



Veri Madenciliği ile Hava Kalitesi Tahmini: İstanbul Örneği

Araştırma Makalesi/Research Article

 Nesibe MANAV MUTLU^{1*},  Eyüp Burak ATAHANLI²

¹İstanbul Nişantaşı Üniversitesi (Yönetim Bilişim Sistemleri (İng) Bölümü, İstanbul Nişantaşı Üniversitesi, İstanbul, Türkiye)

²İstanbul Nişantaşı Üniversitesi (Yönetim Bilişim Sistemleri Bölümü, İstanbul Nişantaşı Üniversitesi, İstanbul, Türkiye)

nesibe.manavm@gmail.com, eyupatahanli@gmail.com

(Geliş/Received:28.01.2024; Kabul/Accepted:12.05.2024)

DOI: 10.17671/gazibtd.1426942

Özet— Sanayi devrimi ile başlayan ve günümüzde tehlikeli boyutlara ulaşan hava kirliliği sorununun boyutunu ortaya çıkarmak, olası nedenleri arasındaki ilişkiyi belirlemek ve çözüm önerileri sunmak amacıyla İstanbul'un farklı ilçelerinden alınan çeşitli hava kirliliği verileri kullanılarak istatistiksel testler gerçekleştirilmiştir. Ayrıca önümüzdeki yıllarda oluşabilecek hava kirliliğinin büyüklüğünü tahmin etmek için makine öğrenmesi teknikleri de kullanılmıştır. Alınan verilere dayanarak 2023-2025 yılları için ilçeler bazında Prophet modeli kullanarak tahminler yapılmıştır. Bu tahminlere dayanarak önerilerde bulunulmuştur.(Özet)

Anahtar Kelimeler— hava kirliliği tahmini, veri madenciliği, hava kalitesi analizi

Air Quality Prediction with Data Mining: Istanbul Example

Abstract— Statistical tests were carried out using various air pollution data from different districts of Istanbul in order to reveal the extent of the air pollution problem, which started with the industrial revolution and has reached dangerous levels today, to determine the relationship between possible causes and to offer solution suggestions. In addition, machine learning techniques have been used to predict the magnitude of air pollution that may occur in the coming years. Based on the data received, predictions were made using the Prophet model on a district basis for the years 2023-2025. Recommendations have been made based on these predictions.(Abstract)

Keywords— air pollution prediction, data mining, air quality analysis.

1. GİRİŞ (INTRODUCTION)

Hava kirliliği, sadece hava kalitesini değil, su ve toprağı da etkilediği için tüm canlıların hayatını doğrudan etkileyen en önemli problemlerden biridir. Sadece iklim değişikliğine değil, ekonomiden göçe kadar pek çok sonuca da yol açar [1]. Buna karşı yıllardır alınmaya çalışılan önlemlerin yanı sıra insanların sahip olduğu farkındalık düzeyi de bu önlemlerin etkili olmasındaki önemli bir etkidir.

Hava kirliliği ile ilgili farkındalık düzeylerinin belirlenmesi amacıyla son yıllarda yerel ve uluslararası pek çok çalışma yapılmıştır [2] [3] [4][5] [6][7]. Genel anlamda temiz enerji kullanımının özendirilmesi, çeşitli tasarruf planları, doğru şehir yapılanmaları ve ağaçlandırma çalışmaları gibi global çözüm önerileri problemin çözümüne büyük katkı sağlasa da daha kapsamlı ve özel öneriler geliştirmek gereklidir [8] [9] [10] [11]. Zira hava kirliliğini etkileyen etmenler bir evin farklı odalarında bile çok değişiklik gösterebilir. Hal böyleyken gerek coğrafi koşullar gerekse sosyal yerleşim farklılıkları birbirinden çok ayrışan bölgeleri aynı havuz içinde değerlendirmek problemi uzun vadeli ve geniş bir ölçekte ele almamızı sağlasa da yerel yönetim birimleri probleme karşı yaklaşımlarını daha özel alanda yapmalı, birey olarak bizler de yaşadığımız çevrenin potansiyel risklerini daha dikkatli incelemeliyiz. Bu kapsamda büyük şehirleri mümkün olduğunca küçük birimler halinde incelemeli, hava kirliliğini arttıran etmenleri yerleşim yerine özel olarak tespit edebilmeli ve buna uygun önlemler alınabilmelidir. Çünkü hava kirliliği dünyadaki en başta gelen ölüm nedenlerinden olan kalp hastalıklarının tetikleyen en önemli etmendur [12] [13] [14]. Dünya Sağlık Örgütü 2021 verilerine göre, dünya üzerindeki her 10 kişiden 9'u hava kirliliğinden etkilenmektedir, bunun yanında her yıl hava kirliliğine bağlı olarak 4,2 milyon kişi de hayatını kaybetmektedir. Hava kirliliğinin neden olduğu hastalıklar ve ölüm oranları ülkelerin sağlık harcamalarını yani ekonomilerini da doğrudan etkilemektedir [15] [16] [17]. Sadece Türkiye'de hava kirliliği sonucu asgari 25,845 milyon TL, azami 52,492 milyon TL sağlık harcaması olduğu tespit edilmiştir [18]. UNEP(United Nations Environment Programme)'in dikkat çektiği üzere Dünya Bankası verilerine göre, hava kirliliği dünya ekonomisinde yıllık ortalama 5 trilyon dolara mal olmakta ve 225 milyar dolarlık da gelir kaybına neden olmaktadır. Bu nedenle hava kirliliği riskinin tahmin edilmesi halk sağlığı için önlemlerin önceden alınabilmesi devletlerin bütçesinde iyileştirme sağlayacağı gibi insanların da karbon ayak izi konusunda daha bilinçli hale gelmesine yardımcı olacaktır [19] [20]. UNEP bu konuda farkındalık yaratmada teknolojiyi kullanarak World Environment Situation Room (WESR) isimli bir dijital platformu 2022'de hayata geçirmiştir.

1.1. Alan Yazın (Related Works)

Literatürde hava kirliliğini makine öğrenmesi yöntemlerine dayanarak inceleyen pek çok çalışma

bulunmaktadır. Bunlardan birkaçına aşağıda yer verilmiştir.

Öklü ve Canbay tarafından yapılan çalışmada [21], City Data web sitesi üzerinden toplanan verilere hava kalitesi belirleyici değerlere göre 8 farklı sınıflandırma algoritması uygulanmıştır. Veri setleri 5- katlı çapraz doğrulama ile işleme alınarak başarı sonuçları 3 sınıflı ve 5 sınıflı veri seti için ayrı olarak sunulmuştur. Hava kalitesini sınıflandırmak için 3 ve 5 sınıflı bir sınıflandırma sistemi kullanılmıştır. Hava Kalitesi İndeksinin (AQI) hesaplanan sağlık etkilerine göre sınıfları belirlenmiştir. Makine öğrenimi teknikleri hava kalitesini 3 sınıflı modellerde %87, 5 sınıflı modellerde ise %82 doğrulukla tahmin etmiştir.

Çetin Yağmur'un çalışması [22], trafik yoğunluğu, meteorolojik koşullar ve NOX, SO₂, PM₁₀ hava kirlitcilerinin, çapı 2,5 mikrondan küçük olan PM_{2.5} parçacıklarını nasıl etkilediği incelemiştir. İstanbul Büyükşehir Belediyesi'nin veri paylaştığı açık veri portalı, Beşiktaş hava kalitesi izleme istasyonunun verilerinden yararlanılan çalışmada, Çoklu Doğrusal Regresyon (ÇDR), Rassal Orman (RO), Destek Vektör Makineleri (DVM) ve Yapay Sinir Ağları (YSA) modelleri atmosferik parçacık maddesini tahmin etmek için kullanılmıştır. Karşılaştırıldığında, algoritmaların performans kriterleri modele göre farklılık göstermekle birlikte ancak RO en iyi performans ortalamasına, ÇDR ise en kötü performans ortalamasına sahip bulunmuştur. Sonuçta, R² değeri 0,76 olan iyi bir tahmin modeli geliştirilmiştir.

Oğuz ve Pekin [23], PM₁₀ kirliliğini tahmin etmek için makine öğrenme teknikleri ve meteorolojik faktörler kullandıkları çalışmada, 39.9727 enlem, 32.8637 boylam ve 891 metre yükseklikte yer alan Meteoroloji Genel Müdürlüğü Ankara Bölge İstasyonundan elde edilen meteorolojik verilerden yararlanılmışlardır. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, Ankara Keçiören-Sanatoryum hava kalitesi istasyonunda PM₁₀ kirliliğine ilişkin veri toplamışlardır. Makine öğreniminin çalışma aşamasında, karar ağacı regresyonu, destek vektör regresyonunun, losso regresyonunun ve yapay sinir ağı uygulanmasını ile çalışmışlardır. Bu teknikler ile sıcaklık, çiğlenme noktası sıcaklığı, yağış, bağıl nem, rüzgar hızı, basınç, bulut örtüsü gibi çeşitli faktörleri ve önceki güne ait PM₁₀ ölçümlerini analiz edilmiştir. Sonuç olarak, yapay sinir ağı yöntemi önceki algoritmalarından daha iyi performans göstererek 0,6'lık bir belirleme katsayısı, 18'lik bir ortalama karekök hata ve 12'lik bir ortalama mutlak hata vermiştir.

Çelik ve Arıcı, Kovid-19 salgını öncesinde ve sırasında elde edilen kirlitici konsantrasyonlarının çoklu ölçümlerini hesaba katarak hava kalitesini tahmin etmek için makine öğrenimi tekniklerini kullandıkları çalışmada [24], gelişmiş sanayileri ve yüksek hava kirliliği seviyeleriyle tanınan Zonguldak ilindeki kirlitici konsantrasyonlarını içeren bir veri seti kullanmışlardır. Veriler Çevre ve Şehircilik Bakanlığı Hava İzleme İstasyonları internet sitesinden elde edilmiştir. Çalışma, tümü yüksek düzeyde

tahmin doğruluğu gösteren beş farklı makine öğrenimi tekniğini içermekle birlikte bu yöntemler arasında Karar Ağacı algoritması en olumlu sonucu vermiştir. Hastalığın yayılmasını azaltmak için uygulanan karantina döneminde kirlilik seviyelerinde gözle görülür bir azalma ve hava kalitesinde gözle görülür bir iyileşme yaşandığı tespit edilmiştir. Deneysel sonuçlar, önerilen modelin hava kalitesini etkili bir şekilde tanımlayabildiğini ve gelecekteki kirlilik miktarlarını tahmin edebildiğini göstermiştir.

Akgün ve Barlık çalışmalarında [25], hava kalitesi indeksi (AQI) ile ortalama sıcaklık ve yağış dahil olmak üzere meteorolojik veriler arasındaki ilişkiyi incelemişlerdir. Bağımsız değişkenler olarak bulutluluk, bağıl nem, basınç ve rüzgar hızı ile çoklu doğrusal regresyon analizi ve makine öğrenme tekniklerini kullanarak AQI'yi tahmin etmeyi amaçlamışlardır. Çoklu doğrusal regresyon analizi kullanılarak tahmin etme doğruluğu %41,8 olarak belirlenmiş, Naive Bayes algoritmasının doğruluğu %68,91 ve Karar Ağacı algoritması %70,29'luk bir doğruluğa ulaşırken, Derin öğrenme algoritması %73,05, k-NN algoritması %73,35 ve son olarak Rastgele Orman algoritması %74,89'luk bir AQI tahmin doğruluğuna ulaşmıştır.

Gültepe [26], meteorolojik verileri kullanarak Kastamonu'daki hava kirliliğini simüle ettiği çalışmada, hava kirliliğini tahmin etmek için meteoroloji ve çevre uygulamalarında kullanılan makine öğrenimi teknolojilerinden yararlanmıştır. Sistem performansına ilişkin normalleştirme yaklaşımlarını değerlendirmek için veri kümesi değişkenleri Minimum-Maksimum (Min-Maks) normalleştirme kullanılarak normalleştirilmiştir. Çalışma, performans değerlerini önceki çalışmalarla karşılaştırmış ve zorluk için en iyi sınıflandırma algoritmasını bulmuştur. Yapay Sinir Ağları (YSA), Rastgele Orman, K-En Yakın Komşuluk, Lojistik Regresyon, Karar Ağacı, Doğrusal Regresyon ve Naive Bayes ayrı ayrı oluşturulmuş ve incelenmiştir. Verilerin analizinde Naive Bayes kullanılmıştır. Eğitim ve test verileri veri setinin %70 ve %30'unu oluşturur. Tahmin edildiği gibi yapay sinir ağları %91 başarı oranıyla en doğru tahmin yaklaşımı olmuştur. Doğrusal Regresyon %30 ile en düşük sınıflandırma başarı oranına sahipken, Rastgele Orman ve Karar Ağacı %99 başarı oranına sahip olmuştur.

Ayrıca Wang, Pereira ve Hung [27], hava kirliliği için kontrol süreçlerinin altında yatan ilkeleri ve uygulamaları eleştirel bir şekilde araştırıp bunları uygulamacı mühendisler için bir dizi ayrıntılı tasarım örneğiyle açıklayan ve pratik tasarım hesaplama yöntemlerini çok sayıda sayısal hesaplama ile gösterdikleri bir kitap çalışması yapmışlardır.

Tsai, Zeng ve Chang'ın çalışmalarında [28], PM2.5 konsantrasyonunu tahmin etmek için Uzun-Kısa Süreli Belleğe sahip bir RNN (Recurrent Neural Network-Tekrarlı Sinir Ağları) kullanılmıştır. Python tabanlı Keras, ve Tensorflow'dan yararlanılmıştır. Ağ eğitimi için

2012'den 2016'ya kadar olan Tayvan Çevre Koruma İdaresi verileri kullanılmıştır. Tahmin edilen test verileri 2017 yılına ait verilerden seçilmiştir. PM2.5 konsantrasyonunun tahmin gücü dört saat boyunca Tayvan'daki 66 istasyonda test edilmiştir. Sonuçlar önerilen tekniğin PM2.5'i tahmin edebildiğini göstermiştir.

