiğini ifade eder yani standart hata gibi gözükür (Hodson, 2022). Formülü ise aşağıdaki gibidir;

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2} \tag{3.3}$$

3.3.4 Ortalama Mutlak Hata

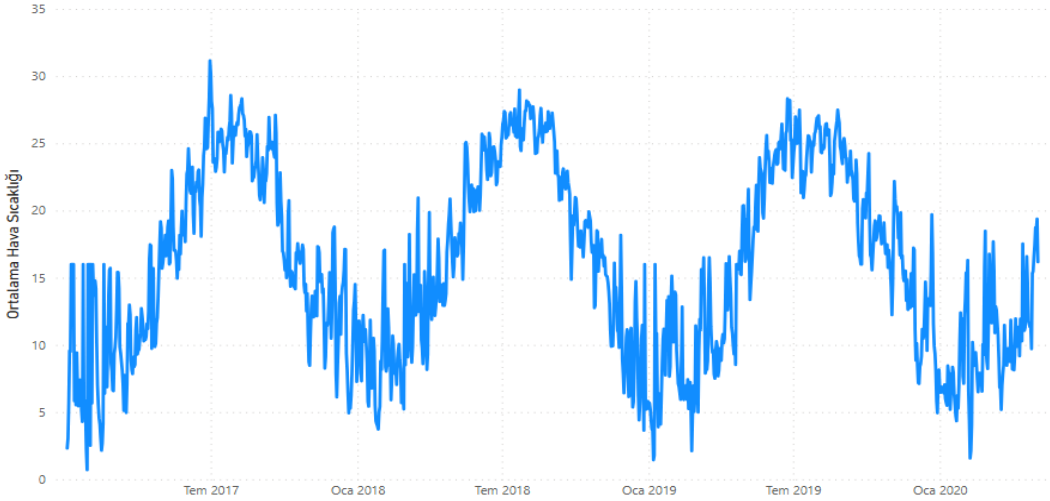
Ortalama mutlak hata (MAE-Mean Absolute Error), tahmin edilen değerler ile gerçek değerler arasındaki mutlak farkların ortalamasıdır, tahmin hatalarının ortalama büyüklüğünü doğrudan ölçer ve birimi tahmin edilen değişken ile aynıdır (James ve ark., 2021).Bazı kaynaklarda hataların mutlak değerlerinin ortalaması olarak da belirtilmektedir.

Formülü ise aşağıdaki gibidir;

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.4}$$

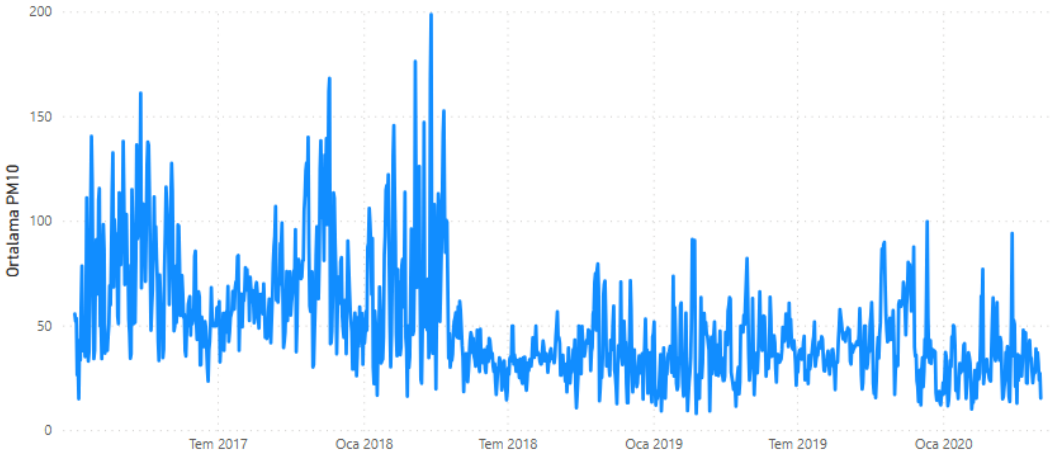
4
BULGULAR

Veri seti, 01.01.2018 - 30.09.2020 yılları arasında 2 yıllık PM_{10}, SO_2, CO, NO_2 NO_X, O_3 parametlerine ilişkin günlük ortalama karışımları (µg/m3) ile rüzgâr yönü (o), rüzgâr hızı (km/h), hava sıcaklığı (°C), basınç ve bağıl nem değerlerini kapsamaktadır. Analiz sürecinde R programından ve PowerBI programlarından faydalanılmıştır.



