

HAVA KİRLİLİĞİ DEĞERLERİNİN MAKİNE ÖĞRENMESİ VE HİBRİT YÖNTEMLER İLE TAHMİNİ: İSTANBUL ÖRNEĞİ

Prediction of Air Pollution Levels Using Machine Learning and Hybrid Methods: The Case of Istanbul

Serkan METİN*

*Dr. Öğr. Üyesi, Malatya Turgut Özal Üniversitesi, serkan.metin@ozal.edu.tr, ORCID: 0000-0003-1765-7474

Araştırma Makalesi / Research Article

ÖZ

Geliş Tarihi:

20.06.2025

Kabul Tarihi:

18.07.2025

Anahtar Kelimeler:

Makine Öğrenme, Hibrit Metod, Hava kirliliği, PM10, Derin Öğrenme

JEL Kodları:

Q53, C45, R15

Benzerlik Oranı: %4

Bu çalışma, İstanbul'daki hava kirliliği seviyelerinin (PM10, SO₂, NO₂, NO_x ve NO) makine öğrenmesi ve hibrit yöntemlerle tahmin edilmesini amaçlamaktadır. Çalışmada, T.C. Çevre Bakanlığı'na ait Sürekli İzleme Merkezi'nden 2013-2025 tarihleri arasında toplanan 4500 veri noktası kullanılmıştır. Veri ön işleme aşamasında eksik değerler medyan ile tamamlanmış, kirlenmeler arası ilişkiler korelasyon analizi ve görselleştirme teknikleriyle incelenmiştir. Modellemede Lineer Regresyon, Random Forest, SVR, XGBoost, hibrit (Voting ve Stacking) ve derin öğrenme modelleri test edilmiştir. Performans metrikleri (MSE, MAE, MAPE, R²) kapsamında en başarılı sonuçlar XGBoost ve Random Forest'tan elde edilmiştir. NO tahmininde Random Forest R²=0,96 ile en iyi performansı gösterirken, SO₂ tahmininde tüm modeller düşük başarı sergilemiştir. Hibrit modeller, PM10 gibi karmaşık değişkenlerde dengeli tahminler sunarken, derin öğrenme bazı durumlarda aşırı öğrenme nedeniyle sınırlı kalmıştır. Sonuçlar, doğrusal olmayan ilişkilerin ağaç tabanlı modellerle daha iyi yakalandığını, ancak SO₂ gibi yüksek varyanslı değişkenlerde ek özellik mühendisliği gerektiğini vurgulamaktadır. Çalışma, hava kalitesi yönetimi için hibrit modellerin etkinliğini ortaya koymakta ve gelecek araştırmalara meteorolojik veri entegrasyonu, hiperparametre optimizasyonu ve zaman serisi modellerinin kullanımını önermektedir.

ABSTRACT

Received Date:

20.06.2025

Acceptance Date:

18.07.2025

Keywords

Machine Learning, Hybrid Methods, Air pollution, PM10, Deep Learning

JEL Codes:

Q53, C45, R15

Similarity Rate: %4

This study aims to predict air pollution levels (PM10, SO₂, NO₂, NO_x, and NO) in Istanbul using machine learning and hybrid methods. A total of 4,500 data points collected between 2013 and 2025 from the Continuous Monitoring Center of the Ministry of Environment of the Republic of Turkey were used. During the data preprocessing phase, missing values were imputed using the median, and the interrelationships among pollutants were examined through correlation analysis and visualization techniques. The modeling phase involved testing various approaches including Linear Regression, Random Forest, Support Vector Regression (SVR), XGBoost, hybrid models (Voting and Stacking), and deep learning models. According to performance metrics (MSE, MAE, MAPE, and R²), XGBoost and Random Forest yielded the most accurate results. For NO prediction, Random Forest achieved the best performance with R² = 0,96, whereas all models showed limited success in predicting SO₂ levels. While hybrid models provided balanced predictions for complex variables such as PM10, deep learning models were occasionally constrained due to overfitting. The findings indicate that tree-based models are more effective in capturing non-linear relationships; however, additional feature engineering is required for high-variance variables like SO₂. Overall, the study demonstrates the effectiveness of hybrid models in air quality management and suggests the integration of meteorological data, hyperparameter optimization, and the adoption of time series models for future research.

Atıf / Citation: Metin, S. (2025). Hava Kirliliği Değerlerinin Makine Öğrenmesi Ve Hibrit Yöntemler İle Tahmini: İstanbul Örneği. *Malatya Turgut Özal Üniversitesi İşletme ve Yönetim Bilimleri Dergisi*, 6(2),190-203.

1. GİRİŞ

Hava kirliliği, atmosferdeki bazı zararlı maddelerin belirli bir konsantrasyonu aşması olarak adlandırılmaktadır (Bai vd., 2018). Bu maddeler, insan sağlığı, hayvanlar, bitkiler ve genel olarak çevre üzerinde olumsuz etkilere neden olabilir. Hava kirliliğine, volkanik patlamalar, orman yangınları gibi doğal olaylarında etkisi olmakla beraber en yoğun etki insan kaynaklı faaliyetlerden meydana gelmektedir. Özellikle 18. yüzyılda sanayi devrimi ile birlikte kömür ve petrol gibi fosil yakıtların yoğun bir şekilde kullanılması sonucu hava kalitesinde düşmeler başlamıştır. 1952 yılında Londra’da meydana gelen "Büyük Sis" olayı, kömür yakıtının çok yoğun bir şekilde kullanılmasından kaynaklanan ölümcül bir hava kirliliği olayıydı. Büyük Sis beş gün sürdü ve 4.000 ila 12.000 arası kişinin ölümüne yol açtığı tahmin ediliyordu. Bu olay ülkelerin hava kirliliğine bakışımı da değiştirdi (Polivka, 2018). Bu olaydan sonra Londra’da ilk kez kömür kullanımına kısıtlamalar uygulanmıştır (Kılıç vd., 2014). 1990 yılından bu yana megakentlerin hızla genişlemesi, sanayi üretiminin küreselleşmesi, pestisit ve toksik kimyasalların artması ve motorlu taşıtların daha fazla kullanımı nedeniyle çevresel hava kirliliği kaynaklı ölümler dünya genelinde artmaktadır ve bu artışlar, en hızlı sanayileşen ülkelerde görülmektedir (Brunekreef & Holgate, 2002).

