



Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi

<https://dergipark.org.tr/tr/pub/yyufbed>



Araştırma Makalesi

Makine Öğrenmesi Algoritmaları ile Malatya ilinde PM₁₀ ve SO₂ Konsantrasyonlarının Tahmini

✉ Ayşegül DEMİR YETİŞ^{*1}, ✉ Nagehan İLHAN², ✉ Songül KARAKUŞ³

¹Bitlis Eren Üniversitesi, Sağlık Hizmetleri Meslek Yüksekokulu, Tıbbi Hizmetler ve Teknikler Bölümü, 13000, Bitlis, Türkiye

²Harran Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 63300, Şanlıurfa, Türkiye

³Bitlis Eren Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, 13000, Bitlis, Türkiye

*Sorumlu yazar e-posta: guldem83@gmail.com

Öz: Çalışma, Malatya ilinin 2020-2024 yılları arası PM₁₀ ve SO₂ parametrelerine ait hava kalitesinin değerlendirilmesi ve tahmin edilmesini amaçlamaktadır. Bu kapsamda veriler, Ulusal Hava Kalitesi İzleme Ağı'ndan temin edilmiştir. Veriler ulusal ve uluslararası mevzuattaki sınır değerlerle mukayese edilerek değerlendirilmiştir. 2020-2023 yıllarının verileri kullanılarak, 2024 yılının PM₁₀ ve SO₂ kirletici konsantrasyonları lineer regresyon (LR), rastgele orman (RF), XGBoost, LightGBM, destek vektör regresyonu (SVR), k-en yakın komşu (k-NN) ve yapay sinir ağları (ANN) olmak üzere 7 farklı makine öğrenmesi (ML) algoritması ile tahmin edilmiştir. Model geliştirme aşaması 2024 yılı için tekrarlanarak, algoritmaların ürettiği sonuçlar gerçek sonuçlarla karşılaştırılmış, algoritmaların performansları ve hata oranları belirlenmiştir. Sonuç olarak; 2020 yılından 2023 yılına kadar PM₁₀ ve SO₂ kirletici parametreleri için değerlerde bir artış, 2024 yılında ise düşüş görülmektedir. Hava kalitesinin değerlendirilmesi kapsamında PM₁₀ parametresi için 2023 yılında 24 saatlik ortalama süre sınır değeri %79 oranla, yıllık ortalama süre sınır değeri ise %91 oranla aşılmıştır. SO₂ parametresi için ise 2021 yılında sadece yıllık ortalama süre sınır değeri %21'lik oranla aşılmıştır. Performans metrikleri değerlendirildiğinde, PM₁₀ ve SO₂ tahminlerinde en yüksek doğruluk XGBoost ve LightGBM algoritmaları ile elde edilmiş; Random Forest ve ANN modelleri de yüksek başarı göstermiştir. Linear Regression ve SVR ise daha mütevazı ancak dengeli performans sunarken, k-NN orta düzeyde bir başarı sağlamıştır.

Anahtar Kelimeler: Hava kirliliği, Makine öğrenmesi, Malatya, PM₁₀, SO₂, Yapay zeka

Predicting PM₁₀ and SO₂ Concentrations in Malatya Province with Machine Learning Algorithms

Abstract: The study aims to evaluate and predict the air quality of Malatya province regarding PM₁₀ and SO₂ parameters between 2020-2024. In this context, the data 2020-2024 have been obtained from the National Air Quality Monitoring Network. The data have been evaluated by comparing the limit values in national and international legislation. Using data from 2020 to 2023, PM₁₀ and SO₂ pollutant concentrations for 2024 have been estimated with 7 different machine learning (ML) algorithms such as linear regression (LR), random forest (RF), XGBoost, LightGBM, support vector regression (SVR), k-nearest neighbor (k-NN) and artificial neural network (ANN). The model development phase has been repeated for 2024, the results produced by the algorithms have been compared with the actual results, and the performances and error rates of the algorithms have been determined. As a result; there is an increase in the values for PM₁₀ and SO₂ pollutant parameters from 2020 to 2023, and a decrease in 2024. Within the scope of air quality assessment, the 24-hour average time limit value for the PM₁₀ parameter has been exceeded by 79% in 2023, and the annual average time limit value has been exceeded by 91%. For SO₂, only the annual average limit was exceeded in 2021, at a rate of 21%. In terms of performance metrics, the highest accuracy in predicting PM₁₀ and SO₂ was achieved with XGBoost and

Gönderilme Tarihi: 18.02.2024

Kabul Tarihi: 09.05.2025

Nasıl atıf yapılır: Demir Yetiş, A., İlhan, N., & Karakuş, S. (2025). Makine öğrenmesi algoritmaları ile Malatya ilinde PM10 ve SO2 konsantrasyonlarının tahmini. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 30(2), 621-635. <https://doi.org/10.53433/yyufbed.1642106>

LightGBM algorithms, while the Random Forest and ANN models also demonstrated strong performance. Linear Regression and SVR provided more modest but balanced results, and k-NN achieved moderate performance.

Keywords: Air pollution, Artificial intelligence, Malatya, Machine learning, PM₁₀, SO₂

1. Giriş

Günümüzde iş, eğitim, sosyal ve kültürel yaşama ulaşma kolaylığı gibi birçok nedenden dolayı insanların büyük bir kısmı nüfusun yoğun olduğu şehirlerde yaşamaktadır (Bozdağ ve ark., 2020; Mao ve ark., 2021). Bu durum da kentleşmenin artmasına paralel olarak hava kirliliği gibi çevre ve insan sağlığını tehdit eden sorunları beraberinde getirmektedir. Sanayi Devrimi ile başlayan makineleşme ve yeni fabrikaların işletmeye alınması, kentlerdeki nüfus artışı ile birleşerek bu sorunu daha da ciddi seviyelere taşımış ve iklim değişikliği açısından büyük bir endişe kaynağı haline gelmiştir (Yağmur, 2022; Rakholia ve ark., 2023).

Hava kirliliği; havanın doğal olarak bileşiminde yer alan katı, sıvı ve gaz gibi maddelerin insan sağlığına, canlı yaşamına ve ekolojik dengeye zarar verebilecek konsantrasyon ve sürede bulunması olarak tanımlanabilmektedir (Tırıs ve ark., 1993). Hava kirliliğine sebep olan parametreler; partikül madde (PM), azot oksitler (NO_x), ozon (O₃), kükürt oksitler (SO_x), karbonmonoksit (CO), hidrokarbonlar (HC), uçucu organik bileşikler (VOC) ve polisiklik aromatik hidrokarbonlar (PAH) gibi kirleticilerdir (Müezzinoğlu, 2005; Öztürk & Bayram, 2019). Hava kirliliğinin kaynakları doğal ve antropojenik nedenlerdir. Doğal nedenler; orman yangınları, çöl tozları ve yanardağların çevresinde oluşan gaz ve toz bulutları; antropojenik (insandan kaynaklı) nedenler ise çoğunlukla ısınma, ulaşım ve sanayi kaynaklarından oluşmaktadır (Kırımhan, 2006; Behcet & Yakın, 2020).

Dünya Sağlık Örgütü (WHO) verileri; dünya nüfusunun neredeyse tamamının hava kirleticileri için belirlenen limit değerleri aştığını ve yüksek düzeyde kirleticileri içeren havayı soluduğunu ortaya çıkarmıştır. WHO'ya göre düşük ve orta gelirli ülkelerde yaşayan insanların 10 kişiden 9'u yüksek düzeyde hava kirleticilerine maruz kalmaktadır (Aladag, 2023; WHO, 2025). Ayrıca WHO hava kirliliğinden kaynaklı her yıl 8 milyon insanın erken öldüğü; bunun 4.2 milyonunun dış ortam (sanayi, trafik ve ısınma gibi) ve 3.8 milyonunun ise iç ortam (evlerde ısınma ve yemek amaçlı kullanılan yakıtlar gibi) hava kirliliği nedeniyle gerçekleştiğini bildirmiştir (WHO, 2024; WHO, 2025). Temiz Hava Hakkı Platformu'nun (THHP) 2024 yılı raporuna göre Türkiye'de nüfusun yüzde 92'sinden fazlası WHO standartlarına göre kirli hava solumaktadır (THHP, 2024).

Hava kirleticilerinin atmosferdeki farklı bileşimleri insan sağlığı üzerinde çok çeşitli etkilere neden olmaktadır. Bu etkiler maruz kalınan kirleticinin türü ve süresine göre değişmektedir. Ayrıca etkilenen kişinin yaşı, beslenme durumu, sağlık durumu ve genetik yatkınlığı gibi çeşitli duyarlılık faktörlerine göre etkilenme düzeyi değişmektedir. Bunlara ek olarak, hava kirliliğine kısa veya uzun vadeli maruziyetler azalan yaşam süresi ile ilişkilendirilmiştir (Bozdağ ve ark., 2020; Akgün & Barlık, 2023). Daha düşük sosyoekonomik gruplar, yaşlılar, çocuklar ve önceden bir sağlık sorunu olanlar hava kirleticilerine karşı daha duyarlıdır (WHO, 2024). 18 yaş altı kişilerde her yıl 1 200'den fazla ölümün hava kirliliğinden kaynaklandığı tahmin edilmektedir (EEA, 2024).