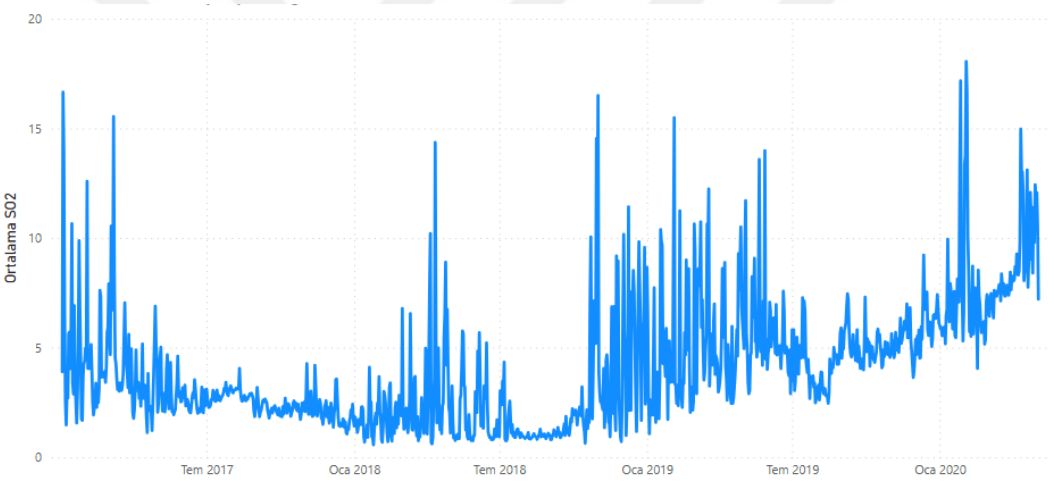
Şekil 4.1 Hava sıcaklığının ortalama değeri

Ortalama hava sıcaklığını aylara göre grafiği Şekil 4.1'de verilmiştir. Hava sıcaklığı trend grafiğinde görülebileceği gibi yaz aylarında artan sıcaklıklar kış aylarında ise azalmaktadır.



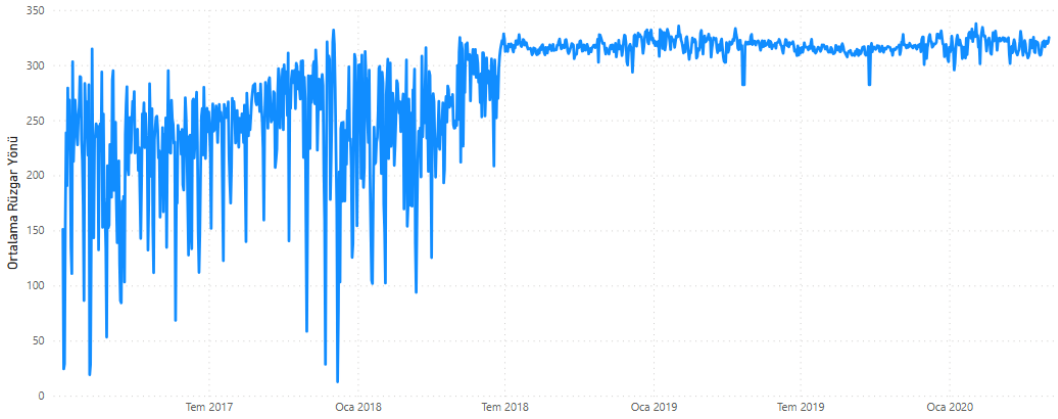
Şekil 4.2 PM_{10} ortalama değeri

PM_{10} trend grafiği ise Şekil 4.2'de yer almaktadır. 2018 mayıs ayından sonra değerlerin daha az seyrettiği gözlemlenmektedir.



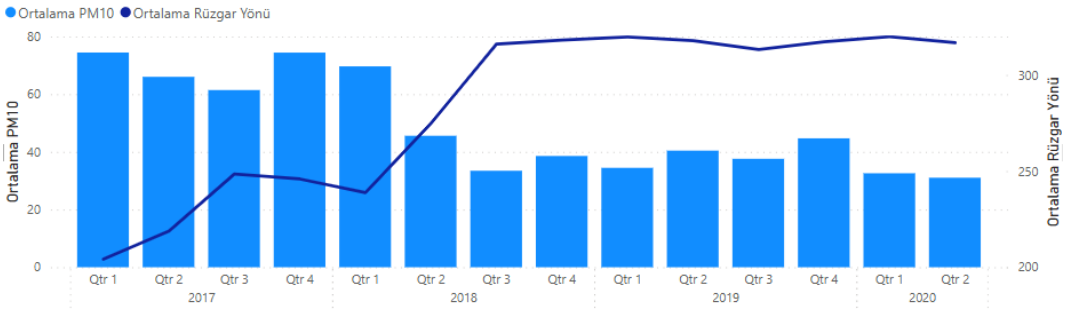
Şekil 4.3 SO_2 ortalama değeri

SO_2 için aylara göre ortalama değeri trend grafiği Şekil 4.3'de yer almaktadır. Burada düzenli bir etki görülememektedir. Diğer değişkenlerle arasında da bir bağlantı olduğunu söylemek zor olmaktadır ancak 2018 yılı Haziran ayına kadar azalan bir eğilim gösteren değerler 2018 Haziran ayından sonra az da olsa yukarı yönlü bir artış göstermektedir denilebilir.