Günümüzde dünyanın çoğu yerinde hava kirliliği çok daha karmaşık bir sorun haline gelmiştir. Sanayi, tarım, enerji üretimi ve ulaşım gibi faaliyetler, hava kalitesini etkileyen çok sayıda kirleticinin atmosfere salınmasına neden olmaktadır. Temel olarak hava kalitesini düşüren etkenler iki gruba ayrılmaktadır (Partigöç & Çubukçu, 2017). Birinci grup katı partiküller olarak da adlandırılan PM10 ve PM2.5’tür. İkinci grup ise karbon dioksit (CO₂), karbon monoksit (CO), metan (CH₄), azot oksitler (NO_x), kükürt dioksit (SO₂), amonyak (NH₃) ve ozon (O₃) gibi gazlar ve kimyasal maddelerden oluşmaktadır. ABD Çevre Koruma Ajansı (EPA), beş büyük hava kirleticisi olan Ozon (O₃), PM2.5 ve PM10 dahil olmak üzere partikül kirliliği, karbon monoksit (CO), kükürt dioksit (SO₂) ve azot dioksit (NO₂) için bir Hava Kalitesi İndeksi(AQI) seviyesi belirlemişlerdir (Horn & Dasgupta, 2024). Bu kirleticilerden PM10 ve PM2.5; havadaki sıvı damlacıkları ile parçacıkların karışımı partikül maddeyi oluşturur. PM10, 10 mikrondan küçük veya ona eşit bir boyuta sahip parçacıklardır, PM2.5 ise 2.5 mikrondan küçük veya ona eşit bir boyuta sahip ultra ince parçacıklardır (Zeydan, 2021). PM10, oksitlerden ve toz parçacıklarından, PM2.5 ise karbonlardan, sülfatlardan, nitratlardan ve silikattan oluşur (Chew vd., 2020). Bu partiküller solunduğunda solunum sisteminin daha derinlerine nüfuz ederek astım, öksürük gibi solunum yolu rahatsızlıklarına neden olabilmektedir. ABD-EPA'ya göre 24 saat için güvenli maruz kalma seviyeleri PM10 için 0-54 µg/m³ ve PM 2.5 için 0-12 µg/m³’dür. Karbon Monoksit (CO), karbon içeren yakıtların eksik yanması sonucu üretilir. Kentsel yerleşim yerlerinde, benzinli araçlar en önemli CO üreten kaynaklardır (Georgoulis vd., 2002). CO gazına maruz kalmak, insanlarda karbonmonoksit zehirlenmesine, göğüs ağrısına, kalp rahatsızlıklarına, zihinsel yeteneklerin azalmasına, görme bozukluklarına neden olur. ABD-EPA'ya göre güvenli maruz kalma ölçüsü 8 saat için 0-9,4 ppm’dir. Ozon(O₃), Dünya'nın hem üst atmosferinde (stratosfer) hem de yer seviyesinde bulunan bir gazdır. Stratosferik ozon, ultraviyole ışınlarla karşı bir bariyer işlevi gördüğü için iyi olarak kabul edilirken, troposferde ve yer seviyesinde, güneş radyasyonu ve ozon öncüllerini içeren bir dizi karmaşık fotokimyasal reaksiyon yoluyla oluşan ikincil bir hava kirleticisi olmakla birlikte önemli bir sera gazı olarak, iklim değişikliğine de sebep olmaktadır (Zoran vd., 2020). O₃, endüstrilerden, otomobil emisyonlarından, benzin buharlarından, çözücülerden, kimyasallardan ve elektronik cihazlardan salınır. O₃, insanlar tarafından aşırı solunduğunda, akciğer fonksiyonunda azalma, solunum yollarında iltihaplanma ve gözlerde, burunda ve boğazda tahrişe sebep olmaktadır. ABD-EPA'ya göre O₃ için 8 saatte güvenli maruz kalma seviyesi 0-0,054 ppm'dir. Azot Dioksit (NO₂), dünya genelinde kentsel hava kirleticiler arasında en yaygın olanıdır. NO₂ içeren havayı solumak, insan solunum sisteminde tahrişe neden

olabilir. Dünya Sağlık Örgütü'nün tanımladığı NO₂ konsantrasyon limiti, yıllık ortalamalar için yaklaşık 10 µg/m³ ve günlük ortalamalar için 25 µg/m³'tür (Amritha vd., 2024). NO₂, otomobil emisyonları, elektrik üretimi, yakıt yakımı, fosil yakıt yanması ve çeşitli endüstriyel prosesler sonucu çevreye salınır. Kükürt Dioksit (SO₂), doğası gereği asidik ve aşındırıcıdır ve atmosferde diğer bileşiklerle reaksiyona girerek sülfürik asit ve diğer kükürt oksitleri oluşturabilir. SO₂'nin ana kaynakları, kükürt içeren malzemeleri işleyen endüstriyel faaliyetlerdir. Doğal kaynaklar arasında volkanlar, okyanuslar, biyolojik bozulma ve orman yangınlarından salınımlar bulunur. SO₂'ye maruz kalmanın temel yolu solunmadır ve solunan kükürt dioksitin %90'a kadarı burunda emilir (Thriel vd., 2010). SO₂, insanlarda solunum rahatsızlığına, astıma, göz, burun ve boğaz tahrişine, hava yollarında iltihaplanmaya ve kalp hastalıklarına neden olur. Hava kirleticilerinin maruz kalma sürelerine ait risk değerleri Tablo 1'te verilmiştir.

Tablo 1. Hava Kirleticilerine Maruz Kalma Risk Aralıkları

Risk Aralığı	PM10 (ug/m3)	PM2.5 (ug/m3)	CO (ppm)	SO ₂ (ppb)	NO ₂ (ppb)	O ₃ (ppm)
Risksiz	0-54	0-12.0	0-4,4	0-35	0-53	0-0.54
Az Riskli	55-154	12.1-35,4	4,5-9,4	36-75	54-100	0,055-0,070
Hassas Grupla İçin Riskli	155-254	35.5-55,4	9,5-12,4	76-185	101-360	0,071-0,085
Riskli	255-354	55.5-150,4	12,5-15,4	186-304	361-649	0,086-0,105
Yaygın Risk	355-424	150.5-250,4	15,5-30,4	305-604	650-1249	0,106-0,200
Tehlikeli	425-604	250.5-500,4	30,5-50,4	605-1004	1250-2049	–

Bu çalışmanın amacı, makine öğrenmesi ve derin öğrenme yöntemlerini kullanarak temel 5 hava kalitesini düşüren maddelerin günlük değerlerini tahmin etmektir. Kullanılan modeller ile elde edilen sonuçları kapsamında ileriye dönük halk sağlığı açısından oluşabilecek riskler önceden tahmin edilip gerekli önlemlerin alınması için yöneticilere karar vermelerinde yardımcı olabilecektir. Çalışmanın diğer çalışmalardan farkı ise, PM10, SO₂, NO₂, NO_x ve NO değerlerinin tahmininde hibrit modeller kullanılmıştır. Makalenin ikinci bölümde, konu ile ilgili literatür taraması yapılmıştır. Üçüncü bölümde çalışmaya ait veri ve yöntemler hakkında bilgi verilmiştir. Dördüncü bölümde, çalışmadan elde edilen bulgular ait bilgiler ile yorumları verilmiştir. Son bölümde ise sonuç ve öneriler bulunmaktadır.

2. LİTERATÜR TARAMASI

Literatür incelemesi yapıldığında hava kirliliği ile birçok ulusal ve uluslararası çalışmaların yapıldığı görülmüştür.