Rybarczyk ve Zalakeviciute çalışmalarında [29], meteorolojik faktörlerin bir kombinasyonuna dayalı olarak ince partikül madde tahminine yönelik yöntem belirlemeyi amaçlamışlardır. Ekvador'un Quito şehrinin uzun yıllara dayanan meteorolojik verileri, makine öğrenimi metodolojisini kullanarak modeller oluşturmak için kullanılmıştır. Makalede, ince parçacıkların konsantrasyonlarını etkili bir şekilde iki kategoriye ayıran bir karar ağacı tekniği sunulmuştur.

Bu çalışmada, yukarıdaki örnekler doğrultusunda İstanbul'un halka açık verileri [30] kullanılarak istatistiksel yöntemler ve makine öğrenmesi algoritmaları yardımıyla, İstanbul'un hava kalitesinin ölçüldüğü ilçelerinde geçmişten bugüne analizler yapılmış ve önümüzdeki yıllarda kirliliğin varabileceği boyutlar makine öğrenmesi modeliyle tahmin edilmiştir. Bu sayede tüm çevreyi etkileyen bu kirlilik krizinin araştırılması sayesinde potansiyel salgın hastalıklar, tarımsal salgınlar ve belki hayvanlardan insanlara bulaşabilen salgınlar bile önlenilebilecektir [31] [32] [33] [34]. Bulgular sayesinde, sürdürülebilir çevre politikalarının belirlenmesinde farkındalık geliştirilmesi, insanların bireysel günlük faaliyetlerinin bile hava kirliliği konusundaki etkisi hakkında bilinçlenmeleri, daha sağlıklı bir toplum için önleyici hizmetlerin hayati etkisi belirlenebilecektir.

Sürdürülebilirlik için hava kirliliği üzerine yapılacak bu çalışma üniversite düzeyinden başlayarak tüm öğrenim düzeylerindeki gençlerin dikkatini çekecek sürdürülebilir bir yatırıma dönüşecektir. Aynı zamanda İstanbul Belediyesinin ilçeler bazında alacağı önlemlere yahut yeni geliştireceği projelerde rapor niteliğinde olacaktır, kurumun sürdürülebilir yatırımlara öncelik vermesine katkı sağlayacaktır.

1.2. Hava kirliliği metriklerinin tanımlanması (Defining air pollution metrics)

Hava kirliliği verilerinde kullanılan temel değerler ve tanımları aşağıdaki şekilde listelenmiştir:

PM10 ve PM2.5: Partikül madde (PM) kirliliği, hava kirliliğinin ana kaynaklarından biridir. PM10, havadaki partiküllerin çapının 10 mikrometreden küçük olan bir bölümünü ifade ederken, PM2.5, çapı 2.5 mikrometreden küçük olan partiküllerin oranını ifade eder. Birçok çalışma PM10 seviyelerindeki artışın astım başta olmak üzere çeşitli hastalıkların semptomlarını arttırdığını göstermektedir [35] [36].

NO₂: Nitrojen dioksit, fosil yakıtların yakılması ve endüstriyel faaliyetler gibi insan kaynaklı etkenlerden kaynaklanan bir hava kirliliği türüdür [37].

SO₂: Kükürt dioksit, kömür yakma gibi endüstriyel faaliyetlerden kaynaklanan bir hava kirliliği türüdür [37].

CO: Karbon monoksit, fosil yakıtların yakılması gibi kaynaklardan salınan bir hava kirliliği türüdür [37].

Daha detaylı çalışmalarda partikül maddenin detaylı analizi ve havada bulunan kurşun, civa, gibi özel elementler incelenebilir. Bu verilerin yanında rüzgâr, nem, sıcaklık gibi meteorolojik veriler de kullanılabilir. Bu değerler, hava kalitesi indeksleri gibi farklı metriklerde de kullanılabilir ve hava kirliliğinin etkilerini ölçmek için birlikte değerlendirilir.

AQI (Air Quality Index) yani Hava Kalitesi İndeksi [38], hava kirliliğini ölçmek ve raporlamak için kullanılan bir ölçüttür. AQI, çeşitli hava kirleticilerinin konsantrasyonlarını ölçer ve bu kirlilik seviyelerini belli aralıklara göre altı farklı kategoriye ayırır:

- İyi (0-50 AQI)
- Orta (51-100 AQI)
- Hassas Gruplar İçin Kötü (101-150 AQI)
- Kötü (151-200 AQI)
- Çok Kötü (201-300 AQI)
- Tehlikeli (301-500 AQI)

AQI, PM_{2.5}, PM₁₀, NO₂, SO₂, O₃ ve CO gibi hava kirleticilerinin konsantrasyonlarını temel alarak hesaplanır ve hava kirliliği seviyesini belirlemek için kullanılır [21]. AQI'nin kullanımı, halk sağlığı açısından önemli bir araçtır ve hava kirliliği seviyelerinin ölçümünü ve takibini kolaylaştırır.

1.3. Veriye Erişim Yolları (Ways to Access Data)

Bu çalışmada veriye erişim sağlamak için İstanbul Büyükşehir Belediyesi (İBB) 'nin **api** araçları kullanılmıştır. Bu yöntemle İBB' nin veri havuzunda belirtilen yönergeler doğrultusunda hava kirliliği verisi de dahil birçok halka açık veriye erişim sağlanabilir [30]. Bu doğrultuda **Python** programlama dili ve **Json** request kütüphaneleri kullanılarak veriler çekilmiştir. Veriye erişim sürecinde, ölçüm istasyonu bulunan ilçeler ve semtler seçilerek ve istenilen tarih aralığı belirtilerek veriler alınmıştır. Çalışma kapsamında 2010 yılından 2023 yılına kaydedilen 430 MB'lık veri kullanılmıştır. Eğer seçilen konumda ve tarih aralığında ölçüm yapılmamışsa, içerik boş değeri temsil eden *NaN* değeri olarak dönmüştür. Veri, json formatında alınmıştır. Veri json formatında alındıktan sonra veri tipi dönüşümleri yapılmıştır. Bu dönüşüm çalışmanın içeriğindeki matematiksel işlemler için gereklidir. Sonrasında çekilen dosyaların formatı json'dan csv'e dönüştürülmüştür.

2. CRISPDM METODUYLA PROBLEMİ ÇÖZME (SOLVING THE PROBLEM WITH THE CRISP DM METHOD)

Endüstriler genelinde veri madenciliği süreçlerini standartlaştırmak için 1999 yılında yayınlanan bu yöntem, o zamandan beri veri madenciliği, analitik ve veri bilimi projeleri için en yaygın metodoloji olarak kullanılmaktadır. Bu makalenin ilgilendiği sorunun çözülmesinde her aşama modelin ilgili aşamasına göre uyarlanmış olarak aşağıda sırasıyla görülmektedir.

2.1. Problemi Anlamak (Understanding the Problem)

Hava kirliliği sorunu ele alınırken geniş coğrafyalarda ölçüm yaparak değerlendirmek yerine küçük birimler halinde incelenmek istenmiştir. Bu doğrultuda İstanbul' un halka açık verileri kullanılarak geçmiş yıllardaki hava kirliliği verileri analiz edilecek ve gelecek için birtakım tahminler ve önerilerde bulunulacaktır.

2.2. Veriyi Anlamak (Understanding Data)

İlgili kaynaktan edinilen veri gerekli düzenlemeler yapıldıktan sonra bir Pandas veri tipi olan Dataframe yapısına dönüştürülmüştür. Pandas Python programlama dilinde veri manipülasyonu ve analizi için kullanılan güçlü bir kütüphanedir. Dataframe ise veri analizi ve veri işleme için kullanılan bir veri yapısıdır. Genellikle tablo benzeri bir yapıya sahip olan dataframe, sütunlar ve satırlardan oluşur. Türkçeye veri çerçevesi olarak çevrilsede çalışma boyunca yaygın kullanım olan DataFrame olarak nitelendirilecektir. Bütün istasyonlar için ayrıca Dataframe oluşturulduktan sonra ortak yapılacak işlemler için bu veri yapıları bir sözlük yapısı içinde toplanmıştır. Sözlük yapısı ise anahtar-değer çiftlerinden oluşan bir koleksiyonu temsil eder. Burada anahtarlar benzersiz seçilmelidir, ancak anahtarların temsil ettiği değerler herhangi bir Python nesnesi olabilir.

Sonuç olarak 'dfs' sözlük yapısı elde edilmiştir. Bu sözlüğün içinde 'Arnavutköy', 'Beşiktaş', 'Kartal', 'Bağcılar', 'Göztepe', 'Sancaktepe', 'Tuzla', 'Sultangazi1', 'Sultangazi2', 'Sultangazi3', 'Üsküdar1', 'Yenibosna', 'Mobil', 'Sarıyer', 'Ümraniye1', 'Esenler', 'Aksaray', 'Alibeyköy', 'Avcılar', 'Kadıköy', 'Kağıthane1', 'Selimiye', 'Beylikdüzü' isimli anahtar değerler eklenmiştir. Görüldüğü üzere bazı ilçeler birden fazla veri setiyle temsil edilmektedir, veri kaynağı olan İBB o ilçede birden fazla ölçüm istasyonu bulundurduğu için ayrı ayrı değerler numaralandırılarak saklanmayı tercih edilmiştir. Ayrıca bir yere bağlı olmayan, sürekli konumu değişen Mobil ölçüm istasyonu da sözlük yapısının içinde yer almaktadır.

Buna göre ilçe ve semtlerin kısa-orta-uzun zaman dilimlerinde ortalama AQI değişimleri analiz edilmiştir. Uzun zaman dilimi için yıllık ortalama AQI değerleri, orta zaman dilimi için mevsimlere göre ortalama AQI değerleri, kısa zaman için ise günlük değişimler görselleştirilmiştir.

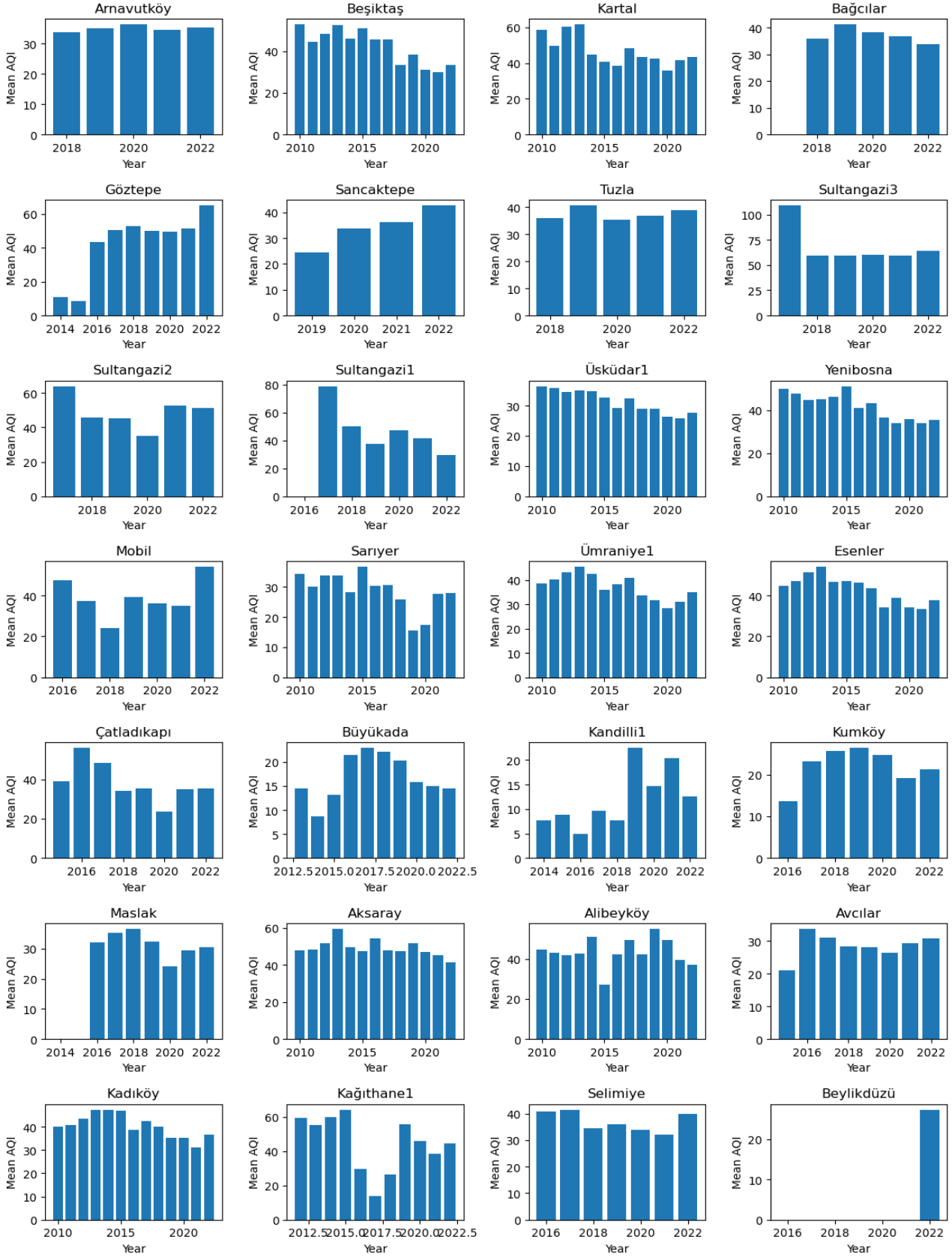
Bu görselleştirmelerden biri aşağıda yer almaktadır. Şekil 1'de yer alan grafiklere göre;

- Yıllara göre ölçüm değeri azalma eğiliminde olan yerler Beşiktaş, Kartal, Bağcılar, Üsküdar, Yenibosna, Ümraniye, Esenevler, Büyükkada, Kadıköy, Sultangazi (1.bölge) ilçe ve semtleridir. Bu yerleşim yerlerinde ölçüm yapılan yıllardan itibaren düzenli düşüş görülmesi de iyileşmeler olduğu ortadadır.
- Ölçüm değerleri artış eğiliminde olanlar ise Göztepe, Sancaktepe, Kandilli şeklindedir. Burada düzenli bir değişim görülmemekle birlikte ölçüme başlanan ilk yıllara göre hava kirliliğinde ölçülen değerlerde bir artış gözlenmektedir. Özellikle Göztepe'de son yıllarda oluşan kirliliği arttıran etmenler başka bilimsel araştırmaların konusu olabilir.
- Esenler, Beşiktaş, Üsküdar, Ümraniye, Kadıköy gibi şehrin eski yerleşim yerlerinde 2010 yılına göre ölçümlerde iyileşme görülmektedir. Bu tarih aralığında söz konusu ilçelerde bu iyileşmenin olası sebepleride araştırılıp ölçüm değerlerinde düşüğe neden olan tüm uygulamalar yaygınlaşmaması faydalı görülmektedir.