Şekil 4.4 Rüzgar yönü ortalama değeri

Rüzgar yönü ortalama değerinin aylara göre trend grafiği Şekil 4.4'te verilmiştir. Buna göre, 2018 mayıs ayından sonra değerlerin daha yüksek seyrettiğini gözlenmektedir. Şekil 4.4'de PM_{10} değeri incelediğinde 2018 Haziran ayında PM_{10} değerinin azaldığından bahsedilmişti, dolayısıyla bu iki değer arasında bir korelasyon ilişkisi gözlemlenebilir.



Şekil 4.5 Rüzgar yönü ve PM_{10} değişkenleri

Şekil 4.5'de verilen ortalama rüzgar yönü ile ortalama PM_{10} arasındaki grafik incelendiğinde aralarında çok yüksek bir ilişki olmayabileceği görülmektedir.

Analizin bu aşamasında, n_splits=10 olarak belirlenen çapraz doğrulama (cross-validation) yöntemi kullanılmış ve veri seti, modelin güvenilirliğini artırmak amacıyla 10 farklı katlamayla eğitim ve doğrulama alt kümelerine ayrılmıştır.

Tablo 4.1incelendiğinde veri setinin zaman aralığının eğitim verisinin 1 Ocak 2018 ile 30 Haziran 2020 arasında incelendiği ve gözlem sayısının 21.884 olduğu görülüyor. Test verisinde ise, zaman aralığı 1 Temmuz 2020 ile 30 Eylül 2020 arasında incelenirken, gözlem sayısının 2.188 olduğu görülmektedir.

PM_{10} değişkeni için ise eğitim verisinde ortalama 47.59 µg/m³, test verisinde ise ortalama 35.34 µg/m³ olarak ölçümlenmiştir. Test döneminde PM_{10} seviyesindeki görülen azalmanın sebebi artan sıcaklık, atmosferik hareketlilik, pandemi etkileri ya da kısıtlamalarla bağlantılı olabilir.

Öne çıkan kirleticilerdeki değişkenlerin eğitim ve test verisindeki değişimleri dikkate alındığında yüksek oranlı düşüşler gözlemlenmektedir. Özellikle pandemi nedeniyle motorlu taşıt kullanımının azalması ya da meteorolojik koşulların etkisi ile açıklanabileceği söylenebilir.

O_3 incelendiğinde ise; eğitim döneminde 58.62 seviyesindeyken, test döneminde 92.50 seviyesinde gözlemlenmiştir. Ozon, güneş ışığı ve NOx etkileşimi ile oluştuğundan, yaz aylarında ozon düzeyi artma eğilimindedir.

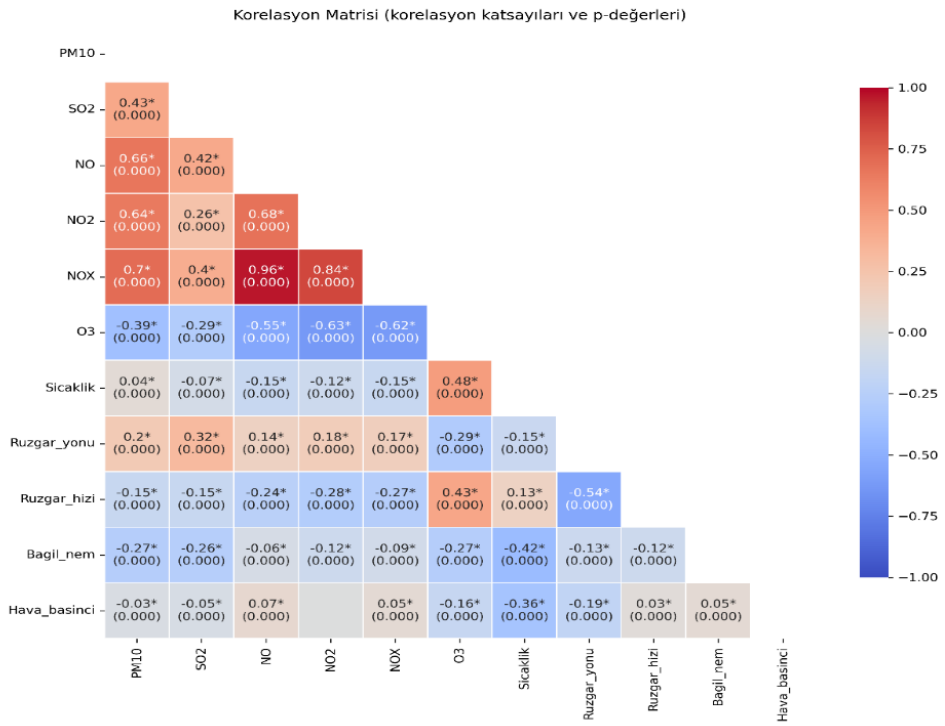
Tablo 4.1 Günlere göre değişken ortalamaları