Kurt vd. (2008), bu çalışmada, İstanbul Bölgesi için çevrimiçi bir hava kirliliği tahmin sistemi geliştirilmiştir. Sistem, sinir ağları kullanarak üç hava kirliliği göstergesi SO₂, PM10 ve CO seviyesini tahmin etmek için kullanılmıştır. Kılıç vd. (2014), Boğazlar ve Marmara Denizi'ni kapsayan bölgede gemi trafiğinin NO_x ve SO₂ emisyonları arasındaki ilişkiyi incelemişlerdir. Çalışmada %50'ye varan artışa sebep olduklarını belirlemişlerdir. Domańska & Wojtylak (2014), makalede, PM10, SO₂, O₃ için emisyon konsantrasyonlarını tahmin model önerilmişlerdir. Önerilen model (e-APFM), Hava Kirliliği Tahmin Modeli'nin (APFM) bir uzantısıdır. Önerilen modelin deneysel doğrulaması, Polonya'daki Meteoroloji ve Su Yönetimi Enstitüsü'nden alınan veriler üzerinde iki yıllık bir süre boyunca (Ocak 2011 ile Aralık 2012 arasında) yürütülmüştür. Deneyler, e-APFM yönteminin ölçülen ve öngörülen konsantrasyonlar arasında APFM yöntemine kıyasla ilk gün için benzer sapmalara sahip olduğunu, ikinci ve üçüncü gün (günlük değerler için) için çoğunlukla daha kötü olduğunu göstermiştir. Gümüş vd. (2015), bu çalışmada, Marmara bölgesinde ölçüm istasyonlarından elde edilen PM10 ve PM2.5 değerleri üzerinde istatistiksel analizler yapılmıştır. Uygulana t-testi ve Anova sonuçları incelendiğinde kırsal kesim ölçüm

istasyonlarının ölçüm değerleri benzer çıkarken kentsel yerleşim alanlarındaki ölçüm değerlerinin sanayileşme, trafik ve nüfus yoğunluğundan etkilenmesi sonucu farklılık gösterdiği tespit edilmiştir. Yılmaz (2018), Manisa iline ait 2009-2017 yılları arasındaki verilerdeki PM10 değerlerinin değişimleri incelenmiş ve PM10 değerinin rüzgâr hızı ile olan ilişkisini belirlemek için bir regresyon denklemi oluşturulmuştur. Gültepe (2019), çalışmada Kastamonu iline ait hava kirliliğini tahmin etmek için Min-Max normalizasyon tekniği ile birlikte YSA, Lineer Regresyon, Lojistik Regresyon, Bayes Ağı, K-En Yakın Komşu, Rastgele Orman ve Karar Ağacı yöntemleri kullanılmıştır. Çalışma sonucunda, Karar Ağacı ve Rastgele Orman yöntemleri %99'luk doğruluk oranına ulaşmıştır.

Tütmez (2020), PM10 analizleri için istatistiksel öğrenmeye dayalı Elastic-net, Lasso ve Ridge algoritmalarını kullanmıştır. Modellerin başarı oranları karşılaştırıldığında elastic-net modeli en uygun model olarak tespit edilmiştir. Altunkaynak vd. (2020), KNN algoritması için veri ön işleme adımıyla dalgacık dönüşüm tekniği kullanılarak 2016-2018 yılları arasında Erzincan ili için elde edilen 651 ölçüm değerinden hava kirliliği tahmini yapmışlardır.

Aladağ (2023), hava kirliliği tahmini için ARIMA ile Ampirik Mod Ayrıştırması (EMD)'den oluşan hibrit bir model önermiştir. Oluşturulan EMD-ARIMA yöntemi ile Van iline ait 2019-2020 kış dönemini PM10 ve SO₂ değerlerinden tahminler yapılmıştır. Bulgular EMD-ARIMA yönteminin tahmin yeteneğinin yüksek olduğunu göstermiştir. Can & Özsoy (2023), Karabük ilindeki hava kirliliğinin tespit etmek için yeni bir cihaz tasarlanmışlardır. Cihazdan elde edilen veriler üzerinde farklı istatistiksel analizler yapmışlardır. Analiz sonuçlarında PM2.5, SO₂, NO_x ve CO değerlerinin ölçüm aralıkları boyunca çok farklılaştıkları belirlenmiştir. Bağcı (2023), bu çalışmada, Van ilindeki PM10 seviyeleri incelenmiştir. Yöntem, lognormal, Weibull ve Gamma dağılımları kullanılarak modellenmiştir. Elde edilen sonuçlara göre Gamma dağılımının Van ilindeki PM10 konsantrasyonlarının modellenmesinde diğer iki dağılımdan daha iyi performans gösterdiği ve aşım değerlerini doğru bir şekilde tahmin ettiği görülmüştür.

Petrić vd. (2024), PM10, NO, NO₂ ve O₃ gibi kentsel ortamlarda yaygın olarak izlenen hava kirlleticilerinin saatlik konsantrasyonlarını modellemek için LSTM, CNN ve MLP yöntemleri kullanılmıştır. Çalışmasını sonucunda Avusturya'nın Graz kentindeki beş sahada saatlik NO, NO₂, PM10 ve O₃ değerleri için 0,83 ile 0,87 arasında değişen R2 katsayılarına ulaşan doğruluk elde etmişlerdir. Ramadan vd. (2024), Nesnelerin İnterneti (IoT) sensörleri ve AI yaklaşımları tarafından desteklenen sistemler aracılığıyla toplanan NH₃, CO, NO₂, CH₄, CO₂, SO₂, O₃, PM2.5 ve PM10 gerçek zamanlı verilerinden kirlilik seviyelerini tahmin etmek için LSTM, Random Forest ve Lineer Regresyon modellerini kullanmışlardır. LSTM modeli, sıcaklık ve nem tahmini için %99'luk bir varyasyon katsayısı (R2) ve 0,33'lük bir ortalama mutlak yüzde hatası (MAE) elde etti. PM2.5 için, Random Forest modeli diğerlerinden daha iyi performans göstererek %84'lük bir R2 ve 10,11'lik bir MAE değeri elde etmiştir. Yang vd. (2024), Kalman filtreleme tekniğini kullanarak ve mevsimsel kapılı tekrarlayan birimler (SGRU) derin öğrenme modelini kullanarak Tayvan'daki hava kirliliği seviyelerini tahmin etmek için etkili bir yöntem önermişlerdir. Bu çalışmada, 2005 ile 2021 yılları arasında Tayvan'daki çeşitli izleme istasyonlarından elde edilen Azot Dioksit (NO₂) ve Kükürt Dioksit (SO₂) değerleri kullanılmıştır. Önerilen SGRU modeli, Shanhua'da azot dioksit için 0.006'lık en düşük MAPE değerini, Keelung için 1,08'lik en yüksek değeri ve Taitung'da kükürt dioksit için 0.005'lik en düşük değeri, Cianjhen için 2,47'lik en yüksek değeri elde etmiştir. Ergin vd. (2024), 6 Şubat 2023 Kahramanmaraş depreminden en fazla etkilenen Kahramanmaraş, Gaziantep, Hatay, Malatya ve Adıyaman illerinin PM 10, PM 2.5, SO₂, NO, NO₂, NO_x, CO ve O₃ değerlerini 2022 Şubat, Mart, Nisan ve 2023 Şubat, Mart, Nisan aylarına göre kıyaslayarak için depremin partiküler madde miktarlarının artışı üzerinde araştırmışlardır. Sonuçlar ölçüm yapılan bazı istasyonlarda değerlerde artış olduğunu göstermiştir. Balcılar (2024), Eskişehir 'in 2023 yılındaki PM10, PM2.5 ve SO₂

