

Regresyon Analizi ile Hava Kalitesinin Sağlık Üzerindeki Etkisinin İncelenmesi

Hakan Yüksel^{1*}

¹Bilgisayar Teknolojileri / Teknik Bilimler Meslek Yüksekokulu, Isparta Uygulamalı Bilimler Üniversitesi, Türkiye

*hakanyuksel@isparta.edu.tr

(Received: 04 December 2024, Accepted: 06 December 2024)

(3rd International Conference on Recent Academic Studies ICRAS 2024, December 03-04, 2024)

ATIF/REFERENCE: Yüksel, H. (2024). Regresyon Analizi ile Hava Kalitesinin Sağlık Üzerindeki Etkisinin İncelenmesi. *International Journal of Advanced Natural Sciences and Engineering Researches*, 8(11), 245-255.

Özet –Küresel ısınma ve iklim değişikliğinin yoğun bir şekilde yaşandığı günümüzde hava kalitesi hem akut hem de kronik sağlık durumlarını etkileyen kritik bir rol üstlenmektedir. Bu makale, hava kirliliği ile sağlık etkileri arasındaki ilişkiyi analiz etmekte ve çeşitli regresyon analizleriyle en yüksek doğruluk oranını tespit etmektedir. Partikül madde (PM_{2.5} ve PM₁₀), nitrojen dioksit (NO₂), kükürt dioksit (SO₂) ve ozon (O₃) gibi temel kirlleticilerin hava kalitesi ölçümlerinin etkisini ortaya çıkarmaktadır. Aynı zamanda hava koşullarının da hava kalitesine etkisi göz önünde bulundurulmaktadır. Hem hava kalitesi ölçümleri hem de hava koşulları sağlık etkisi üzerinden değerlendirilmiştir. Makine öğrenmesi regresyonu, bağımsız değişkenler veya özellikler ile bağımlı bir değişken veya sonuç arasındaki ilişkiyi araştırmak için kullanılan bir tekniktir. Bu çalışmada aynı zamanda sağlık etkisi ölçümler için doğrusal regresyon, rasgele orman regresyon, k-en yakın komşu regresyon ve gradient boosting regresyonu kullanılmıştır. Sonuçlar incelendiğinde, rasgele orman Regresyon ve Gradient Boosting Regresyonun en iyi sonucu verdikleri, gradient boosting Regresyonun ise rasgele orman Regresyona göre daha hızlı sonuç ürettiği gözlenmiştir. Veriler incelendiğinde sağlık risklerini azaltmak için hava kirliliğini azaltmak ve küresel ve yerel aksiyonların alınması gerekliliği ortaya koymaktadır.

Anahtar Kelimeler –Hava Kalitesi, Sağlık, Regresyon, Analiz, Makine Öğrenmesi.

I. GİRİŞ

Hava kalitesi, insan refahı için derin etkileri olan çevre sağlığının temel bir yönüdür. Sanayileşme ve kentleşme küresel olarak hızlandıkça, hava kirliliği özellikle yoğun nüfuslu bölgelerde morbidite ve mortaliteye önde gelen bir katkıda bulunan faktör olarak ortaya çıkmıştır [1] [2]. İnce partikül madde (PM_{2.5} ve PM₁₀), azot oksitler (NO₂), kükürt dioksit (SO₂) ve ozon (O₃) gibi kirleticiler, solunum ve kardiyovasküler sistemler üzerindeki olumsuz etkileri nedeniyle yaygın olarak belgelenmiştir [3] [4]. Bu kirleticilere kronik maruziyet, astım ve kronik obstrüktif akciğer hastalığı (KOAH) gibi önceden var olan rahatsızlıkları kötüleştirmekle kalmaz, aynı zamanda akciğer kanseri ve felç gibi yeni sağlık sorunlarına da katkıda bulunmaktadır. Çocuklar, yaşlılar ve düşük gelirli topluluklar gibi savunmasız nüfuslar, orantısız risklerle karşı karşıyadır. Bu da halk sağlığı zorluklarını daha da kötüleştirmektedir [5].

Regresyon analizi, hava kalitesi ile sağlık sonuçları arasındaki karmaşık ilişkiyi anlamak için temel bir metodoloji haline gelmiştir. Bu istatistiksel araç, araştırmacıların sosyoekonomik faktörler, yaşam tarzı davranışları ve hava koşulları gibi kafa karıştırıcı değişkenleri hesaba katarak bireysel kirleticilerin sağlık üzerindeki etkisini izole etmelerini sağlamaktadır [6]. Zaman serisi regresyonu ve dağıtılmış gecikme modelleri gibi son gelişmeler hem anlık hem de uzun vadeli sağlık etkilerine dair daha derin içgörüler sunmaktadır. Çalışmalar incelendiğinde PM2.5 konsantrasyonlarında 10 $\mu\text{g}/\text{m}^3$ lük bir artışın kardiyovasküler hastalıklar için hastane yatışlarında önemli bir artışla ilişkili olduğunu göstererek kirlilik seviyelerini azaltmanın aciliyetini vurgulamıştır [7].

Hava kalitesi ile sağlık arasındaki etkileşim, kentsel yayılma ve iklim değişikliği de dahil olmak üzere küresel çevresel değişiklikler tarafından daha da karmaşık hale gelmiştir. Yükselen sıcaklıklar yer seviyesindeki ozon oluşumunu şiddetlendirirken, iklim değişikliğiyle bağlantılı aşırı hava olayları kirlilik olaylarını yoğunlaştırarak sağlık sonuçları olan bir geri bildirim döngüsü oluşturmaktadır [8].

Bu çalışma, hava kalitesi ile sağlık arasındaki etkileşimli müdahalelere öncelik vermek için sağlam istatistiksel modeller kullanmanın önemini vurgulamaktadır. Ayrıca, regresyon analizi kullanarak hava kalitesinin sağlık üzerindeki etkilerini araştırılmıştır. Veri odaklı içgörülerden yararlanarak, hava kalitesi ve halk sağlığının etkisine katkıda bulunmayı amaçlamaktadır.

II. LİTERATÜR ÖZETİ

Hava kalitesi ile sağlık arasındaki ilişki, hava kirliliğinin çeşitli ve geniş kapsamlı etkilerini vurgulayarak çevresel sağlık araştırmalarının odak noktası olmuştur. Çok sayıda çalışma, partikül madde (PM2.5 ve PM10), ozon (O3), azot oksitler (NO₂) ve kükürt dioksit (SO₂) gibi havadaki kirleticilerin solunum ve kardiyovasküler hastalıkları şiddetlendirme, hastane yatışlarını artırma ve erken ölümlere katkıda bulunmadaki rolünü vurgulamaktadır [4] [9].