Eğitim Verisi (01.01.2018 - 30.06.2020) – n=21884											
Gün	**PM10**	**SO2**	**NO**	**NO2**	**NOX**	**O3**	**Sicaklik**	**Ruzgar_yonu**	**Ruzgar_hizi**	**Bagil_nem**	**Hava_basinci**
Pazartesi	50.72	4.87	18.90	26.57	55.32	55.33	15.47	110.19	2.52	80.92	998.57
Salı	48.35	4.68	14.29	24.53	46.09	58.13	15.59	114.69	2.46	82.07	998.03
Çarşamba	49.04	4.45	14.36	24.26	45.96	57.17	15.21	115.22	2.45	83.29	998.22
Perşembe	47.00	4.48	13.07	23.78	43.50	59.05	14.94	114.95	2.41	83.22	998.41
Cuma	48.70	4.53	15.07	24.78	47.43	56.57	14.89	120.06	2.31	82.54	998.94
Cumartesi	48.91	4.69	13.29	22.42	42.49	59.65	15.16	112.25	2.45	81.71	999.22
Pazar	40.41	4.38	10.25	18.73	34.07	64.46	15.31	110.25	2.38	81.57	999.16
Toplam	**47.59**	**4.58**	**14.18**	**23.59**	**44.99**	**58.62**	**15.23**	**113.94**	**2.43**	**82.19**	**998.65**
Test Verisi (01.07.2020 - 30.09.2020) - n=2188											
Gün	**PM10**	**SO2**	**NO**	**NO2**	**NOX**	**O3**	**Sicaklik**	**Ruzgar_yonu**	**Ruzgar_hizi**	**Bagil_nem**	**Hava_basinci**
Pazartesi	36.81	4.99	7.12	2.66	14.40	91.61	23.82	68.23	2.74	80.81	996.79
Salı	37.26	5.52	6.17	3.61	13.96	90.75	23.50	60.71	2.89	80.10	996.71
Çarşamba	35.41	5.71	7.52	3.49	16.04	90.18	23.53	73.28	2.74	78.35	996.92
Perşembe	34.33	5.17	7.69	2.90	14.64	92.07	23.89	65.74	2.98	81.05	996.76
Cuma	35.70	4.81	7.86	2.66	15.43	88.63	23.92	72.78	3.01	80.06	996.12
Cumartesi	35.36	5.02	5.49	2.99	12.33	94.52	24.10	66.82	3.42	79.56	995.51
Pazar	32.47	4.87	3.78	2.34	11.47	99.92	23.78	61.75	2.89	81.55	995.80
Toplam	**35.34**	**5.16**	**6.53**	**2.95**	**14.06**	**92.50**	**23.79**	**67.11**	**2.95**	**80.19**	**996.38**

4.1 Değişkenler Arası Korelasyonların İncelenmesi

Şekil 4.6'de yer alan korelasyon matrisinde değişkenler arası anlamlılıkları (p-değeri) ve korelasyon katsayıları incelenebilmektedir.

PM_{10} ile NO_X (0.70), NO (0.66), NO_2 (0.64) yüksek pozitif korelasyon göstermektedir. Bu değişkenler PM_{10} tahmin edilmesinde daha etkili olabilecekleri söylenebilmektedir.



Şekil 4.6 Korelasyon matrisi

PM_{10} ile negatif korelasyon gösteren değişkenler ise O_3 (-0.39): Ozonla PM_{10} arasında negatif korelasyon mevcut. Bu, güneşli ve rüzgârlı havalarda ozon artarken PM_{10}'un azalması anlamına gelmektedir. Bağıl nem ile PM_{10} arasında negatif (-0.27) ve Rüzgâr hızıyla PM_{10} arasında da yine negatif korelasyon (-0.15) mevcut, bu da rüzgarla kirliliğin farklı bölgelere yayılması ve nemin partikülleri bastırması gibi fiziksel nedenlerle açıklanabilmektedir.

Korelasyon matrisinde bulunan diğer değişkenler arasındaki ilişkiler incelendiğinde NO_X ile NO (0.96) ve NO_X ile NO_2 (0.84) arasında çok güçlü pozitif korelasyon olduğu görülmektedir. Bu, bu değişkenlerin birbirine çok benzer bilgi

taşıdığını ve modelde çoklu doğrusal bağlantı (multicollinearity), aşırı uyum (overfitting) gibi sorunlara yol açabilmektedir. Gerektiğinde boyut indirgeme ya da sadece biriyle çalışmak tercih edilebilir dolayısıyla analizde PM10 değişkeni ile korelasyonu en yüksek olan NO_X değişkeni bırakılmış ve NO ve NO_2 değişkenleri çalışmadan çıkarılmıştır. Daha sonrasında kalan değişkenler Min-Max yöntemine göre normalize edilmiştir.