Partikül madde (PM), toz, kir, is, duman ve sıvı damlacıkları dahil olmak üzere havada asılı kalan ve solunabilen partiküllerdir. Dizel araçlar ve kömürle çalışan elektrik santralleri ve endüstriler, fosil yakıtların yanması, tarım ve ikincil kimyasal reaksiyonlar gibi kaynaklardan ileri gelmektedir. PM₁₀; çapı 10 mikrometreden küçük olan partiküller, solunabildikleri ve solunum sisteminde birikebildikleri için astım gibi solunum yolu hastalıkları, felç ve kalp rahatsızlıklarına neden olan ciddi sağlık sorunları oluştururlar (Cekim, 2020; Mao ve ark., 2021; WHO, 2024). Partikül maddelere kısa süreli maruz kalma akciğer hastalıklarına neden olur ve düşük konsantrasyonlara uzun süreli maruz kalma kansere ve bebek ölümlerine yol açar. WHO'ya göre PM₁₀ miktarı 24 saatlik ortalama 50 µg/m³'ü geçmemelidir. Ayrıca PM₁₀ düzeyindeki 10 µg/m³'teki artışın her yaş için ölüm oranlarında artışa neden olduğu ifade edilmektedir. Yine önerilen seviyelerin üzerindeki partikül maddeye maruz kalmanın 2022'de tahmini olarak 239.000 kişinin erken ölümüne neden olduğunu bildirmiştir (Bozdağ ve ark., 2020; Akgün & Barlık, 2023; EEA, 2024).

Kükürt dioksit (SO₂); kükürt oranı yüksek malzemelerin ve fosil yakıtların (özellikle kömür) yakılması ve kükürt oranı yüksek olan metallerin eritilmesiyle ortaya çıkmaktadır. Ayrıca ısınmada,

sanayide ve trafikte oluşan bir kirletici olup, insan sağlığına ve ekolojiye zarar vermektedir (Rakholia ve ark., 2023; Mutlu & Atahanlı, 2024; Utku & Can, 2024).

Sağlık sorunlarının yanı sıra, hava kirliliği artan sağlık hizmetleri maliyetleri, azalan yaşam beklentisi ve sektörler arası iş günü kayıpları nedeniyle ekonomiyi önemli ölçüde etkileyebilmektedir. Ayrıca bitki örtüsüne ve ekosistemlere, su ve toprak kalitesine ve yerel ekosistemlere zarar vermektedir (Mutlu & Atahanlı, 2024; EEA, 2024). Hava kirliliği, insan sağlığı üzerinde zararlı olmasının yanında insan hayatını ve faaliyetlerini de olumsuz yönde etkilemektedir. Yoğun hava kirliliği, trafikte görüş mesafesini azaltarak trafik kazası olasılığını artırmakta, pus oluşumuna neden olarak insanların normal seyahatlerini geciktirebilmektedir. Diğer olumsuz etkilerini ise, asit yağmurları, ozon tabakasının incelmeye, küresel ısınma olarak sıralamak mümkündür. Tüm bu olumsuz etkileri önlemek ve daha sağlıklı, sürdürülebilir yaşam alanları oluşturabilmek için hava kirliliğine neden olan faktörler incelenmeli ve gerekli önlemler alınmalıdır (Ünalı & Yalçın, 2022; Yağmur, 2022; Oğuz & Pekin 2022).

Şehirlerde hava kalitesinin izlenmesi amacıyla çeşitli noktalarda hava kalitesi izleme istasyonları kurulmuştur. Bu istasyonların kurulması yüksek maliyet ve sonrasında ise pahalı bakım masrafları gerektirebilmektedir. Bu nedenle, hava kirliliğine ait tahmin çalışmalarının önemi artmıştır. Hava kirliliğinin tahminine yönelik literatürde yer alan çalışmalarda, yapay zeka (AI) teknolojileri kullanılmıştır (Bozdağ ve ark., 2020; Ünalı & Yalçın, 2022). AI'nın gelişmesiyle birlikte, tahmin hedefi ile dış değişkenler arasındaki doğrusal olmayan ilişkiyi hesaba katan yapay sinir ağı modelleri (ANN), makine öğrenmesi (ML), derin öğrenme (DL) algoritmalarının uyarlanabilirlikleri ve sağlamlıkları nedeniyle tahmin görevlerinde mükemmel performansa sahiptir (Mao ve ark., 2021). ML bilgisayar sistemlerinin açıkça programlanmadan belirli bir görevi gerçekleştirmek için kullanıldığı ve bilgisayarın doğrudan verilerden bilgi edinme ve böylece problemleri çözmeyi öğrenme yollarını içeren algoritmaların ve istatistiksel modellerin kullanıldığı bir AI uygulamasıdır (Mahesh, 2020; Aylak ve ark., 2021). AI yöntemleri, sensörler aracılığıyla elde edilen karmaşık ve büyük miktardaki verilerdeki desenleri çıkarabilir (Utku & Can, 2024).

Son zamanlarda literatürde hava kalitesi ile ilgili Coğrafi Bilgi Sistemi (CBS)'ne dayalı istatistiksel analizlerin (Righini ve ark., 2014; Demirarslan & Akıncı, 2018; Kalıpcı & Başer, 2019; Jia, 2019; Vural, 2021; Verma ve ark., 2023) yanında tahmin modellerine dayalı çeşitli (LASSO Regresyonu, Destek Vektör Makineleri (SVM), Rastgele Orman (RF) ve k-en yakın komşu (k-NN) algoritmaları gibi) ML yaklaşımlarının kullanıldığı (Cekim, 2020; Bozdağ ve ark., 2020; Ay & Ekinci, 2022; Çevik ve ark., 2022; Gökçek ve ark., 2022; Oğuz & Pekin, 2022; Kumar & Pande, 2023; Akgün & Barlık, 2023; Aladağ, 2023; Moezzi ve ark., 2024) çalışmaların yapıldığı görülmüştür.

Literatürdeki çalışmalardan farklı olarak bu çalışmada iki kirleticiye ait güncel verilere ML yaklaşımının uygulanması çalışmanın önemini artırmaktadır. Çünkü birden fazla ve farklı hava kirleticileri için çeşitli tahmin modelleri oluşturmak daha uzun sürmektedir. Ayrıca her kirleticiye ait ML hattı oluşturmak daha fazla çalışmayı gerektirmekte ve modelleri dağıtmak, sürdürmek ve izlemek için etkili bir model performansı gerekmektedir.

Çalışma kapsamında Malatya ilinin 2020-2024 yıllarına ait PM₁₀ ve SO₂ verileri Çevre, Şehircilik ve İklim Değişikliği Bakanlığı'nın Ulusal Hava Kalitesi İzleme Ağı (UHKİA)'ndan temin edilmiştir. 2020-2023 yıllarının verileri kullanılarak, 2024 yılının PM₁₀ ve SO₂ kirletici konsantrasyonları lineer regresyon (LR), RF, XGBoost, LightGBM, destek vektör regresyonu (SVR), k-NN ve ANN olmak üzere 7 farklı ML algoritması ile tahmin edilmiştir. Model geliştirme aşaması 2024 yılı için tekrarlanarak, algoritmaların ürettiği sonuçlar gerçek sonuçlarla karşılaştırılmış, algoritmaların performansları ve hata oranları belirlenmiştir.

2. Materyal ve Yöntem

2.1. Çalışma alanı

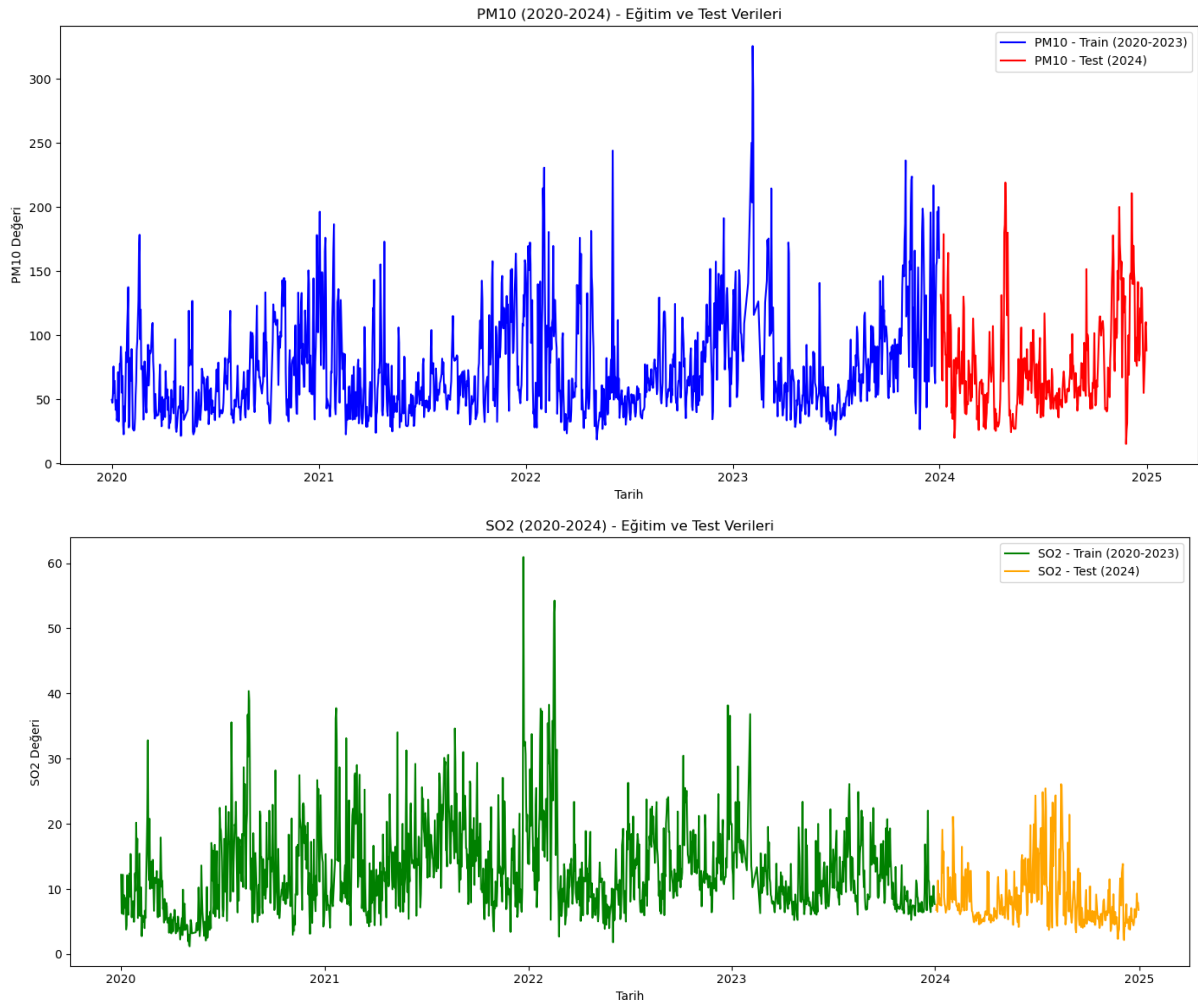
Malatya ili, Doğu Anadolu Bölgesi'nin Yukarı Fırat havzasında 35°34' - 39°03' K enlemleri ile 38°45' - 39°08' D boylamları arasında yer almaktadır (Karabaş, 2019). İl 12.313 km²'lik yüzey alanına, 900 m rakıma ve 800 165 nüfusa sahiptir. Malatya, Doğu, Güneydoğu ve Orta Anadolu arasında yer alan bir ova olup, kuzeyden güneye doğru hafif bir eğimi vardır (ÇŞİB, 2022). İl, kayısı üretimi bakımından ülkemiz için önemli bir merkezdir. Meyvecilik, hayvancılık, mermer-madencilik, tekstil ve