Partikül madde (PM), olumsuz sağlık etkilerinin belgelenmiş bir kanıtıdır. İnce parçacıklar (PM2.5) solunum sisteminin derinliklerine nüfuz ederek iltihaplanmaya ve oksidatif strese neden olmaktadır. Çalışmalar, PM maruziyeti ile astım, kronik obstrüktif akciğer hastalığı (KOAH) ve kardiyovasküler olaylar gibi durumların yüksek riskleri arasında sürekli olarak güçlü bir ilişki olduğunu göstermiştir. Regresyon modellerini kullanan bir çalışma, artan PM2.5 seviyelerini solunum sıkıntısı nedeniyle acil servise yapılan ziyaretlerin daha yüksek sıklığıyla ilişkilendirdi [10] [11].

Benzer şekilde, ozon ve nitrojen oksitler gibi gaz halindeki kirleticiler akut ve kronik sağlık etkileriyle ilişkilidir. NO_x ve uçucu organik bileşikler (VOC'ler) içeren fotokimyasal reaksiyonlar yoluyla oluşan ozon, özellikle daha sıcak aylarda zararlıdır [12]. Zaman serisi regresyon analizi kullanan çalışmalar, yüksek ozon seviyeleri ile astım ataklarındaki artışlar ve solunum sorunları nedeniyle hastaneye yatışlar arasında önemli bir korelasyon olduğunu ortaya koymaktadır.

Regresyon analizi, hava kirliliğinin sağlık üzerindeki etkilerini ölçmede önemli bir araç olarak ortaya çıkmıştır [7] [13]. Geleneksel doğrusal regresyon modelleri genellikle kirletici-sağlık ilişkilerini belirlemek için kullanılırken, dağıtılmış gecikmeli doğrusal olmayan modeller (DLNM'ler) gibi gelişmiş teknikler araştırmacıların gecikmiş sağlık etkilerini yakalamasına olanak tanımaktadır. Dikkat çekici bir örnek, kısa vadeli maruz kalma etkilerini değerlendirmek için günlük hava kalitesi verilerini sağlık sonuçlarıyla bütünleştiren çalışmaları içermektedir. Kısa süreli kirletici artışlarının bile ölçülebilir sağlık sonuçlarına sahip olabileceğini ortaya koymaktadır [7] [14] [15].

Araştırmalar ayrıca farklı demografik gruplar arasında hava kirliliğine maruz kalma ve bunun sağlık sonuçlarında önemli farklılıkları da vurgulamaktadır [16] [17]. Gelir ve sağlık hizmetlerine erişim gibi sosyoekonomik faktörler genellikle belirli popülasyonların savunmasızlığını daha da kötüleştirmektedir. Çocuklar, yaşlılar ve önceden var olan rahatsızlıkları olan bireyler daha yüksek risklerle karşı karşıyadır ve bu da acil müdahaleleri gerektirmektedir [18].

Hava kirliliğine atfedilen küresel hastalık yükü, sıkı hava kalitesi standartlarına ve etkili azaltma stratejilerine olan ihtiyacı vurgulamaktadır [19] [20]. Bölgeler arası karşılaştırmalı analizler, emisyon

azaltımları ve yeşil kentsel girişimler gibi sağlam izleme sistemleri ve proaktif politikalara sahip ülkelerin daha iyi halk sağlığı sonuçları gösterdiğini göstermektedir [21].

Mevcut literatür, çevre bilimi, halk sağlığı ve istatistiksel modellemeyi birleştiren disiplinler arası araştırmalara olan kritik ihtiyacı vurgulamaktadır [22] [23] [24]. Araştırmacılar, regresyon analizinden yararlanarak hava kirliliğini azaltmayı ve sağlık etkilerini hafifletmeyi amaçlayan politikaları bilgilendirmek için eyleme geçirilebilir içgörüler sağlayabilirler. Kirletici dinamikleri ve sağlık sonuçları üzerindeki hava kalitesi etkileri de dahil olmak üzere ortaya çıkan zorlukları ele almak için daha fazla analiz çalışmalarında bulunmak gereklidir.

III. MATERYAL VE YÖNTEM

Bu kısımda çalışmada kullanılan veriseti, verisetine ait önışlem adımları ve kullanılan yöntemin detayları bahsedilmiştir.

A. Veri Seti

Bu çalışmada, Rabie El Kharoua tarafından hazırlanmış “Hava Kalitesi ve Sağlık Etkisi Veri Seti” kullanılmıştır [25]. Bu veri seti, hava kalitesi ile sağlık arasındaki ilişkiye dair kapsamlı bir görünüm sunarak araştırma, tahmin, modelleme ve istatistiksel analiz için tercih edilmektedir.

Bu veri seti, 5811 kayıt için hava kalitesi ve sağlık üzerindeki etkisi hakkında kapsamlı bilgi barındırmaktadır. Farklı kategoriler altında 15 farklı özellik bulunmaktadır. En önemli özellik olan hava kalitesi endeksi (AQI) nin yanı sıra çeşitli kirleticilerin konsantrasyonları, hava koşulları ve sağlık etkisi ölçümleri gibi değişkenleri içerir. Hedef değişken, sağlık etkisini hava kalitesine ve diğer ilgili faktörlere göre kategorize eden sağlık etkisi sınıfı olarak verilmiştir. Veriseti özelliklerine ilişkin detaylar şu şekildedir.

Hava Kalitesi Ölçümleri

- AQI: Hava Kalitesi Endeksi, havanın şu anda ne kadar kirli olduğunu veya ne kadar kirli olacağını tahmin edildiğini ölçen bir ölçüdür.
- PM10: Çapı 10 mikrometreden küçük partikül madde konsantrasyonu ($\mu\text{g}/\text{m}^3$).
- PM2_5: Çapı 2,5 mikrometreden küçük partikül madde konsantrasyonu ($\mu\text{g}/\text{m}^3$).
- NO2: Azot dioksit konsantrasyonu (ppb).
- SO2: Kükürt dioksit konsantrasyonu (ppb).
- O3: Ozon konsantrasyonu (ppb).