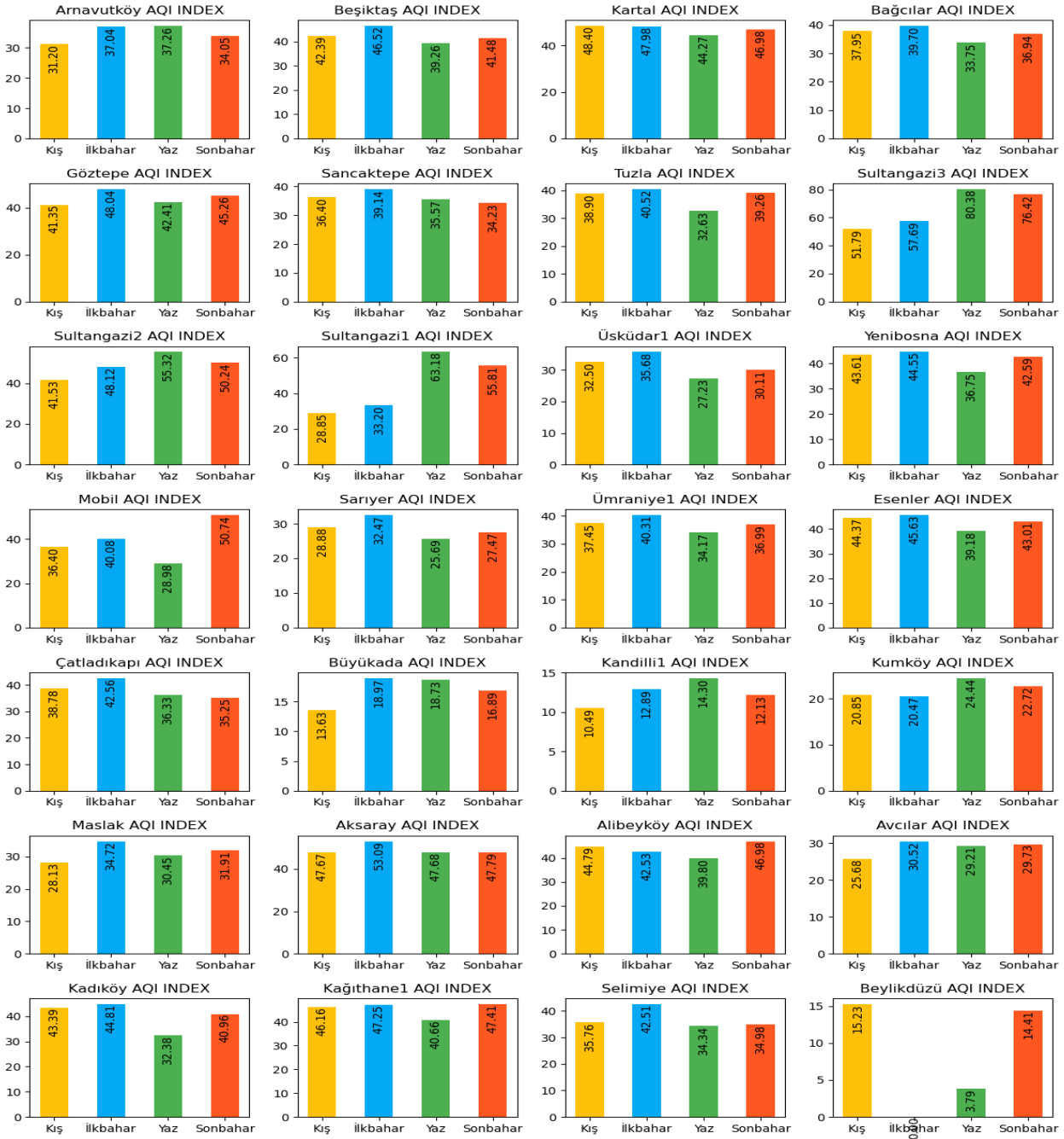
• Ayrıca, grafiklere bakıldığında bir diğer dikkat çeken durum Sarıyer istasyonunda 2019-2020 yıllarında yaşanan kayda değer düşüş olmuştur. İlk bakışta bu dönemin yaşanan Covid-19 pandemisinin etkisi söz konusu olabilir, zira bölge uzaktan çalışmaya da elverişli olan bir çok iş kolunda çalışmalar yapan kurumlara ev sahipliği yapmaktadır. Uzaktan çalışma ve esnek saatler uygulamalarının yaygınlaşmasının ardından trafik ve dolayısıyla kirliliğin azalması bu sonuçlara neden olmuş olabilir.

• Sarıyer istasyonunda olduğu gibi Kağıthane1 istasyonu da 2017'de kaydettiği dip değerler ile dikkat çekmektedir. Buradaki düşüşün %70'e varan oranda olması istasyonun ölçümlerinin incelenmesi gerekliliğine işaret etmektedir. Bu durumun, öncesi ve sonrası olarak başka yıllarda da yaşandığı göz önünde bulundurulmalıdır.

• Sancaktepe istasyonu, 2019'dan itibaren veri kaydetmeye başlamış ve takip eden yıllarda ölçüm değerleri sürekli artış göstermiştir. Bu nedenle burada artışın nedenlerinin araştırılması faydalı olacaktır.



Şekil 1. Geçmiş yıllara göre ilçe-semtlerin AQI değerleri (AQI values of districts and districts according to previous years)



Şekil 2. 2010-2022 yılları arası Mevsim kırılımında AQI ortalamaları (AQI averages by season)

Yukarıdaki Şekil 2’de istasyonların 2010-2022 yılları arasında ölçülen AQI değerlerinin mevsim kırılımında ortalamaları gösterilmiştir. Bazı istasyonlar 2010 yılından sonra veri kaydetmeye başlamıştır. Bu grafikler ilk ölçüm tarihinden 2022 yılına kadar olan sürede değerlendirilmiştir.

Grafiğe göre istasyonların bulunduğu yerleşim yerlerinde ortak bir örgü bulunmamaktadır. Bu durum hava kirliliği probleminin bir çok etmene bağlı [39], küçük mesafelerde bile hızlı bir şekilde değişebilen [40], ayrıca alınan önlemlerin sonuçlarının da hızlı bir şekilde gözlemlenebilen [41] bir problem olduğunu bir kez daha

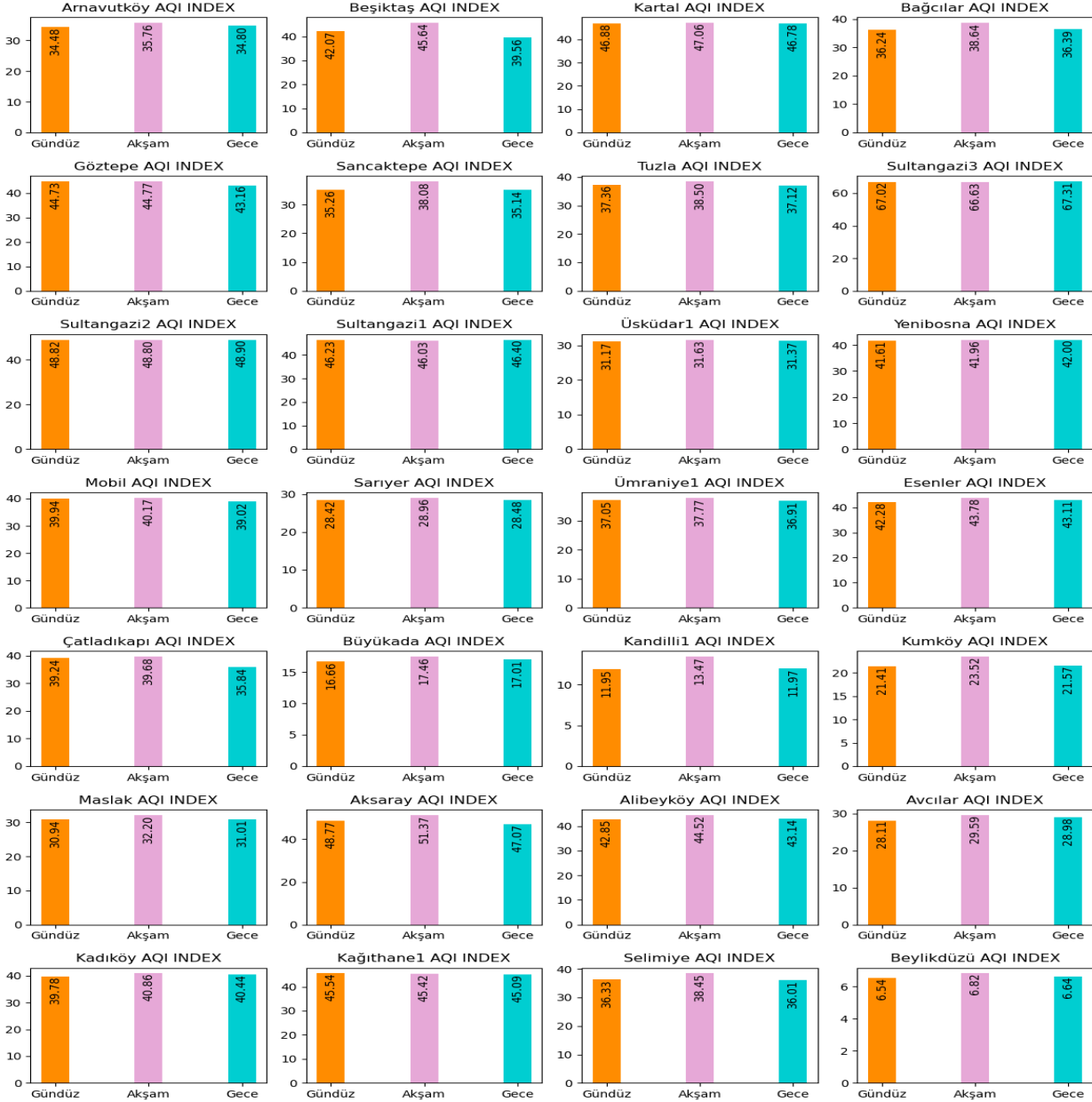
göstermiştir. Bütün bunlar bu çalışmanın önemini daha iyi bir şekilde vurgulamaktadır.

En çok dikkat çeken 3 ölçüm istasyonu da Sultangazi ilçesinde bulunmaktadır. Bu istasyonların ölçümleri mevsimler arası büyük değişim göstermektedir. İstasyonların ölçümleri yaz aylarında %110’a varan bir artış göstererek artış düzeyini gözler önüne sermektedir. Bölgedeki taş ocaklarının bu artışa sebep olup olmadığı, olası diğer etmenler başka bilimsel çalışmaların konusu olarak araştırılabilir.

Öte yandan yaz aylarında kirliliğin azaldığı, kışın %35’ e varan artışların olduğu ölçüm istasyonları da

bulundurmaktadır (Beylikdüzü kısıtlı veri sebebiyle bu analize dahil edilmemiştir). Bunların ortak özelliği diğer yerleşim yerlerine göre daha eski bir tarihi olan, nüfus yoğunluğu fazla ve konutlarında kışın ısınma enerji harcamasını düşürecek yeni nesil altyapısı bulunmayan bölgeler olduğu dikkat çekmektedir.

Yaz aylarında normalde 20 binden az olan toplam Adalar nüfusunun 50-70 binlere çıkması nedeniyle [42] Büyükada'da değerlerdeki %50'ye yakın artışın turizm kaynaklı olduğu tahmin edilmektedir. Bu artışa rağmen adada taşıt trafiği olmaması, geniş ormanlık alanları sayesinde ölçüm değerleri iyidir.