verilerini inceleyerek değerlerin sınırlar üzerinde olduğunu belirlemiştir. 3 farklı istasyondan alınan verilerin konumlarına göre analizler yapıldığında istasyonların bulunduğu konumdaki trafik ve kentleşme yoğunluğunun bu değerler üzerindeki doğrudan etkisi gözlenmiştir. Oğuz & Pekin (2024), PM10, SO₂, NO₂, NO_x kirletici konsantrasyonlarını tahmin etmek için önceki saatin sıcaklık, bağıl nem, rüzgâr, basınç ve hava kirletici ölçümlerini dikkate alarak geliştirilen MLPNN modeli ile tahmin yapılmıştır. Sonuçlar değerlendirildiğinde yöntemin başarılı olduğu tespit edilmiştir.

3. MATERYAL VE METOT

Halk sağlığı açısından riskli olan hava kirleticilerinin günlük değerlerini tahmin etmek amacıyla gerçekleştirilen bu çalışmada, öncelikle tahminlerin yapılacağı bölge belirlenmiştir. Bu kapsamda nüfusun en yoğun olduğu il olarak İstanbul seçilmiştir.

3.1. Verilerin Toplanması

Çalışmada kullanılan veri seti T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı ait Sürekli İzleme Merkezi Veri Bankası (SİM)'den çekilmiştir. SİM veri tabanı uygulamasında bulunan filtreleme seçeneklerinden İstasyon Grubu Marmara THM, Şehir İstanbul, İstasyon Kandilli-MTHM ve ölçü parametreleri olarak da PM10 (µg/m³), SO₂ (µg/m³), NO₂ (µg/m³), NO_x (µg/m³) ve NO (µg/m³) seçilerek 15.01.2013-10.05.2025 tarihleri arasındaki toplam 4500 veri sistemden çekilmiştir.

3.2. Veri Analizi

Verilerin analizi için Python programlama dili kullanılmıştır. Python programlama dili, özellikle derin öğrenme ve makine öğrenmesi uygulamaları için geliştirilen TensorFlow, Keras, PyTorch ve Scikit-learn gibi güçlü kütüphaneler, model kurulumundan eğitime, hiperparametre ayarlamasından model değerlendirmesine kadar tüm süreci sistematik ve esnek bir biçimde yönetmeye olanak tanımaktadır. Açık kaynaklı yapısı, veri işleme ve görselleştirme için Pandas, NumPy ve Matplotlib gibi kütüphanelerle entegre çalışması, araştırmacılara uçtan uca çözüm imkânı sağlar.

Veri analizi için geliştirilen iki hibrit yöntem kullanılmıştır. Birinci olarak, Voting Hybrid Regressor (Oylama Tabanlı Hibrit Regresyon) modeli kullanılmıştır. Bu model, birden fazla regresyon modelinin tahmin sonuçlarının ortalamasını alarak nihai sonucu üretir. Bu yöntem, bireysel model hatalarını telafi ederek genelde daha dengeli ve istikrarlı tahminler üretir. İkinci yöntem ise, Stacking Hybrid Regressor (Yığınlama Tabanlı Hibrit Regresyon)dır. Stacking, farklı regresyon modellerinin tahmin çıktıları bir araya getirilerek, bu çıktılar üzerine ikinci bir model (meta-model) eğitilmesidir. Meta-model, hangi temel modelin hangi koşullarda daha iyi çalıştığını öğrenerek nihai tahmini üretir. Bu modelin avantajları, karmaşık ilişkileri daha iyi yakalar ve genelde en yüksek doğruluk oranlarını sağlar. Hibrit modelleri oluşturmak için Random Forest, Support Vector Regression, Lineer Regresyon ve XGBoost modelleri birleştirilerek oluşturulmuştur. Ayrıca derin öğrenme yöntemi de tahmin için kullanılmıştır.

Random Forest (RF), bir ensemble öğrenme algoritmasıdır ve birden çok karar ağacının bir araya gelerek daha güçlü bir model oluşturmasını sağlar. RF, her bir karar ağacını bağımsız olarak eğitir. Ağaçlar, belirli bir veri seti üzerinde eğitilirken, hem verilerden rastgele alt kümeler alınır hem de verinin her bir özelliği için rastgele seçim yapılır. Karar ağaçları her düğümde Gini veya entropi gibi ölçütler kullanılır. Hesaplama için kullanılan yöntemler Denklem 1 ve 2'de verilmiştir.

$$Gini(D) = 1 - \sum_{i=1}^C p_i^2 \quad (1)$$

$$Entropy(D) = - \sum_{i=1}^C p_i \log_2(p_i) \quad (2)$$

Lineer Regresyon (LR), istatistiksel ve makine öğrenmesi alanlarında en yaygın kullanılan ve temel yöntemlerden biridir. Amacı, bağımlı bir değişken ile bir veya daha fazla bağımsız değişken arasındaki doğrusal ilişkiyi modellemektir. LR, bir doğrusal fonksiyon kullanarak verileri modellemeye çalışır. LR modelinde, bağımsız değişkenler $x_1, x_2, \dots, x_n$ ile bağımlı değişken y arasındaki ilişki doğrusal olarak kabul edilir ve genel formülde Denklem 3'deki şekilde yazılır.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon \quad (3)$$

XGBoost, gradient boosting algoritmalarının en verimli ve güçlü versiyonlarından biridir. Bu algoritma, her adımda hatayı minimize ederek bir model inşa eder. Her yeni model, önceki modelin hatalarını düzeltmeye çalışır. XGBoost, karar ağaçları ile ensemble bir model oluşturur. Her ağaç, önceki ağaçların hatalarını öğrenir. Her ağaç için optimize edilmiş formül Denklem 4'de verilmiştir.

$$f_t(x) = f_{t-1}(x) + \eta \cdot \Delta f_t(x) \quad (4)$$

Support Vector Regression (SVR), Support Vector Machines (SVM) algoritmasının regresyon versiyonudur. SVR, doğrusal olmayan ilişkilerde bile iyi performans gösterebilir. SVR, verileri yüksek boyutlu bir uzaya yansıtarak doğrusal olmayan ilişkileri doğrusal hale getirir. SVR, bir doğrusal regresyon modeline benzer şekilde çalışır, ancak amaçlanan hedef, tahmin edilen değer ile gerçek değer arasındaki hatanın, belirli bir epsilon değeri ile sınırlı olmasını sağlamak ve hata terimini minimize etmektir. SVR'nin başarısı, kernel fonksiyonları kullanarak doğrusal olmayan veri noktaları için de başarılı tahminler yapabilmesindendir. Kernel fonksiyonu genellikle Denklem 5'de verilen RBF (Radial Basis Function) kullanır.