Hava Koşulları

- Temperature: Santigrat derece ($^{\circ}\text{C}$) cinsinden sıcaklık.
- Humidity: Nem yüzdesi (%).
- WindSpeed: Saniye başına metre (m/s) cinsinden rüzgar hızı.

Sağlık Etki Ölçümleri

- RespiratoryCases: Bildirilen solunum vakası sayısı.
- CardiovascularCases: Bildirilen kardiyovasküler vaka sayısı.
- HospitalAdmissions: Bildirilen hastaneye yatırılanların sayısı.

Sağlık Etki Sınıfı

- HealthImpactScore: Hava kalitesi ve diğer ilgili faktörlere dayalı genel sağlık etkisini gösteren, 0 ile 100 arasında değişen bir puan.
- HealthImpactClass: Sağlık etkisi puanına dayalı sağlık etkisinin sınıflandırılmasıdır.

B. Veri Önışleme

Veriseti analizi ve eğitimi öncesinde, optimum performansı sağlamak için bir ön işleme aşamasından geçmiştir. Ön işleme hattı birkaç önemli adımı kapsamaktadır. 5811 kayıt ve 15 özellik içeren veri setinde

hem boş geçilen hem de tekrar edilen veriler incelendiğinde herhangi bir sonuca ulaşamamıştır. Ayrıca özelliklerin tamamı int64 veri tipinde saklanmaktadır.

Veriseti analizinden sonra regresyon analizleri gerçekleştirilmiştir. Bu regresyon analizlerinde detaylı bir model değerlendirmesi sağlamak için, veri setinin %80'ini eğitim için, %20'ini ise test için ayıran bir veri bölümü uygulanmıştır.

C. Regresyon

Doğrusal Regresyon, bir veya daha fazla bağımsız değişkene dayalı sürekli bir bağımlı değişkeni tahmin etmek için kullanılan istatistiksel ve makine öğrenimi yöntemidir. Giriş özellikleri ile hedef değişken arasında,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

ϵ hata terimini temsil etmekte. Formül doğrusal bir ilişki olduğunu varsayar. Model, kareli artıkların toplamını en aza indirmek için β katsayılarını optimize ederek gözlenen verilere en iyi uyumu sağlamaktadır.

Rastgele Orman Regresörü, tahmin doğruluğunu artırmak için birden fazla karar ağacı kullanan bir topluluk öğrenme yöntemidir. Ormandaki her ağaç, verilerin rastgele bir alt kümesi üzerinde eğitilir ve her düğümde bölünmek üzere rastgele özellikler seçmektedir. Tahminler, tüm ağaçların çıktılarının ortalaması alınarak yapılır, aşırı uyumu azaltır ve sağlamlığı artırılır. Bu parametrik olmayan yöntem, verilerdeki karmaşık, doğrusal olmayan ilişkileri yakalamak için özellikle etkilidir.

K-En yakın komşu *Regresörü*, özellik alanındaki k-en yakın komşuların hedef değerlerinin ortalamasını alarak tahminler yapan bir tür örnek tabanlı öğrenme yöntemidir. Örnekler arasındaki benzerlik genellikle Öklid mesafesi gibi mesafe ölçümleri kullanılarak ölçülmektedir. Algoritma, tahminler tüm veri setinin taranmasını gerektirdiğinden büyük veri setleri için hesaplama açısından yoğun olsa da oldukça esnektir ve doğrusal olmayan desenlere uyum sağlayabilmektedir.

Gradient Boosting *Regresörü*, her yeni modelin bir öncekinin kalıntı hatalarını düzelttiği, modelleri sıralı olarak oluşturan gelişmiş bir topluluk tekniğidir. Belirtilen bir kayıp fonksiyonunu optimize ederek zayıf öğrenenleri aşama aşama birleştirir. Yöntem, bu kaybı en aza indirmek için gradyan inişini kullanmakta ve bu da onu verilerdeki karmaşık ilişkileri yakalamak için oldukça verimli hale getirmektedir. Gradient Boosting, yüksek doğruluğu nedeniyle yaygın olarak kullanılır ancak aşırı uyumu önlemek için dikkatli bir ayarlama gerektirmektedir.

IV. BULGULAR

Hava Kalitesi ve Sağlık Etkisi veri seti'ne ait ilk 5 kayıt Tablo 1'de gösterilmektedir. Özellikler ve çıktı sütunlarının veri seti bölümünde açıklanan ifadelerle ait değerler olduğu tabloda açıkça görülmektedir.

Tablo 1. Veri Setinin Değerleri

RecordID	AQI	PM10	PM2_5	NO2	SO2	O3	Temperature	Humidity	WindSpeed	RespiratoryCases	CardiovascularCases	HospitalAdmissions	HealthImpactScore	HealthImpactClass
1	187.270	295.853039	13.038560	6.639263	66.161150	54.624280	5.150335	84.424344	6.137755	7	5	1	97.244041	0.0
2	475.357	246.254703	9.984497	16.318326	90.499523	169.621728	1.543378	46.851415	4.521422	10	2	0	100.000000	0.0
3	365.996	84.443191	23.111340	96.317811	17.875850	9.006794	1.169483	17.806977	11.157384	13	3	0	100.000000	0.0
4	299.329	21.020609	14.273403	81.234403	48.323616	93.161033	21.925276	99.473373	15.302500	8	8	1	100.000000	0.0
5	78.0093	16.987667	152.111623	121.235461	90.866167	241.795138	9.217517	24.906837	14.534733	9	0	1	95.182643	0.0