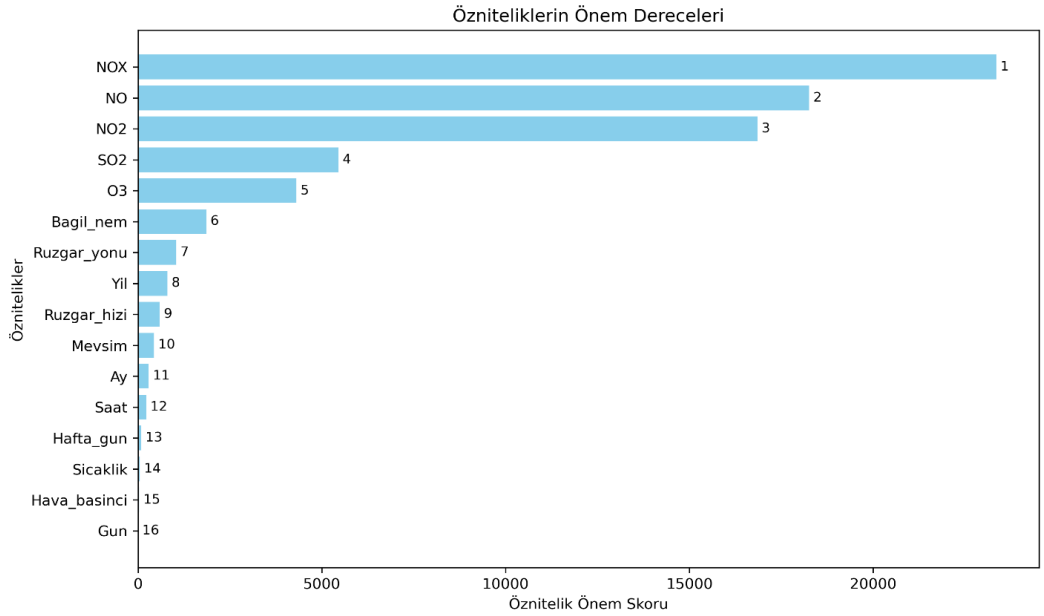
O_3, rüzgâr hızı, bağıl nem gibi değişkenler PM_{10} üzerinde negatif etki yaratsa dahi bu değişkenler modelde bulunmalıdır ki tüm yönlü etkiler de tahminleme de yer alabilsin.

Hava basıncı ve sıcaklık gibi değişkenlerin yüksek bir korelasyon ilişkisi olduğu görülmüyor ancak bu tarz değişkenler kirlilik türlerinin tahmininde etkili olabilir ve modellere katkı sağlayabilir.

4.2 Özniteliklerin Önem Derecelerinin İncelenmesi

NO_X, en yüksek önem skoruna sahip olduğundan PM_{10} tahmini için en güçlü belirleyicidir.

NO ve NO_2 değişkenlerinin de korelasyonları tahmin gücüne büyük katkı sağlıyor. SO_2 Hava kirliliğinin önemli bir göstergesi ve PM_{10} ile doğrudan ilişkili olduğu görülmektedir.



Şekil 4.7 Öznitelik önem skoru

O_3, negatif korelasyon göstermesine rağmen, tahmin modelinde anlamlı katkısı olduğu görümektedir. Modelin sadece pozitif etkiler değil, negatif ilişkileri de başarıyla yakaladığını göstermektedir.

Bu uyumlar modelin korelasyon yapısıyla doğru orantıda çalıştığını ve öznitelik önem tablosunun doğru değişkenleri öne çıkardığını söylemeyi mümkün kılmaktadır.

Burada öznitelik seçimi yapılırken, mevsim, yıl, ay, hafta, gün ve saat gibi zamansal değişkenler de dahil edilmiştir. Fakat PM_{10} üzerinde çok etkisi olmadığı görüldüğünden dolayı değişken olarak analizde yer verilmemiştir.

4.3 Modellerin Veri Setlerindeki Performansların Karşılaştırılması

Tablo 4.2 incelendiğinde XGB, GBR, DTR, MLP, RFR makine öğrenmesi yöntemlerinin R^2, MSE, RMSE ve MAE metriklerine göre eğitim ve test veri setlerindeki performanslarını karşılaştırmaktadır.

Tablo 4.2 Yöntemlerin başarı oranlarının karşılaştırılması

Model	R2			MSE			RMSE			MAE		
	Eğitim	**Test**	**Tüm**	**Eğitim**	**Test**	**Tüm**	**Eğitim**	**Test**	**Tüm**	**Eğitim**	**Test**	**Tüm**
XGB	0.96	0.72	0.96	0.0002	0.0003	0.0002	0.0155	0.0183	0.0158	0.0107	0.0110	0.0107
GBR	0.89	0.73	0.89	0.0007	0.0003	0.0007	0.0266	0.0179	0.0259	0.0158	0.0107	0.0154
DTR	0.91	0.68	0.91	0.0006	0.0004	0.0005	0.0237	0.0197	0.0233	0.0132	0.0123	0.0131
MLP	0.85	0.72	0.85	0.0009	0.0003	0.0009	0.0305	0.0183	0.0296	0.0175	0.0115	0.0169
RFR	0.91	0.75	0.91	0.0005	0.0003	0.0005	0.0234	0.0173	0.0229	0.0126	0.0101	0.0124

XGBoost, eğitim verisindeki en yüksek skoru veren model olduğu için, veriyi en iyi öğrenen model olduğu söylenebilir. Buna karşılık Random Forest Regressor (RFR) yönteminin genelleme kabiliyetinin daha yüksek olduğunu göstermektedir.

