

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**GAZİANTEP'TE PM_{2.5} KONSANTRASYONUNUN ZAMANSAL VE MEKÂNSAL
TAHMİNİNE YÖNELİK TRANSFER ÖĞRENME DESTEKLİ
HİBRİT YAPAY ZEKA MODELLERİ**

YÜKSEK LİSANS TEZİ

Türkan ZENGİN GÖMLEKSİZ

İklim Bilimi ve Meteoroloji Mühendisliği Anabilim Dalı

Atmosfer Bilimleri Programı

MAYIS 2025

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**GAZİANTEP'TE PM_{2.5} KONSANTRASYONUNUN ZAMANSAL VE MEKÂNSAL
TAHMİNİNE YÖNELİK TRANSFER ÖĞRENME DESTEKLİ
HİBRİT YAPAY ZEKA MODELLERİ**

YÜKSEK LİSANS TEZİ

**Türkan ZENGİN GÖMLEKSİZ
(511221013)**

İklim Bilimi ve Meteoroloji Mühendisliği Anabilim Dalı

Atmosfer Bilimleri Programı

Tez Danışmanı: Prof. Dr. Hüseyin TOROS

MAYIS 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**SPATIO-TEMPORAL ESTIMATION OF PM_{2.5} CONCENTRATIONS IN GAZIANTEP
USING TRANSFER LEARNING-BASED
HYBRID ARTIFICIAL INTELLIGENCE MODELS**

M.Sc. THESIS

**Türkan ZENGİN GÖMLEKSİZ
(511221013)**

Department of Climate Science and Meteorological Engineering

Atmospheric Science Programme

Thesis Advisor: Prof. Dr. Hüseyin TOROS

MAY 2025

İTÜ, Lisansüstü Eğitim Enstitüsü'nün 511221013 numaralı Yüksek Lisans Öğrencisi Türkan ZENGİN GÖMLEKSİZ, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "GAZİANTEP'TE PM_{2.5} KONSANTRASYONUNUN ZAMANSAL VE MEKÂNSAL TAHMİNİNE YÖNELİK TRANSFER ÖĞRENME DESTEKLİ HİBRİT YAPAY ZEKA MODELLERİ" başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Prof. Dr. Hüseyin TOROS**
İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Dr. Öğr. Üyesi Ceyhan KAHYA**
İstanbul Teknik Üniversitesi

Prof. Dr. Mete TAYANÇ
Marmara Üniversitesi

Teslim Tarihi : **25 Mart 2025**
Savunma Tarihi : **13 Mayıs 2025**





Aileme,



ÖNSÖZ

Tez çalışmam süresince bilgi ve deneyimiyle bana yol gösteren, sabrı ve desteğiyle her zaman yanımda olan değerli danışmanım Prof. Dr. Hüseyin TOROS'a; bu zorlu süreçte motivasyon kaynağım olan, anlayışı, sevgisi ve sonsuz desteğiyle yanımda duran kıymetli eşime; hayatım boyunca maddi ve manevi desteklerini hiçbir zaman esirgemeyen sevgili aileme en içten teşekkürlerimi sunarım.