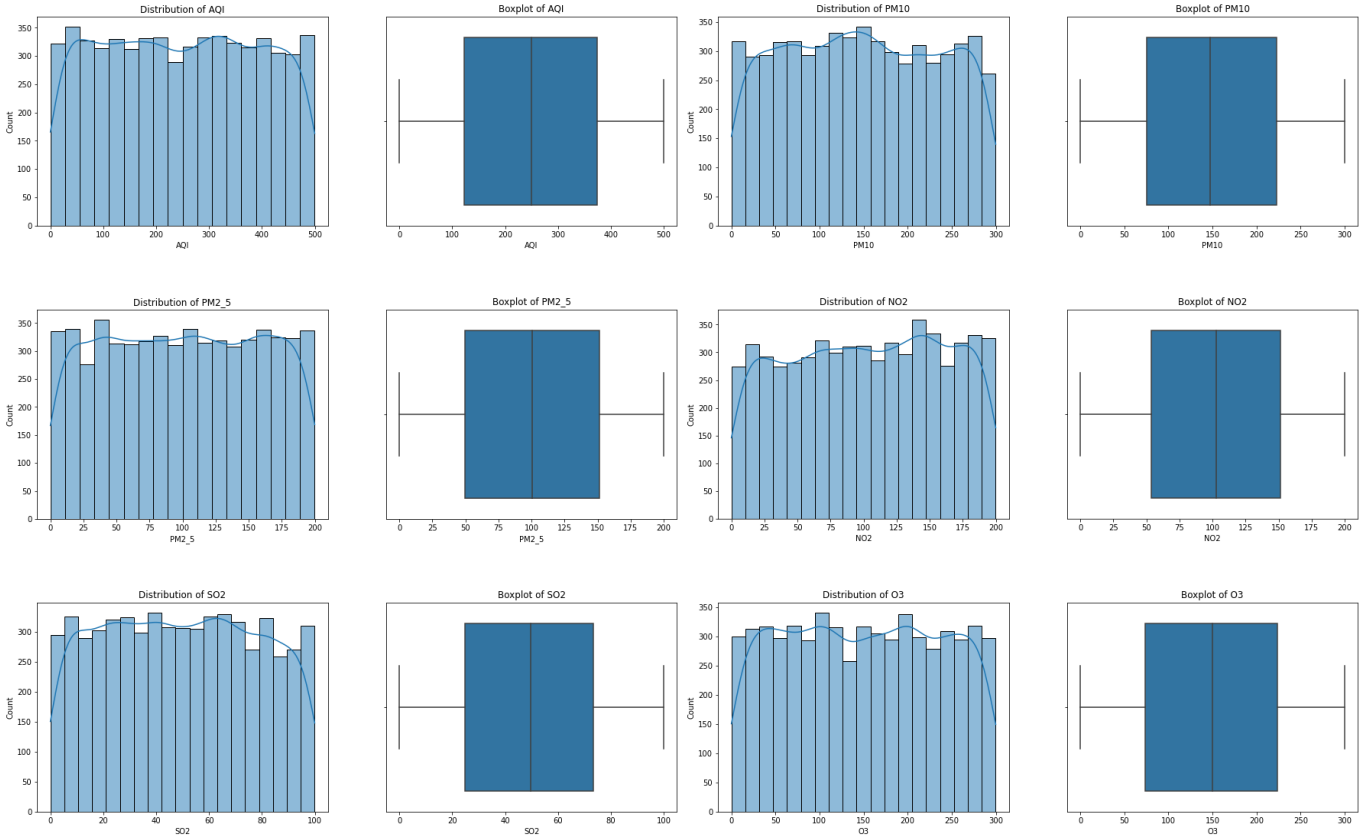
Bir veri kümesinin dağılımının merkezi eğilimini, dağılımını ve durumu özetleyen tanımlayıcı istatistikler üretmek için temel istatistik bilgilerinin alınması gereklidir. Python da veri setimiz üzerinde bu işlemleri describe() metodu ile gerçekleştirmekteyiz.

Hem sayısal hem de nesne serilerini ve ayrıca karışık veri türlerinin DataFrame sütun kümelerini analiz etmektedir. describe() metodu, veri setimiz içerisindeki sayısal değerlerin yüzdeler oranları, ortalama, standart sapma vb. gibi bazı temel istatistiksel bilgilerini almak için kullanılmaktadır. Veri setimize ait temel istatistik değerler Tablo 2’ de gösterilmektedir.

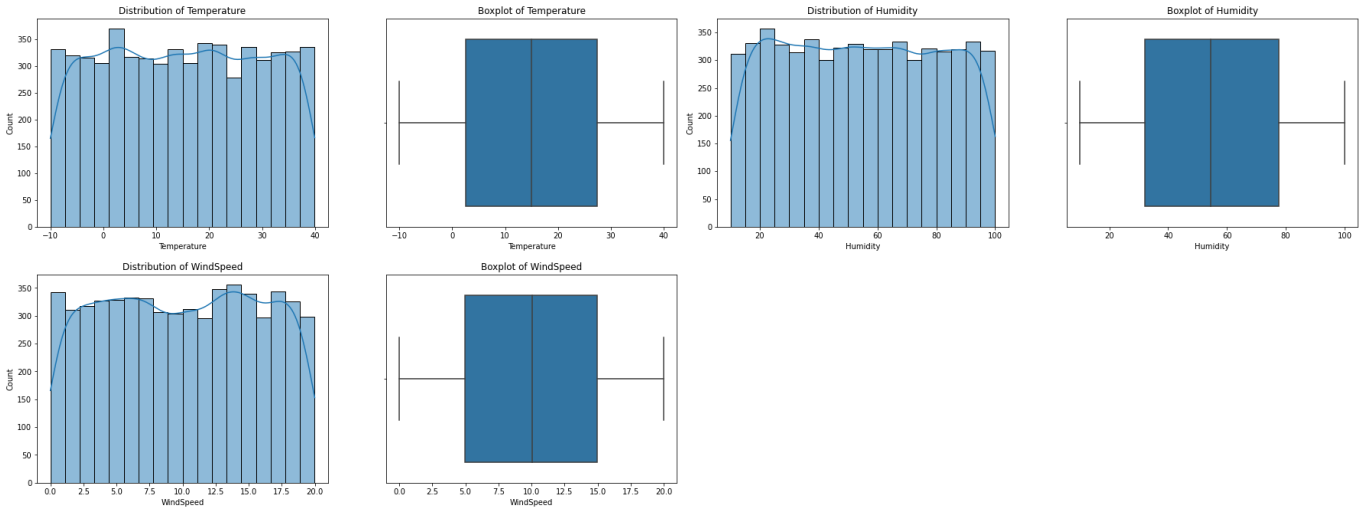
Tablo 2. Veri Setinin Temel İstatistik Değerleri

