



Makine Öğrenmesi Algoritmaları Kullanılarak Hava Kalitesi İndeksinin Tahmini

Mustafa Akgün^{1*}, Necla Barlık²

^{1*} Ardahan Üniversitesi, Mühendislik Fakültesi, Çevre Mühendisliği Bölümü, Ardahan, Türkiye, (ORCID: 0000-0002-7172-1855), mustafaakgun@ardahan.edu.tr

² Ardahan Üniversitesi, Mühendislik Fakültesi, Çevre Mühendisliği Bölümü, Ardahan, Türkiye, (ORCID: 0000-0003-0000-1903), neclabarlik@ardahan.edu.tr

(İlk Geliş Tarihi 23 Ocak 2023 ve Kabul Tarihi 17 Mayıs 2023)

(DOI: 10.31590/ejosat.1241170)

ATIF/REFERENCE: Akgün, m., Barlık, N. (2023). Makine Öğrenmesi Algoritmaları Kullanılarak Hava Kalitesi İndeksinin Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, (51), 97-107.

Öz

Bu çalışmada; 15/09/2019 – 15/05/2020 ve 15/09/2020 – 15/05/2021 tarihleri arasındaki Ardahan, Kars, Erzurum ve Iğdır illeri şehir merkezlerindeki Hava Kalitesi İndeksi (HKİ) ve meteorolojik parametreler (ortalama sıcaklık, yağış, bulutluluk, nispi nem, basınç ve rüzgâr hızı) arasında çoklu doğrusal regresyon analizi ve makine öğrenmesi yöntemleri kullanılarak HKİ'nin tahmin edilmesi amaçlanmıştır. Çoklu doğrusal regresyon analizi ile yapılan HKİ'ni tahminleme oranı %41,8 bulunmuştur. Makine öğrenmesi yöntemlerinden Naive Bayes algoritmasının HKİ'ni %68,91 doğrulukla tahmin ettiği, Karar Ağacı algoritmasının HKİ'ni %70,29 doğrulukla tahmin ettiği, Derin öğrenme algoritmasının HKİ'ni %73,05 doğrulukla tahmin ettiği, k-NN algoritmasının HKİ'ni %73,35 doğrulukla tahmin ettiği ve Rastgele Orman algoritmasının HKİ'ni %74,89 doğrulukla tahmin ettiği sonucuna ulaşılmıştır.

Anahtar Kelimeler: Hava kalitesi indeksi, Makine öğrenmesi, Meteorolojik parametreler.

Estimation of Air Quality Index Using Machine Learning Algorithms

Abstract

In this study, it is aimed to estimate air quality index (AQI) in the city centers of Ardahan, Kars, Erzurum and Iğdır between 9/15/2019 – 5/15/2020 and 9/15/2020 – 5/15/2021 in Turkey. AQI has been estimated between AQI and meteorological parameters (average temperature, precipitation, cloudiness, relative humidity, pressure, wind speed) by using multiple linear regression analysis and machine learning methods. The rate of estimating AQI by multiple linear regression analysis was found to be 41.8%. It has been concluded that the Naive Bayes algorithm predicts the HKI with 68.91% accuracy, the Decision Tree algorithm predicts the HKI with 70.29% accuracy, the Deep learning algorithm predicts the HKI with 73.05% accuracy, the k-NN algorithm predicts the HKI with 73.35% accuracy and the Random Forest algorithm predicts the HKI with 74.89% accuracy from machine learning methods.

Keywords: Air quality index, Machine learning, Meteorological parameters.

* Sorumlu Yazar: mustafaakgun@ardahan.edu.tr

1. Giriş

Hava kirliliğinin insan sağlığı üzerinde hem akut hem de kronik etkileri vardır. Bu etkiler bir dizi farklı sistemi ve organı etkileyerek minör üst solunum yolu tahrişinden, kronik solunum sistemi hastalıklarına, akciğer kanserine ve kalp hastalıklarına neden olabilir. Çocuklarda akut solunum yolu enfeksiyonlarına, yetişkinlerde kronik bronşite, önceden var olan kalp ve akciğer hastalıklarının veya astım krizlerinin şiddetlenmesine kadar uzanan olumsuz etkileri olabilir. Karbonmonoksit (CO), kükürt dioksit (SO₂), nitrojen oksitler (NO_x), uçucu organik bileşikler (VOC'ler), ozon (O₃), ağır metaller ve solunabilir partikül maddeler (PM_{2.5} ve PM₁₀) gibi hava kirlleticilerinin kimyasal bileşimleri, reaksiyon özellikleri, emisyonları, parçalanma süreleri ve uzun veya kısa mesafelerde yayılma yetenekleri farklılık gösterir. Bu kimyasallar bir takım doğal ve/veya antropojenik faaliyetlerle çevreye sızarak çevre ve insan sağlığı üzerinde olumsuz etkilere neden olabilir. Bunlara ek olarak, hava kirliliğine kısa veya uzun vadeli maruziyetler azalan yaşam süresi ile ilişkilendirilmiştir (Kampa & Castanas, 2008). Ayrıca PM₁₀ düzeyindeki 10 µg/m³'teki artışın her yaş için ölüm oranlarında artışa neden olduğu bildirilmiştir (Katsouyanni, Touloumi, Samoli, Gryparis, Le Tertre, Monopolis, Rossi, Zmirou, Ballester, Boumghar, Anderson, Wojtyniak, Paldy, Braunstein, Pekkanen, Schindler, and Schwartz 2001). Ayrıca, genellikle toz veya dumana maruziyetin KOAH gelişimi ile ilişkili olduğu bildirilmiştir (T.C. Sağlık Bakanlığı, 2017). HKİ; hava kirliliği konsantrasyonunu 0 ila 500 aralığında bir değere dönüştürmektedir. Bu değerlerin anlamları ve sağlık için endişe seviyeleri Tablo 1'de görülmektedir. Tablo 1'deki 0, 50, 100, 150, 200, 300, 500 arasındaki HKİ değerleri kesme noktaları olarak adlandırılır. Bu değerler ve hesaplamada kullanılan parametreler ise Tablo 2'de verilmiştir ("T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı", 2022). SO₂, NO₂, CO, O₃ ve PM₁₀ parametrelerine ait ulusal ve Avrupa birliği ülkelerine ait sınır değerler ise Tablo 3'te verilmiştir.