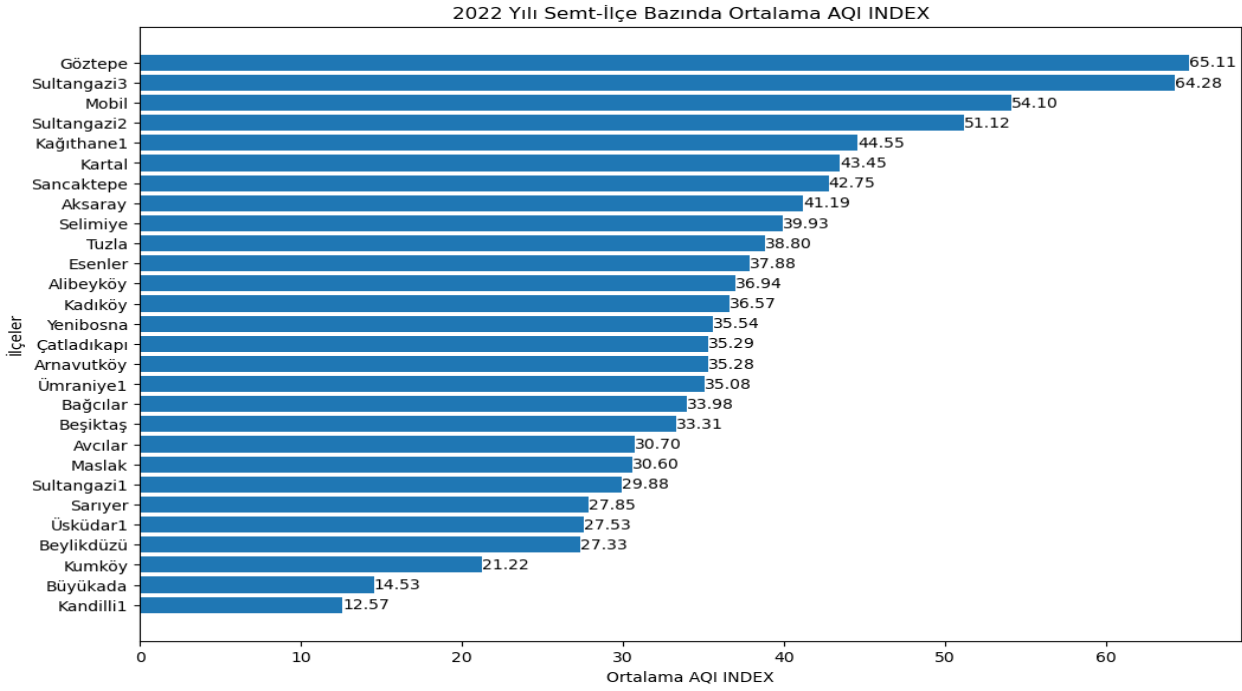
Şekil 3. Gün içi saat dilimlerinde AQI ortalama (AQI average across time zones during the day)

Yukarıdaki Şekil 3, istasyonların gün içinde gündüz (08:00 – 16:00) akşam (16:00 - 00:00) ve gece (00:00 – 08:00) olmak üzere 3 zaman diliminde saatlik AQI değerlerinin ortalamasını göstermektedir. Genel eğilime baktığımızda gün içinde değişimlerin sınırlı olduğu ortaya çıkmaktadır. Bu değişim akşam saatlerinde artan kirlilik olarak kendini göstermektedir. Özellikle ciddi trafik problemi olan Beşiktaş, Maslak, Üsküdar Kandilli bölgesinde artış daha

belirgin olarak görülmektedir [43][44]. Kirliliğin önemli unsurlarından olan partikül madde artışı trafikte dizel araçların kullanımıyla doğrudan ilişkilidir [45]. İstanbul'da dizel araçların oranı Avrupa birliği ülkelerine göre daha yüksek seviyelerdedir[46]. Söz konusu ülkelerde dizel araçların sıkı emisyon ölçüm denetimlerine tabi tutulması, bazı ülkelerde yasaklar uygulanması dizel araçlara gösterilen ilgiyi azaltmıştır. Bu durum tüm dünyada dizel

araç üretiminin azalması ve yeni nesil motor teknolojilerinin gelişmesine sebep olmuştur, öyle ki ülkemizde de trafiğe kaydı yapılan yeni araçlarda dizel araçların oranları

gittikçe düşmekte, alternatif olarak elektrikli araçların kullanımı da yaygınlaşmaktadır [47].



Şekil 4. İlçe-Semtlerin Ortalama AQI değerleri (Average AQI values of districts)

Şekil 4'e göre 2022 yılında en yüksek AQI değerlerini ölçen istasyonlar; Göztepe, Sultangazi (3.bölge), Mobil istasyonudur.

Mobil istasyon Şekil 1'e göre en yüksek değerini 2022 yılında kaydetmiştir. Bu istasyonun ölçüm yaptığı yerler kaydedilmeli ve bunlara özel analizler yapılmalıdır. Veri setinde böyle bir veri olmadığı için böyle bir yorumda bulunmak bu çalışma kapsamında mümkün olmamıştır.

Öte yandan 2022 yılı içinde en düşük AQI değerleri kaydeden istasyonlar; Kumköy, Büyükada ve Kandilli'dir.

Kumköy, İstanbul'un Sarıyer ilçesine bağlı bir sahil semtidir. Genellikle yaz aylarında yerli ve yabancı turistlerin tercih ettiği bir destinasyondur. Kilyos olarak da bilinen Kumköy, Karadeniz kıyısında yer alır ve güzel plajlarıyla ünlüdür. Kumköy daha az sanayileşmiş ve kalabalık olmayan bir bölge olduğu için, yakın çevresinde ormanlık alanlara sahip olduğu için genellikle İstanbul şehir merkezine göre daha temiz bir havaya sahiptir.

Büyükada ise İstanbul'un en popüler adalarından biridir ve Marmara Denizi'nde yer alır. Büyükada, İstanbul'un Anadolu Yakası'na bağlı olan ve İstanbul Boğazı'ndan uzakta bulunan bir adadır. Adanın toplam yüzölçümü yaklaşık 5,4 kilometrekaredir ve nüfusu oldukça sınırlıdır. Büyükada, doğal güzellikleri ve tarihi atmosferiyle ünlüdür. Ayrıca adanın doğal alanları ve ormanlık

bölgeleri, oksijen bakımından zengin bir ortam oluşturur. Büyükada'nın araç trafiğine kapalı olması ve bisiklet veya fayton gibi alternatif ulaşım araçlarının kullanılması, adanın temiz bir atmosfer sunmasını sağlar.

Kandilli ise İstanbul'un Üsküdar ilçesine bağlı bir semttir. İstasyon Boğaziçi Üniversitesi'nin Kandilli yerleşkesinde bulunmaktadır ve ormanlık alanlarla çevrilidir. Yerleşke İstanbul'un diğer bölgelerine göre daha temiz bir hava kalitesine sahiptir.

2.3. Veriyi Hazırlamak (Preparing the Data)

Verinin içinde bulunan ReadTime isimli değişken ölçüm yapılan zamanı temsil etmektedir. Temsili formatı ise; 2022-12-31 23:59:00 şeklindedir.

Bu değişkenin türü veri kaynaklarından metin türü(string) olarak gelmektedir. Ancak çalışmayı kolay yapabilmek ve gerekli istatistiksel çıkarımları görebilmek için bu string yapısını yine Python'da bulunan Datetime (zaman biçimi) türüne çevrilmiştir. Ek olarak her ilçenin hava kirliliği verilerini barından Datetime yapısına özel tarih değişkenleri eklenmiştir. Bunlar; gün, ay, yıl, saat, mevsim, günün bölümleri şeklinde olmuştur.

Bu sayede hem yapılan istatistiksel çıkarımları değerlendirmek daha kolay olmuş, hem de kurulan modele bu bağımsız değişkenleri ekleyerek bağımlı değişkenle arasındaki bağlantıyı öğrenmesi sağlanmıştır.

2.4. Modelleme (Modelling)

Bu bölümde öncelikle zaman serisi problemi modellemede kullanılan bazı modeller tanıtılmış sonrasında ise bu çalışmanın konusu olan modele yer verilmiştir.

Zaman serisi modellemede oldukça fazla alternatif mevcuttur. Bununla ilgili birkaç model aşağıda verilmiştir:

ARIMA: Zaman serilerindeki otokorelasyonu ele alan ve durağanlık olmayan bileşenleri modellemek için kullanılan bir yöntemdir. Geçmiş değerlerin kombinasyonu ile mevcut değerleri tahmin ederek gelecekteki değerleri belirleyebilir.

SARIMAX: Arima modeline benzer şekilde çalışır, ancak mevsimsel bileşenlere sahip zaman serilerini daha etkili bir şekilde modelleyebilir. Ayrıca Arima yöntemine dışsal değişkenlerin eklenmesiyle genişletilmiş bir yöntemdir. Bu model, mevsimsel örüntülerin yanı sıra dışsal faktörlerin etkisini de hesaba katarak tahminleri yapabilir.

LSTM: Derin öğrenme tabanlı bir yapay sinir ağı modelidir ve zaman serilerinde uzun vadeli bağımlılıkları ele almak için etkilidir. Geçmiş gözlemlerin hafızasını kullanarak gelecekteki değerleri tahmin etme yeteneğine sahiptir. Model başarısını birçok regresyon probleminde kanıtlamış olmasına rağmen, derin öğrenme yapısı bunun gibi birden fazla model barındıran projelerde eğitim sürelerinin uzamasına yol açmaktadır. Veri setinin saatlik olduğu düşünülürse bu dezavantaj daha da gün yüzüne çıkacaktır. Çeşitli ölçeklendirme yöntemleriyle kullanıldığında başarılı olma ihtimali yüksek bir modeldir.

Prophet: Meta (eski adıyla Facebook) tarafından geliştirilen açık kaynak kodlu bir zaman serisi tahminleme algoritmasıdır. Bu model, zaman değişkeni (ds) ve bağımlı değişken (y) olarak iki girdi alır. Prophet, bu verileri kullanarak tahmin değeri (yhat) ile birlikte tahminin üst sınırını (yhat_upper) ve alt sınırını (yhat_lower) belirleyen bir aralık tahmini yapar. Bu çalışmada, kullanım kolaylığı, hızlı ve basit yapısı ile Prophet modeli tercih edilmiştir.

Prophet modeli, zaman serilerinin yapısını analiz eder ve trendleri, mevsimsel desenleri ve diğer bileşenleri hesaba katarak gelecekteki değerleri tahmin etmeye çalışır. Tahminlerde, yhat değeri genel tahmin değerini temsil ederken, yhat_upper ve yhat_lower değerleri ise tahminin üst ve alt sınırlarını gösterir. Bu aralık tahmini, tahminin belirsizlik düzeyini ifade eder ve güven aralığı sağlar [48].

Hava kirliliği ölçümü yapan istasyonların tümünde elementlerin tamamı ölçülmediği için bazı veri setlerinde eksiklikler bulunmaktadır. Bu modelin kurulma aşamasında hata alınmasına neden olur. Bundan kaçınmak için tüm veri setinde bulunan AQI verisinin analiz edilmesi daha doğru olmuştur.

Model yapay zeka ve derin öğrenme projeleri için Google tarafından piyasaya sürülen **Google Colab** programlama ortamında Python dili ile gerçekleştirilmiştir. Colab, bulut sunuculara çalışan bilgisayar sistemleri sayesinde yalnızca internet bağlantısı ve bir tarayıcı yardımıyla

modeller kurmaya yardımcı olur. Model hızlı ve düşük bellek ihtiyacından dolayı gerekli kütüphanelerin kurulumu sonrasında giriş düzeyi donanım özellikleriyle kısa sürede sonuç alınabilmektedir. Bir sonraki bölümde bu çalışmanın çıktıları grafiklerle verilmiştir.

2.5. Değerlendirme (Evaluation)

Makine öğrenmesi modellerinin başarısını ölçmek için birden fazla hesaplama yöntemi vardır. Bu iş probleminde amacımız sürekli bir değişkeni modellemek olduğu için **MAPE** (Mean absolute percentage error) metriği seçilmiştir. Literatürde MAPE ile model doğruluğu örnekleri mevcuttur (bkz. [49] [50] [51] [52]). MAPE Türkçe karşılığı ile Ortalama Mutlak Yüzde Hatası, tahmin modellerinin performansını değerlendirmek için kullanılan bir metriktir. MAPE, tahmin edilen değerlerin gerçek değerlere göre yüzde olarak ne kadar sapma gösterdiğini ölçer [49]. Denklemi aşağıdaki eşitlikle ifade edilmiştir:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right|$$

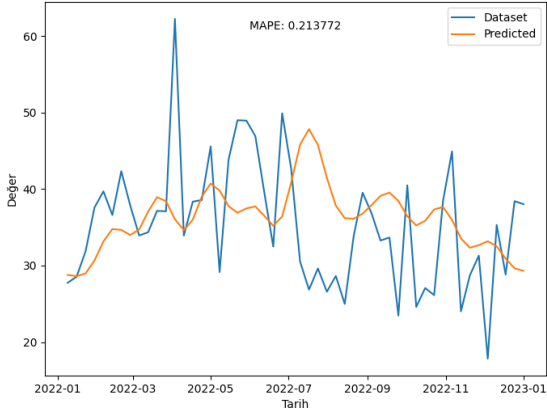
Burada Y_i gerçek değerdir ve \hat{Y}_i tahmin değeridir. Farkları Y_i 'nin gerçek değerine bölünür. Bu oranın mutlak değeri, zaman içinde tahmin edilen her nokta için toplanır ve uygun noktaların n sayısına bölünür. MAPE değeri yüzde olarak ifade edilir ve genellikle 0-1 arasında bir değer alır. MAPE'nin düşük olması, model tahminlerinin gerçek değerlerden daha az sapmış olduğunu gösterir. MAPE, temel içsel parametreleri kullanarak zaman serilerini büyük bir doğrulukla tahmin etmemizi sağlar [52].

2.6. Bulgular ve tartışma (Findings and Discussion)

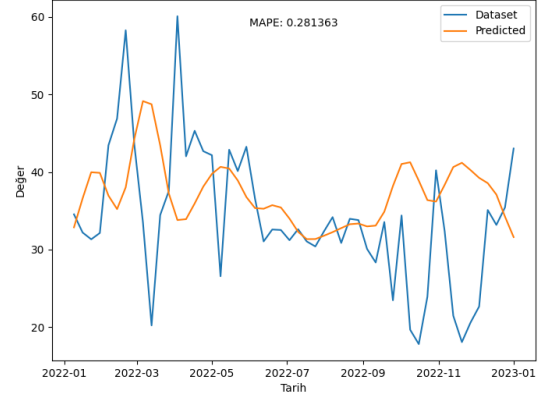
Modelleme sonucu elde edilen bulgular devamında verilmiştir.

Şekil 5'te istasyonlar için kurulan modellerin 2022 yılı içinde yaptığı tahminler ve ölçülen gerçek değerler karşılaştırılmıştır. Grafik üzerindeki MAPE değeri de model başarısını sayısal olarak ölçmek için eklenmiştir. Kurulan modellerin eğitim setinde 2022 yılına ait verilerin çıkarıldığı unutulmamalıdır. Model başarısı ancak ve ancak daha önce görmediği verileri tahmin ederek objektif olarak değerlendirilebilir.