	RecordID	AQI	PM10	PM2_5	NO2	SO2	O3	Temperature	Humidity	WindSpeed	RespiratoryCases	CardiovascularCases	HospitalAdmissions	HealthImpactScore	HealthImpactClass
count	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000	5811.00000
mean	2906.00000	248.438476	148.654997	100.223714	102.293445	49.456838	149.312431	14.975500	54.776853	9.989177	9.974187	4.988986	2.001033	93.785223	0.281191
std	1677.635538	144.777638	85.698500	58.096612	57.713175	28.530329	86.534240	14.483067	26.020786	5.776950	3.129234	2.216791	1.398794	13.318904	0.714075
min	1.000000	0.005817	0.015848	0.031549	0.009625	0.011023	0.001661	-9.990900	10.001506	0.002094	1.000000	0.000000	0.000000	22.448488	0.000000
25%	1453.500000	122.951293	75.374954	49.435171	53.538538	24.887264	73.999665	2.481500	31.995262	4.952343	8.000000	3.000000	1.000000	98.203057	0.000000
50%	2906.000000	249.127841	147.634997	100.506337	102.987736	49.530165	149.559871	14.942428	54.543904	10.051742	10.000000	5.000000	2.000000	100.000000	0.000000
75%	4358.500000	373.630668	222.436759	151.340260	151.658516	73.346617	223.380126	27.465374	77.641639	14.971840	12.000000	6.000000	3.000000	100.000000	0.000000
max	5811.000000	499.858837	299.901962	199.984965	199.980195	99.969561	299.936812	39.963434	99.997493	19.999139	23.000000	14.000000	12.000000	100.000000	4.000000

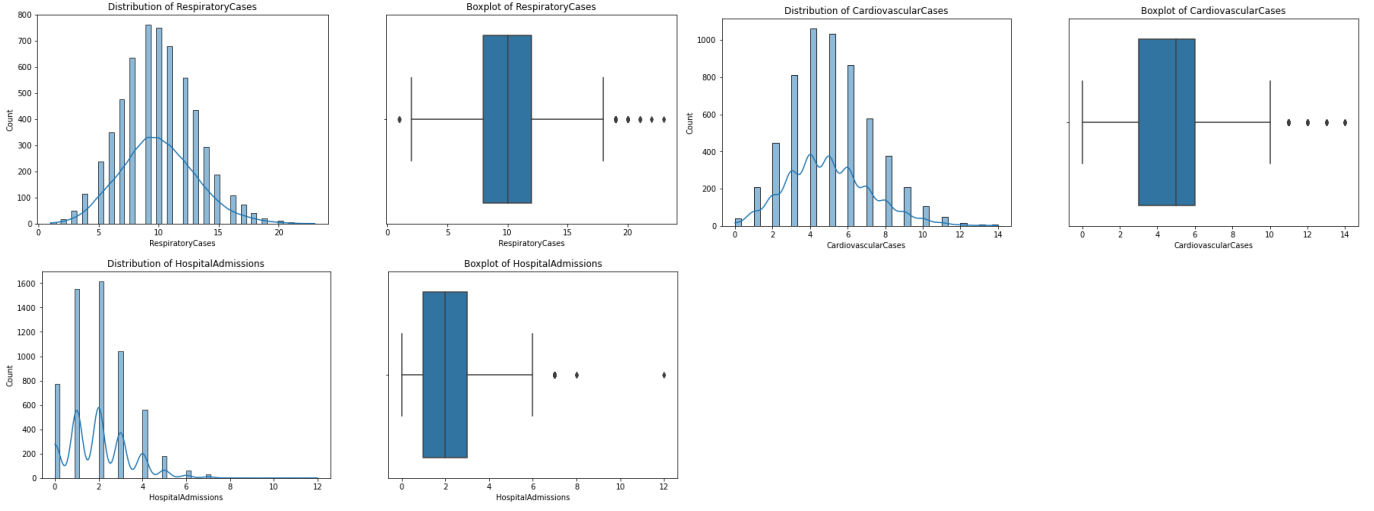
Histogram, verileri aralıklara gruplandırır ve ardından verilerin her aralıkta meydana gelme sıklığını çizmektedir. Kutu grafiği ise, birden fazla veri grubunu karşılaştırmak için kullanılır ve verilerin medyanını, çeyreklik aralığını ve maksimum ve minimum değerlerini göstermektedir. Veri seti’ nin histogram eğrileri ve kutu grafikleri Hava Kalitesi Ölçümleri Şekil 1, Hava Koşulları Şekil 2, Sağlık Etki Ölçümleri Şekil 3 ve Sağlık Etki Sınıfı Şekil 4’ te gösterilmiştir.



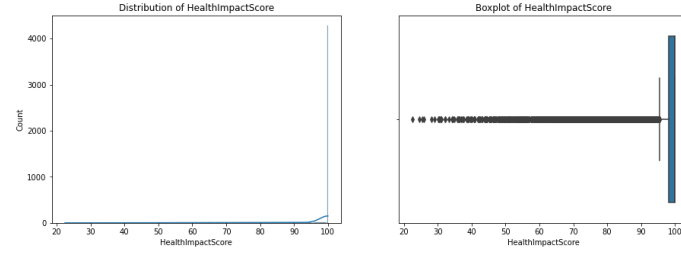
Şekil 1. Hava Kalitesi Ölçümlerinin Histogram ve Kutu Grafikleri