RMSE değerleri incelendiğinde, Random Forest, test setinde en düşük RMSE değeri ile en kararlı tahmini sunmaktadır.

RFR, test verisinde en düşük ortalama mutlak hatayı vererek tahminlerde en az sapmayı göstermiştir.

5
SONUÇ

Bu çalışmada kullanılan beş farklı makine öğrenmesi algoritması (XGB, GBR, DTR, MLP ve RFR), PM10 konsantrasyonlarının tahmininde gösterdikleri başarıya göre karşılaştırılmıştır. Tablo 4.2 incelendiğinde, özellikle XGBoost (XGB) ve Random Forest Regressor (RFR) algoritmalarının diğer modellere kıyasla daha yüksek doğruluk ve daha düşük hata oranlarına ulaştığı görülmektedir. XGB, eğitim verisinde $R^2 = 0.96$ ile en yüksek başarıyı sergilerken, test verisinde de 0.72'lik bir değerle güçlü performansını sürdürmüştür. Benzer şekilde, RFR modeli test verisinde $R^2 = 0.75$ değeri ile tüm modeller arasında test setindeki en iyi sonucu vermiştir. RMSE ve MAE gibi hata metrikleri açısından da bu iki modelin oldukça düşük değerlere sahip olması, tahminlerinin istikrarlı ve güvenilir olduğunu ortaya koymaktadır.

Ancak bu sonuçların değerlendirilmesinde dikkat edilmesi gereken önemli bir bağlamsal faktör bulunmaktadır: Test verisi pandemi dönemine denk gelmektedir. 2020 yılında küresel ölçekte etkili olan COVID-19 pandemisi nedeniyle, birçok ülkede olduğu gibi Türkiye'de de sokağa çıkma yasakları, üretim ve ulaşım faaliyetlerinde ciddi azalmalar yaşanmıştır. Bu durum, hava kirliliği düzeylerinde belirgin bir düşüşe neden olmuş, özellikle PM10 gibi kirleticilerin atmosferdeki yoğunluğu normal koşullara göre önemli ölçüde azalmıştır.

Bu bağlamda, test verisindeki düşük PM10 seviyeleri, bazı modellerin "aşırı uyum" (overfitting) yaptığı izlenimini verse de, aslında bu durum bir yapısal veri değişikliğine işaret etmektedir. Pandemi, hava kalitesine dair olağan dışı bir ortam oluşturmuş ve öğrenilen modeller bu düşük kirletici seviyelerine dair örüntüleri tanımakta zorluk çekmiş olabilir. Bu nedenle test verisinde elde edilen model başarıları, pandemi döneminin etkisiyle farklı bir veri dağılımına sahip olan yapının

doğru şekilde yansıtıldığını da göstermektedir. Bu yönüyle, model performansları yalnızca sayısal metriklerle değil, veri bağlamı gözetilerek de değerlendirilmelidir.

KAYNAKÇA

Acito, F. (2023). *Predictive analytics with KNIME: Analytics for citizen data scientists*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-45630-5

Akbulut Zencirci, S. (2017). Hava Kirliliği. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Halk Sağlığı Dergisi*, 2(2), 24-36. https://dergipark.org.tr/tr/pub/estudamhsd/issue/39507/466079

Aksoy, S. (2023). Aksaray ilinde karayolu kaynaklı karbon ayak izinin belirlenmesi. [Yüksek lisans tezi, Aksaray Üniversitesi]. YÖK Tez Merkezi. https://acikerisim.aksaray.edu.tr/server/api/core/bitstreams/f0ad8981-8dad-48f5-b90c-788841568b09/content

Anderson, F. T., McDaniel, M. A., & Einstein, G. O. (2017). Remembering to remember: An examination of the cognitive processes underlying prospective memory. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp. 451–463). Academic Press. https://doi.org/10.3758/s13421-018-0849-x

Andriy Burkov, (2019). The Hundred-Page Machine Learning, Andriy Burkov Self-Published.

Atkinson, R. (2000). Atmospheric chemistry of VOCs and NOx. *Atmospheric Environment*, 34(12-14), 2063-2101. https://doi.org/10.1016/S1352-2310(99)00460-4

Awad, M., & Khanna, R. (2015). Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, 1-248. Springer. https://doi.org/10.1007/978-1-4302-5990-9/COVER

Bıyık, İ. (2010). *Hava kirliliğinin yapay zeka teknikleri ile belirlenmesi* [Yüksek lisans tezi, Bahçeşehir Üniversitesi]. Açık Bilim. https://hdl.handle.net/20.500.14719/312Açık Bilim

Bishop, C. M. (2006). *Pattern recognition and machine learning* (1st ed.). Springer. https://doi.org/10.1007/978-0-387-45528-0

Botchkarev, A. (2018). Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. Interdisciplinary *Journal of Information, Knowledge, and Management,* 14, 45-76. https://doi.org/10.28945/4184