sağlık sektörleri açısından öne çıkmaktadır. İlde, ortalama yıllık sıcaklık 13.6 °C ve ortalama yıllık yağış miktarı 376 mm'dir. İklim özellikleri; kışları soğuk ve yağışlı, yazları ise sıcak ve kuraktır (Varol ve ark., 2021).

2.2. Veri seti ve veri ön işleme

Bu çalışmada, Çevre, Şehircilik ve İklim Değişikliği Bakanlığı'nın UHKİA'dan temin edilen Malatya ili için 1 Ocak 2020 ile 31 Aralık 2024 yılları arası olmak üzere toplam 5 yıllık günlük PM₁₀ ve SO₂ kirletici verilerine ait veri seti kullanılmıştır (UHKİA, 2025). Veri seti tarih, PM₁₀ ve SO₂ alanlarından oluşmaktadır. Veri setinde toplam 1827 adet veri yer almaktadır. Temin edilen verilerden 74 tanesinin çeşitli nedenlerden dolayı ölçümler sırasında eksik/ kayıp olduğu tespit edilmiştir. ML algoritmalarından daha sağlıklı sonuçlar alabilmek için eksik veriye bir ön işleme uygulanması gerekmektedir. Ön işleme için kullanılan yöntemlerden biri eksik ve kayıp verilerin ortalama değerler yoluyla tamamlanmasıdır. Yağmur (2022), Kumar & Pande (2023) tarafından yapılan çalışmalarda olduğu gibi eksik verileri tamamlamak için öncelikle ilgili ayların ortalamaları hesaplanmıştır. Elde edilen ortalama değerler ilgili aydaki eksik olan yerlere eklenmiştir. Çalışmada Python programlama dili kullanılmıştır. Veri ön işleme ve analiz için pandas ve numpy, makine öğrenmesi modellerinin uygulanması için scikit-learn, xgboost ve lightgbm, görselleştirmeler için matplotlib ve seaborn kütüphaneleri kullanılmıştır. Tüm analizler Jupyter Notebook ortamında gerçekleştirilmiştir.

Sonraki aşamada 2020-2023 yıllarına ait olan veriler eğitim için, 2024 yılına ait olan veriler ise test için ayrılmıştır. Şekil 1'de, 2020-2023 yılları eğitim dönemi olarak mavi (PM₁₀) ve yeşil (SO₂) renklerle; 2024 yılı test dönemi ise kırmızı (PM₁₀) ve turuncu (SO₂) renklerle gösterilmiştir.



Şekil 1. Eğitim ve test verilerinin zamana bağlı değerleri.

Çalışmada kullanılan PM₁₀ ve SO₂ değişkenlerine ilişkin sırasıyla ortalama, medyan, minimum, maksimum, standart sapma, çarpıklık, basıklık ve Jarque-Bera normallik testi gibi betimleyici istatistikler hesaplanmış ve Çizelge 1’de sunulmuştur. PM₁₀ değişkeninde ortalama 74.76, medyan 63.84, minimum 15.13 ve maksimum 325.53 olarak hesaplanmıştır. SO₂ değişkeninde ise ortalama 12.10, medyan 10.49, minimum 1.20 ve maksimum 60.94’tür. Standart sapma değerleri PM₁₀ için 38.59 ve SO₂ için 6.71 bulunmuştur; bu durum PM₁₀ değişkeninde SO₂’ye göre daha yüksek varyasyon olduğunu göstermektedir.

Çarpıklık değerleri PM₁₀ için 1.45 ve SO₂ için 1.58 olup her iki serinin de sağa çarpık (pozitif çarpık) dağıldığını göstermektedir. Basıklık değerleri PM₁₀ için 2.86 ve SO₂ için 4.37 bulunmuştur. Bu da özellikle SO₂ serisinde sivri uçlu bir dağılım eğilimi olduğunu göstermektedir. Jarque-Bera normallik testi sonuçlarına göre, her iki seri de 0.00 p-değerine sahip olmakta ve bu durum serilerin normal dağılıma uymadığını göstermektedir.

Çizelge 1. Betimleyici istatistikler

| | PM ₁₀ | SO ₂ |
|----------------------|------------------|-----------------|
| Ortalama | 74.76 | 12.10 |
| Medyan | 63.84 | 10.49 |
| Minimum | 15.13 | 1.20 |
| Maksimum | 325.53 | 60.94 |
| Standart Sapma | 38.59 | 6.71 |
| Çarpıklık (Skewness) | 1.45 | 1.58 |
| Basıklık (Kurtosis) | 2.86 | 4.37 |
| Jarque-Bera p-değeri | 0.00 | 0.00 |

2.3. ML algoritmaları

Çalışmada, 2020-2023 yıllarına ait günlük PM₁₀ ve SO₂ verileri kullanılarak 2024 yılının PM₁₀ ve SO₂ değerlerini tahmin etmek için LR, RF, XGBoost, LightGBM, SVR, k-NN ve ANN algoritmaları kullanılmıştır.

ML algoritmalarından biri olan LR algoritması, değişkenler arasındaki ilişkiyi belirlemek için kullanılan bir tahmin yöntemidir. İki değişken arasındaki ilişkiyi regresyon doğrusunu verilere uydurarak tanımlamaktadır. Bu iki değişkenden biri bağımlı değişken diğeri ise bağımsız değişkendir. Modelleme yapmadan önce bağımlı ve bağımsız değişken arasında bir ilişki olduğundan emin olmak gerekmektedir. Bunun için de saçılım grafiği kullanılarak değişkenler arasındaki ilişkinin gücüne bakılabilir (Denklem 1) (Srivani & Bala, 2020).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

RF algoritması hem sınıflandırma hem de regresyon problemleri için kullanılan bir topluluk öğrenmesi algoritmasıdır (Singh ve ark., 2021). Bu algoritmanın temelini karar ağaçları oluşturmaktadır. Algoritma ile giriş parametreleri kümesinden rastgele parametreler seçilerek çok sayıda karar ağacı oluşturulmaktadır. Bu karar ağaçlarından elde edilen sonuçlar kullanılarak algoritmanın sınıflandırma çıktısı hesaplanmaktadır (Denklem 2) (Bozdağ ve ark., 2020).

$$\hat{y} = (1/M) \sum_{m=1}^M T_m(x) \quad (2)$$

XGBoost algoritması gradyan destekli karar ağaçları kullanılarak hız ve performans için tasarlanmıştır (Dhaliwal ve ark., 2018). Bu algoritma eğitim verilerinin kullanıldığı denetimli öğrenme problemleri için kullanılmaktadır ve tüm senaryolarda ölçeklenebilirdir (Denklem 3) (Wang ve ark., 2017).

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F \quad (3)$$

LightGBM algoritması, birden fazla karar ağacı kullanarak daha güçlü bir model oluşturan bir yöntemdir ve regresyon, sıralama vb. birçok ML uygulamalarında kullanılmaktadır. LightGBM karar ağacı algoritmasına dayandığı için en uygun yaprak bazlı stratejiyi kullanmaktadır. Daha hızlı eğitim, daha düşük bellek ve büyük veri işleme yeteneği gibi özelliklere sahiptir (Li, 2018).