Mayıs 2025

Türkan ZENGİN GÖMLEKSİZ



İÇİNDEKİLER

	<u>Sayfa</u>
ÖNSÖZ	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SEMBOLLER	xv
ÇİZELGE LİSTESİ	xviii
ŞEKİL LİSTESİ	xxi
ÖZET	xxiii
SUMMARY	xxvii
1. GİRİŞ	1
1.1 Tezin Amacı	3
1.2 Literatür Araştırması	4
1.2.1 PM _{2.5} tahminine yönelik makine öğrenmesi, derin öğrenme ve transfer öğrenme çalışmaları	4
1.2.2 Gaziantep ilini konu alan hava kirliliği çalışmaları	7
2. PM_{2.5} KİRLİLİĞİ VE METEOROLOJİK FAKTÖRLER	9
2.1 PM _{2.5} Kaynakları ve Oluşumu	9
2.2 PM _{2.5} ve Sağlık Üzerine Etkileri	11
2.3 Meteorolojik Faktörlerin PM _{2.5} Üzerindeki Etkisi	15
3. GAZİANTEP İLİ'NİN COĞRAFİK KOŞULLARI VE İKLİMİ	19
3.1 Gaziantep'in Coğrafi Koşulları	20
3.2 Gaziantep'in İklimi	21
4. HAVA KİRLİLİĞİ TAHMİNİNDE KULLANILAN YÖNTEMLER	25
4.1 Karar Ağacı (Decision Tree)	25
4.2 Rastgele Orman (Random Forest)	26
4.3 Gradyan Artırma (Gradient Boosting)	27
4.4 Hafif Gradyan Artırma Makinesi (LightGBM - Light Gradient Boosting Machine) Algoritması	28
4.5 Uzun Kısa Süreli Bellek (LSTM)	29
4.6 Geçitli Yinelemeli Birimler (GRU - Gated Recurrent Units)	30
4.7 Evrişimli Sinir Ağları (CNN - Convolutional Neural Networks)	31
4.8 Grafik Sinir Ağları (GNN - Graph Neural Networks)	32
4.9 Gizli Bağımlılık Faktörü (LDF - Latent Dependency Factor)	32
4.10 Performans Değerlendirme Metrikleri	33
5. VERİ VE YÖNTEM	35
5.1 Veriler Hakkında Bilgiler	35
5.1.1 PM _{2.5} sensör verisi	35
5.1.2 Meteorolojik veriler	39
5.1.3 Copernicus Atmosphere Monitoring Service (CAMS) yeniden analiz verileri	40
5.1.4 MODIS MAIAC Aerosol Optik Derinlik (AOD) verisi	44
5.1.5 Arazi kullanımı verisi	46
5.1.6 Nüfus yoğunluğu verisi	47
5.1.7 NDVI (Bitki Örtüsü İndeksi) verisi	48
5.1.8 Rakım verisi	49
5.1.9 Yol yoğunluğu verisi	49
5.1.10 Sanayi ağırlıklı mesafe verisi	51

5.2 Veri Ön İşleme	53
5.2.1 Eksik veri tamamlanması	53
5.2.1.1 Eksik sensör verilerinin tamamlanması	53
5.2.1.2 Eksik NDVI (Bitki Örtüsü İndeksi) verilerinin tamamlanması	57
5.2.2 Aykırı veri tespiti	57
5.2.3 Özellik seçimi	61
5.2.4 Veri normalizasyonu	63
5.3 Yöntem	64
5.3.1 Zamansal modelleme	65
5.3.1.1 Makine öğrenmesi modelleri	65
5.3.1.2 Derin öğrenme modelleri	68
5.3.2 Mekânsal modelleme	75
5.3.2.1 Makine öğrenmesi modelleri	76
5.3.2.2 Mekânsal regresyon ve grafik tabanlı modeller	78
5.3.3 Zamansal modelden elde edilen verinin mekânsal modelde kullanımı ..	81
5.3.4 Transfer learning destekli hibrit modelleme	87
5.3.4.1 Autoencoder modeli ile gizli bağımlılık faktörü çıkarımı (LDF) ...	87
5.3.4.2 Transfer learning ile veri tamamlama	88
5.3.4.3 Transfer learning ile PM _{2.5} tahmini	89
6. SONUÇ VE ÖNERİLER	93
KAYNAKLAR	97
ÖZGEÇMİŞ	105