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (5)$$

Derin öğrenme (Deep Learning) modelleri, çok katmanlı yapay sinir ağları kullanarak karmaşık veri ilişkilerini öğrenmeyi amaçlar. Derin öğrenme modelleri, özellikle büyük veri setlerinde ve yüksek doğruluk gereksinimlerinde oldukça başarılıdır. Her bir katman, bir doğrusal ilişkiyi modelleyerek belirli bir aktivasyon fonksiyonu üzerinden geçer. Derin öğrenme modelinin genel formülü Denklem 6'daki gibidir.

$$y = f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2) \quad (6)$$

X: Girdi verisi

W_1, W_2 : Ağırlıklar

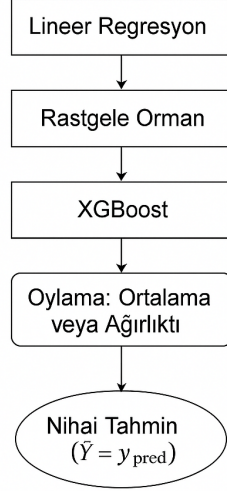
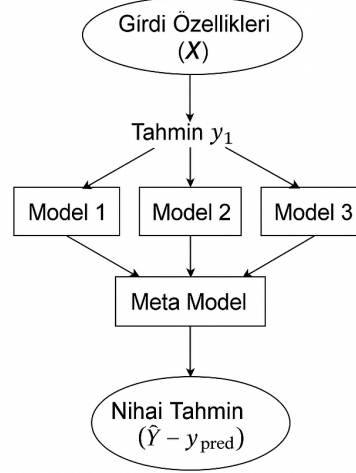
b_1, b_2 : Bias terimleri

f: Aktivasyon fonksiyonu

y: Çıktı

Bu formül, her katman için hesaplamaları temsil eder. İlk katmanda, giriş verisi ağırlıklarla çarpılır, bias eklenir, ardından aktivasyon fonksiyonu uygulanır. Sonraki katmanlar için aynı işlem tekrarlanır ve sonunda tek bir sürekli değer elde edilir.

Makine öğrenmesi modelleri ile oluşturulan hibrit modellere ait akış diyagramı Grafik 1'de verilmiştir.

Oylama Hibrid Regresyon**Yığınlama Hibrid Regresyon****Grafik 1. Model Akış Diyagramı****4. BULGULAR VE TARTIŞMA**

Bu bölümde ilk olarak veri setini uygulanacak modellere uygun hale getirmek için gerekli veri ön işleme adımları uygulanmıştır. Makine öğrenmesi ve derin öğrenme yöntemleri eksik veri setleri üzerinde istenilen sonuçları vermemektedir. Bunu için öncelikle elde edilen veri seti içerisindeki eksik verilerin ön işlemeyen geçirilmesi gerekmektedir. Veri setini oluşturan parametrelere ait eksik veri sayısı Tablo 2’de verilmiştir.

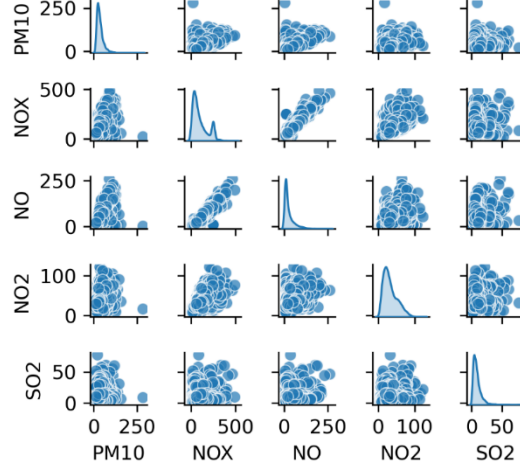
Tablo 2. Veri Seti İçindeki Eksik Veri Sayısı

Parametreler	Eksik Veri Sayısı
PM10 ($\mu\text{g}/\text{m}^3$)	284
SO ₂ ($\mu\text{g}/\text{m}^3$)	366
NO ₂ ($\mu\text{g}/\text{m}^3$)	243
NO _x ($\mu\text{g}/\text{m}^3$)	514
NO ($\mu\text{g}/\text{m}^3$)	368

Tablo 2, incelendiğinde veri setini oluşturan parametrelerdeki veri eksikliği fazla olduğundan eksik verilerin silinmesi yöntemi yerine verilerin tamamlama işlemi yapılmıştır. Eksik veriler Ortanca (median) yöntemi ile tamamlanmıştır. Ortanca (median) yöntemi, bir veri kümesindeki değerler küçükten büyüğe sıralandığında tam ortada kalan değeri alır. Verilerin merkezi eğilimini temsil eder ve aykırı değerlerin etkisine karşı dayanıklıdır. Bu özelliğinden dolayı bu yöntem eksik veri tamamlamada tercih edilmiştir.

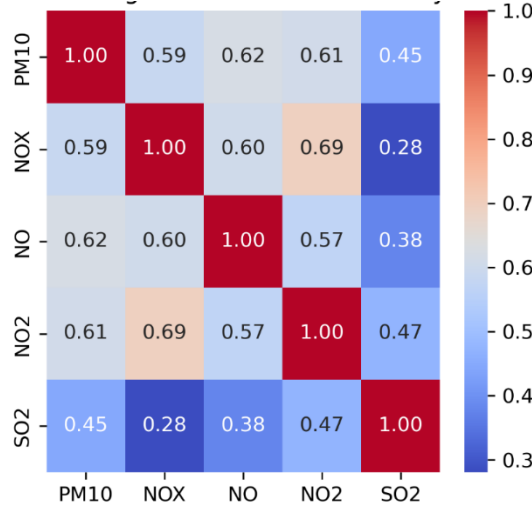
Grafik 2’de, kirleticilerin PM10, NO_x, NO, NO₂, SO₂ arasındaki çift değişkenli ilişkileri ve her bir değişkenin tek değişkenli dağılımını göstermektedir. Tek Değişkenli Dağılım grafiklerin diyagonalinde, her kirleticinin konsantrasyonunun frekans dağılımı verilmiştir. Çift Değişkenli İlişkiler (Scatterplotlar), iki değişken arasındaki potansiyel doğrusal veya doğrusal olmayan ilişkileri incelemeye olanak sağlar. Grafik 2 incelendiğinde PM10 ile NO_x, NO, NO₂ arasında pozitif bir ilişki olduğundan PM10’un azot oksitlerle benzer kaynaklardan etkilenebileceğini

göstermektedir. Bunun ile birlikte PM10 ve SO₂ arasındaki ilişki daha zayıf olduğundan bu iki maddenin farklı kaynaklardan veya atmosferik süreçlerden etkilendiğini göstermektedir.