SVR algoritması, doğrusal olmayan problemlerin çözümünde yaygın olarak kullanılmaktadır. 1995 yılında Vapnik ve arkadaşları tarafından geliştirilmiştir (Hong ve ark., 2015; Tien ve ark., 2017). SVR’de veri seti; uyum sağlayan örnekler ve tahmin edilen örnekler olmak üzere iki gruba ayrılmaktadır. Bu algoritma verileri, yüksek boyutlu bir özellik uzayına dönüştürülerek karmaşık olan ilişkileri daha iyi modelleyebilmektedir. Modelin performansını değerlendirmek için doğrulama örnekleri kullanılmaktadır (Denklem 4) (Jiang ve ark., 2020).

$$f(x) = \langle w, \phi(x) \rangle + b \quad (4)$$

k-NN algoritması parametrik olmayan bir öğrenme algoritması olup regresyon ve sınıflandırma problemlerinde kullanılmaktadır (Suguna & Thanushkodi, 2010; Zhang ve ark., 2018; Demirezen ve ark., 2021; Ataş ve ark., 2022). k-NN algoritması sınıfı bilinen örnekler, örnekler arası mesafenin hesaplanmasında kullanmak için bir uzaklık metriği ve en yakın komşu sayısını ifade eden k parametresinden oluşmaktadır (Denklem 5) (Gök, 2017).

$$\hat{y}(x) = (1/k) \sum_{i \in N_k(x)} y_i \quad (5)$$

ANN ise insan beyninin çalışma şeklinden yola çıkılarak geliştirilen ve yapay nöronlardan oluşan bir makine öğrenmesi yöntemidir (Demir Yetiş ve ark., 2021; İlhan ve ark., 2022). Bir nöron ağırlıklarla çarpılmış girdiler ve çıktılardan oluşmaktadır. Yapay sinir ağları temelde girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç katmandan oluşmaktadır. Girdi katmanı giriş verilerinin alındığı ilk katmandır. Gizli katman ise verilerin işlendiği ve özelliklerin çıkarıldığı katmandır. Çıktı katmanı da modelin tahmin veya sınıflandırma sonuçlarının üretildiği katman olarak bilinmektedir. Aşağıda yer alan formülasyon (Denklem 6) yoluyla hesaplanmaktadır.

$$a_j = \sigma(\sum_i w_{ij} a_i + b_j) \quad (6)$$

Burada a_i önceki katmandan gelen çıktılar, w_{ij} ağırlıklar, b_j bias terimi ve σ aktivasyon fonksiyonudur.

Bu çalışmada RF için 50 karar ağacı, XGBoost ve LightGBM için 0,1 öğrenme oranı, SVR için radyal tabanlı fonksiyon kullanılarak $C=100$ ve gamma parametresi 0,1, k-NN için komşu sayısı 5 ve ANN için her birinde 3 nöron bulunan 3 gizli katman kullanılmış, aktivasyon fonksiyonu olarak relu, iterasyon sayısı 500 ve ağırlık optimizasyonu için Adam (Adaptive Moment Estimation) algoritması seçilmiştir.

2.4. Performans metrikleri

Çalışmada kullanılan modellerin performansını test edebilmek için dört metrik kullanılmıştır. Bu metrikler korelasyon katsayısı (R^2), ortalama karesel hata (MSE), kök ortalama karesel hata (RMSE), ortalama mutlak hata (MAE)’dir. R^2 eğitim verileri ile hazırlanan modelin test verileri üzerindeki başarımını gösteren bir metriktir. R^2 değerinin 1’e yakın olması oldukça başarılı olduğunu göstermektedir. MSE ve RMSE, hata karelerinin ortalaması ve karekökünü temsil eder; bu değerlerin sıfıra yaklaşması modelin doğruluğunun arttığını ve tahminlerin gözleme daha yakın olduğunu gösterir. MAE, tahmin ve gerçek değerler arasındaki mutlak farkların ortalamasını verir ve yine sıfıra yakın değerler daha iyi performansı işaret eder. MAPE ise yüzde cinsinden hata oranını hesaplar; düşük MAPE değerleri modelin yüksek doğrulukla tahmin yaptığını gösterir. Denklem 7- 10’da kullanılan performans metriklerine yer verilmiştir. Denklem 7- 10’da yer alan y gerçek değeri, \hat{y} y ’nin tahmin edilen değeri, \bar{y} y değerlerinin ortalamasını ve n ise örnek sayısını ifade etmektedir (Demir Yetiş ve ark., 2024).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

3. Bulgular ve Tartışma

Bu çalışmada, Malatya ili için UHKİA'dan elde edilen 2020-2024 yıllarına ait PM₁₀ ve SO₂ verilerinin ulusal ve uluslararası mevzuata göre kalite değerlendirmesi yapılmıştır. Ayrıca bu parametrelerin 2020- 2023 verileri kullanılarak 2024 yılı tahminine yönelik ML tekniklerinden LR, RF, XGBoost, LightGBM, SVR, k-NN ve ANN algoritmalarının R², MSE, RMSE ve MAE metriklerine göre alınan sonuçlar karşılaştırmalı olarak analiz edilmiştir.

Çalışma kapsamında hava kalitesi parametreleri PM₁₀ ve SO₂ için 2020-2024 arası yıllara ait tanımlayıcı istatistiksel değerler Çizelge 2'de görülmektedir. Buna göre bütün yıllar baz alındığında PM₁₀ için minimum (min) değer 2024 yılında 15.13 µg/m³ ve maksimum (mak) değer 2023 yılında 325.53 µg/m³ olarak ölçülmüştür. SO₂ değeri için ise min değer 2020 yılında 1.20 µg/m³ ve mak değer 2021 yılında 60.94 µg/m³ olarak ölçülmüştür. Yine Çizelge 2'de min, mak ve ortalama (ort) değerleri yıl bazlı değerlendirildiğinde 2020 yılından itibaren 2023 yılına kadar PM₁₀ ve SO₂ için genel bir artış gözlemlenirken, 2024 yılında düşüş görülmektedir. Karabekiroğlu ve ark. (2023) tarafından Malatya'da yapılan çalışmada da hava kirleticileri (PM₁₀ ve SO₂) için 2018 yılından itibaren 2022 yılına kadar topoğrafya ve meteorolojik şartların yanında fabrikalar, ulaşım ve ısınma kaynaklı oluşan kirleticilere bağlı olarak değişebilmesine rağmen değerlerde hep bir artış gözlenmiştir. Ayrıca çizelgede görüldüğü üzere 6 Şubat 2023 yılında yaşanan büyük deprem felaketinin yıkıcı etkisi nedeniyle hava kirleticilerinde de büyük oranda artış olmuştur. 2024 yılında gözlemlenen azalma ise, deprem sonrası yaşanan ölümlere ve ilden göçlerin artışı nedeniyle hava kalitesinde iyileşmeye bağlanabilir. Benzer şekilde Avrupa Çevre Ajansı (EEA) (2024) hava kirliliği emisyonlarında son yirmi yılda dünya genelinde azalma yaşandığını, ancak bu iyileşmeye rağmen, hava kirliliğinin Avrupa'daki en büyük çevresel sağlık riski olmaya devam ettiğini bildirmiştir.

Çizelge 2. Malatya ilinin 2020-2024 yıllarına ait PM₁₀ ve SO₂'nin min, mak ve ort değerleri

| Yıl | PM ₁₀ (µg/m ³) | | | SO ₂ (µg/m ³) | | |
|------|---------------------------------------|--------|-------|--------------------------------------|-------|-------|
| | Min | Mak | Ort | Min | Mak | Ort |
| 2020 | 21.4 | 178.31 | 66.75 | 1.20 | 40.38 | 10.99 |
| 2021 | 22.62 | 196.29 | 70.66 | 3.42 | 60.94 | 14.98 |
| 2022 | 18.64 | 243.97 | 75.39 | 1.82 | 54.26 | 14.18 |
| 2023 | 21.96 | 325.53 | 88.57 | 5.24 | 36.84 | 11.53 |
| 2024 | 15.13 | 219.08 | 75.99 | 2.17 | 26.09 | 8.2 |

Hava kalitesini değerlendirmeye yönelik "Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği" (HKDYY, 2008), Dünya Sağlık Örgütü'nün Hava Kalitesi Rehberi - Global Air Quality Guidelines (WHO, 2021) ve Avrupa Birliği Direktifi'nin Avrupa Hava Kalitesi Durumu - Europe's Air Quality Status (EU, 2024) gibi ulusal ve uluslararası mevzuatta yer alan ortalama süreye göre PM₁₀ ve SO₂ kirleticileri için sınır değerler Çizelge 3'de gösterilmiştir. Çizelgede PM₁₀ parametresi için EU ve Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği (HKDYY)'nin izin verdiği 24 saatlik ortalama sürede sınır değerler 50 µg/ m³ ve yıllık ortalama sürede 40 µg/ m³'tür. Çizelge 4'te sınır değerlerini aşan gün sayısına göre; 24 saatlik ortalama süreye ait sınır 50 µg/m³ değeri en düşük 2020 yılında % 64 oranında (231 gün), en yüksek ise 2023 yılında %78 (263 gün) oranında aşılmıştır. Yıllık ortalama

süreye ait sınır $40 \mu\text{g}/\text{m}^3$ değeri en düşük 2020 yılında % 81 oranında (291 gün), en yüksek ise 2023 yılında % 90 oranında (303 gün) aşılmıştır. Çizelge 3'te sınır değerlerini aşan gün sayısına göre; SO₂ parametresi için ise, 24 saatlik ortalama sürede EU ve HKDYY'nin izin verdiği $125 \mu\text{g}/\text{m}^3$, yıllık ortalama sürede de sadece HKDYY'nin izin verdiği $20 \mu\text{g}/\text{m}^3$ sınır değerleri verilmiştir. Saatlik ve 24 saatlik ortalama süreye ait sınır değerler olan $350 \mu\text{g}/\text{m}^3$ ve $125 \mu\text{g}/\text{m}^3$ değerleri bütün yıllarda aşılmamış olup, yıllık ortalama $20 \mu\text{g}/\text{m}^3$ sınır değerleri 2021 yılında % 22'lik oranla 78 gün aşılmıştır. Yılmaz (2022) tarafından yapılan çalışmada da 2021 yılına ait yıllık ortalama SO₂ konsantrasyonları benzer şekilde HKDYY (2008)'de yer alan sınır değeri ($20 \mu\text{g}/\text{m}^3$) aşmamıştır. Ancak Sari ve Esen (2019) tarafından yapılan çalışmada ise HKDYY'de yer alan bu sınır değerinin birçok ilde aşıldığı bildirilmiştir. Hava kirlenmelerinin; emisyonları, kimyasal bileşimleri, parçalanma süreleri, reaksiyon özellikleri, kısa ve uzun mesafelere ulaşımı ve yayılma yetenekleri birbirinden farklı olduğu için kirlenici kaynakların baskısı altında olmayan illerde bile mevzuat değerleri aşılmakta veya çeşitli etkilerle tam tersi durumlar da yaşanabilmektedir (Akgün & Barlık, 2023).