KISALTMALAR

AOD	: Aerosol Optik Derinlik
ARIMA	: AutoRegressive Integrated Moving Average
Bi-LSTM	: Bidirectional Long Short-Term Memory
CAMS	: Copernicus Atmosphere Monitoring Service
CNN	: Convolutional Neural Networks (Evrişimli Sinir Ağları)
CSB	: Coğrafi Bilgi Sistemleri
DANN	: Domain Adversarial Neural Network
DL	: Deep Learning (Derin Öğrenme)
ECMWF	: European Centre for Medium-Range Weather Forecasts
EPA	: Environmental Protection Agency (Çevre Koruma Ajansı)
ERA	: ECMWF Reanalysis (ECMWF Yeniden Analiz)
EU	: European Union (Avrupa Birliği)
GASKİ	: Gaziantep Su ve Kanalizasyon İdaresi
GCN	: Graph Convolutional Network (Grafik Evrişim Ağı)
GEE	: Google Earth Engine
GHSL	: Global Human Settlement Layer
GNN	: Graph Neural Networks (Grafik Sinir Ağları)
GRU	: Gated Recurrent Unit (Kapılı Tekrarlayan Birim)
IFS	: Integrated Forecasting System (Entegre Tahmin Sistemi)
LDF	: Latent Dependency Factor (Gizli Bağımlılık Faktörü)
LSTM	: Long Short-Term Memory (Uzun Kısa Vadeli Bellek)
MAE	: Mean Absolute Error (Ortalama Mutlak Hata)
MGM	: Meteoroloji Genel Müdürlüğü
ML	: Machine Learning (Makine Öğrenmesi)
MMD	: Maximum Mean Discrepancy (Maksimum Ortalama Farklılık)
OSB	: Organize Sanayi Bölgesi
PM	: Partiküler Madde
PMF	: Pozitif Matris Faktörizasyonu
RMSE	: Root Mean Squared Error (Kök Ortalama Kare Hata)
RNN	: Recurrent Neural Network (Tekrarlayan Sinir Ağı)
SİM	: Sürekli İzleme Merkezi
SO₂	: Kükürt Dioksit
SVR	: Support Vector Regression (Destek Vektör Regresyonu)
TL	: Transfer Learning (Transfer Öğrenme)
UHKİA	: Ulusal Hava Kalitesi İzleme Ağı
WHO	: World Health Organization (Dünya Sağlık Örgütü)



SEMBOLLER

T	: Sıcaklık (°C)
RH	: Bağıl Nem (%)
WS	: Rüzgar Hızı (m/s)
WD	: Rüzgar Yönü (°)
P	: Basınç (hPa)
t	: Zaman (gün)
x, y	: Konumsal Koordinatlar (enlem, boylam)
z	: Yükseklik
N	: Veri sayısı
μ	: Ortalama değer
σ	: Standart sapma
R^2	: Belirlilik katsayısı
PM_{2.5}	: İnce partikül madde (2.5 mikron altı)
PM₁₀	: İnce partikül madde (10 mikron altı)



ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 2.1: PM _{2.5} için DSÖ, AB, Ulusal ve EPA limit değerleri ($\mu\text{g}/\text{m}^3$ cinsinden).	13
Çizelge 5.1: Atapark istasyonuna ait değişkenlerin istatistiksel özeti (01.01.2021–31.12.2023).....	36
Çizelge 5.2: Beydilli istasyonuna ait değişkenlerin istatistiksel özeti (01.01.2021–31.12.2023).....	36
Çizelge 5.3: Nizip istasyonuna ait değişkenlerin istatistiksel özeti (01.01.2021–31.12.2023).....	37
Çizelge 5.4: Fevzi Çakmak istasyonuna ait değişkenlerin istatistiksel özeti (01.01.2021–31.12.2023).....	37
Çizelge 5.5: Meteoroloji istasyonuna ait değişkenlerin istatistiksel özeti (01.01.2021–31.12.2023).....	37
Çizelge 5.6: GASKİ istasyonuna ait değişkenlerin istatistiksel özeti (01.01.2021–31.12.2023).....	38
Çizelge 5.7: Gaziantep’te yer alan UHKİA istasyonlarına ait konum bilgileri. ...	38
Çizelge 5.8: İstasyonlara ait yapısal alan, tarım alanı, orman ve su kütlesi yüzdeleri.....	46
Çizelge 5.9: PM _{2.5} istasyonlarına ait 1 km ² alandaki nüfus yoğunluğu değerleri (kişi/km ²).....	47
Çizelge 5.10: İstasyonlara ait 5 km yarıçaplı bölgelerde hesaplanan yol yoğunluğu değerleri.....	50
Çizelge 5.11: İstasyonlara ait PM _{2.5} verilerindeki eksik gözlem sayısı ve oranları.	54
Çizelge 5.12: İstasyonlara göre aykırı değerlerin sayısı.	60
Çizelge 5.13: XGBoost modelinin istasyon bazlı performans metrikleri.	66
Çizelge 5.14: Random Forest modelinin istasyon bazlı performans metrikleri.	66
Çizelge 5.15: Decision Tree modelinin istasyon bazlı performans metrikleri.....	66
Çizelge 5.16: LightGBM modelinin istasyon bazlı performans metrikleri.	67
Çizelge 5.17: LSTM ve GRU modellerinin performans karşılaştırması.	69
Çizelge 5.18: Zamansal modelleme sonuçları: tüm modellerin performans karşılaştırması.	70
Çizelge 5.19: XGBoost mekânsal model performans sonuçları.....	76
Çizelge 5.20: Random forest mekânsal model performans sonuçları.....	76
Çizelge 5.21: LightGBM mekânsal model performans sonuçları.....	77
Çizelge 5.22: GWR modeli ile hesaplanan istasyon bazlı R ² değerleri.....	79
Çizelge 5.23: GCN modeli performans sonuçları (mekânsal modelleme).....	80
Çizelge 5.24: Mekânsal modelleme sonuçlarının karşılaştırılması.	80