Şekil 2. Hava Koşullarının Histogram ve Kutu Grafikleri



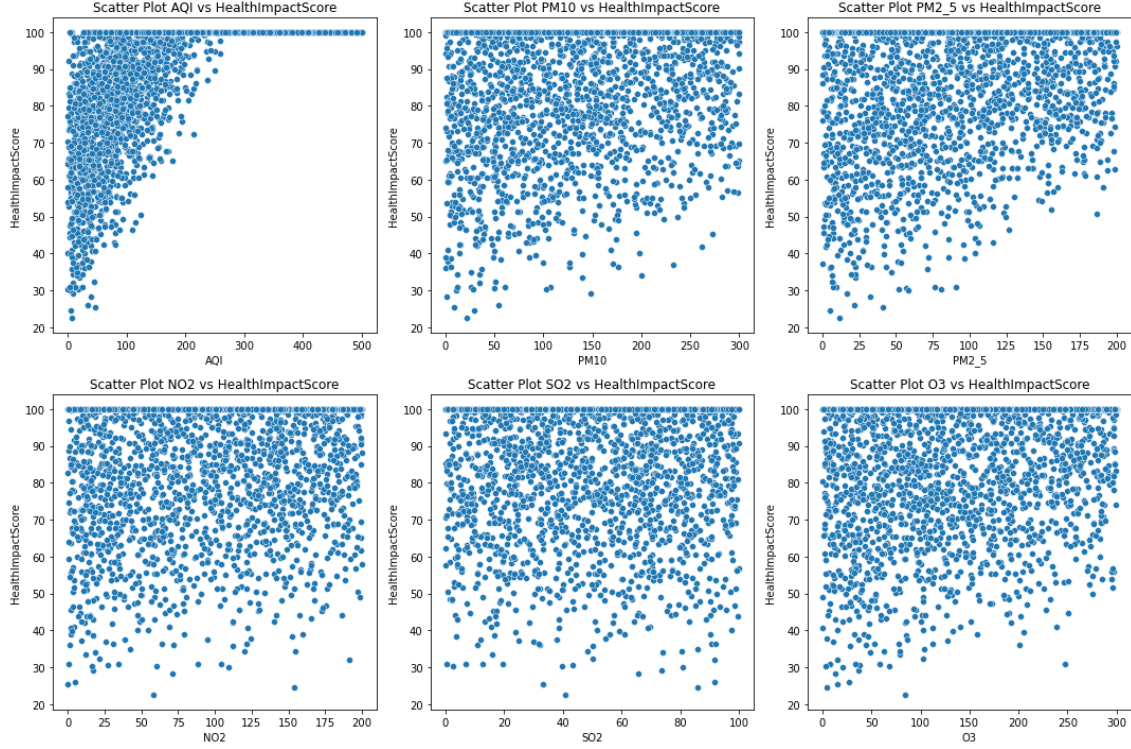
Şekil 3. Sağlık Etki Ölçümlerinin Histogram ve Kutu Grafikleri



Şekil 4. Sağlık Etki Sınıfının Histogram ve Kutu Grafikleri

Dağılım grafikleri çeşitli hava kalitesi göstergeleri (AQI, PM10, PM2.5, NO2, SO2 ve O3) ile bir "Sağlık Etki Puanı" arasındaki ilişkiyi göstermektedir (Şekil 5). Tüm kirleticilerde, düşük kirletici konsantrasyonlarında yüksek Sağlık Etki Puanlarının belirgin bir kümelenmesi vardır ve bu, kirliliğin minimum olduğu durumlarda daha iyi sağlık sonuçları olduğunu göstermektedir. Kirletici seviyeleri arttıkça, Sağlık Etki Puanları daha fazla değişkenlik gösterir ve azalma eğilimindedir, bu da kötüleşen sağlık etkilerini gösterir. Ancak, eğilimler kesinlikle doğrusal değildir. AQI ve partikül madde (PM10 ve PM2.5), gazlara (NO2, SO2, O3) kıyasla daha net azalan Sağlık Etki Puanı kalıpları göstermektedir ve bu da partiküllerden kaynaklanan daha güçlü acil sağlık etkilerini göstermektedir.

Bu sonuçlar, bu ilişkileri doğru bir şekilde ölçmek ve modellemek için daha fazla regresyon analizine ihtiyaç olduğunu vurgulamaktadır.

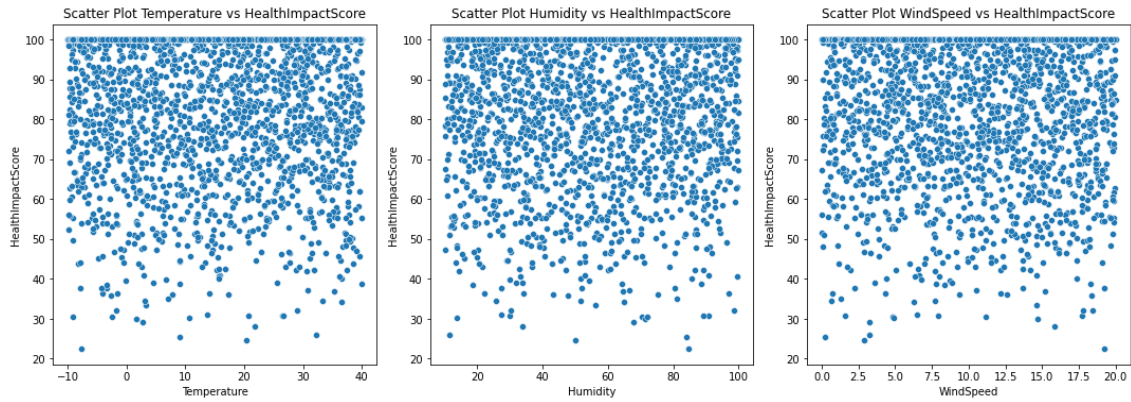


Şekil 5. Hava Kalitesi Ölçümlerinin Sağlık Etki Skoruna göre dağılım grafiği

Dağılım grafikleri, çevresel değişkenler (Sıcaklık, Nem ve Rüzgar Hızı) ile Sağlık Etki Puanı arasındaki ilişkiyi incelemektedir (Şekil 6).

Sıcaklık ve Sağlık Etki Puanı arasında güçlü bir eğilim gözlenmez, ancak daha yüksek sıcaklıklar puanlarda biraz daha geniş bir yayılma gösterir ve bu da aşırı sıcaklıkların sağlığı farklı şekilde etkileyebileceğini göstermektedir. Nem, dağılım nem seviyeleri arasında geniş bir şekilde dağılmış durumdadır ve bu da net bir doğrusal ilişki olmadığını gösterir. Ancak puanlardaki değişkenlik, diğer faktörlerle olası etkileşimleri göstermektedir. Rüzgar Hızı ise nemle benzer şekilde, rüzgar hızı Sağlık Etki Puanı ile belirgin bir desen göstermez, ancak orta hızlar puanların biraz daha sıkı kümelenmesini gösterir.

Bu sonuçlar, diğer değişkenlerle etkileşimleri yakalamak için regresyon modellemesi yoluyla daha fazla analiz gerektirebilecek dolaylı veya doğrusal olmayan ilişkileri gösterir.