Tablo 1. EPA hava kalitesi indeksi ("T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı", 2022) (Table 1. EPA air quality index)

Hava Kalitesi İndeksi Değerleri	Sağlık Endişe Seviyeleri	Renkler	Anlamı
0-50	İyi	Yeşil	Hava kalitesi memnun edici ve hava kirliliği az riskli veya hiç risk teşkil etmiyor.
51-100	Orta	Sarı	Hava kalitesi uygun fakat alışılmadık şekilde hava kirliliğine hassas olan çok az sayıdaki insanlar için bazı kirleticiler açısından orta düzeyde sağlık endişesi oluşabilir.
101-150	Hassas	Turuncu	Hassas gruplar için sağlık etkileri oluşabilir. Genel olarak kaminun etkilenmesi olası değildir.
151-200	Sağlıksız	Kırmızı	Herkes sağlık etkileri yaşamaya başlayabilir, hassas gruplar için sağlık etkileri söz konusu olabilir.
201-300	Kötü	Mor	Sağlık açısından acil durum oluşturabilir. Nüfusun tamamının etkilenme olasılığı yüksektir.
301-500	Tehlikeli	Kahverengi	Sağlık alarmı: Herkes daha ciddi sağlık etkileri ile karşılaşabilir.

HKİ'nin 0-50 aralığında olması hava kalitesinin memnun edici olduğu ve hava kirliliğinin az riskli veya hiç risk teşkil etmediği, 51-100 aralığında olması hava kalitesinin uygun fakat alışılmadık şekilde hava kirliliğine hassas olan çok az sayıdaki insanlar için bazı kirleticiler açısından orta düzeyde sağlık endişesi oluşturabileceği durumu, 101-150 aralığında olması hassas gruplar için sağlık etkileri oluşabileceği fakat genel olarak kaminun etkilenmesinin olası olmadığı durumu, 151-200 aralığında olması herkesin sağlık etkileri yaşamaya başlayabileceği, hassas gruplar için sağlık etkilerinin söz konusu olabileceği durumu, 201-300 aralığında olması bu durumun sağlık açısından acil durum oluşturabileceği ve nüfusun tamamının etkilenme olasılığı yüksek olduğu durumu ve 301-500 aralığında olması ise herkesin ciddi sağlık etkileri ile karşılaşabilir bir durumda olduğu anlamına gelmektedir.

Tablo 2. Ulusal Hava Kalitesi İndeksi Kesme Noktaları (“T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı”, 2022)

(Table 2. National Air Quality Index Breakpoints)

İndeks	HKİ	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (µg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)
		1 Sa Ort	1 Sa Ort	8 Sa Ort	8 Sa Ort	24 Sa Ort
İyi	0-50	0-100	0-100	0-5500	0-120	0-50
Orta	51-100	101-250	101-200	5501-10000	121-160	51-100
Hassas	101-150	251-500	201-500	10001-16000	161-180	101-260
Sağlıksız	151-200	501-850	501-1000	16001-24000	181-240	261-400
Kötü	201-300	851-1100	1001-2000	24001-32000	241-700	401-520
Tehlikeli	301-500	>1101	>2001	>32001	>701	>501

Tablo 2 incelendiğinde HKİ hesaplanırken kullanılan hava kirletici parametreleri ve bu parametrelerin 1 saatlik ve 8 saatlik ortalamalarına karşılık gelen HKİ değerlerini görülmektedir.

Tablo 3. İndeks Hesaplanan Parametrelerin Sınır Değerleri (“T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı”, 2022)

(Table 3. Limit Values of Index Calculated Parameters)

Parametre	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (µg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)
	1 Sa Ort	1 Sa Ort	8 Sa Ort	8 Sa Ort	24 Sa Ort
Ulusal Sınır Değer	410	270	10000	120	70
AB Üye Ülkeleri sınır Değeri	350	200	10000	120	50

Tablo 3’te HKİ hesaplanırken kullanılan hava kirletici parametrelerin ulusal ve AB üye ülkelerindeki sınır değerleri verilmiştir.

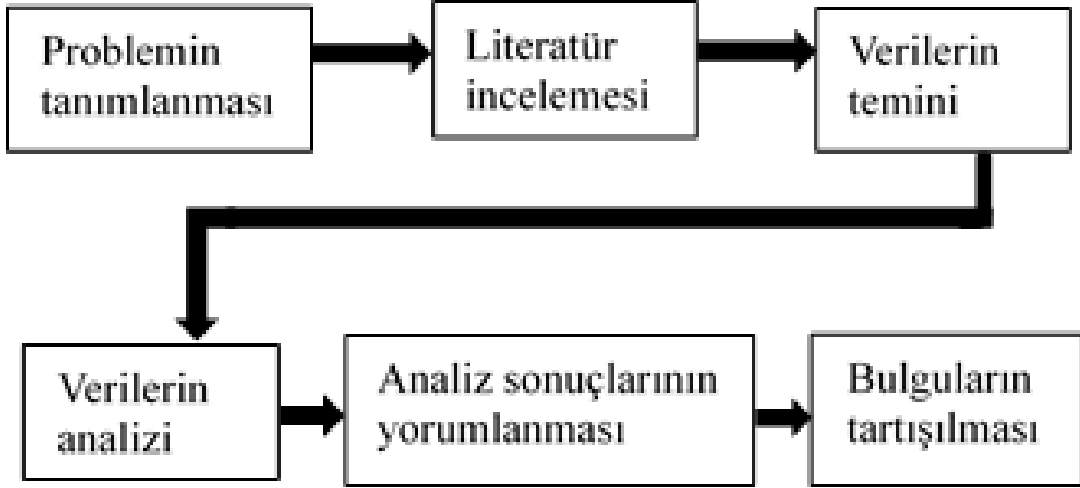
Sonuç olarak, günümüzde hava kirliliği önemli bir çevre ve sağlık sorunudur. Literatürde Doğu Anadolu bölgesinin Erzurum-Kars bölümünün (Ardahan, Erzurum, Iğdır ve Kars) hava kalitesi ile ilgili çok fazla çalışmaya rastlanılmamıştır. Yapılan bir çalışmada Ardahan merkezinde Çevre ve Şehircilik Bakanlığı’nın kurduğu Ulusal Hava Kalitesi İzleme Sistemi’ne bağlı ölçüm istasyonundan çeşitli veriler alınmış, kirlilik parametrelerinin (SO₂, NO₂, NO, NO_x, PM₁₀) değerleri Türkiye Hava Kalitesi Değerlendirme ve Yönetimi Yönetmeliği’ndeki limit kirletici değerlerinin altında olduğu saptanmıştır. Ancak Ardahan merkezde kış aylarındaki kirletici parametrelerin konsantrasyonlarının, yaz aylarındaki değerlerden fazla olduğu ve PM₁₀ değerinin 4 kat, SO₂’nin ise 7 kat yüksek olduğu belirtilmiştir. Ardahan ilinde 10/2017 – 09/2018 tarihleri arasında kullanılan kömürün yanması sonucunda yaklaşık 788 ton SO₂ oluştuğu belirtilmiştir (Guliyev & Akgün, 2020). Bu sebeple çalışma alanının hava kalitesinin izlenmesi ve bazı meteorolojik veriler ile makine öğrenmesi algoritmaları kullanılarak tahminlenmesi sonucu alınabilecek önlemlerle bu bölüm halkının solunum yolu ile ilgili sağlık sorunları yaşamasının önlenmesi halk sağlığı açısından oldukça önemlidir. Bu nedenle bu çalışmada 2019-2021 yılları arasında Ardahan, Erzurum, Iğdır ve Kars illerindeki hava kirliliğinin değerlendirilmesi amaçlanmıştır. Bunun için meteorolojik parametreler ile (ortalama sıcaklık, yağış, bulutluluk, nispi nem, basınç, rüzgâr hızı) HKİ’nin tahmin edilip edilemeyeceğini belirlemek için çoklu doğrusal regresyon analizi ve makine öğrenmesi yöntemleri kullanılmıştır.