Çizelge 3. PM₁₀ ve SO₂ için ulusal ve uluslararası mevzuatta kirlenici sınır değerleri

| Kirlenici | Ortalama süre | WHO (2021) | EU (2024) | HKDYY (2008) |
|------------------|---------------|-----------------------------|------------------------------|------------------------------|
| PM ₁₀ | 24 saatlik | $45 \mu\text{g}/\text{m}^3$ | $50 \mu\text{g}/\text{m}^3$ | $50 \mu\text{g}/\text{m}^3$ |
| | Yıllık | $15 \mu\text{g}/\text{m}^3$ | $40 \mu\text{g}/\text{m}^3$ | $40 \mu\text{g}/\text{m}^3$ |
| SO ₂ | Saatlik | - | $350 \mu\text{g}/\text{m}^3$ | $350 \mu\text{g}/\text{m}^3$ |
| | 24 saatlik | $40 \mu\text{g}/\text{m}^3$ | $125 \mu\text{g}/\text{m}^3$ | $125 \mu\text{g}/\text{m}^3$ |
| | Yıllık | - | - | $20 \mu\text{g}/\text{m}^3$ |

Çizelge 4. PM₁₀ ve SO₂ parametrelerinin yıl bazlı ulusal ve uluslararası mevzuatlara göre müsaade edilen sınır değerlerini aşan gün sayısı

| Yıl | Veri alınan gün sayısı | | PM ₁₀ ($\mu\text{g}/\text{m}^3$) | | SO ₂ ($\mu\text{g}/\text{m}^3$) |
|------|---|--|---|--|---|
| | PM ₁₀ ($\mu\text{g}/\text{m}^3$) | SO ₂ ($\mu\text{g}/\text{m}^3$) | 24 saat için ortalama sürenin aşıldığı yüzdeler ve gün sayısı ($50 \mu\text{g}/\text{m}^3$) | Yıllık için ortalama sürenin aşıldığı yüzdeler ve gün sayısı ($40 \mu\text{g}/\text{m}^3$) | Yıllık için ortalama sürenin aşıldığı yüzdeleri ve gün sayısı ($20 \mu\text{g}/\text{m}^3$) |
| 2020 | 361 | 344 | 64 (231) | 81 (291) | 10 (33) |
| 2021 | 362 | 361 | 65 (235) | 83 (302) | 22 (78) |
| 2022 | 362 | 361 | 73 (264) | 85 (309) | 17 (62) |
| 2023 | 337 | 354 | 78 (263) | 90 (303) | 5 (19) |
| 2024 | 361 | 362 | 74 (266) | 88 (317) | 4 (14) |

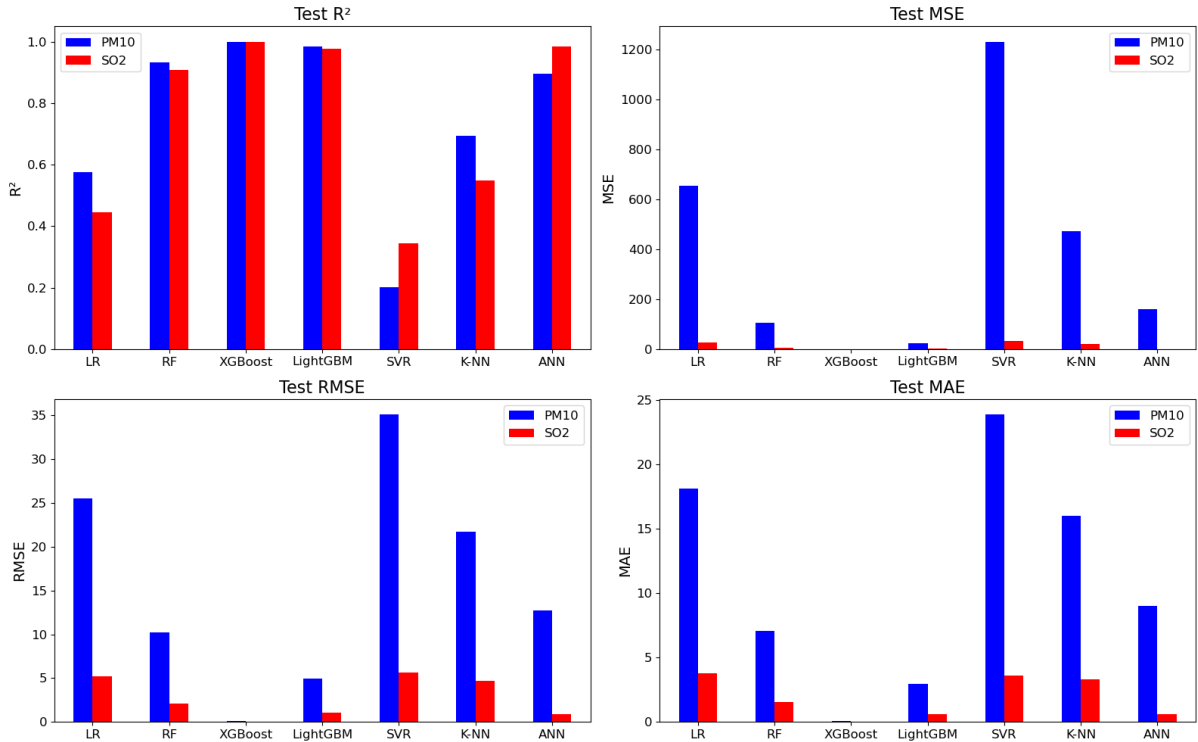
PM₁₀ ve SO₂ kirlenici parametrelerin 2020-2023 verileri kullanılarak 2024 yılı tahminine yönelik ML tekniklerinden LR, RF, XGBoost, LightGBM, SVR, k-NN ve ANN algoritmalarının karşılaştırmalı bir analizine yer verilmiştir. Makine öğrenmesi modellerinde bağımsız değişkenler ilgili parametrelerin geçmiş gözlemlerinden türetilen gecikmeli (lag) değişkenlerden oluşmaktadır. Özellikle 90 günlük lag yapılandırmasında, PM₁₀ ve SO₂ değişkenlerinin her biri için bir önceki 90 günün değerleri bağımsız değişken olarak kullanılmıştır (örneğin: PM_{10_lag_1}, PM_{10_lag_2}, ..., PM_{10_lag_90} ve SO_{2_lag_1}, SO_{2_lag_2}, ..., SO_{2_lag_90}). Bu yapılandırma, modellerin geçmiş dönem eğilimlerini ve mevsimsel desenleri öğrenebilmesine olanak sağlamış ve model performansını güçlendirmiştir. Her bir algoritmanın R², MSE, RMSE ve MAE metrikleri hesaplanarak bir karşılaştırma yapılmıştır. Çizelge 5 ve Şekil 1'de PM₁₀ ve SO₂ özelliklerine göre LR, RF, XGBoost, LightGBM, SVR, k-NN ve ANN algoritmalarının R², MSE, RMSE ve MAE metriklerine göre elde edilen deneysel sonuçlar görülmektedir.

PM₁₀ ve SO₂ hava kirliliği parametrelerinin tahmini için çeşitli makine öğrenmesi algoritmaları uygulanmış ve farklı gecikme (lag) değerlerinin model performansına etkisi incelenmiştir. Zaman serisi analizlerinde lag değişkenlerinin kullanımı, geçmiş gözlemlerin mevcut ve gelecekteki değerler üzerindeki etkisinin modellenmesine olanak tanımakta ve özellikle çevresel verilerde sıklıkla gözlenen mevsimsel ve kısa vadeli desenlerin yakalanmasına önemli katkılar sağlamaktadır. Çalışmada 30, 90, 120 ve 367 günlük lag yapılandırmaları denenmiş ve en kararlı ve yüksek performanslı sonuçların 90 günlük lag ile elde edildiği belirlenmiştir. Bu durum, 90 günlük geçmiş bilgisinin hem kısa vadeli

dalgalanmaları hem de mevsimsel etkileri modellemeye yeterli düzeyde katkı sağladığını ortaya koymaktadır.