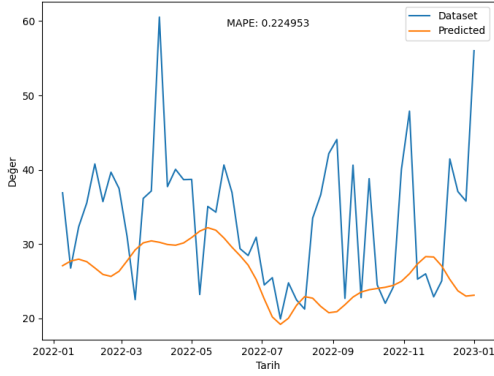
Aşağıdaki Şekillerde 5'ten 26'ya kadar olan grafiklerde mavi çizgiler ölçülen AQI değerlerini, turuncular ise modelin geçmiş verilerden yola çıkarak tahminlediği AQI değerlerini temsil etmektedir. Eğitim ve tahminler aylık ortalamalar üzerinden yapılmış, her ölçüm istasyonunun 2010 yılına kadar ölçüm yaptığı tüm zamanlardan 2021 aralık ayına kadarki kaydettiği değerler eğitim seti olarak modele sunulmuş ve sonrasında 2022 yılı için tahmin yapılması istenmiştir.



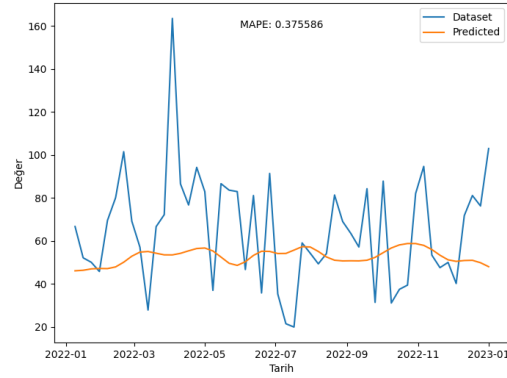
Şekil 5. Arnautköy istasyonu model değerlendirme (Arnautköy station model evaluation)



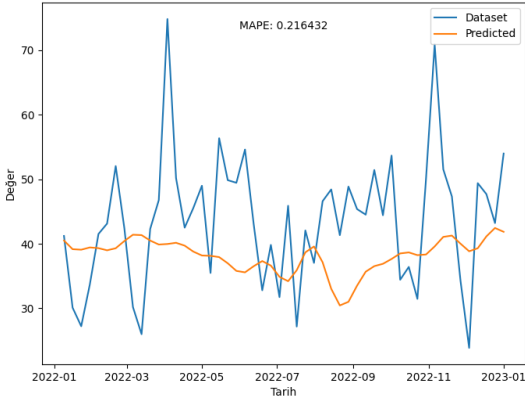
Şekil 8. Bağcılar istasyonu model değerlendirme (Bağcılar station model evaluation)



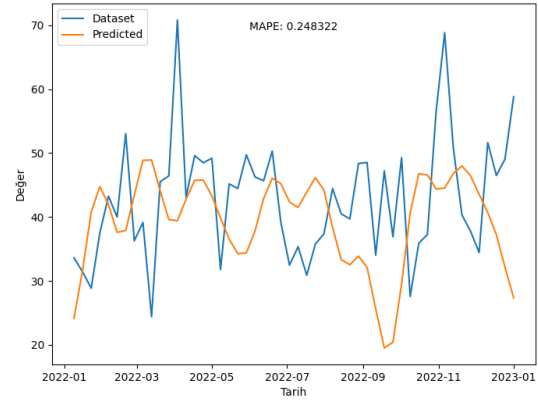
Şekil 6. Beşiktaş istasyonu model değerlendirme (Beşiktaş station model evaluation)



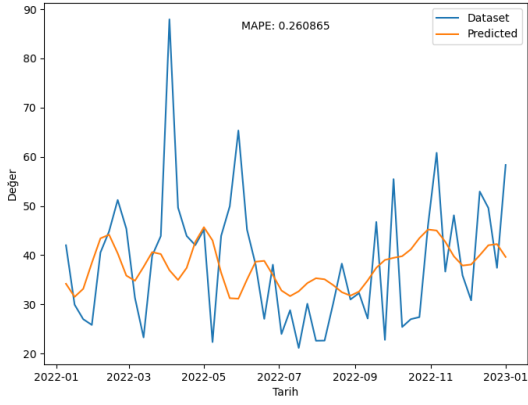
Şekil 9. Göztepe istasyonu model değerlendirme (Göztepe station model evaluation)



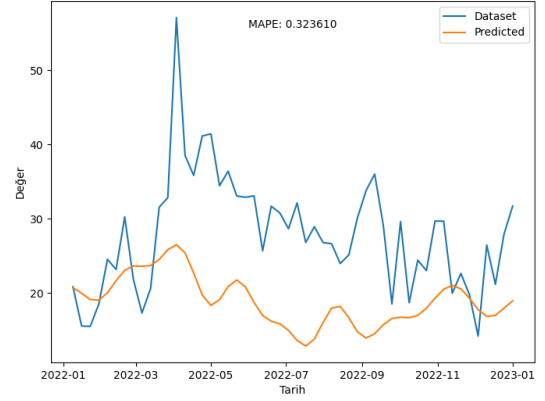
Şekil 7. Kartal istasyonu model değerlendirme (Kartal station model evaluation)



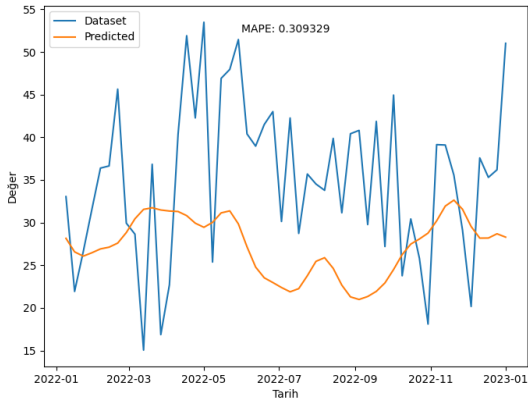
Şekil 10. Sancaktepe istasyonu model değerlendirme (Sancaktepe station model evaluation)



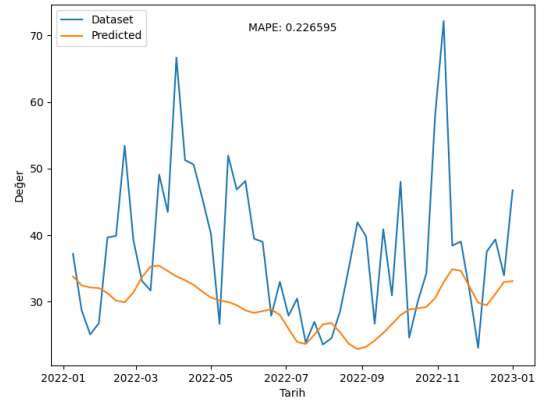
Şekil 11. Tuzla istasyonu model değerlendirme (Tuzla station model evaluation)



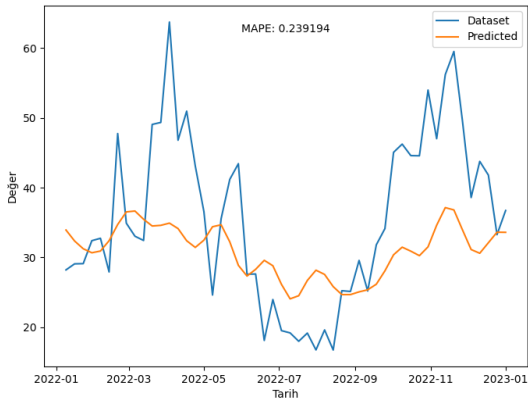
Şekil 14. Sarıyer istasyonu model değerlendirme (Sarıyer station model evaluation)



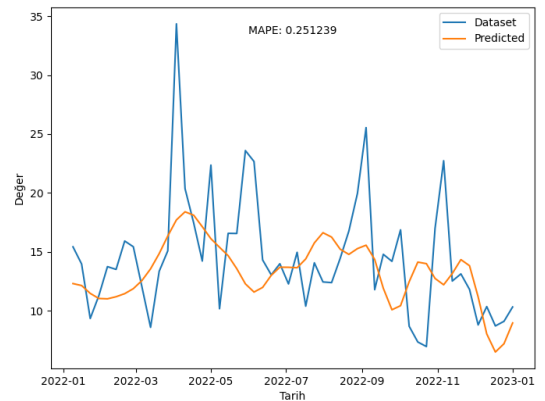
Şekil 12. Ümraniye istasyonu model değerlendirme (Ümraniye station model evaluation)



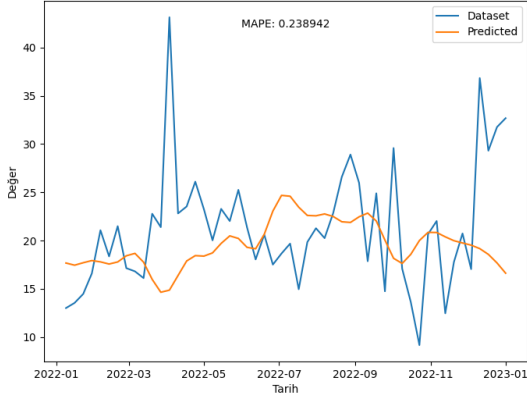
Şekil 15. Esenler istasyonu model değerlendirme (Esenler station model evaluation)



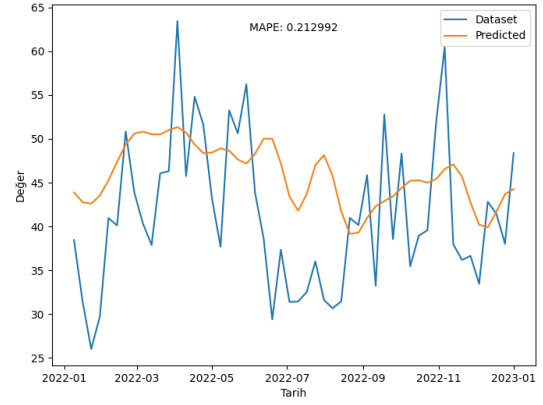
Şekil 13. Yenibosna istasyonu model değerlendirme (Yenibosna station model evaluation)



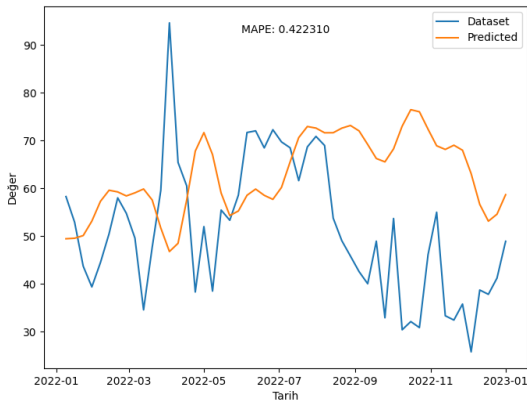
Şekil 16. Büyükada istasyonu model değerlendirme (Büyükada station model evaluation)



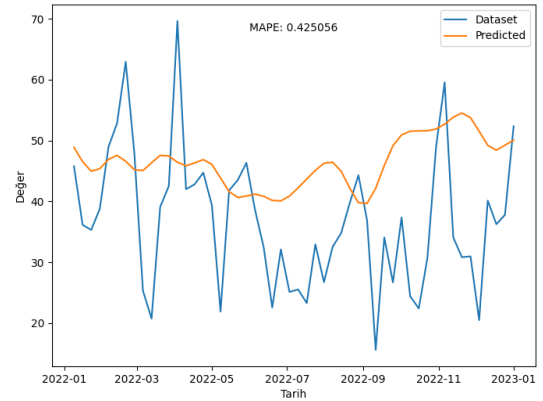
Şekil 17. Kumköy istasyonu model değerlendirme (Kumköy station model evaluation)



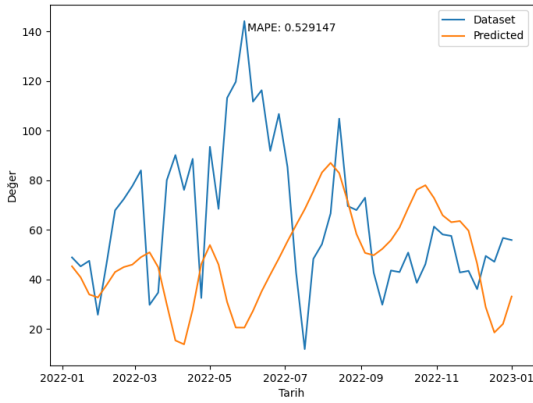
Şekil 20. Aksaray istasyonu model değerlendirme (Aksaray station model evaluation)



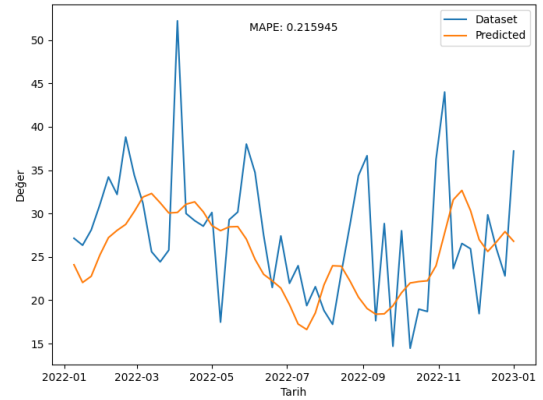
Şekil 18. Sultangazi2 istasyonu model değerlendirme (Sultangazi2 station model evaluation)



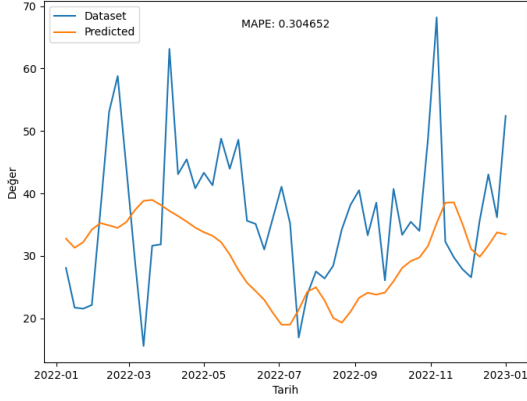
Şekil 21. Alibeyköy istasyonu model değerlendirme (Alibeyköy station model evaluation)