Bu çalışmanın giriş bölümünde hava kirleticiler ve HKİ’nin genel hatları, materyal metot kısmında çalışmada kullanılan verilerin nereden alındığı ve hangi yöntemlerle analiz edildiği, bulgular ve tartışma kısmında analiz sonuçlarının literatür ile karşılaştırılması ve sonuç kısmında ise yapılacak benzer çalışmalar için ileriye yönelik tavsiyeler verilmiştir.

2. Materyal ve Metot

2.1. Materyal

Kullanılan fosil yakıtlar HKİ'ni doğrudan etkilemektedir (Akan & Morcalı, 2017). Bu sebeple kış aylarının soğuk geçtiği ve ısınmak için fazla yakıt kullanılan, Doğu Anadolu bölgesinin Erzurum-Kars bölümünde yer alan iller çalışma için seçilmiştir. Çalışma için hava sıcaklığının genellikle 15 oC'nin altında kaydedildiği 15 Eylül – 15 Mayıs tarihleri arası ısınma periyodu olarak belirlenmiştir. Ardahan, Erzurum, Iğdır ve Kars şehir merkezleri için 2019 – 2021 yılları arasındaki ısınma periyodunda saatlik ortalama kirletici konsantrasyonları Ulusal Hava Kalitesi İzleme Ağı üzerinden sağlanmıştır. Aynı dönem içerisindeki ortalama günlük hava sıcaklığı (oC), yağış, bulutluluk, basınç (hPa), nispi nem (%) ve rüzgâr hızı (m/s)'ndan oluşan meteorolojik parametreler ise Meteoroloji Genel Müdürlüğü'ne bağlı Meteorolojik Veri Bilgi Satış ve Sunum Sistemi (MEVBİS) üzerinden sağlanmıştır. Takip edilen metodoloji Şekil 1 'de gösterilmiştir.



Şekil 1. Takip edilen metodoloji (Figure 1. followed methodology)

2.2. Metot

Bir bölgedeki kirliliğin seviyesini anlamak için uygun bir aracın geliştirilmesi, o bölgede yaşayanların sağlıklarının korunması için gerekli önlemlerin alınmasında yetkililer açısından büyük önem taşımaktadır (Kyrkilis, Chaloulakou, & Kassomenos, 2007). Bu amaçla, geliştirilen standart değerler, gerek uyarıcı ve anlaşılabilir olması gerekse de kullanımı açısından yaygın olarak bir indekse çevrilebilir sunulabilmektedir. Belli bir bölgedeki hava kalitesinin karakterize edilmesi için ülkelerin kendi sınır değerlerine göre dönüştürdükleri ve kirlilik sınıflandırılmasının yapıldığı bu indekse Hava Kalitesi İndeksi (HKİ) adı verilmektedir. Bu indeks; beş temel kirletici parametre (PM₁₀, SO₂, NO₂, CO ve O₃) üzerinden hava kalitesini hesaplanmaktadır. HKİ hesaplamasında 1 numara olarak verilen denklem kullanılmaktadır (Fang, Liu, Li, Sun, & Miao, 2015). Denklem 1'e göre her bir kirletici parametre için ayrı ayrı kesme noktaları dikkate alınarak bir parametre indeksi (I_p) tanımlanmaktadır. Kirletici parametre indeksleri arasındaki maksimum değer, HKİ değeri (Denklem 2) olarak kabul edilmektedir.

$$I_p = \left[\frac{(I_{Hi} - I_{Lo})}{(BP_{Hi} - BP_{Lo})} \right] (C_p - BP_{Lo}) + I_{Lo} \quad 1$$

$$HKI = \max(I_{PM10}, I_{SO2}, I_{NO2}, I_{CO}, I_{O3}) \quad 2$$

I_p; p kirleticisi için indeks,

C_p; p kirleticisi için kırılmış ortalama konsantrasyon,

BP_{Hi}; C_p' ye eşit veya büyük kesme noktası konsantrasyonu,

BP_{Lo}; C_p' ye eşit veya küçük kesme noktası konsantrasyonu,

I_{Hi} ve I_{Lo}; sırasıyla BP_{Hi} ve BP_{Lo}'ye karşılık gelen HKİ değerleridir.

2.2.1. Çoklu Doğrusal Regresyon Modeli

Çoklu regresyon analizi, bağımlı değişkenle ilişkili olan iki ya da daha çok bağımsız değişkene dayalı olarak, bağımlı değişkenin tahmin edilmesine yönelik bir analiz türüdür. Çoklu doğrusal regresyon analizi için matematiksel model, n tane bağımsız değişken için şu şekilde yazılabilir:

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_n x_n \quad 3$$

Bu denklemde y , bağımsız değişkenler $X_1=x_1, X_2=x_2, \dots, X_n=x_n$ değerlerini aldığında Y değişkeninin beklenen değerini göstermektedir. a , kesişim noktası ve b_1, b_2, \dots, b_n regresyon katsayılarıdır.

Bağımsız değişkenlerin bağımlı değişken üzerindeki ortak etkisi R^2 ile incelenir. Bu değer tüm değişkenlerin birlikte y 'de açıkladıkları varyans oranını verir. R^2 değerinin 1'e yaklaşması regresyon denkleminin ifade ettiği ilişkinin deterministik bir ilişkiye yaklaştığını gösterir (Büyüköztürk, 2013).

2.2.2. Makine Öğrenmesi

Yapay sinir ağları; sınıflandırma, modelleme ve tahmin gibi birçok günlük hayat probleminin çözümünde başarılı sonuç veren bir yöntemdir (Arı and Berberler 2017). Makine öğrenmesi algoritmaları, gözlemlenen verilerin genellikle istatistiksel bir kayıp ya da hata dağılımları işlevini en aza indirerek gerçekleştirilen sınıflandırma modellerinde eğitimini yönlendirir (Kuncheva 2014; Marsland 2014). Eğitim için kullanılan veriler, sınıflandırılmış verileri tahmin etmek için benzer girdi değişkenlerine uygulanır (Hastie, Tibshirani, and Friedman 2008; Witten and Frank 2005). Makine öğrenme algoritmalarında, mevcut verilerin tipi ve istenen analiz göz önüne alındığında performanslarını optimize etmek için ayarlanmış bir veya daha fazla algoritmadan bir tanesinin seçilmesi gerekmektedir (Guyon, 2009).