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140. https://doi.org/10.1007/BF00058655/METRICS

Breiman, L. (2001). Random forests. *Machine Learning,* 45(1), 5-32. https://doi.org/10.1023/A:1010933404324/METRICS

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785-794. https://doi.org/10.1145/2939672.2939785/SUPPL_FILE/KDD2016_CHEN_BOOSTING_SYSTEM_01-ACM.MP4

Çağlayan Akay, E. (2018). Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi. *Sosyal Bilimler Araştırma Dergisi*, 7(2), 41-53. https://www.acarindex.com/pdfs/1116151

Çelik, B. D. (2022). *Makine öğrenmesi yöntemleri ile hava kirliliği tahmini: Manisa ve Zonguldak örneği* [Yüksek lisans tezi, Gazi Üniversitesi]. AVESİS. https://avesis.gazi.edu.tr/yonetilen-tez/f7c12e6c-ab62-4ada-98f9-60a2ae3c8376/makine-ogrenmesi-yontemleri-ile-hava-kirliligi-tahmini-manisa-ve-zonguldak-ornegi

European Environment Agency. (2022). *Air quality in Europe 2022*. https://doi.org/10.2800/488115

EPA. (2023). Health and Environmental Effects of Particulate Matter. https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm

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

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning,* 63(1), 3-42. https://doi.org/10.1007/S10994-006-6226-1/METRICS

Giniş, T., Gürsoy, C., Mısırlıoğlu, E. D., Kocabaş, C. N. (2014). The Evaluation of Carbon Monoxide Poisoning in a Tertiary Children's Hospital During an 8-Year Period. *Türkiye Çocuk Hastalıkları Dergisi*, 8(2), 55-58.

Grahame, T. J., & Schlesinger, R. B. (2009). Cardiovascular health and particulate vehicular emissions: a critical evaluation of the evidence. *Air Quality, Atmosphere, & Health*, 3(1), 3. https://doi.org/10.1007/S11869-009-0047-X

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182. https://doi.org/10.1162/153244303322753616

Güven, A. (2022). *Hava kirliliğinin makine öğrenmesi yöntemleriyle tahmini* [Yüksek lisans tezi, Bursa Uludağ Üniversitesi Fen Bilimleri Enstitüsü]. http://hdl.handle.net/11452/25516

Harita Genel Müdürlüğü. (t.y.). Yalova Mülk İdare İl Haritası. Erişim tarihi: 15 Mart 2025, https://www.harita.gov.tr/uploads/files/products/yalova-mulk-idare-il-haritasi-1325.pdf

Hasan, T., Aytekin, U., Üniversitesi, İ. İ. B. F., & Sistemleri, Y. B. (2024). İşletmeler İçin makine öğrenimi hizmet stratejisine genel bakış. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi,* 14(4), 1901-1923. https://doi.org/10.30783/NEVSOSBILEN.1521972

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer. https://doi.org/10.1007/978-0-387-84858-7

Hava Kalitesi Raporu - Sürekli İzleme Merkezi. (2024). *Çevre ve Şehircilik Bakanlığı*. https://ced.csb.gov.tr/laboratuvar-olcum-ve-izleme-i-82320

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development, 15(14), 5481-5487. https://doi.org/10.5194/GMD-15-5481-2022

İçöz, E. (2019). *Development of a smart sensor system to detect air pollution* [Yüksek lisans tezi, Abdullah Gül Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik ve Bilgisayar Mühendisliği Anabilim Dalı]. Ulusal Tez Merkezi. Erişim adresi: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=tYnz9LMmFMfNXL1J05mWpQ&no=iTDz2DmqMC-OJ2IeAuIfiw

İlkılıç, C., & Behçet, R. (2006). Hava Kirliliğinin İnsan Sağlığı ve Çevre Üzerindeki Etkisi. *Fırat Üniversitesi Doğu Araştırmaları Dergisi*, 5(1), 66-72. https://dergipark.org.tr/tr/pub/fudad/issue/47092/592365

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning. Springer. https://doi.org/10.1007/978-1-0716-1418-1

Kaya, E. (2007). *Azot oksit (NOx) tutucu adsorbent geliştirme ve karakterizasyonu* [Yüksek lisans tezi, Gazi Üniversitesi]. https://avesis.gazi.edu.tr/yonetilen-tez/462b485b-d4a8-42f1-9215-0eea61221968/azot-oksit-nox-tutucu-adsorbent-gelistirme-ve-karakterizasyonu

Kaya, K. (2019). *Veri madenciliği yöntemleri kullanarak hava kirliliği tahmini* [Yüksek lisans tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü]. https://research.itu.edu.tr/tr/studentTheses/veri-madencili%C4%9Fi-y%C3%B6ntemleri-kullanarak-hava-kirlili%C4%9Fi-tahmini

Kayri, M., & Boysan, M. (2008). Assesment of relation between cognitive vulnerability and depression's level by using classification and regression tree analysis. *Hacettepe Universitesi Egitim Fakultesi Dergisi-Hacettepe University Journal Of Education*, 34. https://avesis.yyu.edu.tr/yayin/129ff741-2429-4ff3-9036-0915c1e07bd1/assesment-of-relation-between-cognitive-vulnerability-and-depressions-level-by-using-classification-and-regression-tree-analysis

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *In Advances in Neural Information Processing Systems*, 3146-3154. https://github.com/Microsoft/LightGBM.