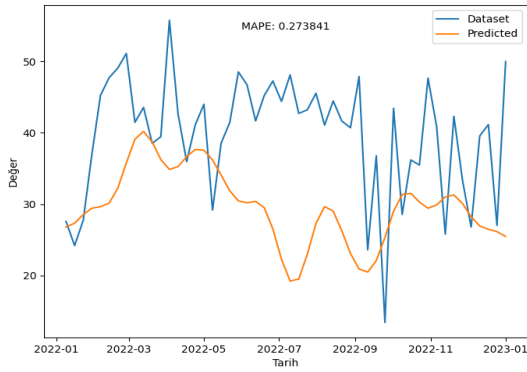
Şekil 19. Sultangazi3 istasyonu model değerlendirme (Sultangazi3 station model evaluation)



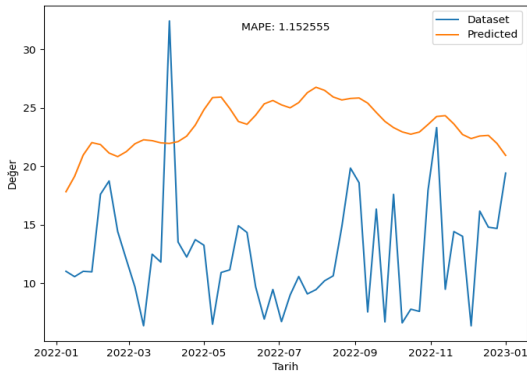
Şekil 22. Üsküdar1 istasyonu model değerlendirme (Üsküdar1 station model evaluation)



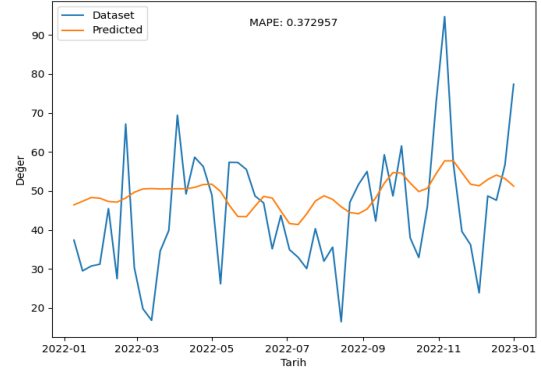
Şekil 23. Kadıköy istasyonu model değerlendirme (Kadıköy station model evaluation)



Şekil 24. Selimiye istasyonu model değerlendirme (Selimiye station model evaluation)

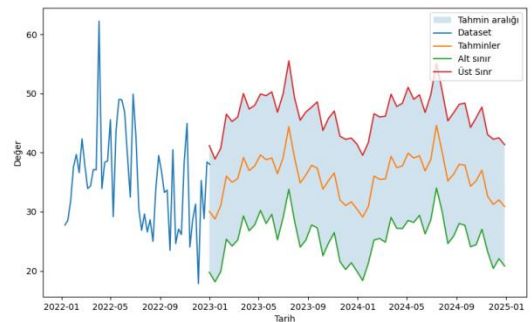


Şekil 25. Kandilli istasyonu model değerlendirme (Kandilli station model evaluation)

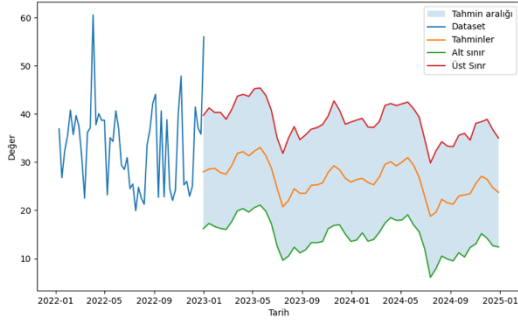


Şekil 26. Kağıthane1 istasyonu model değerlendirme (Kağıthane1 station model evaluation)

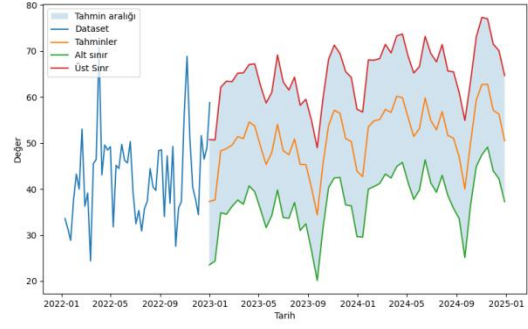
Aynı zamanda her model için MAPE değeri hesaplanmış, grafiklerin üzerinde eklenmiştir. Bazı istasyonlarda yaklaşık %20 hata oranıyla tahmin yapılırken, %100'ü aşan hata oranları da görülebilmektedir. Örneğin, sadece Şekil 25'te MAPE değeri 1.152555 olarak gelmiştir. Bunun anlamı, tahmin edilen değerlerin gerçek değerlere göre ortalama yüzde olarak %115.2555 hata ile sapma göstermesidir. Yani, tahminler gerçek değerlerden ortalama olarak yaklaşık %15 daha fazla çıkmıştır. Bu, tahmin modelinin Kandilli1 veri seti üzerinde iyi performans sergilemediğini göstermektedir. Bu fark birçok etmene bağlı olarak ortaya çıkabilir. Çalışmanın bir amacı da hem veri analizi hem modelleme aşamasında ortaya çıkan bu farka dikkat çekmektir. Nispeten düşük hata oranıyla tahmin yapan model için tamamiyle başarılı demek ya da aksi durumda başarısız demek doğru olmayacaktır. Birçok semtte veriler 5 yıldan az olurken 2010 yılından itibaren aralıksız kayıt alan istasyonlar da vardır ve bilindiği gibi veri miktarı makine öğrenmesi projelerinde başarıyı etkileyen önemli bir unsurdur. Ek olarak tahminlenen 2022 yılında ilgili istasyonda yaşanabilecek özel durumları modelin tahmin etmesi oldukça güçtür zira yalnızca bağımlı değişken ve zaman değişkeni ile eğitim ve test aşamaları tamamlanmıştır. Hem model başarısını daha doğru ölçmenin hem de çeşitli optimizasyonlarla başarıyı artırmanın yolları oldukça fazladır. Çalışmanın devamında eğitim setine 2022 yılı dahil edilmiş ve her istasyon için yeniden model kurulmuş, 2023-2025 yılı için tahminlerde bulunulmuştur. Aşağıda bu modeller görülebilir:



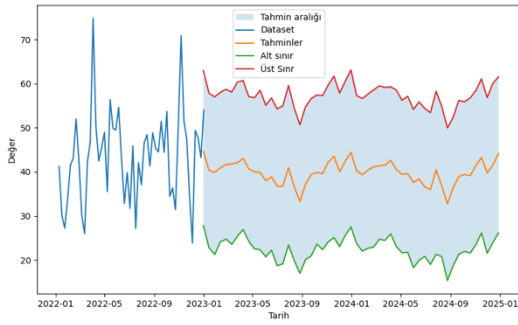
Şekil 27. Arnavutköy istasyonu 2023-2025 tahmini (Arnavutköy station 2023-2025 forecast)



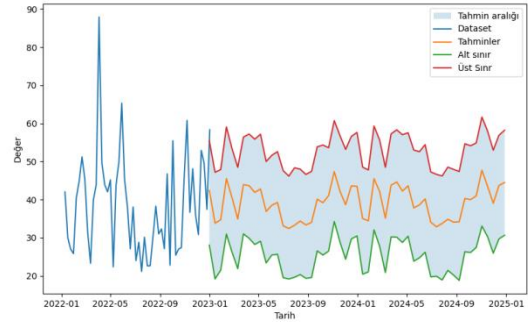
Şekil 28. Beşiktaş istasyonu 2023-2025 tahmini (Beşiktaş station 2023-2025 forecast)



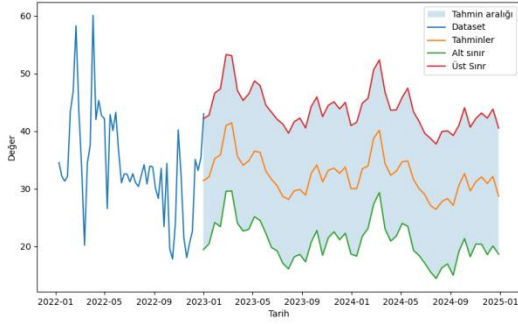
Şekil 32. Sancaktepe istasyonu 2023-2025 tahmini (Sancaktepe station 2023-2025 forecast)



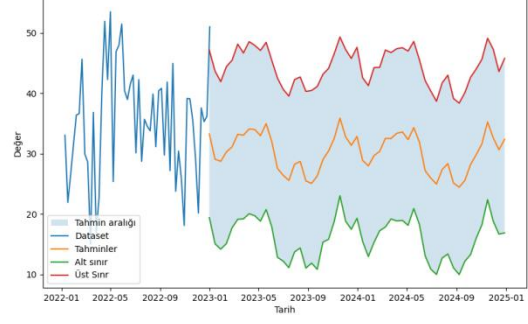
Şekil 29. Kartal istasyonu 2023-2025 tahmini (Kartal station 2023-2025 forecast)



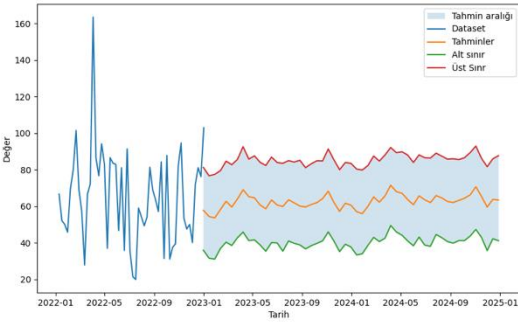
Şekil 33. Tuzla istasyonu 2023-2025 tahmini (Tuzla station 2023-2025 forecast)



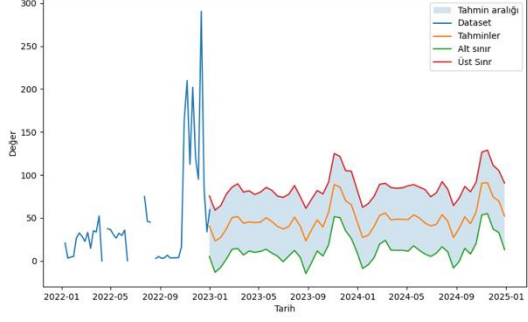
Şekil 30. Bağcılar istasyonu 2023-2025 tahmini (Bağcılar station 2023-2025 forecast)



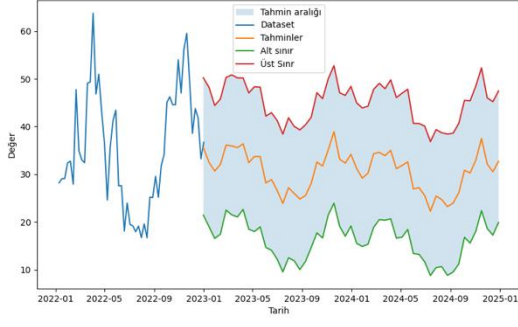
Şekil 34. Ümraniye1 istasyonu 2023-2025 tahmini (Ümraniye1 station 2023-2025 forecast)



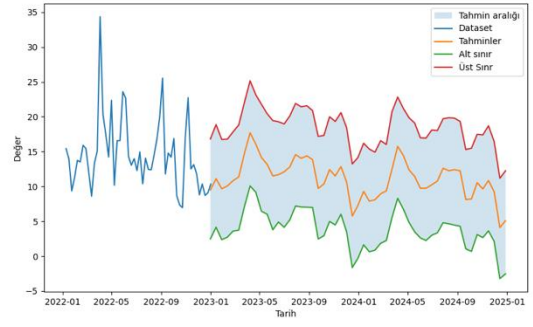
Şekil 31. Göztepe istasyonu 2023-2025 tahmini (Göztepe station 2023-2025 forecast)



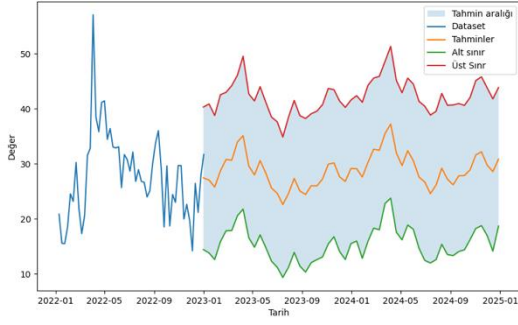
Şekil 35. Mobil istasyonu 2023-2025 tahmini (Mobil station 2023-2025 forecast)



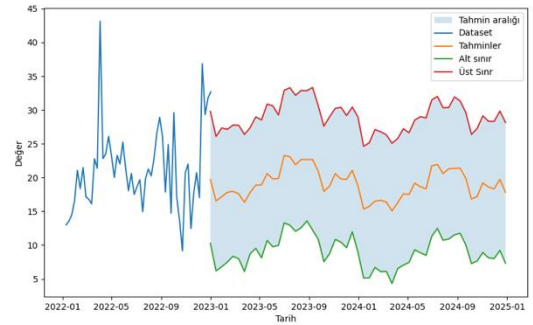
Şekil 36. Yenibosna istasyonu 2023-2025 tahmini (Yenibosna station 2023-2025 forecast)



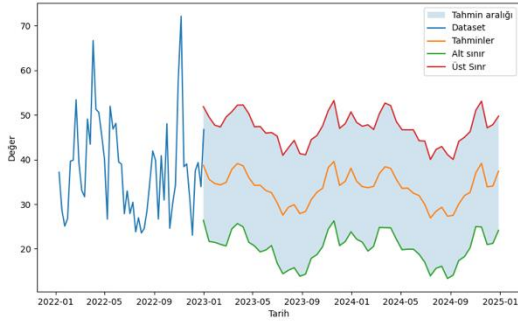
Şekil 40. Büyükada istasyonu 2023-2025 tahmini (Büyükada station 2023-2025 forecast)