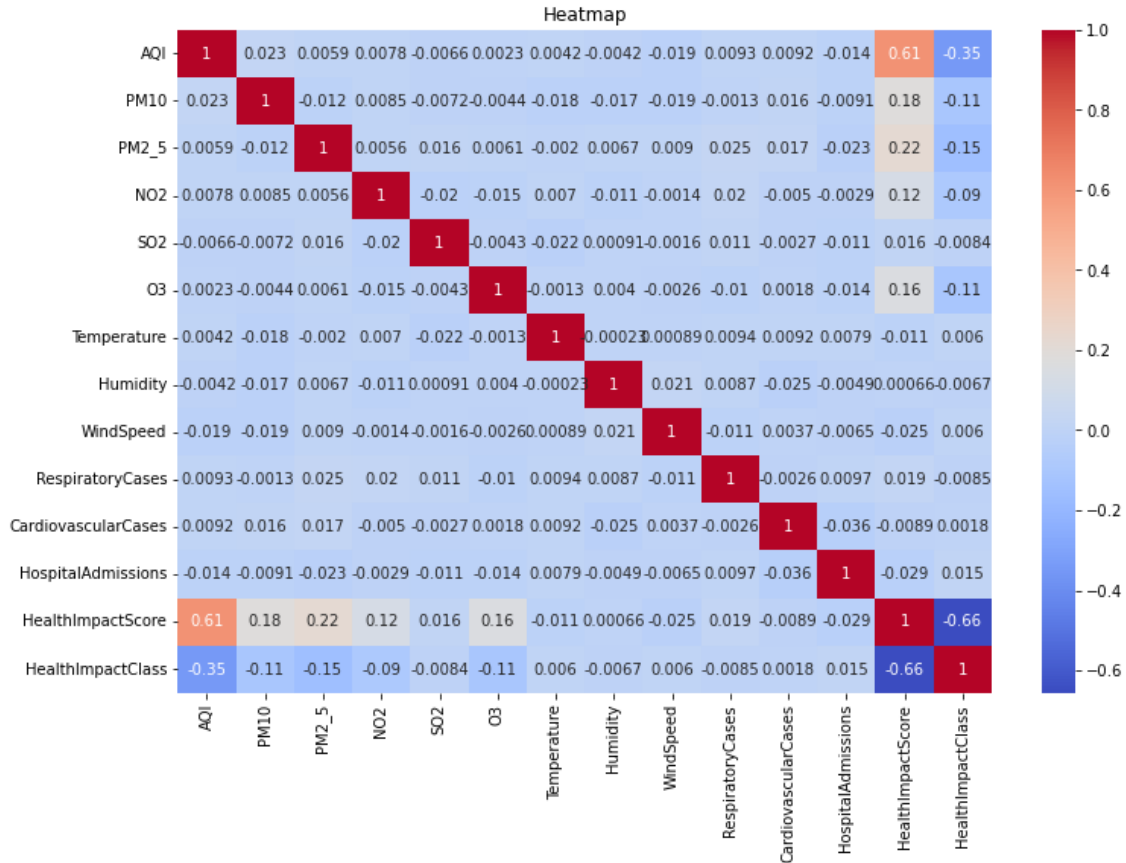


Şekil 6. Hava Koşullarının Sağlık Etki Skoruna göre dağılım grafiği

Karmaşıklık matrisi çeşitli hava kalitesi, çevre ve sağlıkla ilgili değişkenler arasındaki ilişkiyi göstermektedir (Şekil 7). Hava Kalitesi Endeksi, Sağlık Etki Puanı ile güçlü bir pozitif korelasyon (0,61)

göstermektedir ve bu da kötüleşen hava kalitesinin daha kötü sağlık sonuçlarıyla ilişkili olduğunu göstermektedir. Partikül madde (PM10, PM2,5) de Sağlık Etki Puanı ile orta düzeyde pozitif korelasyonlara sahiptir. Sağlık Ölçümleri, Kardiyovasküler vakalar ve solunum vakaları daha zayıf korelasyonlar göstermektedir ve bu muhtemelen birden fazla değişkene olan karmaşık bağımlılıklarını yansıtmaktadır. Çevresel Faktörler, sıcaklık ve nem, Sağlık Etki Puanı ile minimal veya hiç doğrudan korelasyon göstermemektedir ve bu da rollerinin dolaylı veya bağlama bağlı olabileceğini düşündürmektedir.

Genel olarak, veriler Hava Kalitesi Endeksi ve partikül maddeyi sağlık etkilerinin temel itici güçleri olarak vurgularken, sıcaklık ve rüzgar hızının ikincil rolleri olduğu görülmektedir. Korelasyonlar, hava kirliliğini ele almak için hedefli müdahalelere olan ihtiyacı vurgulamaktadır.



Şekil 7. Karmaşıklık matrisi

Tablo 3, tahmin performansına ve hesaplama verimliliğine dayalı olarak dört regresyon modelinin karşılaştırmalı değerlendirmesini sağlar. Temel ölçütler arasında Ayarlanmış R-Kare, R-Kare, Kök Ortalama Karesel Hata (RMSE) ve Alınan Zaman verilmiştir.

Tablo 3. Farklı Regresyon Modellerinden Elede Edilen Parametrik Değerler

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
LinearRegression	0.67	0.67	7.59	0.02
RandomForestRegressor	0.96	0.96	2.79	2.73
KNeighborsRegressor	0.82	0.82	5.58	0.26
GradientBoostingRegressor	0.96	0.96	2.69	1.50

Gradient Boosting ve Random Forest, hesaplama açısından daha yoğun olmalarına rağmen en yüksek doğruluğu (0.96) ve en düşük RMSE'yi (sırasıyla 2,69 ve 2,79) elde etmiştir. Doğrusal Regresyon en

düşük doğruluğa (0.67) sahiptir ancak en hızlı modeldir. KNeighbors, hız ve doğruluk arasında bir denge olarak orta düzeyde performans sağlamaktadır.

Karmaşık, doğrusal olmayan veriler için, yüksek doğruluk ve düşük hata nedeniyle Gradient Boosting ve Random Forest tercih edilirken, Linear Regression hesaplama hızının önceliklendirildiği basit ilişkiler için uygundur. KNeighbors, daha hızlı çalışma zamanı ile orta düzeyde doğruluk için orta yol çözümü sunmaktadır.