Çizelge 5.25: Özellik entegrasyonu öncesi ve sonrası R^2 performans sonuçları. ...	82
Çizelge 5.26: Transfer learning modeli performans sonuçları	90



ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 2.1: Partikül Madde Boyutlarının Karşılaştırılması	9
Şekil 2.2: Türkiye'nin en yüksek PM _{2.5} seviyelerine sahip 10 ili (2023 verilerine göre)	14
Şekil 2.3: Türkiye'de 2022 Yılında Hava Kirliliğine Atfedilen Ölümün Oransal Olarak En Fazla Olduğu İlk 10 İli	15
Şekil 3.1: Gaziantep haritası kısa açıklama.....	19
Şekil 3.2: Gaziantep haritası kısa açıklama.....	20
Şekil 3.3: Gaziantep ilinin fiziki haritası.....	21
Şekil 3.4: Gaziantep ilinde 1940–2024 yılları arasında gözlemlenen rüzgar yönlerinin dağılımını gösteren rüzgar diyagramı.....	22
Şekil 4.1: Bir karar ağacının yapısı: kök düğüm, iç düğümler ve yaprak düğümler.	26
Şekil 4.2: Rastgele orman (Random Forest) algoritmasının temel çalışma prensibi.	27
Şekil 4.3: XGBoost ve LightGBM algoritmasının temel işleyişi.	29
Şekil 4.4: Temel Tekrarlayan Sinir Ağı (RNN) hücresinin yapısı.	30
Şekil 4.5: LSTM hücresinin detaylı yapısı: unutma, giriş ve çıkış kapıları ile hücre durumu.	30
Şekil 4.6: Bir Konvolüsyonel Sinir Ağı (CNN) mimarisinin temel yapısı: öznetelik çıkarımı (evrişim ve havuzlama katmanları) ve sınıflandırma (tam bağlantılı katmanlar).	31
Şekil 4.7: GNN yapısının basit gösterimi.	32
Şekil 5.1: Çalışmada izlenen veri analizi ve modelleme süreci.	35
Şekil 5.2: Hava kalitesi ölçüm istasyonlarının konumları.	38
Şekil 5.3: 2021–2023 yılları arasında Atapark istasyonuna ait ortalama meteorolojik değişkenlerin zamansal değişimi.	41
Şekil 5.4: 2021–2023 yılları arasında Beydilli istasyonuna ait ortalama meteorolojik değişkenlerin zamansal değişimi.	42
Şekil 5.5: 2021–2023 yılları arasında Nizip istasyonuna ait ortalama meteorolojik değişkenlerin zamansal değişimi.	43
Şekil 5.6: MODIS uydu verileri kullanılarak oluşturulan Gaziantep ve çevresine ait ortalama AOD (Aerosol Optical Depth) dağılım haritası.	45
Şekil 5.7: Atapark, Beydilli ve Nizip istasyonlarının 5 km çevresindeki yol ağlarının mekânsal dağılımı.	50
Şekil 5.8: Fevzi Çakmak, Meteoroloji ve GASKİ istasyonlarının 5 km çevresindeki yol ağlarının mekânsal dağılımı.	50