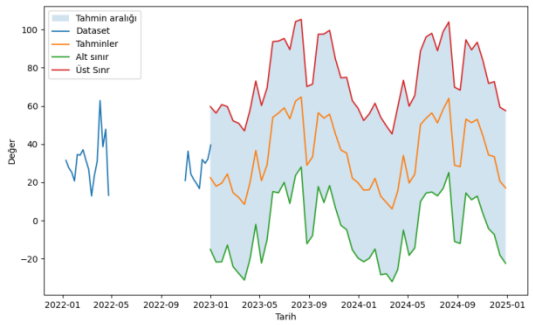
Şekil 37. Sarıyer istasyonu 2023-2025 tahmini (Sarıyer station 2023-2025 forecast)



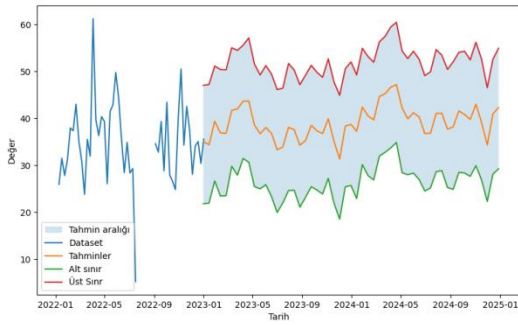
Şekil 41. Kumköy istasyonu 2023-2025 tahmini (Kumköy station 2023-2025 forecast)



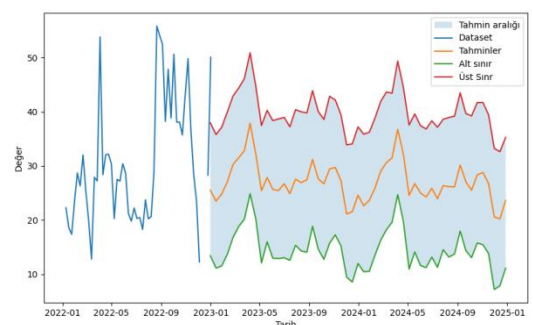
Şekil 38. Esenler istasyonu 2023-2025 tahmini (Esenler station 2023-2025 forecast)



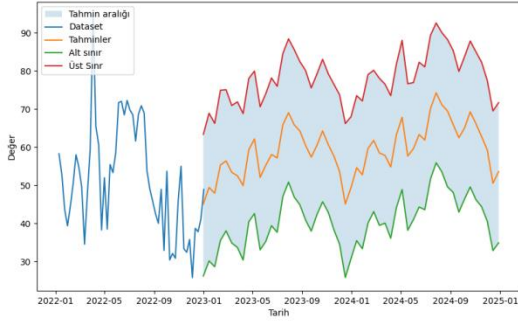
Şekil 42. Sultangazi istasyonu 2023-2025 tahmini (Sultangazi station 2023-2025 forecast)



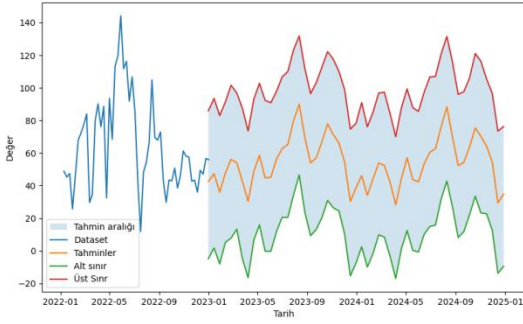
Şekil 39. Çatladıkapı istasyonu 2023-2025 tahmini (Çatladıkapı station 2023-2025 forecast)



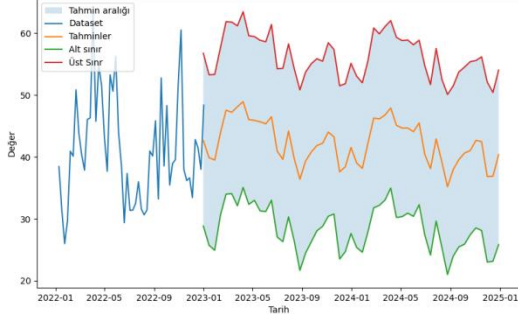
Şekil 43. Maslak istasyonu 2023-2025 tahmini (Maslak station 2023-2025 forecast)



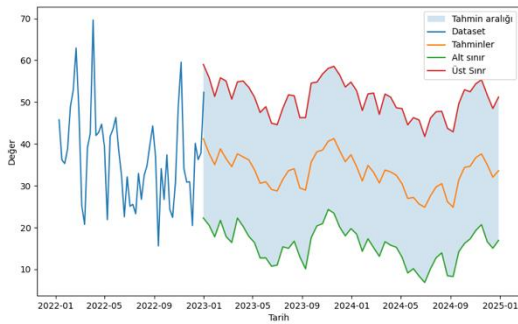
Şekil 44. Sultangazi2 istasyonu 2023-2025 tahmini (Sultangazi2 station 2023-2025 forecast)



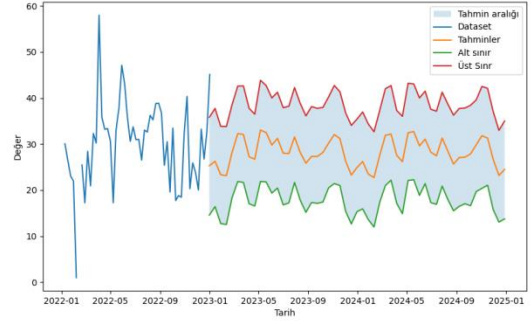
Şekil 45. Sultangazi3 istasyonu 2023-2025 tahmini (Sultangazi3 station 2023-2025 forecast)



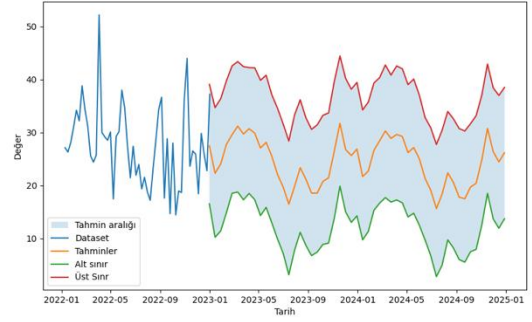
Şekil 46. Aksaray istasyonu 2023-2025 tahmini (Aksaray station 2023-2025 forecast)



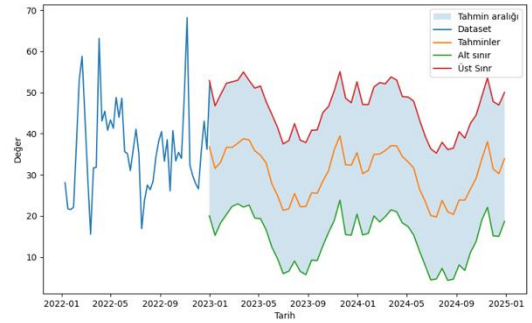
Şekil 47. Alibeyköy istasyonu 2023-2025 tahmini (Alibeyköy station 2023-2025 forecast)



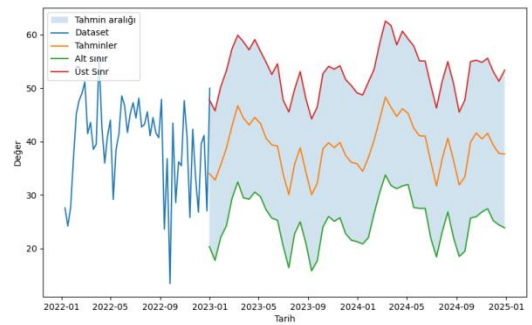
Şekil 48. Avcılar istasyonu 2023-2025 tahmini (Avcılar station 2023-2025 forecast)



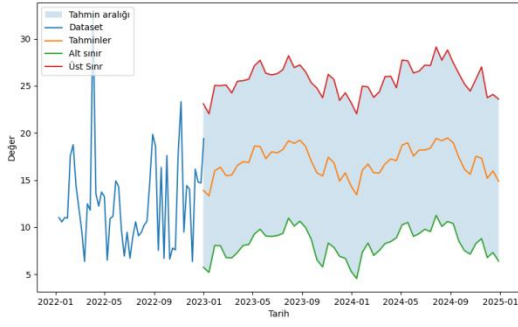
Şekil 49. Üsküdar1 istasyonu 2023-2025 tahmini (Üsküdar1 station 2023-2025 forecast)



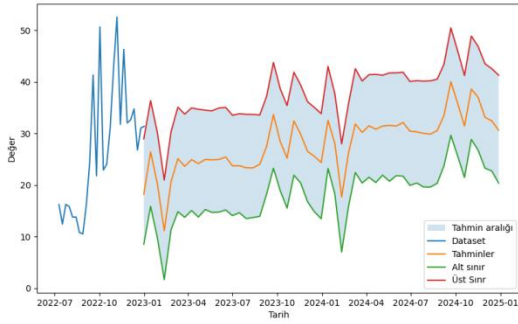
Şekil 50. Kadıköy istasyonu 2023-2025 tahmini (Kadıköy station 2023-2025 forecast)



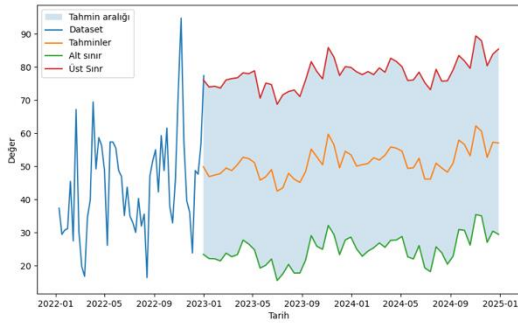
Şekil 51. Selimiye istasyonu 2023-2025 tahmini (Selimiye station 2023-2025 forecast)



Şekil 52. Kandilli istasyonu 2023-2025 tahmini (Kandilli station 2023-2025 forecast)



Şekil 53. Beylikdüzü istasyonu 2023-2025 tahmini (Beylikdüzü station 2023-2025 forecast)



Şekil 54. Kağıthane1 istasyonu 2023-2025 tahmini (Kağıthane1 station 2023-2025 forecast)

BÖLÜM 3: SONUÇ VE ÖNERİLER (CONCLUSION AND RECOMMENDATIONS)

Bu bölüm, çalışmanın özgün sonuçlarının paylaşıldığı, tartışıldığı bölümdür.

3.1. Model Sonuçlarının Yorumu (Interpretation of Model Results)

Çalışmada her ilçe için yapıla her modele eklenen MAPE değerlerinin ortalaması 0.316495 olarak bulunmuştur ki bu da alanda yapılan başka çalışmaların MAPE değerleri ile karşılaştırıldığında kabul edilebilir bir modelleme olduğunu göstermektedir [53]. Ancak literatürde uzun

sürekli verilerle çalışılmış ve uzun süreli tahmin yapmış kapsamlı bir örnek bulunmadığı ve tahminlemede daha çok diğer değişkenlerin ayrı ayrı değişimine odaklanıldığı için tam bir karşılaştırma mümkün olamamıştır.

Saatlik, günlük, haftalık zaman dilimlerinde kurulan modellerin hata oranı oldukça yüksek olabilmektedir. Çalışma kapsamında veriler bir yıllık süreyle aylık olarak tahmin edilmiştir. Böylelikle kısa zaman dilimlerinde hava kirliliğini etkileyebilecek fazla sayıda unsur ile model performansında oluşabilecek olası olumsuz etkilerden korunmuştur. Her ilçe için yapılmış olan modelleme tek tek yukarıda şekillerle gösterilmiştir. Çalışma gerektiği taktirde daha fazla veri eklenerek daha iyi tahminler elde edilebilir. Böylece karar vericilerin alacağı önlemler ile gelecek zaman dilimlerinde model tahmini ile ölçülen değerler kıyaslanabilecektir.

Sultangazi, Mobil, Çatladıkapı istasyonlarında eksik verilerin varlığı göze çarpmaktadır. Prophet modeli az miktarda eksik veriler barından veri setlerinde de başarılı tahmin yapabilecek modellerden biridir. Yine de bu çevre problem ele alınırken geçmiş verilerin önemi göz ardı edilmemeli, ölçüm istasyonlarının sayısı ve kalitesi sürekli artırılmalıdır.

Mobil istasyonu sürekli veriye sahip olmadığı ve sabit konumda olmadığı için bölgesel fikirler verememektedir. Ancak, nedenselliği sorgulamak, olası özel durumlarda hava kalitesini ölçebilmek için bunun gibi istasyonlar şarttır ve sayısı artırılmalıdır. Ölçüm yaptığı lokasyonların tam konumu da ilgili verilerde paylaşılmalıdır.

Tahmin sonuçlarına göre AQI değeri artış trendine giren bir istasyon bulunmamaktadır. Bu durum en azından kirliliğin agresif bir şekilde artmamasını öngörse de, bazı bölgeler (Sultangazi, Sarıyer, Üsküdar) için yılın belirli zamanlarında yükseliş beklentisi göze çarpmaktadır. Bu artışın sebepleri o zaman diliminde detaylıca incelenmelidir. Özellikle bu zamanlarda yerel yönetimlerin yapacağı farkındalık çalışmaları ve alınacak önlemlerle bu artışın önüne geçilmeye çalışılmalıdır.

Modelin çıktı olarak ürettiği; alt sınır, üst sınır değerleri gözetilmeli, özellikle üst sınırın üzerine çıkılan zamanlarda bir sonraki bölümde verilen bölgesel öneriler dikkate alınabilir.

Bu çalışmada verilerin ve değerlendirmenin il geneli için kapsayıcı olması nedeniyle başka modellerin denenip karşılaştırılması ve model doğruluğu konusunda başka metriklere yer verilmesi mümkün olmamıştır.