Grafik 2. Kirlenmeler Arasındaki Konsantrasyon İlişkisi

PM10, NO_x, NO, NO₂, SO₂ kirlilik değerleri arasındaki ilişkileri gösteren korelasyonlar Grafik 3’de verilmiştir. Grafik 3, hava kalitesi parametreleri arasındaki ilişkileri anlamak için önemlidir. Orta ve yüksek korelasyonlar, bileşiklerin benzer kaynaklardan salındığını veya birbirleriyle kimyasal etkileşimde bulunduğunu gösterir. Grafik 3’deki korelasyon değerleri incelendiğinde PM10, NO_x (0,59), NO (0,62) ve NO₂ (0,61) ile orta derecede pozitif bir korelasyon sahip olduğunu göstermektedir. Bu, PM10 konsantrasyonlarının bu gazlarla ilişkili olduğunu, ancak birebir nedensellik ilişkisi olmadığını gösterir.



Grafik 3. Kirlilik Değerleri Arasındaki Korelasyon

Çalışmada kullanılan her bir modele ait sonuçların ayrı ayrı ve birleşik modellerin performanslarını ölçmek için denklem 7,8,9,10 ve 11’deki matematiksel formüller kullanılmıştır.

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

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

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

$$\text{Accuracy}(\%) = 100 - \text{MAPE} \quad (11)$$

Mean Squared Error (MSE), tahmin edilen değerler ile gerçek değerler arasındaki farkların karesinin ortalamasını ifade eder. Düşük MSE, modelin daha iyi tahmin yaptığını gösterir. Mean Absolute Error (MAE), tahmin edilen ve gerçek değerler arasındaki farkların ortalamasını ifade eder. Daha düşük MAE, tahminlerin daha doğru olduğunu gösterir. Mean Absolute Percentage Error (MAPE), gerçek değerlere kıyasla tahmin hatasının yüzdesini ifade eder. MAPE değerinin düşük olması modelin daha iyi performans gösterdiğini ifade eder. R^2 (Determinasyon Katsayısı), modelin veri varyansını ne kadar açıkladığını gösterir. 1'e yakın bir R^2 değeri, modelin çok iyi bir performans gösterdiğini ifade eder. Accuracy (Doğruluk Yüzdesi), tahminlerin ne kadar doğru olduğunu yüzdelik olarak ifade eder.

Tablo 3 incelendiğinde PM10 tahmininde en iyi sonucu Derin Öğrenme modeli elde etmiştir ($R^2=0,55$), ancak doğruluk yüzdesi VotingHybrid modeline göre daha düşüktür. Bu fark, derin öğrenme modelinin karmaşık yapısından kaynaklı olarak aşırı uyum (overfitting) yapma eğiliminden kaynaklanabilir. VotingHybrid ($R^2=0,55$, %74,81 doğruluk) ise daha dengeli bir modelleme kapasitesi sergileyerek istatistiksel olarak da tutarlı sonuçlar üretmiştir. Linear Regression en düşük başarıyı gösteren yöntemdir; bu durum, PM10'un doğrusal olmayan dağılım yapısına sahip olabileceğine işaret etmektedir. SVR modeli ise düşük MAPE oranı ile kısa vadeli tahminlerde daha iyi performans gösterebilir.

Tablo 3. PM10 Değeri İçin Model Tahmin Performans Değerleri

Model	MSE	MAE	MAPE	R2	Accuracy (%)
LinearRegression	182,58	9,09	27,79	0,48	72,20
RandomForest	166,96	8,72	26,59	0,52	73,40
SVR	184,87	8,64	25,24	0,47	74,75
XGBoost	164,94	8,48	25,75	0,53	74,24
VotingHybrid	157,69	8,34	25,19	0,55	74,80
StackingHybrid	157,87	8,41	25,66	0,55	74,33
DeepLearning	156,63	8,56	27,20	0,55	72,79

SO2 tahmininde tüm modellerde başarı düşüktür; Tablo 4'de en yüksek R^2 değeri Derin Öğrenme modeli tarafından elde edilmiştir ($R^2=0,33$). Random Forest ve XGBoost gibi güçlü algoritmalar dahi bu değişkende düşük doğruluk göstermiştir. Bu durum SO2 verisinin yüksek varyansa sahip olması, eksik veri problemleri veya mevsimsel etkilerin yeterince temsil edilmemesinden kaynaklanıyor olabilir. Hibrit modeller de burada klasik modellere yakın performans göstermiş, beklenen üstünlüğü sağlayamamıştır. Bu nedenle, SO2 tahmini için ileri düzey özelliklik mühendisliği, dışsal faktörlerin (meteorolojik değişkenler vb.) entegrasyonu önemlidir.

Tablo 4. SO₂ Değeri İçin Model Tahmin Performans Değerleri

Model	MSE	MAE	MAPE	R2	Accuracy (%)
LinearRegression	33,50	3,64	54,97	0,29	45,02
RandomForest	36,07	3,71	55,58	0,24	44,41
SVR	35,86	3,55	47,28	0,24	52,71
XGBoost	35,43	3,71	55,46	0,25	44,53
VotingHybrid	34,15	3,55	51,52	0,28	48,47
StackingHybrid	33,24	3,58	53,88	0,30	46,11
DeepLearning	31,66	3,56	56,12	0,33	43,87

Tablo 5'deki değerlere göre NO₂ tahmini açısından en yüksek başarıyı XGBoost modeli elde etmiştir. Bu modelin R² skoru 0,772, MAPE değeri %21,58, ve doğruluğu %78,41 olarak kaydedilmiştir. Bunu sırasıyla Random Forest (R²: 0,746) ve StackingHybrid (R²: 0.769) takip etmiştir. Bu üç modelin benzer performans göstermesi, NO₂ verisinde ağaç tabanlı yöntemlerin güçlü öznetelik öğrenme kabiliyetinin etkili olduğunu göstermektedir. Özellikle Linear Regression modelinin düşük R² değeri (0,518) ve yüksek MAPE değeri (%42,5) ile başarısız bir performans sergilemesi, verinin doğrusal olmayan yapılar içerdiğini ve bu nedenle basit doğrusal ilişkilerle modellemenin yetersiz kaldığını ortaya koymaktadır. SVR ise orta düzeyde bir performans göstermiştir (R²: 0,67), ancak doğruluğu %68,77 ile XGBoost'a kıyasla daha düşüktür. Hibrit modellerden VotingHybrid ve StackingHybrid, temel modellerin başarılarını ortalamaya alarak istikrarlı bir genel performans sergilemiştir.

Tablo 5. NO₂ Değeri İçin Model Tahmin Performans Değerleri

Model	MSE	MAE	MAPE	R2	Accuracy (%)
LinearRegression	169,10	9,43	42,52	0,51	57,47
RandomForest	88,82	5,62	21,58	0,74	78,41
SVR	114,19	7,42	31,22	0,67	68,77
XGBoost	79,92	5,58	21,58	0,77	78,41
VotingHybrid	85,41	5,93	23,85	0,75	76,14
StackingHybrid	80,77	5,70	22,44	0,76	77,55
DeepLearning	83,44	5,37	21,54	0,76	78,40

Tablo 6'da NO_x tahmininde en yüksek başarı yine XGBoost modeline aittir. R² skoru 0,84 ve doğruluğu %78,70 ile bu model, karmaşık ilişkilere sahip NO_x verisinde öne çıkmıştır. Onu çok yakın performansla Random Forest (R²: 0,83) ve StackingHybrid (R²: 0,84) takip etmiştir. Bu sonuçlar, XGBoost'un yüksek öğrenme kapasitesi ve genel hataları minimize etme becerisi sayesinde ağaç temelli modeller arasında öne çıktığını göstermektedir. Linear Regression modelinin yine çok düşük başarı göstermiştir (R²: 0,53, MAPE: %44,2). Derin öğrenme modeli, NO_x tahmininde beklenenin altında bir performans göstermiştir (R²: 0,71, doğruluk: %65,27). Bu durum, modelin aşırı öğrenmeye uğramış veya veri boyutunun derin öğrenme için yetersiz kalmış olabileceğini düşündürmektedir.