Şekil 5.9:	Atapark, Beydilli ve Nizip istasyonlarının en yakın OSB'ler arasındaki mekânsal mesafelerin görselleştirilmesi.	52
Şekil 5.10:	Fevzi Çakmak, Meteoroloji ve GASKİ istasyonlarının en yakın OSB'ler arasındaki mekânsal mesafelerin görselleştirilmesi.	53
Şekil 5.11:	Atapark istasyonu için eksik ve tamamlanmış PM _{2.5} verilerinin zaman serisi gösterimi.	55
Şekil 5.12:	Beydilli istasyonu için eksik ve tamamlanmış PM _{2.5} verilerinin zaman serisi gösterimi.	55
Şekil 5.13:	Fevzi Çakmak istasyonu için eksik ve tamamlanmış PM _{2.5} verilerinin zaman serisi gösterimi.	56
Şekil 5.14:	Nizip istasyonu için eksik ve tamamlanmış PM _{2.5} verilerinin zaman serisi gösterimi.	56
Şekil 5.15:	Atapark istasyonunda tamamlanan PM _{2.5} verilerinin histogram analizi.	57
Şekil 5.16:	Beydilli istasyonunda tamamlanan PM _{2.5} verilerinin histogram analizi.	58
Şekil 5.17:	Fevzi Çakmak istasyonunda tamamlanan PM _{2.5} verilerinin histogram analizi.	59
Şekil 5.18:	Nizip istasyonunda tamamlanan PM _{2.5} verilerinin histogram analizi.	60
Şekil 5.19:	Farklı istasyonlar için PM _{2.5} zaman serisi (2021–2023).	61
Şekil 5.20:	Atapark İstasyonu için değişkenler arası Pearson korelasyon matrisi.	62
Şekil 5.21:	Beydilli İstasyonu için değişkenler arası Pearson korelasyon matrisi.	63
Şekil 5.22:	Nizip İstasyonu için değişkenler arası Pearson korelasyon matrisi.	64
Şekil 5.23:	Zamansal modelleme için LightGBM ile elde edilen özellik önem skoru.	68
Şekil 5.24:	Atapark istasyonu için gerçek ve LightGBM modeli ile tahmin edilen PM _{2.5} zaman serisi.	71
Şekil 5.25:	Beydilli istasyonu için gerçek ve LightGBM modeli ile tahmin edilen PM _{2.5} zaman serisi.	71
Şekil 5.26:	Nizip istasyonu için gerçek ve LightGBM modeli ile tahmin edilen PM _{2.5} zaman serisi.	72
Şekil 5.27:	Atapark istasyonu için gerçek ve tahmin edilen PM _{2.5} değerlerinin saçılım grafiği.	72
Şekil 5.28:	Beydilli istasyonu için gerçek ve tahmin edilen PM _{2.5} değerlerinin saçılım grafiği.	73
Şekil 5.29:	Nizip istasyonu için gerçek ve tahmin edilen PM _{2.5} değerlerinin saçılım grafiği.	73
Şekil 5.30:	Atapark istasyonu için LightGBM modeline ait hata dağılımı (residual) yoğunluk grafiği.	74
Şekil 5.31:	Beydilli istasyonu için LightGBM modeline ait hata dağılımı (residual) yoğunluk grafiği.	74
Şekil 5.32:	Nizip istasyonu için LightGBM modeline ait hata dağılımı (residual) yoğunluk grafiği.	75