Makine öğrenmesi algoritmaları ile meteorolojik parametreler kullanılarak HKİ'ni tahmin etmek için yapılan bu çalışma üç aşamada gerçekleştirilmiştir. Bunlar; (i) verilerin ön işlenmesi, (ii) sınıflandırma ve (iii) tahminleme aşamalarıdır. Verilerin ön işlenmesi aşamasında, Ulusal Hava Kalitesi İzleme Ağı üzerinden alınan HKİ'leri ve meteorolojik parametrelerdeki eksik veriler çıkarılmış ve ısınma periyodu olarak belirlenen dönemler için tarihe göre sıralanmıştır. İkinci aşama olan sınıflandırma aşamasında ise HKİ'leri bağımlı değişken olarak belirlenmiş ve T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı tarafından belirtildiği şekilde iyi, orta, hassas, sağlıksız, kötü ve tehlikeli olmak üzere altı ayrı sınıfa ayrılmıştır. Ayrıca meteorolojik parametreler bağımsız değişken olarak belirlenmiş ve reel sayı olarak kullanılmıştır. Üçüncü aşama olan tahminleme aşamasında ise bu veriler RapidMiner 9.10 versiyonunda farklı makine öğrenmesi (Naive Bayes, Derin Öğrenme, Rastgele Orman, k-NN, Karar Ağacı) algoritmaları ile analiz edilmiştir. Bu algoritmalar ile analiz edilirken verilerin %70'i eğitim verisi olarak, %30'u test verisi olarak kullanılmıştır. En başarılı algoritma tespit edilirken doğruluk ve Kappa değerleri göz önüne alınmıştır.

Naive Bayes

Naive Bayes, nesnelere sınıflandırmak için Bayes teoremini kullanan olasılıksal bir algoritmadır. "Toy" olarak adlandırılır çünkü girdinin tüm özelliklerinin birbirinden bağımsız olduğu varsayımını yapar ki bu genellikle gerçek dünya verilerinde geçerli değildir. Bu sınırlamaya rağmen Naive Bayes, uygulaması nispeten basit olduğundan ve az miktarda veriyle iyi çalışabildiğinden sınıflandırma için güçlü bir araç olabilir. Genel olarak Naive Bayes, çeşitli sınıflandırma görevleri için kullanılabilen basit ama güçlü bir algoritmadır. Bağımsızlık varsayımlarına rağmen, özellikle veri miktarı sınırlı olduğunda pratikte iyi çalışabilir (Vembandasamy et al., 2015).

Karar Ağacı

Karar ağacı algoritması, regresyon ve sınıflandırma problemlerinin çözümünde kullanılır. Dal, yapraklar ve kök düğümünden oluşur ve görünümü ters ağaçta benzerdir. Düğümlerde değişkenler temsil edilirken, dallarda ise değişkenlerin yer aldığı değerler temsil edilir. Yapraklarda ise nihai sonuç temsil edilir. Kök düğümünden yaprağa ulaşana kadar ardışık düğümler takip edilir ve karar alma işlemi gerçekleşmiş olur (Ünalı & Yalçın, 2022).

Derin Öğrenme

Derin öğrenme algoritmaları, çok sayıda dağıtılmış temsil düzeyini keşfetmeyi amaçlayan makine öğrenimi algoritmalarının bir alt kümesidir (Guo et al., 2016). Derin öğrenme, beynin sinir ağlarının yapısından ve işlevinden ilham alan makine öğreniminin bir alt kümesidir. Kalıpları tanımak ve karar vermek için birbirine bağlı düğümler veya "nöronlar" katmanlarından oluşan yapay sinir ağlarının eğitilmesini içerir. Derin öğrenme algoritmaları, görüntü ve konuşma tanıma, doğal dil işleme ve bilgisayar görüşü gibi büyük miktarlarda karmaşık, yapılandırılmamış veri içeren görevler için özellikle uygundur. Derin öğrenme modelleri, açık programlamaya ihtiyaç duymadan kendi başlarına öğrenip gelişebilir ve bu da onları çok çeşitli sorunları çözmek için güçlü bir araç haline getirir.

k-NN

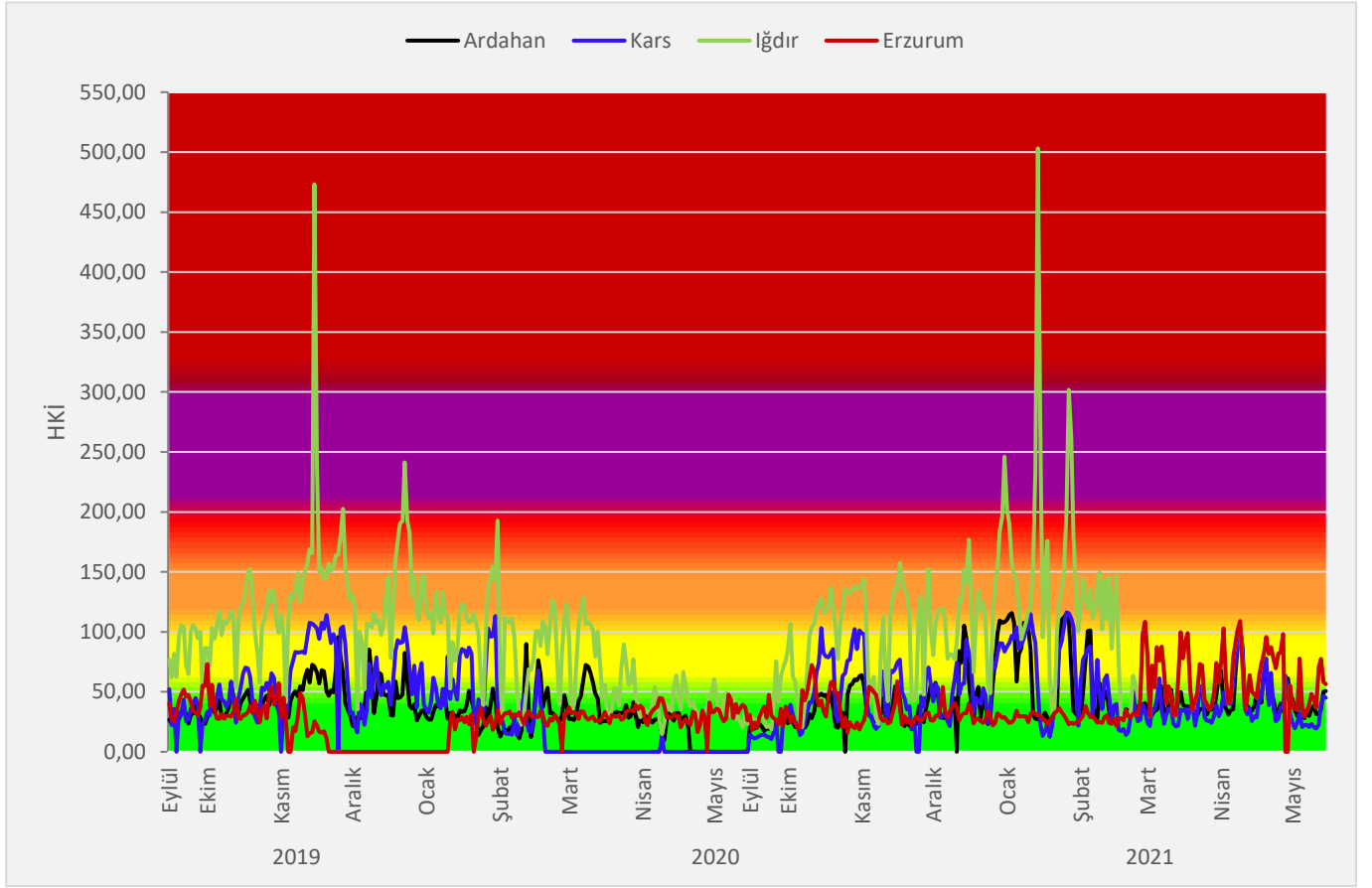
kNN, en basit makine öğrenmesi yöntemi olarak kabul edilmektedir. Model oluşturmak için bir eğitim sürecine ihtiyaç duyulmaz ve bellek tabanlıdır. K adet en yakın (en benzer) komşu ile regresyon ve sınıflandırma işlemleri yapılır. KNN algoritmasında, uzaklık ölçüğü olarak Öklid mesafesi kullanılır (Ünalı & Yalçın, 2022).