V. SONUÇLAR

Bu çalışmada, regresyon analizi kullanarak hava kalitesi ile sağlık arasındaki kritik bağlantıyı incelenmiştir. Havadaki kirleticilerin çeşitli sağlık sonuçları üzerindeki önemli etkisi analizlerle ortaya konmuştur. PM2.5, PM10, NO₂, SO₂ ve O₃ gibi kirleticiler, solunum ve kardiyovasküler hastalık risklerinin artmasıyla tutarlı bir şekilde ilişkilidir. Özellikle AQI'nın belirlenmesinde etkili olan havadaki kirleticiler sağlık üzerinde önemli bir rol oynamaktadır. doğrusal regresyon, rasgele orman regresyon, k-en yakın komşu regresyon ve gradient boosting regresyonu olmak üzere gelişmiş regresyon teknikleri, bu karmaşık ilişkilerin sağlam bir şekilde modellenmesine olanak tanımıştır. Farklı regresyon çeşitleri hem doğrusal hem de doğrusal olmayan kalıpları barındırır ve tahmin doğruluğunu artırmıştır. Sonuçlarda bu durumu göstermektedir. Rasgele orman Regresyon ve Gradient Boosting Regresyonun en iyi sonucu verdikleri, gradient boosting Regresyonun ise rasgele orman Regresyona göre daha hızlı sonuç ürettiği gözlenmiştir.

Gelecekteki araştırmalar, tahmini modelleri daha da geliştirmek için iklim değişikliği dinamikleri de dahil olmak üzere daha geniş değişkenleri entegre etmeye odaklanmalıdır. Küresel sıcaklıklar arttıkça, hava kirliliği ve sağlık arasındaki etkileşimin yoğunlaşması beklenmekte ve bu da koordineli, veri odaklı çözümlerin aciliyetini vurgulamaktadır. Bu çalışma, hava kalitesini analiz etmek ve sağlık ilişkisini ortaya koymak için makine öğrenimi regresyonları kullanmış ve hava ölçüm ve hava koşullarının sağlık üzerindeki etkisini ortaya çıkarmıştır.

KAYNAKLAR

- [1] Pekdoğan, T. (2024). Yapay Zekâ Tabanlı Hava Kalitesi İyileştirme Stratejilerinin Değerlendirilmesi. *EMO Bilimsel Dergi*, 14(2), 87-97.
- [2] Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: a review. *Frontiers in public health*, 8, 14.
- [3] Bourdrel, T., Bind, M. A., Béjot, Y., Morel, O., & Argacha, J. F. (2017). Cardiovascular effects of air pollution. *Archives of cardiovascular diseases*, 110(11), 634-642.
- [4] Zoran, M. A., Savastru, R. S., Savastru, D. M., & Tautan, M. N. (2020). Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy. *Science of the total environment*, 738, 139825.
- [5] Lu, X., Zhang, S., Xing, J., Wang, Y., Chen, W., Ding, D., ... & Hao, J. (2020). Progress of air pollution control in China and its challenges and opportunities in the ecological civilization era. *Engineering*, 6(12), 1423-1431.
- [6] Yu, R., Yang, Y., Yang, L., Han, G., & Move, O. A. (2016). RAQ—A random forest approach for predicting air quality in urban sensing systems. *Sensors*, 16(1), 86.
- [7] Refenes, A. N., Zapranis, A., & Francis, G. (1994). Stock performance modeling using neural networks: a comparative study with regression models. *Neural networks*, 7(2), 375-388.
- [8] Gao, M., Yin, L., & Ning, J. (2018). Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. *Atmospheric Environment*, 184, 129-139.
- [9] Horn, S. A., & Dasgupta, P. K. (2024). The Air Quality Index (AQI) in historical and analytical perspective a tutorial review. *Talanta*, 267, 125260.
- [10] Hamanaka, R. B., & Mutlu, G. M. (2018). Particulate matter air pollution: effects on the cardiovascular system. *Frontiers in endocrinology*, 9, 680.
- [11] Jin, T., Di, Q., Réquia, W. J., Yazdi, M. D., Castro, E., Ma, T., ... & Schwartz, J. (2022). Associations between long-term air pollution exposure and the incidence of cardiovascular diseases among American older adults. *Environment international*, 170, 107594.
- [12] Banach, M., Długosz, R., Talaśka, T., & Pedrycz, W. (2022). Air pollution monitoring system with prediction abilities based on smart autonomous sensors equipped with ANNs with novel training scheme. *Remote Sensing*, 14(2), 413

- [13] Tsai, Y. T., Zeng, Y. R., & Chang, Y. S. (2018, August). Air pollution forecasting using RNN with LSTM. In 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress
- [14] Gupta, V., Kapadia, S., & Bhadane, C. (2023, February). Time Series Analysis and Forecasting of Air Quality in India. In 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-5). IEEE.
- [15] Mao, Y., & Lee, S. (2019, August). Deep convolutional neural network for air quality prediction. In Journal of Physics: Conference Series (Vol. 1302, No. 3, p. 032046). IOP Publishing.
- [16] Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: a survey. *Artificial Intelligence Review*, 56(9), 10031-10066.
- [17] Atahanlı, E. B., & Mutlu, N. M. (2024). Veri Madenciliği ile Hava Kalitesi Tahmini: İstanbul Örneği. *Bilişim Teknolojileri Dergisi*, 17(3), 139-158.
- [18] QI, Zhongang; WANG, Tianchun Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. (2018). *IEEE Transactions on Knowledge and Data Engineering*. 30, (12), pp.2258-2297
- [19] Rahman, P. A., Panchenko, A. A., & Safarov, A. M. (2017, October). Using neural networks for prediction of air pollution index in industrial city. In *IOP Conference Series: Earth and Environmental Science* (Vol. 87, No. 4, p. 042016). IOP Publishing.
- [20] Lu, X., Zhang, S., Xing, J., Wang, Y., Chen, W., Ding, D., ... & Hao, J. (2020). Progress of air pollution control in China and its challenges and opportunities in the ecological civilization era. *Engineering*, 6(12), 1423-1431.
- [21] Gültepe, Y. (2019). Makine öğrenmesi algoritmaları ile hava kirliliği tahmini üzerine karşılaştırmalı bir değerlendirme. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 8-15.
- [22] Öklü, M., & Canbay, P. (2023). Makine Öğrenmesi Yöntemleri ile Şehirlerin Hava Kalitesi Tahmini. *International Journal of Advances in Engineering and Pure Sciences*, 35(1), 39-53.
- [23] Kök, İ., Şimşek, M. U., & Özdemir, S. (2017, December). A deep learning model for air quality prediction in smart cities. In *2017 IEEE international conference on big data (big data)* (pp. 1983-1990). IEEE.
- [24] Ünalı, S., & Yalçın, N. (2022). Hava kirliliğinin makine öğrenmesi tabanlı tahmini: Başakşehir örneği. *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 4(1), 35-44.
- [25] Kaggle, 2024. <https://www.kaggle.com/datasets/rabieelkharoua/air-quality-and-health-impact-dataset>