Şekil 5.33: Mekânsal modelleme için LightGBM ile elde edilen özellik önem skoru.....	78
Şekil 5.34: Atapark istasyonu için gerçek ve tahmin $PM_{2.5}$ değerlerinin karşılaştırılması (Mekânsal Model).....	83
Şekil 5.35: Beydilli istasyonu için gerçek ve tahmin $PM_{2.5}$ değerlerinin karşılaştırılması (Mekânsal Model).....	83
Şekil 5.36: Nizip istasyonu için gerçek ve tahmin $PM_{2.5}$ değerlerinin karşılaştırılması (Mekânsal Model).....	84
Şekil 5.37: Gerçek vs. tahmin saçılım grafiği - Atapark (Mekânsal).....	84
Şekil 5.38: Gerçek vs. tahmin saçılım grafiği - Beydilli (Mekânsal).....	85
Şekil 5.39: Gerçek vs. tahmin saçılım grafiği - Nizip (Mekânsal).....	86
Şekil 5.40: Atapark istasyonu için density residual (hata dağılımı) Grafiği.....	87
Şekil 5.41: Beydilli istasyonu için density residual (hata dağılımı) Grafiği.....	88
Şekil 5.42: Nizip istasyonu için density residual (hata dağılımı) Grafiği.....	89
Şekil 5.43: Tahmin edilen nihai $PM_{2.5}$ ortalama dağılım haritası.....	91



**GAZİANTEP’TE PM_{2.5} KONSANTRASYONUNUN ZAMANSAL VE
MEKÂNSAL
TAHMİNİNE YÖNELİK TRANSFER ÖĞRENME DESTEKLİ
HİBRİT YAPAY ZEKA MODELLERİ**

ÖZET

Hava kirliliği, günümüzde çevresel sürdürülebilirlik ve insan sağlığı açısından en önemli küresel sorunlardan biri olarak kabul edilmektedir. Sanayi tesislerinden yayılan emisyonlar, araç trafiğinin yoğunluğu ve evlerdeki enerji tüketimi gibi insan faaliyetleri, atmosferde zararlı gazların ve partiküllerin birikmesine yol açarak ekosistemlerin dengesini bozmakta ve ciddi sağlık sorunlarını tetiklemektedir. Bu bağlamda, çapı 2.5 mikrometreden küçük olan PM_{2.5} partikülleri, solunum yoluyla akciğerlere kolayca ulaşabilmesi ve toksik etkileri nedeniyle hava kirliliği araştırmalarında özel bir öneme sahiptir. Türkiye’nin Güneydoğu Anadolu Bölgesi’nde yer alan Gaziantep, sanayi ve ticaret alanındaki lider konumu, hızlı kentleşme süreci ve artan nüfusuyla hava kirliliği riskinin yüksek olduğu bir şehir olarak öne çıkmaktadır. Bu tez çalışması, Gaziantep ilinde PM_{2.5} konsantrasyonlarının zamansal ve mekânsal dağılımını tahmin etmek amacıyla yenilikçi bir yaklaşım geliştirmeyi amaçlamış; makine öğrenmesi, derin öğrenme ve transfer öğrenme tekniklerini birleştirerek hava kalitesi yönetiminde veri odaklı bir çözüm sunmuştur.

Çalışmanın temel hedefi, Gaziantep’teki PM_{2.5} kirliliğinin zaman içindeki değişimlerini ve coğrafi yayılımını anlamak, bu verilere dayanarak güvenilir tahmin modelleri oluşturmak ve eksik veri problemini çözerek PM_{2.5} ölçümü yapılmayan bölgeler için kirlilik tahminleri yapmaktır. Bu amaçla, Gaziantep’teki altı hava kalitesi izleme istasyonu (Atapark, Beydilli, Nizip, Fevzi Çakmak, GASKİ ve Meteoroloji) temel alınmış; 01.01.2021 - 31.12.2023 tarihleri arasındaki günlük veriler analiz edilmiştir. Ancak, GASKİ ve Meteoroloji istasyonlarında PM_{2.5} ölçümleri bulunmaması, bu bölgeler için alternatif bir tahmin yöntemi gerektirmiştir. Çalışma, veri toplama, ön işleme, modelleme ve değerlendirme olmak üzere dört ana aşamada yürütülmüştür. Veri kaynakları arasında, T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı’nın Ulusal Hava Kalitesi İzleme Ağı’ndan elde edilen PM_{2.5} ölçümleri, Open-Meteo’nun sağladığı meteorolojik veriler (sıcaklık, bağıl nem, rüzgar hızı vb.), Copernicus Atmosphere Monitoring Service (CAMS) yeniden analiz verileri, MODIS Aerosol Optik Derinlik (AOD) verileri, CORINE arazi kullanımı verileri, Global Human Settlement Layer’dan nüfus yoğunluğu verileri ve Sentinel-2’den NDVI (bitki örtüsü indeksi) verileri yer almıştır. Eksik veriler, Zincirleme Denklemlerle Çoklu Atama (MICE) yöntemiyle tamamlanmış; aykırı değerler ise Çeyrekler Arası Açıklık (IQR) yaklaşımıyla tespit edilip işlenmiştir.