Rastgele Orman

Rastgele orman, tahminlerde bulunmak için karar ağaçlarının bir kombinasyonunu kullanan bir makine öğrenimi algoritmasıdır. Eğitim sırasında birden fazla karar ağacı oluşturularak ve ardından nihai bir tahmine ulaşmak için her ağaç tarafından yapılan tahminlerin ortalamasını alarak çalışır. Rastgele ormanlar, özellikle sınıflandırma ve regresyon gibi görevler için kullanışlıdır ve genellikle verilerin yüksek boyutlu veya çok sayıda özelliğe sahip olduğu durumlarda kullanılır. Rastgele orman algoritması hem denetimli hem de denetimsiz öğrenme için kullanılabilir ve makine öğrenimindeki en doğru ve sağlam yöntemlerden biri olarak kabul edilir. (Dokuz et al., 2020).

3. Araştırma Sonuçları ve Tartışma

Şekil 2’de 2019 – 2021 yılları ısınma dönemindeki Ardahan, Kars, Iğdır ve Erzurum illerine ait günlük ortalama HKİ değerleri görülmektedir.



Şekil 2. 2019 – 2021 yılları ısınma dönemi günlük ortalama HKİ değerleri (Daily average AQI values for the warm-up period between 2019 and 2021 years)

Şekil 2 incelendiğinde 2019-2021 yılları arasında Ardahan ili HKİ değerinin hassas ve daha aşağı bir değerde olduğu görülmektedir. Benzer şekilde Kars ili için HKİ değerinin 2019-2020 yılları arasında hassas ve daha aşağı bir değerde olduğu görülmektedir. Iğdır ili için HKİ değerinin ise 2019-2020 yılları arasında hassas ve daha üst bir değerde olduğu, yer yer tehlikeli düzeye çıktığı görülmektedir. Erzurum ili için HKİ değerinin 2019-2020 yılları arasında orta ve daha aşağısındaki bir değerde olduğu görülmektedir.

Tablo 4’te (4a, 4b, 4c, 4d) çalışılan illerin 2019-2021 yıllarında ısınma periyodu olarak belirlenen aralıkta sıcaklık, basınç, nispi nem, rüzgâr hızı, bulutluluk, yağış ve HKİ değerlerinin maksimum ve minimum değerleri verilmiştir.

Tablo 4a. 2019 – 2021 yılları ısınma periyodunda Ardahan ili için HKİ ve meteorolojik parametrelere ait bazı istatistikler (Table 4a. Some statistics of AQI and meteorological parameters for Ardahan province in the warming period of 2019 – 2021 years)

	Minimum Değer		Maksimum Değer	
	2019-2020	2020-2021	2019-2020	2020-2021
Sıcaklık (°C)	-21	-23	14	17
Basınç (hPa)	805,06	801,7	829,4	828,2
Nispi nem (%)	35	24	89	95
Rüzgâr hızı (ms ⁻¹)	0,4	0,4	2,7	2,6
Bulutluluk	0	0	7	0
Yağış (mm)	0	0	24,6	0
HKİ	9,81	17,11	95,77	115,67

Tablo 4a'ya göre Ardahan ilinde sıcaklığın, basıncın ve nispi nemin en düşük olduğu dönem 2020-2021 yılları arasındır. Sıcaklığın, nispi nemin ve HKİ'ni en yüksek olduğu dönem 2020-2021 yılları arasındır. Rüzgâr hızı her iki dönemde de en düşük $0,4 \text{ ms}^{-1}$ olarak saptanmıştır. Rüzgâr hızı 2019-2020 yılları arasında ise en yüksek değerdedir. Bulutluluk ve yağış verileri 2019-2020 yılları arasında en yüksek değerini almıştır. HKİ değeri ise en düşük 2019-2020 yılları arasında saptanmıştır.

Tablo 4b. 2019 – 2021 yılları ısınma periyodunda Kars ili için HKİ ve meteorolojik parametrelere ait bazı istatistikler (Table 4b. Some statistics of AQI and meteorological parameters for Kars province in the warming period of 2019 – 2021 years)

	Minimum Değer		Maksimum Değer	
	2019-2020	2020-2021	2019-2020	2020-2021
Sıcaklık (°C)	-20	-10	16	26
Basınç (hPa)	809,8	901	833,4	935,1
Nispi nem (%)	27	31	89	97
Rüzgâr hızı (ms^{-1})	0,8	0,2	5,4	3,4
Bulutluluk	0	0	8	8
Yağış (mm)	0	0	26,4	22,3
HKİ	13,10	18,89	113,95	503,20

Tablo 4b'ye göre Kars ilinde sıcaklığın, basıncın ve nispi nemin en düşük olduğu dönem 2019-2020 yılları arasındır. Sıcaklığın, nispi nemin ve HKİ'ni en yüksek olduğu dönem 2020-2021 yılları arasındır. Rüzgâr hızı 2020-2021 yılları arasında en düşük $0,2 \text{ ms}^{-1}$ olarak saptanmıştır. Rüzgâr hızı 2019-2020 yılları arasında ise en yüksek değerdedir. Yağış verisi 2019-2020 yılları arasında en yüksek değerini almıştır. Bulutluluk ve yağış verileri her iki dönemde de en düşük 0 olarak tespit edilmiştir. Bulutluluk verisinin en yüksek değeri ise her iki dönemde de 8 olarak gözlenmiştir. HKİ değeri ise en düşük 2019-2020 yılları arasında saptanmıştır.