Çizelge 5. Test verisi için makine öğrenmesi algoritmalarının performans sonuçları

| Model | Feature | R2 | MSE | RMSE | MAE |
|------------------|------------------|----------|----------|----------|----------|
| LinearRegression | PM ₁₀ | 0.575948 | 652.7814 | 25.54959 | 18.14836 |
| | SO ₂ | 0.444018 | 26.80713 | 5.177561 | 3.785273 |
| RandomForest | PM ₁₀ | 0.931878 | 104.8665 | 10.24043 | 7.090137 |
| | SO ₂ | 0.908702 | 4.401989 | 2.098092 | 1.525896 |
| XGBoost | PM ₁₀ | 0.999992 | 0.012818 | 0.113216 | 0.076901 |
| | SO ₂ | 0.99998 | 0.000967 | 0.031096 | 0.021497 |
| LightGBM | PM ₁₀ | 0.984083 | 24.50218 | 4.949968 | 2.983875 |
| | SO ₂ | 0.977161 | 1.101217 | 1.049389 | 0.61398 |
| SVR | PM ₁₀ | 0.200161 | 1231.262 | 35.08935 | 23.87817 |
| | SO ₂ | 0.343415 | 31.6578 | 5.626527 | 3.580222 |
| k-NN | PM ₁₀ | 0.692911 | 472.7294 | 21.74234 | 16.01155 |
| | SO ₂ | 0.548647 | 21.76235 | 4.665014 | 3.314497 |
| ANN | PM ₁₀ | 0.895205 | 161.3197 | 12.70117 | 8.999542 |
| | SO ₂ | 0.984125 | 0.765441 | 0.874895 | 0.602135 |



Şekil 2. Test veri setine ait R², MSE, RMSE ve MAE sonuçları.

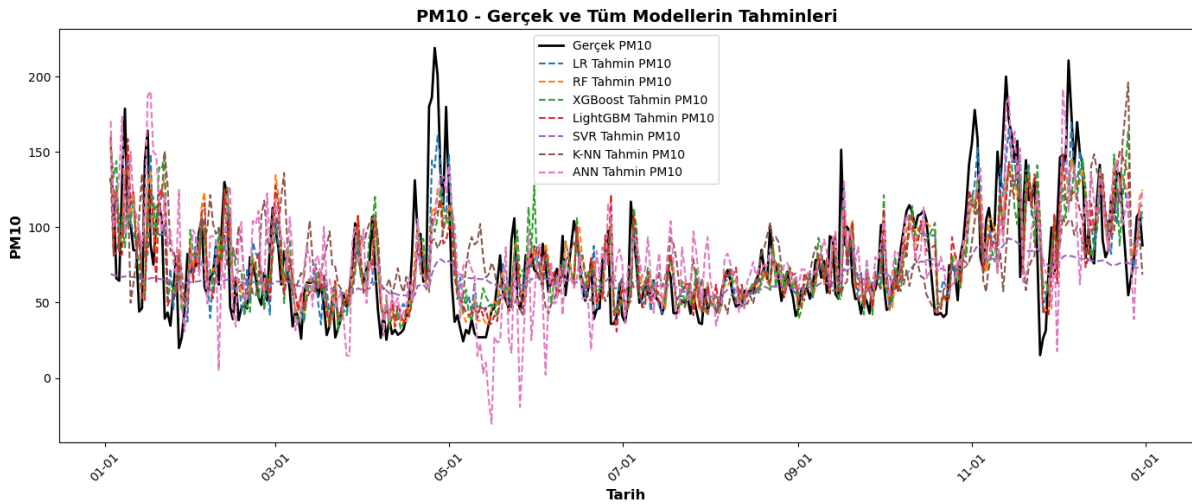
Çizelge 5 ve Şekil 2 incelendiğinde PM₁₀ değişkeni açısından 90 günlük lag ile XGBoost (R² ≈ 0.9999), LightGBM (R² ≈ 0.9841) ve Random Forest (R² ≈ 0.9319) algoritmaları en yüksek determinasyon katsayılarına ulaşmıştır. Özellikle XGBoost modeli, RMSE (≈ 0.11) ve MAE (≈ 0.08) değerlerinde oldukça düşük hata oranları göstererek dikkat çekmiştir. ANN (R² ≈ 0.8952) ve K-NN (R² ≈ 0.6929) orta düzeyde performans sergilerken, Lineer Regresyon (LR, R² ≈ 0.5759) ve SVR (R² ≈ 0.2002) görece düşük başarı performansı göstermiştir. SO₂ değişkeni için yapılan tahminlerde, XGBoost (R² ≈ 0.99998), LightGBM (R² ≈ 0.9772) ve Random Forest (R² ≈ 0.9087) modelleri en güçlü uyumu

sağlamıştır. XGBoost'un RMSE (≈ 0.0311) ve MAE (≈ 0.0215) düzeyinde neredeyse sıfıra yakın hata değerleri elde etmesi, modelin SO₂ serisini çok yüksek doğrulukla öğrendiğini göstermektedir. ANN ($R^2 \approx 0.9841$) SO₂ tahminlerinde dikkate değer bir başarı sergilerken, k-NN ve LR orta seviyede, SVR ise görece sınırlı bir performans göstermiştir.

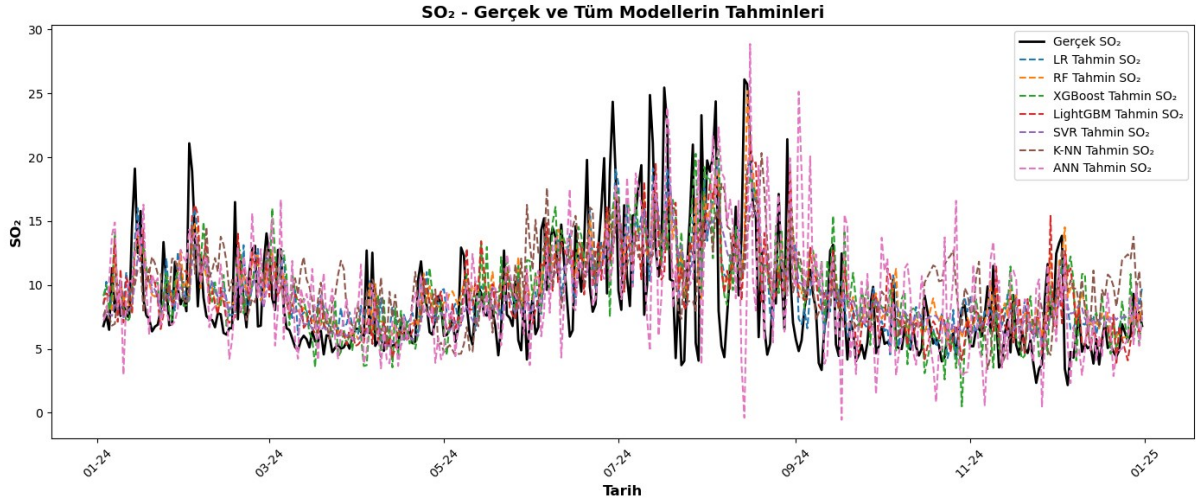
Oğuz & Pekin (2022) tarafından yapılan çalışmada PM₁₀ konsantrasyonunun tahmini için ANN, SVR, RF regresyonu ve LASSO algoritmalarını kullanmışlar ve ANN algoritmasında diğer algoritmalara göre daha iyi sonuç verdiği bildirilmiştir. Gökçek ve ark. (2022) da Kayseri ili için k-NN, SVR, RF, ANN ve LR ML algoritmalarını kullanarak 3 istasyon için PM₁₀ konsantrasyonu bazlı tahmin yapılmış, en iyi sonucun 3 istasyon için de SVR algoritmasında elde edildiğini raporlamışlardır. Bozdağ ve ark. (2020) da Ankara iline ait 7 istasyondan elde edilen PM₁₀ konsantrasyonları LASSO, SVR, RF, k-NN, XGBoost ve ANN gibi ML algoritmaları kullanılarak tahmin edilmiş ve elde ettikleri en iyi tahmin sonuçlarının ANN algoritması olduğunu ifade etmişlerdir.

PM₁₀ konsantrasyonuna ait 2024 yılı için kullanılan ML algoritmalarının gerçek PM₁₀ konsantrasyon değerleriyle tahmin edilen PM₁₀ konsantrasyon değerlerinin karşılaştırması ise Şekil 2'de verilmiştir. Şekil 3'e göre PM₁₀ için Malatya iline ait gerçek 2024 verileri ile tahmin edilen değerler karşılaştırılmıştır. Grafikte görülebileceği üzere gerçek değerler ile tahmin edilen değerler kullanılan yöntemlere göre neredeyse birbirleri ile aynı görünmektedir. Bu da kullanılan ML yöntemlerinin oldukça başarılı sonuçlar verdiğini göstermektedir. Şekil 4'te SO₂ konsantrasyonuna ait 2024 yılı için kullanılan ML algoritmalarının gerçek SO₂ konsantrasyon değerleriyle tahmin edilen SO₂ konsantrasyon değerlerinin karşılaştırmalı bir grafiği verilmiştir.

Şekil 3 ve 4'teki grafiklerden de görülebileceği üzere modellerin genel eğilimleri başarılı şekilde yakaladığı ve ani değişimlerde dahi büyük sapmalar yaşanmadığı görülmektedir. PM₁₀ tahminlerinde LR ve RF, istikrarlı ve dengeli sonuçlar sunarak genel eğilimleri başarılı şekilde takip etmiştir. XGBoost ve LightGBM, özellikle ani yükseliş ve düşüşlerde yüksek hassasiyet göstermiş ancak bu durum zaman zaman aşırı öğrenme (overfitting) yol açmıştır. ANN modeli, bazı dönemlerde agresif tahminler üreterek sapmalar göstermiştir. SO₂ tahminlerinde ise LR ve SVR, en istikrarlı ve dengeli sonuçları verirken, XGBoost ve LightGBM yine hızlı değişimlere iyi yanıt vermiş ancak yer yer aşırı dalgalanmalar gözlenmiştir. Genel olarak, ensemble modeller değişkenlikte daha hassas, doğrusal ve basit modeller ise daha kararlı bir performans sergilemiştir.