Modelleme süreci, zamansal ve mekânsal analizleri birleştiren çok katmanlı bir strateji izlemiştir. Zamansal tahminler için, geçmiş $PM_{2.5}$ değerleri, hareketli ortalamalar ve meteorolojik parametreler kullanılarak XGBoost, LightGBM, Random Forest, Decision Tree, LSTM ve GRU modelleri test edilmiştir. Hiperparametre optimizasyonu, Optuna aracıyla gerçekleştirilerek modellerin performansı en üst düzeye çıkarılmıştır. Mekânsal analizde, istasyonların coğrafi konumları, nüfus yoğunluğu, sanayi mesafesi ve arazi kullanımı gibi değişkenler dikkate alınmış; LightGBM, XGBoost, Random Forest'un yanı sıra Geographically Weighted Regression (GWR) ve Graph Convolutional Network (GCN) modelleri uygulanmıştır. Çalışmanın en yenilikçi yönü, zamansal modelden elde edilen tahminlerin mekânsal modele entegre edilmesi ve transfer öğrenme ile hibrit bir sistem geliştirilmesidir. Bu yaklaşım, hava kirliliği tahmininde genellikle ayrı ayrı ele alınan zamansal ve mekânsal boyutları birleştirerek, hem zaman içindeki değişimleri hem de coğrafi dağılımı eşzamanlı olarak modelleme imkânı sunmuştur.

Bu yenilikçi çerçevenin bir diğer önemli bileşeni, transfer öğrenme ile hibrit bir sistemin geliştirilmesidir. Transfer öğrenme, veri eksikliği gibi pratik zorlukları aşmada etkili bir çözüm sunmuş; özellikle GASKİ ve Meteoroloji istasyonlarında doğrudan $PM_{2.5}$ ölçüm verisi bulunmaması sorununu ele almıştır. Bu süreçte, Atapark, Beydilli, Nizip ve Fevzi Çakmak istasyonlarından elde edilen zengin veri seti, sensörsüz istasyonlara bilgi aktarımı için bir kaynak olarak kullanılmıştır. Transfer öğrenme, yalnızca eksik verilerin tamamlanmasını değil, aynı zamanda modellerin genelleme kapasitesini artırarak farklı coğrafi koşullara uyarlanabilirliğini sağlamıştır. Hibrit sistem, zamansal ve mekânsal modellerin güçlü yönlerini birleştirirken, transfer öğrenmenin bu entegrasyonu optimize etme yeteneği, çalışmanın metodolojik yeniliğini pekiştirmiştir. Bu yaklaşım, hava kirliliği tahmininde veri kısıtlarının üstesinden gelmek için pratik ve ölçeklenebilir bir yöntem olarak öne çıkmaktadır.