Tablo 6. NO_x Değeri İçin Model Tahmin Performans Değerleri

Model	MSE	MAE	MAPE	R2	Accuracy (%)
LinearRegression	2824,84	32,88	44,28	0,53	55,71
RandomForest	1014,66	14,73	19,07	0,83	80,92

SVR	3707,64	29,62	29,49	0,38	70,50
XGBoost	922,18	15,27	21,29	0,84	78,70
VotingHybrid	1290,14	18,81	21,79	0,78	78,20
StackingHybrid	944,23	16,16	22,51	0,84	77,48
DeepLearning	1734,46	23,95	34,72	0,71	65,27

Tablo 7'deki sonuçlar incelendiğinde NO değişkeni için Random Forest modeli $R^2=0,96$ ile neredeyse kusursuz bir tahmin kapasitesi göstermiştir. Bu, değişkenin verisinde bulunan belirgin desenlerin karar ağaçları tarafından etkili biçimde öğrenildiğini göstermektedir. SVR ve Lineer Regresyon ise başarısız olmuş, özellikle Lineer Regresyon'un doğruluk oranı %0,24 gibi oldukça düşük bir seviyede kalmıştır. XGBoost ve StackingHybrid modelleri de yüksek doğruluk sağlamıştır. Derin Öğrenme bu değişkende beklenenden zayıf performans sergilemiştir ($R^2=0,83$), bu durum sınırlı veri hacminde derin yapılar kullanmanın aşırı öğrenmeye neden olabileceğini göstermektedir.

Tablo 7. NO Değeri İçin Model Tahmin Performans Değerleri

Model	MSE	MAE	MAPE	R2	Accuracy (%)
LinearRegression	511,23	13,71	99,75	0,47	0,24
RandomForest	32,09	2,56	15,08	0,96	84,91
SVR	504,34	9,31	48,02	0,47	51,97
XGBoost	38,07	2,85	19,31	0,96	80,68
VotingHybrid	80,94	4,36	25,73	0,91	74,26
StackingHybrid	38,19	2,93	19,83	0,96	80,16
DeepLearning	158,40	5,97	42,32	0,83	57,67

5. SONUÇ, TARTIŞMA VE ÖNERİLER

Bu çalışmada, kullanılan veriseti T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı ait SİM Veri Bankasından çekilmiştir. Sim veri tabanı uygulamasında bulunan filtreleme seçeneklerinden İstasyon Grubu Marmara THM, Şehir İstanbul, İstasyon Kandilli-MTHM ve ölçü parametreleri olarak da PM10 ($\mu\text{g}/\text{m}^3$), SO₂ ($\mu\text{g}/\text{m}^3$), NO₂ ($\mu\text{g}/\text{m}^3$), NO_x ($\mu\text{g}/\text{m}^3$) ve NO ($\mu\text{g}/\text{m}^3$) seçilerek 15.01.2013-10.05.2025 tarihleri arasındaki toplam 4500 veriden oluşmaktadır. Bu çalışma kapsamında, hava kirleticileri olan PM10, SO₂, NO₂, NO_x ve NO üzerine geliştirilen farklı makine öğrenmesi ve derin öğrenme modellerinin performansları kapsamlı bir biçimde karşılaştırılmıştır. Modeller; Ortalama Kare Hata (MSE), Ortalama Mutlak Hata (MAE), Ortalama Mutlak Yüzde Hata (MAPE), R-kare (R^2) ve doğruluk oranı gibi metriklerle değerlendirilmiştir. Elde edilen bulgular, tahmin edilen hedef değişkenin yapısına ve veri desenlerine göre model başarısının ciddi şekilde değiştiğini göstermektedir.

Genel olarak, XGBoost ve Random Forest modelleri, özellikle NO₂, NO_x ve NO değişkenlerinde yüksek R^2 skorları ve düşük hata oranlarıyla öne çıkmıştır. Bu durum, karar ağacı temelli algoritmaların doğrusal olmayan ilişkileri modellemede üstün performans gösterdiğini ortaya koymaktadır. Örneğin, NO tahmininde Random Forest modeli $R^2=0,96$ ile neredeyse kusursuz bir tahmin gerçekleştirmiştir. Buna karşılık, Lineer Regresyon hemen her değişkende düşük başarı sağlamış; özellikle NO tahmininde %0,24 doğruluk ile neredeyse geçersiz sonuçlar üretmiştir. Bu da doğrusal varsayımlar üzerine kurulu modellerin kompleks çevresel verilerde yetersiz kaldığını göstermektedir.

Hibrit modeller (Voting ve Stacking) çoğu zaman klasik yöntemlere göre daha tutarlı ve dengeli performanslar sunmuştur. PM10 gibi değişkenlerde hibrit yapıların hata oranlarını düşürdüğü ve tahmin gücünü artırdığı görülmektedir. Bununla birlikte, hibrit modellerin başarısı, kullanılan temel algoritmaların kalitesine doğrudan bağlıdır. Derin öğrenme modelleri ise bazı değişkenlerde (örneğin PM10 ve SO₂) rekabetçi performans göstermiş, ancak NO_x ve NO gibi değişkenlerde klasik modellere kıyasla geride kalmıştır.

Gelecek çalışmalar için özellikle modellerin düşük performans sergilediği SO₂ gibi değişkenler için dışsal çevresel faktörler sıcaklık, rüzgâr ve nem modele dahil edilmesinin tahmin başarısını artırmada faydalı olacağı düşünülmektedir. Hiperparametre optimizasyonları için grid search veya Bayesian yöntemleri kullanılabilir. Yine zaman serisi için LSTM veya Transformer tabanlı yapay sinir ağlarının hibrit modellere entegrasyonu yapılabilir.