Koyuncugil, A. S., & Özgülbaş, N. (2009). Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. *Bilişim Teknolojileri Dergisi*, 2(2), 21.

Kuhn, M., & Johnson, K. (2019). Feature Engineering and Selection: A Practical Approach for Predictive Models, 1-297. https://doi.org/10.1201/9781315108230/FEATURE-ENGINEERING-SELECTION-MAX-KUHN-KJELL-JOHNSON/RIGHTS-AND-PERMISSIONS

Kumral, C. D., Topal, A., Ersoy, M., Çolak, R., & Yiğit, T. (2022). Random Forest Algoritmasının FPGA Üzerinde Gerçekleştirilerek Performans Analizinin Yapılması. *El-Cezeri*, 9(4), 1315-1327. https://doi.org/10.31202/ECJSE.1134799

Liu, H., & Motoda, H. (1998). Feature Selection for Knowledge Discovery and Data Mining. Springer US. https://doi.org/10.1007/978-1-4615-5689-3

Maleki, H., Sorooshian, A., Goudarzi, G., Baboli, Z., Tahmasebi Birgani, Y., & Rahmati, M. (2019). Air pollution prediction by using an artificial neural network model. *Clean Technologies and Environmental Policy*, 21(6), 1341-1352. https://doi.org/10.1007/S10098-019-01709-W/FIGURES/7

Memnun, D. S. (2013). SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri. *İlköğretim Online*, 12(1), 1-6. https://dergipark.org.tr/tr/pub/ilkonline/issue/8586/106689

Mockenhaupt, A., & Schlagenhauf, T. (2024). Digitalisierung und Künstliche Intelligenz in der Produktion. https://doi.org/10.1007/978-3-658-41935-6

Murphy, K. P. (2012). Machine Learning A Probabilistic Perspective. *MIT Press*. http://mitpress.mit.edu

Ott, W. R., & Hunt, W. F. (1976). A Quantitative Evaluation of the Pollutant Standards Index. *Journal of the Air Pollution Control Association*, 26(11), 1050-1054. https://doi.org/10.1080/00022470.1976.10470358;WGROUP:STRING:PUBLICATION

Power, D. J. (2011). Decision Support Systems Concept. *Encyclopedia of Decision Making and Decision Support Technologies*, 1, 271-283. https://doi.org/10.4018/978-1-59904-843-7.CH027

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2017). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 2018-December, 6638-6648. https://arxiv.org/pdf/1706.09516

Sağlık Bakanlığı Temel Sağlık Hizmetleri Genel Müdürlüğü. (2010). Türkiye'nin Hava Kirliliği Ve Iklim Değişikliği Sorunlarina Sağlik Açisindan Yaklaşim.

Sutton, R. S., & Barto, A. G. (2014). Reinforcement Learning: An Introduction Second edition, in progress.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267-288. https://doi.org/10.1111/J.2517-6161.1996.TB02080.X

Türkoğlu, B. B. (2020). *Predictive analysis of flight traffic at New York JFK Airport on air pollution using machine learning* (Yüksek lisans tezi, Boğaziçi Üniversitesi, Çevre Bilimleri Enstitüsü, Çevre Teknolojisi Bilim Dalı). Ulusal Tez Merkezi. Erişim adresi: https://tez.yok.gov.tr/UlusalTezMerkezi/tezDetay.jsp?id=S6KViT8FTrQI1XGWoMG-Hg&no=Ax4mRVe4tf9oxbtQKj4wTQ

Uğurlu, O. (2021). Konya İlinde Hava Kirliliğine Etki Eden Faktörlerin Mekânsal Analizi. Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü. https://hdl.handle.net/20.500.12452/8013

van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. *Machine Learning*, 109(2), 373-440. https://doi.org/10.1007/S10994-019-05855-6/FIGURES/5

WHO. (2002). The World Health Organization report 2002: reducing risks, promoting healthy life. *World Health Organization*, 232.

Zhang, J. J., Wei, Y., & Fang, Z. (2019). Ozone pollution: A major health hazard worldwide. *Frontiers in Immunology*, 10(OCT), 492681. https://doi.org/10.3389/FIMMU.2019.02518/BIBTEX

Zilyas, D., & Yılmaz, A. (2023). Makine Öğrenmesi Yöntemleri İle Eğitim Başarısının Tahmini Modeli. *DUJE (Dicle University Journal of Engineering),* 14, 437-447. https://doi.org/10.24012/dumf.1322273

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301-320. https://doi.org/10.1111/J.1467-9868.2005.00503.X

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. Valunya T., Parim C., Karaman F.
"Estimating Air Pollution in Istanbul through Machine Learning Techniques"
International Applied Statistic Congress (UYIK)