Tablo 4c. 2019 – 2021 yılları ısınma periyodunda Iğdır ili için HKİ ve meteorolojik parametrelere ait bazı istatistikler (Table 4c. Some statistics of AQI and meteorological parameters for Iğdır province in the warming period of 2019 – 2021 years)

	Minimum Değer		Maksimum Değer	
	2019-2020	2020-2021	2019-2020	2020-2021
Sıcaklık (°C)	-5	-10	23	26
Basınç (hPa)	905,2	901	931,6	935,1
Nispi nem (%)	34	31	99	97
Rüzgâr hızı (ms^{-1})	0	0	3,2	3,4
Bulutluluk	0	0	8	8
Yağış (mm)	0	0	15,9	22,3
HKİ	14,94	18,89	473,23	503,20

Tablo 4c'ye göre Iğdır ilinde sıcaklığın, basıncın ve nispi nemin en düşük olduğu dönem 2020-2021 yılları arasındır. Sıcaklığın ve HKİ'nin en yüksek olduğu dönem 2020-2021 yılları arasındır. HKİ değeri en düşük 2019-2020 yılları arasında saptanmıştır. Nispi nem verisi en yüksek değerini 2019-2020 yılları arasında almıştır. Rüzgâr hızı her iki dönemde de en düşük 0 ms^{-1} olarak saptanmıştır. Rüzgâr hızı 2020-2021 yılları arasında ise en yüksek değerdedir. Yağış verisi 2020-2021 yılları arasında en yüksek değerini almıştır. Bulutluluk ve yağış verileri her iki dönemde de en düşük 0 olarak tespit edilmiştir. Bulutluluk verisinin en yüksek değeri ise her iki dönemde de 8 olarak gözlenmiştir.

Tablo 4d. 2019 – 2021 yılları ısınma periyodunda Erzurum ili için HKİ ve meteorolojik parametrelere ait bazı istatistikler (Table 4d. Some statistics of AQI and meteorological parameters for Erzurum province in the warming period of 2019 – 2021 years)

	Minimum Değer		Maksimum Değer	
	2019-2020	2020-2021	2019-2020	2020-2021
Sıcaklık (°C)	-17	-16	17	20
Basınç (hPa)	799,2	796,1	825,2	821,9
Nispi nem (%)	25	30	98	99
Rüzgâr hızı (ms ⁻¹)	0,1	0,1	4,6	4,4
Bulutluluk	0	0	7	7
Yağış (mm)	0	0	42,5	14,6
HKİ	11,14	15,96	73,07	72,35

Tablo 4d'ye göre Erzurum ilinde sıcaklığın ve nispi nemin en düşük olduğu dönem 2019-2020 yılları arası, en yüksek olduğu dönem 2020-2021 yılları arasındır. Basıncın en düşük olduğu dönem ise 2020-2021 yılları arası, en yüksek olduğu dönem ise 2019-2020 yılları arasındır. Rüzgâr hızı her iki dönemde de en düşük 0,1 ms⁻¹ olarak saptanmıştır. Rüzgâr hızı 2019-2020 yılları arasında ise en yüksek değerdedir. Yağış verisi 2019-2020 yılları arasında en yüksek değerini almıştır. Bulutluluk ve yağış verileri her iki dönemde de en düşük 0 olarak tespit edilmiştir. Bulutluluk verisinin en yüksek değeri ise her iki dönemde de 7 olarak gözlenmiştir. HKİ'nin en yüksek ve en düşük olduğu dönemler ise 2019-2020 yılları arasındır.

3.1. Çoklu Doğrusal Regresyon Analizi

HKİ ve Meteorolojik parametreler arasındaki korelasyon analizi sonuçları Tablo 5'te verilmiştir. Tablo 5'e göre HKİ ile yağış ve rüzgâr hızı arasında negatif yönlü, nispi nem ve basınç arasında pozitif yönlü bir ilişki bulunduğu gözlenmiştir. Ancak ortalama sıcaklık ve bulutluluk arasında anlamlı bir ilişki bulunmadığı belirlenmiştir.

Tablo 5. HKİ ve Meteorolojik parametreler arasındaki korelasyon analizi sonuçları (Table 5. Correlation analysis results between AQI and Meteorological parameters)

Değişkenler	HKİ	Ortalama Sıcaklık	Yağış	Bulutluluk	Nispi Nem	Basınç	Rüzgâr Hızı
HKİ	r	1	,026	-,096**	-,021	,065**	,613**
	p		,230	,000	,329	,003	,000
	n	2177	2177	2177	2128	2144	2168
Ortalama Sıcaklık	r	,026	1	,052*	-,043*	-,520**	,302**
	p	,230		,011	,039	,000	,000
	n	2177	2356	2356	2277	2322	2347
Yağış	r	-,096**	,052*	1	,336**	,285**	-,077**
	p	,000	,011		,000	,000	,000
	n	2177	2356	2356	2277	2322	2347
Bulutluluk	r	-,021	-,043*	,336**	1	,505**	-,011
	p	,329	,039	,000		,000	,606
	n	2128	2277	2277	2277	2243	2275
Nispi Nem	r	,065**	-,520**	,285**	,505**	1	-,079**
	p	,003	,000	,000	,000		,000
	n	2144	2322	2322	2243	2322	2313
Basınç	r	,613**	,302**	-,077**	-,011	-,079**	1
	p	,000	,000	,000	,606	,000	
	n	2168	2347	2347	2275	2313	2347
Rüzgâr Hızı	r	-,225**	,132**	,116**	,296**	-,156**	-,229**
	p	,000	,000	,000	,000	,000	
	n	2177	2356	2356	2277	2322	2347

** P< 0,01

* P< 0,05

Tablo 6. HKİ ve Meteorolojik parametreler arasındaki regresyon analizi sonuçları (Table 6. Results of regression analysis between AQI and Meteorological parameters)

	Katsayılar	Standart Hata	t	R ²	Düzenlenmiş R ²	Standart Hata	F
<i>Meteorolojik Parametreler</i>							
Kesişim	-432,021	14,054	-30,740				
Ort. Sıcaklık	-0,600	0,104	-6,321				
Yağış	-,0689	0,263	-2,616	0,420	0,418	28,961	251,491
Bulutluluk	-0,472	0,340	-1,388				
Nispi Nem	-0,152	0,067	2,272				
Basınç	0,571	0,016	35,567				
Rüzgâr hızı	-1,152	0,935	-1,231				

P < 0,05

Tablo 6'da HKİ ve Meteorolojik parametreler arasındaki regresyon analizi sonuçları verilmiştir. Tablo 6 incelendiğinde, HKİ ve meteorolojik parametreler arasında yapılan çoklu doğrusal regresyon analizine göre geliştirilen modelin HKİ'ni %42 oranında tahmin edeceği söylenebilmektedir.