KAYNAKÇA

- Aladağ, E. (2023). Ampirik Mod Ayırıştırmasına Dayalı ARIMA Modeli Kullanılarak Van İli Hava Kirliliğinin Tahmini. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 28(2), 495-509
- Altunkaynak, A., Başakın, E. E. & Kartal, E. (2020). Dalgacık K-En Yakın Komşuluk Yöntemi İle Hava Kirliliği Tahmini. *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 25(3), 1547- 1556 DOI: 10.17482/uumfd.809938
- Amritha, S. Varikoden, H., Patel, V. K., Kuttippurath, J. & Gopikrishnan, G.S. (2024). Global, regional and city scale changes in atmospheric NO₂ with environmental laws and policies. *Sustainable Cities and Society*, 112, 105617 <https://doi.org/10.1016/j.scs.2024.105617>
- Bagci, K. (2023). Application of statistical distributions to PM10 concentrations: Van, Türkiye. *Van Yüzüncü Yıl University the Journal of Social Sciences Institute*, 60, 87-95
- Bai, L., Wang, J., Ma, X. & Lu, H. (2018). Air Pollution Forecasts: An Overview. *Int. J. Environ. Res. Public Health*, 15, 780; doi:10.3390/ijerph15040780
- Balcılar, İ. (2024). Eskişehir’de hava kirliliği: PM10, PM2.5 ve SO2 konsantrasyonlarının mekânsal-zamansal değişimi ve kaynaklarının değerlendirilmesi. *NÖHÜ Müh. Bilim. Derg.* 13(4), 1115-1126
- Brunekreef, B. & Holgate, S. T. (2002). Air pollution and health. *The Lancet*, 360(9341), 1233-1242
- Can, A. & Özsoy, H. (2023). A different perspective on air pollution measurements, *Journal of Polytechnic*, 26(1), 329-344
- Chew, S., Kolosowska, N., Saveleva, L., Malm, T. & Kanninen, K. M. (2020). Impairment of mitochondrial function by particulate matter: Implications for the brain. *Neurochemistry International*, 135, 104694, <https://doi.org/10.1016/j.neuint.2020.104694>
- Domańska, D. & Wojtylak, M. (2014). Explorative forecasting of air pollution. *Atmospheric Environment*, 92, 19-30, <https://doi.org/10.1016/j.atmosenv.2014.03.041>
- Ergin, Y., Demiray, G., Ergin, B. B. & Ekuklu, G. (2024). 2023 Kahramanmaraş Depremi Sonrası Etkilenen Bazı İllerde Hava Kalitesi Değerlerinin Karşılaştırılması. *Afet ve Risk Dergisi*, 7(3), 697-712

- Georgoulis, L. B., Hanninen, O., Samoli, E., Katsouyanni, K., Kunzli, N., Polanska, L., Bruin, Y. B. Alm, S. & Jantunen, M. (2002). Personal carbon monoxide exposure in five European cities and its determinants. *Atmospheric Environment*, 36, 963–974
[https://doi.org/10.1016/S1352-2310\(01\)00473-3](https://doi.org/10.1016/S1352-2310(01)00473-3)
- Gültepe, Y. (2019). Makine Öğrenmesi Algoritmaları ile Hava Kirliliği Tahmini Üzerine Karşılaştırmalı Bir Değerlendirme. *Avrupa Bilim ve Teknoloji Dergisi*, 16, 8-15
- Gümüş, O., Şahin, Ü. A., Onat, B., Özçelik, R., Gedik, E., Solakoğlu, İ. & Taş, N. (2015). Marmara bölgesi hava kalitesinin istatistiksel yöntemlerle analizi. 6. Ulusal Hava Kirliliği ve Kontrolü Sempozyumu-2015, 780-793
- Horn, S. A. & Dasgupta, P. K., (2024). The Air Quality Index (AQI) in historical and analytical perspective a tutorial review. *Talanta*, 25260, <https://doi.org/10.1016/j.talanta.2023.125260>
- Kılıç, A., Kum, S. Ünal, A. & Kındap, T. (2014). Marmara Bölgesi' ndeki Hava Kirliliğinin Modellenmesi, Kirlilik Azaltımı ve Maruziyet Analizi. *BAÜ Fen Bil. Enst. Dergisi*, 16(1,) 27-46
- Kurt, A., Gulbagci, B., Karaca, F. & Alagha, O. (2008). An online air pollution forecasting system using neural networks. *Environment International*, 34, 592–598, <https://doi.org/10.1016/j.envint.2007.12.020>
- Partigöç, N. S. & Çubukçu, K. M. (2017). Hava Kirliliği ve Kent İlişkisine Ampirik Bakış: Ekolojik Sürdürülebilirlik Ekseninde Bir Değerlendirme. *Akademia Disiplinlerarası Bilimsel Araştırmalar Dergisi*, 3 (2), 28-45
- Petrić, V., Hussain, H., Časni, K., Vuckovic, M., Schopper, A., Andrijić, Z. U., Kecorius, S., Madueno, L., Kern, R. & Lovrić, M. (2024). Ensemble Machine Learning, Deep Learning, and Time Series Forecasting: Improving Prediction Accuracy for Hourly Concentrations of Ambient Air Pollutants. *Aerosol and Air Quality Research*. 24(12), <https://doi.org/10.4209/aaqr.230317>
- Polivka, B.J. (2018). The Great London Smog of 1952. *AJN American Journal of Nursing*, 118(4):57-61 DOI: 10.1097/01.NAJ.0000532078.72372.c3
- Ramadan, M. N. A., Ali, M. A. H., Khoo, S. Y., Alkhedher, M. & Alherbawi, M. (2024). Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment. *Ecotoxicology and Environmental Safety*, 283, <https://doi.org/10.1016/j.ecoenv.2024.116856>
- Oğuz, K., & Pekin, M. A. (2024). Prediction of Air Pollution with Machine Learning Algorithms. *Turkish Journal of Science & Technology*, 19(1), 1-12, <https://doi.org/10.55525/tjst.1224661>
- The European Environment Agency. (2021). Air quality in Europe. doi: 10.2800/549289
- Thriel, C., Schäper, M., Kleinbeck, S., Kiesswetter, E., Blaszkewicz, M., Golka, K., Nies, E., Heimsoth, M. R. & Brüning, T. (2010). Sensory and pulmonary effects of acute exposure to sulfur dioxide (SO₂). *Toxicology Letters*, 196,42–50 doi:10.1016/j.toxlet.2010.03.013
- Tütmez, B. (2020). Air Quality Assessment by Statistical Learning-Based Regularization. *Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 35(2), 271-278
- Yang, C. H., Chen, P. H., Yang, C. C. & Chuang, L. Y. (2024). Analysis and Forecasting of Air Pollution on Nitrogen Dioxide and Sulfur Dioxide Using Deep Learning. *IEEE Access*, 12, 165236 – 165252, DOI: 10.1109/ACCESS.2024.3494263

- Yılmaz, B. (2018). Manisa’da partikül madde (PM10) kirliliğinin değerlendirilmesi. BAUN Fen Bil. Enst. Dergisi, 20(1), 626-633
- Zeydan, Ö. (2021). 2019 Yılında Türkiye’deki Partikül Madde (PM10) Kirliliğinin Değerlendirilmesi. Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 11(1): 106-118
- Zhang, Y., Wu, M. & Guo, B. (2024). Identifying the Effect of Air Quality on Analysts' Forecasts. Available at SSRN: <https://ssrn.com/abstract=4920895>
- Zoran, M. A. Savastru, R. S. Savastru, D. M. & Tautan, M. N. (2020). Assessing the relationship between ground levels of ozone (O3) and nitrogen dioxide (NO2) with coronavirus (COVID-19) in Milan, Italy, Science of the Total Environment, 740,140005, <https://doi.org/10.1016/j.scitotenv.2020.140005>