3.2. Öneriler (Suggestions)

Birçok çalışma büyük şehirlerde hava kirliliğine karşı alınabilecek önlemlerden bahsetmiştir [54][55][56][57]. Buna karşın bu çalışmadan çıkan sonuçlara göre başka iller için benzer çalışmaların aynı ya da başka kategorileri dikkate alarak, aynı ya da başka modeller de kullanılarak

yapılması bu şehirlerde yaşayan insanların hava kirliliğinin etkilerini daha iyi anlamasına yardımcı olacaktır. Çalışmaya rüzgar, nem, yağış gibi meteoroloji ve trafik yoğunluğu gibi veriler de eklenerek tahmin doğruluğu desteklenebilir ve eş zamanlı tahminler yapılabilir ve paydaşlarla eş zamanlı paylaşımına açılabilir.

KAYNAKLAR (REFERENCES)

- [1] Internet: J. Mackenzie, Air Pollution: Everything You Need to Know, NRDC, <https://www.nrdc.org/stories/air-pollution-everything-you-need-know#whatis>, 19.07.2023.
- [2] G. Kara, B. Çiğdem, Y. Çay, "Konya'da Hava Kirliliği Konusunda Farkındalık Düzeyinin Belirlenmesi", *Ulusal Çevre Bilimleri Araştırma Dergisi*, 2(2), pp. 91-94, 2019.
- [3] W. Nazar ve M. Niedoszytko, "Air Pollution in Poland: A 2022 Narrative Review with Focus on Respiratory Diseases", *International Journal of Environmental Research and Public Health*, 19(2), p. 895, 2022.
- [4] H. Zamri, Z. A. Shaffiei, N. A. Daud, D. N. Ahmad, "AirAwareMalaysia: Data Visualization and Air Quality Awareness on Air Pollution in Selangor Using Big Data Analytics", *International Conference on Soft Computing and Data Mining*, 457, pp. 223-233, 2022.
- [5] Internet: D. Kim, Media Portrayal and Public Awareness of Air Pollution in Los Angeles and Seoul, Digital Commons Connecticut College, 2022, <https://digitalcommons.conncoll.edu/sip/39/>, 19.07.2023.
- [6] S. Illari, R. T. Lado, L. Marrodan, "Traffic and Pollution Modelling for Air Quality Awareness: An Experience in the City of Zaragoza", *SN Computer Science*, 3(3), p. 281, 2022.
- [7] G. Kanchelashvili, L. Gulbiani, A. Dekanosidze, L. Kvachantiradze, G. Kamkamidze, L. Sturua, "Knowledge Of Georgian Population Towards Air Pollution And Health Effects Of Lead Contamination", *Georgian Med News*, 322, pp. 58-62, 2022.
- [8] A. Piracha, M. T. Chaudhary, "Urban Air Pollution, Urban Heat Island and Human Health: A Review of the Literature", *Sustainability*, 14(15), p. 9234, 2022.
- [9] M. Banach, R. Dlugosz, T. Talaska, W. Pedrycz, "Air Pollution Monitoring System with Prediction Abilities Based on Smart Autonomous Sensors Equipped with ANNs with Novel Training Scheme", *Remote Sensing*, 14(2), p. 413, 2022.
- [10] C. Giorio, "Energy efficiency vs indoor air quality conundrum and possible solutions", *ECG Bulletin*, p. 20, 2022.
- [11] E. Di Pirro, L. Sollustio, J. A. C. Castellar, G. Sgrigna, M. Marchetti, B. Lasserre, "Facing Multiple Environmental Challenges through Maximizing the Co-Benefits of Nature-Based Solutions at a National Scale in Italy", *Forests*, 13(4), p. 548, 2022.
- [12] A. Bhatnagar, "Cardiovascular Effects of Particulate Air Pollution", *Annual Reviews*, 73, pp. 393-406, 2022.
- [13] T. Jin, Q. Di, W. J. Requia, M. D. Yazdi, E. Castro, T. Ma, Y. Wang, "Associations between long-term air pollution exposure and the incidence of cardiovascular diseases among American older adults", *Environment international*, 170, p. 107594, 2022.
- [14] R. D. Brook, J. R. Brook ve S. Rajagopalan, "Air pollution: The "heart" of the problem", *Current Hypertension Reports*, 5, pp. 32-39, 2003.
- [15] X. Xu, H. Yang, C. Li, "Theoretical Model and Actual Characteristics of Air Pollution Affecting Health Cost: A Review", *International Journal of Environmental Research and Public Health*, 19(6), p. 3532, 2022.
- [16] R. Fuller, P. J. Landrigan, K. Balakrishnan, G. Bathan, M. Brauer, J. Caravanos ve T. Chiles, "Pollution and health: a progress update", *The Lancet Planetary Health*, 6(6), pp. 535-547, 2022.
- [17] B. Helene, S. Gabet, G. Kerneis, A. Provent, C. Rieux, N. B. Salem, H. Dupont, et al., "Designing local air pollution policies focusing on mobility and heating to avoid a targeted number of pollution-related deaths: Forward and backward approaches combining air pollution modeling, health impact assessment and cost-benefit analysis", *Environment International*, 159, p. 107030, 2022.
- [18] Ş. Güzel, P. Özer, "Türkiye'de Hava Kirliliği ve Sağlık Harcamaları", *Sağlık ve Sosyal Refah Araştırmaları Dergisi*, 4(2), pp. 186-202, 2022.
- [19] S. Ünalı, N. Yalçın, "Hava Kirliliğinin Makine Öğrenmesi Tabanlı Tahmini: Başakşehir Örneği", *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 4(1), pp. 35-44, 2022.
- [20] D. Sanjeev, "Implementation of Machine Learning Algorithms for Analysis and Prediction of Air Quality", *International Journal of Engineering Research & Technology (IJERT)*, 10(3), pp. 533-538, 2021.
- [21] M. Öklü ve P. Canbay, "Makine Öğrenmesi Yöntemleri ile Şehirlerin Hava Kalitesi Tahmini". *International Journal of Advances in Engineering and Pure Sciences*, 35(1), pp.39-53, 2023.
- [22] E. Çetin Yağmur, "Atmosferik Partikül Maddelerin Makine Öğrenmesi İle Tahmini: Beşiktaş, İstanbul Örneği", *KONJES*, 10(4), pp. 807-826, 2022. doi: 10.36306/konjes.1082866.
- [23] Oğuz, K., & Pekin, M. A. (2022). Makine Öğrenme Algoritmaları ile PM10 Konsantrasyon Tahmini. *Journal of Advanced Research in Natural and Applied Sciences*, 8(2), 201-213. <https://doi.org/10.28979/jarnas.981202>
- [24] B. Duygu Çelik ve N. Arıcı, "Covid-19 Salgın Sürecinde Hava Kalitesi Tahmini: Zonguldak Örneği", *GJES*, vol. 7, no. 3, pp. 222-232, 2021.
- [25] Akgün, M., & Barlık, N. (2023). Makine Öğrenmesi Algoritmaları Kullanılarak Hava Kalitesi İndeksinin Tahmini. *Avrupa Bilim Ve Teknoloji Dergisi* (51), 97-107. <https://doi.org/10.31590/ejosat.1241170>
- [26] Y. Gültepe, "Makine Öğrenmesi Algoritmaları ile Hava Kirliliği Tahmini Üzerine Karşılaştırmalı Bir Değerlendirme", *European Journal of Science and Technology*, 16, 8-15, 2019.
- [27] L.K. Wang, N.C. Pereira and T.S. Hung, **Air pollution control engineering**, 1, New Jersey, Humana Press, 100-105, 2004.
- [28] Y.T. Tsai, Y. R. Zeng and Y. S. Chang, "Air Pollution Forecasting Using RNN with LSTM" 2018 The 16th IEEE International Conference on Pervasive Intelligence and Computing, 1074-107, 2018.

- [29] Y. Rybarczyk and R. Zalakeviciute, "Machine learning approach to forecasting urban pollution" 2016 IEEE Ecuador Technical Chapters Meeting, ETCM, 2016.
- [30] <https://api.ibb.gov.tr/havakalitesi/OpenDataPortalHandler/GetAQIByStationId>
- [31] H. Gong, "Health Effects Of Air Pollution: A Review of Clinical Studies", *Clinics in Chest Medicine*, 13(2), 201-214, 1992. ISSN 0272-5231, [https://doi.org/10.1016/S0272-5231\(21\)00851-0](https://doi.org/10.1016/S0272-5231(21)00851-0).
- [32] R. Chen, H. Kan, B. Chen, W. Huang, Z. Bai, G. Song, G. Pan, "Association of Particulate Air Pollution With Daily Mortality: The China Air Pollution and Health Effects Study", *American Journal of Epidemiology*, 175(11), 1173-1181, 2012. <https://doi.org/10.1093/aje/kwr425>
- [33] Impactsey and V. Ramanathan, "Recent climate and air pollution impacts on Indian agriculture." *Proceedings of the National Academy of Sciences* 111(46), 16319-16324, 2014.
- [34] J. L. Domingo and R. Joaquim, "Effects of air pollutants on the transmission and severity of respiratory viral infections", *Environmental research*, 187 (2020), 109650, 2020.
- [35] J. Q. Koenig, "Air pollution and asthma", *Journal of allergy and clinical immunology*, 104.4, 717-722, 1999.
- [36] K. Donaldson, M. Gilmour, W. MacNee, "Asthma and PM10", *Respiratory Research*, 1(1), pp. 12-15, 2000.
- [37] C. Mallik, "Anthropogenic sources of air pollution", *Air pollut. sources, impacts control*, 6-25, 2019.
- [38] S. A. Horn and P. K. Dasgupta, "The Air Quality Index (AQI) in historical and analytical perspective a tutorial review", *Talanta*, 125260, 2023.
- [39] W. Krewitt, A. Trukenmüller, T.M. Bachmann, et al., "Country-specific damage factors for air pollutants", *Int J LCA*, 6, pp. 199-210 2001.
- [40] E. Bales, N. Nikzad, N. Quick, et al., "Personal pollution monitoring: mobile real-time air quality in daily life", *Pers Ubiquit Comput*, 23, pp. 309-328, 2019.
- [41] C. Fan, Y. Li, J. Guang, Z. Li, A. Elnashar, M. Allam, G. de Leeuw, "The Impact of the Control Measures during the COVID-19 Outbreak on Air Pollution in China", *Remote Sens.*, 12, 1613, 2020.
- [42] Internet: "Adalar Turizm Raporu Ocak 2020", Şehir Planlama Müdürlüğü, <https://sehirplanlama.ibb.istanbul/wp-content/uploads/2020/07/adalar-turizm-raporu-IBB-icin.pdf>
- [43] Internet: Üsküdar "Şehir Kimliği" Çalıştayı Sonuç Raporu, T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, 2014, <https://cevresehiriklimkutuphanesi.csb.gov.tr/SourcePDF/52dcf1a1-9c69-4149-9cdc-067dfe15de1f>
- [44] Internet: Türkiye News, "Air pollution high in areas with traffic congestion in Istanbul", *Hürriyet Daily News*, 2023. <https://www.hurriyetdailynews.com/air-pollution-high-in-areas-with-traffic-congestion-in-istanbul-187565>
- [45] D. Jandacka, D. Durcanska, M. Bujdos, "The contribution of road traffic to particulate matter and metals in air pollution in the vicinity of an urban road", *Transportation Research Part D: Transport and Environment*, 50, pp.397-408, 2017.
- [46] Internet: H. F. Başer, "İstanbul'da hava kirliliğinin en büyük nedeni "trafik"", Anadolu Ajansı, <https://www.aa.com.tr/tr/cevre/istanbulda-hava-kirliliginin-en-buyuk-nedeni-trafik/2425106>, 19.07.2023.
- [47] Internet: Taşıt-kilometre İstatistikleri, 2019, Türkiye İstatistik Kurumu, <https://data.tuik.gov.tr/Bulten/Index?p=Tasit-kilometre-Istatistikleri-2019-37409>, 19.07.2023.
- [48] S. J. Taylor, B. Letham, "Forecasting at scale", *The American Statistician*, 72(1), pp. 37-45, 2018.
- [49] A. H. Setianingrum, N. Anggraini, and M. F. D. Ikram, "Prophet model performance analysis for Jakarta air quality forecasting." In 2022 10th international conference on cyber and it service management (citsm), 1-7, IEEE, 2022.
- [50] F. Cai, "The Prediction of the Air Quality based on the Prophet Algorithm" *Highlights in Science, Engineering and Technology*, 39, 1056-1060, 2023.
- [51] W.Y. Zhang, W.C. Hong, Y. Dong, G. Tsai, J.T. Sung and G.F. Fan, "Application of SVR with chaotic GASA algorithm in cyclic electric load forecasting" *Energy*, 45(1), 850-858, 2012.
- [52] V. Gupta, S. Kapadia and C. Bhadane, "Time Series Analysis and Forecasting of Air Quality in India", In 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), 1-5, IEEE, 2023.
- [53] S. Ünalı ve N. Yalçın, "Hava kirliliğinin makine öğrenmesi tabanlı tahmini: Başakşehir örneği", *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 4(1), 35-44, 2022.
- [54] S. Ertürk Atabey, E. Eroğlu, "Kentsel Negatif Dışsallık Olarak Trafığe Bağlı Hava Kirliliği: Kentsel Kamu Maliyesi Bağlamında Çözüm Önerisi", *Çağdaş Yerel Yönetimler Dergisi*, 30(1), 71-98, 2021.
- [55] R. Akdur, "Hava Kirliliği ve Çözümler", *Halk Sağlığı ile İlgili Güncel Sorunlar ve Yaklaşımlar*, Ankara Tabipler Odası, 151-155, 2009.
- [56] M. E. Akıner, "Akıllı Şehirlerde Hava Kirliliğine Yönelik Mobil Algılama Sistemleri ve Sürdürülebilir Binalarda İç Hava Kalitesine Genel Bakış", *Teoriden Pratiğe Akıllı Şehirler*, 2024.
- [57] K. B. Schnelle, M. E. Ternes, R. F. Dunn, **Air pollution control technology handbook**, ABD: Crc Press, 2016.