Şekil 3. ML algoritmalarının gerçek PM₁₀ ve tahmin edilen PM₁₀ değerleri ile karşılaştırılması.



Şekil 4. ML algoritmalarının gerçek SO₂ ve tahmin edilen SO₂ değerleri ile karşılaştırılması.

Kullanılan ML modellerinin eğitim ve test kümeleri üzerindeki performansları karşılaştırılarak overfitting analizi yapılmıştır. Analiz sonuçlarına göre, LR, SVR, k-NN, RF ve ANN modellerinde anlamlı bir overfitting eğilimi gözlemlenmemiş, bu modeller eğitim ve test kümelerinde benzer ve dengeli performans sergilemiştir. Buna karşılık, XGBoost ve LightGBM modellerinde eğitim setinde çok yüksek başarı elde edilirken test setinde nispeten daha düşük performans gözlemlenmiş, bu durum hafif düzeyde overfitting işareti olarak değerlendirilmiştir. Genel olarak, çalışmada kullanılan modellerin büyük bölümü aşırı öğrenmeden kaçınırken, XGBoost ve LightGBM’de dikkat edilmesi gereken sınırlı bir overfitting eğilimi belirlenmiştir. Bu sonuçlar, model seçiminde yalnızca eğitim başarımına değil, test setindeki genelleme yeteneğine de odaklanmanın önemini ortaya koymaktadır.

4. Sonuçlar

Bu çalışmada, Malatya ilindeki hava kalitesini değerlendirmek ve makine öğrenmesi teknikleri ile tahmin etmek amacıyla UHKİA’dan 01 Ocak 2020-31 Aralık 2024 tarihlerini kapsayacak şekilde ildeki 1 istasyondan alınan PM₁₀ ve SO₂ kirletici parametre verileri kullanılmıştır. Bu kapsamda hava kirleticilerine ait veriler HKDYY gibi ulusal ile EU ve WHO gibi uluslararası mevzuatta yer alan sınır değerlerle mukayese edilerek değerlendirilmiştir. Ayrıca 2020-2023 yılı verileri dikkate alınarak ML tekniklerine (LR, RF, XGBoost, LightGBM, SVR, k-NN ve ANN) göre hava kirliliğinin 2024 yılı için tahmin ve karşılaştırmalı analizi yapılmıştır.

2020 yılından 2023 yılına kadar PM₁₀ ve SO₂ kirletici parametreleri için değerlerde bir artış gözlemlenirken, 2024 yılında düşüş görülmektedir. 2023 yılında yaşanan büyük deprem felaketinin hava kirleticilerinde büyük oranda artışa neden olduğu, sonrasında ise (2024 yılı) yaşanan ölümler ve ilden göçlerin etkisiyle azalma ve hava kalitesinde iyileşme olduğu söylenebilir. Hava kalitesinin değerlendirilmesi kapsamında PM₁₀ parametresi için 24 saatlik ortalama süre sınır değeri 2023 yılında %79 oranla, yıllık ortalama süre sınır değeri ise 2023 yılında % 91 oranla aşılmıştır. SO₂ parametresi için ise sadece yıllık ortalama süre sınır değeri 2021 yılında % 21’lik oranla aşılmıştır.

Performans metrikleri değerlendirildiğinde, PM₁₀ ve SO₂ için en iyi performans XGBoost ve LightGBM algoritmaları ile elde edilmiş; örneğin PM₁₀ için XGBoost $R^2 = 0.99999$, $MSE \approx 0.01$, $RMSE \approx 0.11$, $MAE \approx 0.08$; SO₂ için XGBoost $R^2 = 0.99998$, $MSE \approx 0.00097$, $RMSE \approx 0.03$ ve $MAE \approx 0.02$ değerleri gözlemlenmiştir. RF ve ANN modelleri de yüksek başarı göstermiştir. LR ve SVR ise daha mütevazı fakat dengeli performans sunarken, k-NN orta düzeyde bir başarı sağlamıştır. Bu çalışma, ilerisi için planlanan çalışmalar açısından farklı bölgelerden alınmış daha fazla izleme sonuçları ve birden fazla kirletici parametre ile çeşitli hava kalitesi endekslerinin tahmininde derin öğrenme tekniklerinin kullanılmasını önermektedir.

Kaynakça

- Akgün, M., & Barlık, N. (2023). Makine öğrenmesi algoritmaları kullanılarak hava kalitesi indeksinin tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, 51, 97-107. <https://doi.org/10.31590/ejosat.1241170>
- Aladağ, E. (2023). Ampirik mod ayrıştırmasına dayalı ARIMA modeli kullanılarak Van ili hava kirliliğinin tahmini. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 28(2), 495-509. <https://doi.org/10.53433/yyufbed.1220578>
- Aladağ, E. (2023). The influence of meteorological factors on air quality in the province of Van, Turkey. *Water, Air, & Soil Pollution*, 234(4), 259. <https://doi.org/10.1007/s11270-023-06265-0>
- Ataş, M., Yeşilnacar, M. İ., & Demir Yetiş, A. (2022). Novel machine learning techniques based hybrid models (LR-KNN-ANN and SVM) in prediction of dental fluorosis in groundwater. *Environmental Geochemistry and Health*, 44(11), 3891-3905. <https://doi.org/10.1007/s10653-021-01148-x>
- Ay, Ş., & Ekinci, E. (2022). Ozon konsantrasyonlarını modellemek için makine öğrenmesi ve derin öğrenme yöntemlerinin karşılaştırılması. *Journal of Intelligent Systems: Theory and Applications*, 5(2), 106-118. <https://doi.org/10.38016/jista.1054331>
- Aylak, B. L., Oral, O., & Yazıcı, K. (2021). Yapay zeka ve makine öğrenmesi tekniklerinin lojistik sektöründe kullanımı. *El-Cezerî Fen ve Mühendislik Dergisi*, 8(1), 74-93. <https://doi.org/10.31202/ecjse.776314>
- Behcet, R., & Yakın, A. (2020). Malatya ili trafik kaynaklı hava kirlleticilerinin emisyon envanteri. *Journal of the Institute of Science and Technology*, 10(4), 2783-2790. <https://doi.org/10.21597/jist.704308>
- Bozdağ, A., Dokuz, Y., & Gökçek, Ö. B. (2020). Spatial prediction of PM10 concentration using machine learning algorithms in Ankara, Turkey. *Environmental Pollution*, 263, 114635. <https://doi.org/10.1016/j.envpol.2020.114635>
- Cekim, H. O. (2020). Forecasting PM10 concentrations using time series models: a case of the most polluted cities in Turkey. *Environmental Science and Pollution Research*, 27(20), 25612-25624. <https://doi.org/10.1007/s11356-020-08164-x>
- Çevik, Z. A., Özbeyaz, A., & Demirci, Y. (2022). Makine öğrenme algoritması kullanarak yenilenebilir enerji ile hava kirliliği arasındaki ilişkinin incelenmesi.
- ÇŞİB. (2022). Malatya ili 2022 yılı çevre durum raporu. Çevre, Şehircilik ve İklim Değişikliği İl Müdürlüğü. <https://webdosya.csb.gov.tr/db/ced/icerikler/malatya-ilcdr-2022-20240726150652.pdf>
- Demir Yetiş, A., İlhan, N., & Kara, H. (2024). Integrating deep learning and regression models for accurate prediction of groundwater fluoride contamination in old city in Bitlis province, Eastern Anatolia Region, Türkiye. *Environmental Science and Pollution Research*, 31(34), 47201-47219. <https://doi.org/10.1007/s11356-024-34194-w>
- Demir Yetiş, A., Yeşilnacar, M. İ., & Atas, M. (2021). A machine learning approach to dental fluorosis classification. *Arabian Journal of Geosciences*, 14(2), 95. <https://doi.org/10.1007/s12517-020-06342-2>
- Demirarslan, K. O., & Akıncı, H. (2018). CBS ve hava kalitesi verileri kullanılarak Marmara Bölgesinin kış sezonunda hava kalitesinin değerlendirilmesi. *Doğal Afetler ve Çevre Dergisi*, 4(1), 11-27. <https://doi.org/10.21324/dacd.344564>
- Demirezen, M. U., Civrizoğlu, A., & Yavanoğlu, U. (2021). Sualtı objelerinin makine öğrenmesi yöntemleri ile tespitinde zaman serisi-görüntü dönüşümü tabanlı yeni yaklaşımlar. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 36(3), 1399-1416. <https://doi.org/10.17341/gazimmfd.826453>
- Dhaliwal, S. S., Nahid, A. A., & Abbas, R. (2018). Effective intrusion detection system using XGBoost. *Information*, 9(7), 149. <https://doi.org/10.3390/info9070149>
- EEA. (2024). Air pollution. European Environment Agency. Erişim tarihi: 11.02.2025. <https://www.eea.europa.eu/en/topics/in-depth/air-pollution?activeTab=fa515f0c-9ab0-493c-b4cd-58a32dfaae0a>
- EU. (2024). Europe's air quality status 2024. European Environment Agency. Erişim tarihi: 17.02.2025. <https://www.eea.europa.eu/publications/europes-air-quality-status-2024>