3.2. Makine Öğrenmesi Yöntemi

Tablo 7'de HKİ ve Meteorolojik parametreler arasındaki makine öğrenmesi analizi sonuçları verilmiştir. Tablo 7 incelendiğinde, makine öğrenmesi yöntemlerinden Naive Bayes'in meteorolojik parametreler kullanarak HKİ'ni tahminlemede %68,91, Karar ağacı yönteminin HKİ'ni tahminlemede %70,29, Derin öğrenme yönteminin HKİ'ni tahminlemede %73,05, k-NN yönteminin HKİ'ni tahminlemede %73,35 ve Rastgele orman yönteminin ise HKİ'ni tahminlemede %74,89 başarılı olduğu bulunmuştur.

Tablo 7. HKİ ve Meteorolojik parametreler arasındaki makine öğrenmesi analizi sonuçları (Table 7. Machine learning analysis results between AQI and Meteorological parameters)

Makina Öğrenmesi Yöntemleri	Doğruluk	Kappa Değerleri
Naive Bayes	%68,91	0,304
Karar Ağacı	%70,29	0,375
Derin Öğrenme	%73,05	0,362
k-NN	%73,35	0,422
Rastgele Orman	%74,89	0,381

Rybarczyk ve Zalakeviciute (2016) karar ağacı algoritmasını kullanarak hava kirliliği tahminine yönelik yaptıkları çalışmada geliştirdikleri modelin %65'in üzerinde bir oranla tahminleme yapabildiğini bildirmişlerdir. Ayturan (2019), havadaki partikül madde konsantrasyonuna yönelik yaptığı çalışmasında geçitli tekrarlayan ünite (GRU) ve devirli sinir ağları (RNN)'nin kombinasyonundan oluşan bir model kullanmış ve bu modelin tahmin performansını %83'e yakın olarak bildirmiştir. Benzer şekilde Irmak ve Aydılek (2019), yaptıkları çalışmalarında Rastgele Orman, Yığın, Karar Ağacı, Örneklemeli Toplam, Eğitimli Arttırıcı, k-NN ve Yapay Sinir Ağları algoritmalarını denemişler ve %76 ile %99 arasında bir oranla geliştirdikleri modelin tahminleme yapabildiğini bildirmişlerdir. Çelik ve Arıcı (2021), ise yaptıkları çalışmada k-NN, Naive Bayes, Rastgele Orman, Karar Ağacı, Destek Vektör Makinası algoritmalarını kullanmışlar ve %92 ile %99 arasında bir doğruluk raporlamışlardır. Gültepe (2019); yapay sinir ağları (YSA), rastgele orman, k-NN, karar ağacı ve naive bayes algoritmalarını ve lineer regresyon metodunu kullanarak yaptığı hava kirliliğinin tahminine yönelik çalışmasında rastgele orman algoritmasını %99, karar ağacı algoritmasının %99 ve YSA'nın %87 doğruluk oranına sahip olduğunu ve lineer regresyon metodunun %30 doğruluğa sahip olduğunu rapor etmiştir. Benzer şekilde yapılan başka bir çalışmada ise rastgele orman algoritması iki farklı sınıflama yapılarak kullanılmış ve %81,61-%86,69 oranlarında başarı bildirilmiştir (Mehtap ve Canbay 2023). Hava kirliliği parametrelerinden PM10 analizine yönelik bir çalışmada ise k-NN, Destek Vektör Regresyonu (DVR), rastgele orman, yapay sinir ağları kullanılmış olup en yüksek doğruluk oranının ise %0,85 ile DVR'na ait olduğu

raporlanmıştır (Gökçek, Şaşa, Dokuz, and Bozdağ 2022). Çalışmada bulunan sonuçlar ile literatürde bulunan çalışmaların sonuçları ile kıyaslanmış ve özet olarak Tablo 8’de verilmiştir.

Tablo 8. Çalışmanın sonuçlarının literatürdeki çalışmalarla kıyaslanması (Table 8. Comparison of the results of the study with the studies in the literature)

Kullanılan model	Performans	Kaynak
Karar Ağacı	>%65	(Rybarczyk & Zalakeviciute, 2016)
Geçitli Tekrarlayan Ünite (GRU), Devirli Sinir ağları (RNN)	~%83	(Ayturan, 2019)
Rastgele Orman, Yığın, Karar Ağacı, Örneklemeli Toplam, Eğitimli Arttırıcı, k-NN, Yapay Sinir Ağları	%76-%99	(Irmak & Aydilek, 2019)
k-NN, Naive Bayes, Rastgele Orman, Karar Ağacı, Destek Vektör Makinası	%92-%99	(Çelik & Arıcı, 2021)
YSA, rastgele orman, k-NN, karar ağacı, naive bayes, lineer regresyon	%30-%99	(Gültepe 2019)
Rastgele orman	%81,61-%86,69	(Mehtap ve Canbay 2023)
k-NN, Destek Vektör Regresyonu (DVR), rastgele orman, yapay sinir ağları	%82-%85	(Gökçek vd., 2022).

4. Sonuç

Bu çalışmada; 15/09/2019 – 15/05/2020 ve 15/09/2020 – 15/05/2021 tarihleri arasındaki Ardahan, Kars, Erzurum ve Iğdır illeri şehir merkezlerindeki HKİ ve meteorolojik parametreler (ortalama sıcaklık, yağış, bulutluluk, nispi nem, basınç, rüzgâr hızı) arasında çoklu doğrusal regresyon analizi ve makine öğrenmesi yöntemleri kullanılarak HKİ'nin tahmini yapılmıştır. Yapılan analizler sonucunda makine öğrenmesi yöntemlerinin çoklu doğrusal regresyon analizine göre daha yüksek doğrulukla HKİ'ni tahmin ettiği sonucuna ulaşılmıştır. Çoklu doğrusal regresyon analizi ile yapılan HKİ'ni tahminleme oranı %41,8 bulunmuştur. Makine öğrenmesi yöntemlerinden Naive Bayes algoritmasının HKİ'ni %68,91, Karar Ağacı algoritmasının HKİ'ni %70,29, Derin öğrenme algoritmasının HKİ'ni %73,05, k-NN algoritmasının HKİ'ni %73,35 ve Rastgele Orman algoritmasının HKİ'ni %74,89 doğrulukla tahmin ettiği sonucuna ulaşılmıştır. Sonuçlardan HKİ tahminine yönelik çalışmalarda makine öğrenmesi algoritmalarının kullanılabilceği anlaşılmıştır. HKİ'ni tahminlemedeki doğruluk oranının artırılmasına yönelik çalışmalarda farklı makine öğrenmesi algoritmalarının denenmesi önerilmektedir. Ayrıca bu çalışmada kullanılan meteorolojik parametrelerin yanı sıra, başka meteorolojik parametrelerin de kullanılarak HKİ'ni tahminleme çalışmalarının yapılması önerilmektedir.