- Gök, M. (2017). Makine öğrenmesi yöntemleri ile akademik başarının tahmin edilmesi. *Gazi University Journal of Science Part C: Design and Technology*, 5(3), 139-148.
- Gökçek, B., Şaşa, N., Dokuz, Y., & Bozdağ, A. (2022). PM10 parametresinin makine öğrenmesi algoritmaları ile mekânsal analizi, Kayseri ili örneği. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 24(70), 65-80. <https://doi.org/10.21205/deufmd.2022247008>
- HKDYY. (2008). Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği. Çevre ve Orman Bakanlığı, Resmî Gazete (06.06.2008). Resmî Gazete Sayısı: 26898.
- Hong, H., Pradhan, B., Xu, C., & Bui, D. T. (2015). Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena*, 133, 266-281. <https://doi.org/10.1016/j.catena.2015.05.019>
- İlhan, N., Demir Yetiş, A., Yeşilnacar, M. İ., & Atasoy, A. D. S. (2022). Predictive modelling and seasonal analysis of water quality indicators: three different basins of Şanlıurfa, Turkey. *Environment, Development and Sustainability*, 24(3), 3258-3292. <https://doi.org/10.1007/s10668-021-01566-y>
- Jia, Q. (2019). Urban air quality assessment method based on GIS technology. *Applied Ecology and Environmental Research*, 17, 9367-9375. http://doi.org/10.15666/aecer/1704_93679375
- Jiang, H., Li, Y., Zhou, C., Hong, H., Glade, T., & Yin, K. (2020). Landslide displacement prediction combining LSTM and SVR algorithms: A case study of Shengjibao Landslide from the Three Gorges Reservoir Area. *Applied Sciences*, 10(21), 7830. <https://doi.org/10.3390/app10217830>
- Kalıpcı, E., & Başer, V. (2019). Coğrafi Bilgi Sistemi (CBS) ve hava kalitesi verileri kullanılarak Türkiye'nin hava kirliliğinin değerlendirilmesi. *Karadeniz Fen Bilimleri Dergisi*, 9(2), 377-389. <https://doi.org/10.31466/kfbd.645698>
- Karabaş, B. (2019). *Use of geographic information system for evaluating the some geotechnical properties in Malatya, Turkey*. Yüksek Lisans Tezi, Hasan Kalyoncu Üniversitesi, Fen Bilimleri Enstitüsü, Gaziantep, Türkiye.
- Karabekiroğlu, B., Zere, H. A., & Toros, H. (2023). Malatya 2018-2022 yılları hava kirliliği verilerinin değerlendirilmesi.
- Kırımhan, S. (2006). *Hava kirliliği ve kontrolü*. Ankara: Turhan Kitabevi.
- Kumar, K., & Pande, B. P. (2023). Air pollution prediction with machine learning: a case study of Indian cities. *International Journal of Environmental Science and Technology*, 20(5), 5333-5348. <https://doi.org/10.1007/s13762-022-04241-5>
- Li, J. (2018). Monthly housing rent forecast based on lightgbm (light gradient boosting) model. *International Journal of Intelligent Information and Management Science*, 7(6), 09-03.
- Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*, 9(1), 381-386.
- Mao, W., Wang, W., Jiao, L., Zhao, S., & Liu, A. (2021). Modeling air quality prediction using a deep learning approach: Method optimization and evaluation. *Sustainable Cities and Society*, 65, 102567. <https://doi.org/10.1016/j.scs.2020.102567>
- Moezzi, S. M. M., Mohammadi, M., Mohammadi, M., Saloglu, D., & Sheikholeslami, R. (2024). Machine learning insights into PM2.5 changes during COVID-19 lockdown: LSTM and RF analysis in Mashhad. *Environmental Monitoring and Assessment*, 196(5), 453. <https://doi.org/10.1007/s10661-024-12567-5>
- Mutlu, N. M., & Atahanlı, E. B. (2024). Veri madenciliği ile hava kalitesi tahmini: İstanbul örneği. *Bilişim Teknolojileri Dergisi*, 17(3), 139-158. <https://doi.org/10.17671/gazibtd.1426942>
- Müezzinoğlu, A. (2005). *Hava Kirliliği ve Kontrolünün Esasları*. Dokuz Eylül Üniversitesi Yayınları.
- Oğuz, K., & Pekin, M. A. (2022). Makine öğrenme algoritmaları ile PM10 konsantrasyon tahmini. *Journal of Advanced Research in Natural and Applied Sciences*, 8(2), 201-213. <https://doi.org/10.28979/jarnas.981202>
- Öztürk, D., & Bayram, T. (2019). Van ili kent merkezinde hava kirliliği. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 8(3), 1142-1153.
- Rakholia, R., Le, Q., Ho, B. Q., Vu, K., & Carbajo, R. S. (2023). Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam. *Environment international*, 173, 107848. <https://doi.org/10.1016/j.envint.2023.107848>

- Righini, G., Cappelletti, A., Ciucci, A., Cremona, G., Piersanti, A., Vitali, L., & Ciancarella, L. (2014). GIS based assessment of the spatial representativeness of air quality monitoring stations using pollutant emissions data. *Atmospheric Environment*, 97, 121-129. <https://doi.org/10.1016/j.atmosenv.2014.08.015>
- Sari, M. F., & Esen, F. (2019). PM10 ve SO2 konsantrasyonları ve meteorolojik parametrelerin konsantrasyonlar üzerine etkileri. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 8(2), 689-697. <https://doi.org/10.28948/ngumuh.598226>
- Singh, V., Gourisaria, M. K., & Das, H. (2021, September). *Performance analysis of machine learning algorithms for prediction of liver disease*. IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 1-7), University of Malaya, Malaysia.
- Sravani, B., & Bala, M. M. (2020, June). *Prediction of student performance using linear regression*. International Conference for Emerging Technology (INCET) (pp. 1-5). IEEE.
- Suguna, N., & Thanushkodi, K. (2010). An improved k-nearest neighbor classification using genetic algorithm. *International Journal of Computer Science Issues*, 7(2), 18-21.
- THHP. (2024). Türkiye'nin hava kalitesi karnesi (Kara Rapor 2024). Erişim tarihi: 12.02.2025. <https://temizhavahakki.org/turkiyenin-hava-kalitesi-karnesi-kara-rapor-2024/>
- Tırıs, M., Kalafatoğlu, E., Okutan, H. (1993) *Hava Kirliliği Kaynakları ve Kontrolü*. TÜBİTAK MAM, Gebze, Kocaeli.
- Tien Bui, D., Tuan, T. A., Hoang, N. D., Thanh, N. Q., Nguyen, D. B., Van Liem, N., & Pradhan, B. (2017). Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides*, 14, 447-458. <https://doi.org/10.1007/s10346-016-0711-9>
- UHKİA. (2025). Ulusal hava kalite izleme ağı. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı. Erişim tarihi: 05.02.2025. https://sim.csb.gov.tr/STN/STN_Report/StationDataDownloadNew
- Utku, A., & Can, Ü. (2024). An effective hybrid model for predicting air quality of Ankara. *NATURENGS*, 5(1), 1-8. <https://doi.org/10.46572/naturengs.1411983>
- Ünalı, S., & Yalçın, N. (2022). Hava kirliliğinin makine öğrenmesi tabanlı tahmini: Başakşehir örneği. *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 4(1), 35-44. <https://doi.org/10.46387/bjesr.1055946>
- Varol, M., Gündüz, K., & Sünbül, M. R. (2021). Pollution status, potential sources and health risk assessment of arsenic and trace metals in agricultural soils: A case study in Malatya province, Turkey. *Environmental Research*, 202, 111806. <https://doi.org/10.1016/j.envres.2021.111806>
- Verma, S., Gangwar, T., Singh, J., Prakash, D., & Payra, S. (2023). Urban air quality monitoring and modelling using ground monitoring, remote sensing, and GIS. In *Geospatial analytics for environmental pollution modeling: analysis, control and management* (pp. 213-247). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-45300-7_9
- Vural, E. (2021). Güneydoğu Anadolu Bölgesi illerinin CBS kullanarak hava kalitesinin mekânsal değişiminin incelenmesi (2007-2019). *Doğal Afetler ve Çevre Dergisi*, 7(1), 124-135. <https://doi.org/10.21324/dacd.718450>
- Wang, W., Shi, Y., Lyu, G., & Deng, W. (2017). Electricity consumption prediction using XGBoost based on discrete wavelet transform. *DEStech Transactions on Computer Science and Engineering*, (aiea).
- WHO. (2021). Global Air Quality Guidelines. Erişim tarihi: 15.02.2025. <https://www.who.int/news-room/questions-and-answers/item/who-global-air-quality-guidelines>
- WHO. (2024). Ambient (outdoor) air pollution. Erişim tarihi: 11.02.2025. [https://www.who.int/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- WHO. (2025). Air pollution. Erişim tarihi: 11.02.2025. https://www.who.int/health-topics/air-pollution#tab=tab_1
- Yağmur, E. Ç. (2022). Atmosferik partikül maddelerin makine öğrenmesi ile tahmini: Beşiktaş, İstanbul örneği. *Konya Journal of Engineering Sciences*, 10(4), 807-826. <https://doi.org/10.36306/konjes.1082866>
- Yılmaz, Z. (2022). Meteorolojik parametrelerin hava kirliliğine etkisinin istatistiksel analizi–Muş ili (2021). *Mühendislik Bilimleri ve Tasarım Dergisi*, 10(4), 1182-1193. <https://doi.org/10.21923/jesd.1100006>

Zhang, S., Cheng, D., Deng, Z., Zong, M., & Deng, X. (2018). A novel kNN algorithm with data-driven k parameter computation. *Pattern Recognition Letters*, 109, 44-54.
<https://doi.org/10.1016/j.patrec.2017.09.036>