Kaynakça

- Akan, D. S., & Morcalı, M. H. (2017). Kahramanmaraş Hava Kirliliği Kaynaklarının İzlenmesi ve Belirlenmesi Monitoring and Determination of Air Pollution Sources in Kahramanmaraş. *KSU Journal of Engineering Sciences*, 20(2), 105–115.
- Arı, A., & Berberler, M. E. (2017). Yapay Sinir Ağları ile Tahmin ve Sınıflandırma Problemlerinin Çözümü İçin Arayüz Tasarımı. *Acta Infologica*, 1(2), 55–73. <https://dergipark.org.tr/en/pub/acin/issue/33868/335553>
- Ayturan, Y. A. (2019). Derin Öğrenme ile Havadaki Partikül Madde Konsantrasyonu Tahmin (Vol. 8, Issue 5). Karatay Üniversitesi.
- Büyüköztürk, Ş. (2013). Sosyal Bilimler İçin Veri Analizi El Kitabı (18th ed.). Pegem Akademi.
- Çelik, D. B., & Arıcı, N. (2021). Covid-19 Salgın Sürecinde Hava Kalitesi Tahmini: Zonguldak Örneği. *Gazi Journal of Engineering Sciences*, 7(3), 222–232. <https://doi.org/10.30855/gmbd.2021.03.05>
- Dokuz, Y., Bozdağ, A., & Gökçek, B. (2020). HavKalitesiParametreleriniTahmini ve Mekansal Dağılım İçin Makine Öğrenmesi Yöntemlerinin Kullanılması. *Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 9(1), 37–47. <https://doi.org/10.28948/ngumuh.654092>
- Fang, C., Liu, H., Li, G., Sun, D., & Miao, Z. (2015). Estimating the impact of urbanization on air quality in China using spatial regression models. *Sustainability (Switzerland)*, 7(11), 15570–15592. <https://doi.org/10.3390/su71115570>
- Guliyev, R., & Akgün, M. (2020). Ardahan’da kullanılan kömürün hava kirliliğine etkisinin incelenmesi. *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22(2), 479–489. <https://doi.org/10.25092/baunfbed.709411>
- Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27–48. <https://doi.org/10.1016/j.neucom.2015.09.116>

- Gökçek, Begüm Öznur, Nuray Şaşa, Yeşim Dokuz, and Aslı Bozdağ. 2022. "PM10 Parametresinin Makine Öğrenmesi Algoritmaları İle Mekânsal Analizi, Kayseri İli Örneği." *Deu Muhendislik Fakultesi Fen ve Muhendislik* 24(70):65–80. doi: 10.21205/deufmd.2022247008.
- Gültepe, Yasemin. 2019. "Makine Öğrenmesi Algoritmaları İle Hava Kirliliği Tahmini Üzerine Karşılaştırmalı Bir Değerlendirme." *European Journal of Science and Technology* (16):8–15. doi: 10.31590/ejosat.530347.
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). El anillo esofágico inferior. In *The Elements of Statistical Learning* (Issue 2).
- Irmak, M. E., & Aydilek, İ. B. (2019). Hava Kalite İndeksinin Tahmin Başarısının Artırılması için Topluluk Regresyon Algoritmalarının Kullanılması. *Academic Platform Journal of Engineering and Science*, 507–514. <https://doi.org/10.21541/apjes.478038>
- Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, 151(2), 362–367. <https://doi.org/10.1016/j.envpol.2007.06.012>
- Katsouyanni, K., Touloumi, G., Samoli, E., Gryparis, A., Le Tertre, A., Monopoli, Y., Rossi, G., Zmirou, D., Ballester, F., Boumghar, A., Anderson, H. R., Wojtyniak, B., Paldy, A., Braunstein, R., Pekkanen, J., Schindler, C., & Schwartz, J. (2001). Confounding and effect modification in the short-term effects of ambient particles on total mortality: Results from 29 European cities within the APHEA2 project. *Epidemiology*, 12(5), 521–531. <https://doi.org/10.1097/00001648-200109000-00011>
- Kuncheva, L. I. (2014). Combining Pattern Classifiers. In *Combining Pattern Classifiers* (2nd ed.). Wiley. <https://doi.org/10.1002/9781118914564.ch9>
- Kyrkilis, G., Chaloulakou, A., & Kassomenos, P. A. (2007). Development of an aggregate Air Quality Index for an urban Mediterranean agglomeration: Relation to potential health effects. *Environment International*, 33(5), 670–676. <https://doi.org/10.1016/j.envint.2007.01.010>
- Marsland, S. (2014). *Machine Learning An Algorithmic Perspective Second Edition*. <https://b-ok.cc/book/2543746/ef80cb>
- Mehtap, Öklü, and Pelin Canbay. 2023. "Makine Öğrenmesi Yöntemleri İle Şehirlerin Hava Kalitesi Tahmini." *International Journal of Advances in Engineering and Pure Sciences* 35(1):39–53. doi: 10.7240/jeps.1175507.
- Rybarczyk, Y., & Zalakeviciute, R. (2016). Machine learning approach to forecasting urban pollution. 2016 IEEE Ecuador Technical Chapters Meeting, ETCM 2016. <https://doi.org/10.1109/ETCM.2016.7750810>
- T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı. (2022). <https://sim.csb.gov.tr/Home/HKI?baslik=HAVZA İZLEME SİSTEMİ>
- T.C. Sağlık Bakanlığı. (2014). Türkiye kronik hava yolu hastalıkları önleme ve kontrol programı (2014-2017).
- Ünalı, S., & Yalçın, N. (2022). Hava Kirliliğinin Makine Öğrenmesi Tabanlı Tahmini: Başakşehir Örneği Prediction of Air Pollution based on Machine Learning Methods: A Case Study for Başakşehir, İstanbul. 4(1), 35–44.
- Vembandasamy, K., Sasipriya, R., & Deepa, E. (2015). Heart Diseases Detection Using Naive Bayes Algorithm. *International Journal of Innovative Science, Engineering & Technology*, 2(9), 441–444.
- Witten, I. H., & Frank, E. (2005). Credibility: Evaluating What's been Learned. In *Data Mining: Practical machine learning tools and techniques*. <http://books.google.com/books?hl=en&lr=&id=QTnOcZJzIUoC&oi=fnd&pg=PR17&dq=Data+Mining+Practical+Machine+Learning+Tools+and+Techniques&ots=3gpDdrWiOc&sig=TZS7G811eXSa2SpAvfD6aBoJ2lw>