Autoencoder tabanlı Latent Dependency Factor (LDF) çıkarımı, bu hibrit sistemin temel taşlarından biri olarak eksik verilerin tamamlanmasında ve GASKİ ile Meteoroloji istasyonlarına bilgi aktarımında kritik bir rol oynamıştır. Autoencoder, ham verilerden sıkıştırılmış ve anlamlı temsiller (LDF) üreterek, $PM_{2.5}$ konsantrasyonlarıyla ilişkili gizli desenleri ortaya çıkarmıştır. Bu süreçte, meteorolojik veriler, mekânsal değişkenler ve mevcut $PM_{2.5}$ ölçümleri Autoencoder'a beslenmiş; model, bu karmaşık veri setinden düşük boyutlu ancak bilgi açısından zengin özellik vektörleri elde etmiştir. LDF'ler, özellikle GASKİ ve Meteoroloji istasyonları gibi sensör verisi olmayan bölgelerde $PM_{2.5}$ seviyelerini tahmin etmek için Inverse Distance Weighting (IDW) yöntemiyle birleştirilmiş; böylece, kaynak istasyonlardan hedef istasyonlara mesafeye dayalı ağırlıklı bir bilgi transferi gerçekleştirilmiştir. Bu yöntem, yerel ölçüm eksikliğini telafi ederken, aynı zamanda istasyonlar arasındaki mekânsal ilişkileri de dikkate alarak tahminlerin güvenilirliğini artırmıştır. LDF çıkarımının başarısı, veri setindeki gürültüyü azaltması ve modellerin öğrenme sürecini kolaylaştırmasıyla da kendini göstermiştir.

Performans değerlendirmesi, Ortalama Mutlak Hata (MAE), Kök Ortalama Kare Hata (RMSE) ve Belirlilik Katsayısı (R^2) metrikleriyle yapılmıştır. Bulgular, LightGBM modelinin diğer yöntemlere kıyasla üstün bir performans sergilediğini göstermiştir. Zamansal modellemede, Atapark istasyonunda $R^2 = 0.9224$, Beydilli'de $R^2 = 0.9560$

ve Nizip'te $R^2 = 0.9631$ gibi yüksek doğruluk oranları elde edilmiş; bu başarı, özellikle son 3 günün hareketli ortalaması ($PM_{2.5_RollingMean_3}$), bir önceki günün $PM_{2.5}$ değeri ($PM_{2.5_Lag_1}$) ve rüzgar hızı gibi değişkenlerin etkili kullanımına dayandırılmıştır. Mekânsal analizde ise LightGBM, Atapark'ta $R^2 = 0.876$, Beydilli'de $R^2 = 0.965$ ve Nizip'te $R^2 = 0.879$ değerleriyle en iyi sonuçları vermiş; nüfus yoğunluğu, yapısal alan yüzdesi ve CAMS $PM_{2.5}$ verileri kirlilik dağılımında en önemli faktörler olarak belirlenmiştir. Zamansal tahminlerin mekânsal modele entegrasyonu, özellikle Nizip'te R^2 değerini 0.879'dan 0.963'e yükselterek model doğruluğunu artırmıştır; ancak Beydilli'de R^2 'de hafif bir düşüş (0.965'ten 0.951'e) gözlenmiştir ki bu, yerel faktörlerin farklı etkilerinden kaynaklanabilir. Transfer öğrenme destekli hibrit yaklaşım, Fevzi Çakmak'ta $R^2 = 0.81$ sonuçlarıyla sensörsüz bölgelerde tahmin yapabilme kapasitesini kanıtlamıştır.

Gaziantep'teki $PM_{2.5}$ seviyeleri, istasyonlar arasında çarpıcı farklılıklar sergilemiştir. Beydilli, $35.4 \mu g/m^3$ ile en yüksek kirlilik seviyesine sahip bölge olarak tespit edilmiş; bu, yoğun kentleşme, yüksek nüfus yoğunluğu (398.2 kişi/ km^2) ve trafik kaynaklı emisyonlarla ilişkilendirilmiştir. Nizip, $22.9 \mu g/m^3$ ile en düşük seviyeyi göstermiş; kentsel alanlardan uzak konumu ve daha az sanayi etkinliği bu durumu açıklamaktadır. Atapark, $24.9 \mu g/m^3$ ile orta seviyede kirlilik sergilemiş; %49.68 ormanlık alan oranı, kirliliği bir ölçüde dengelemiştir. GASKİ ve Meteoroloji istasyonları için transfer öğrenme ile tahmin edilen değerler, kentsel dinamiklere yakınlık nedeniyle orta-yüksek seviyelerde çıkmıştır. Meteorolojik faktörlerin etkisi de analiz edilmiş; rüzgar hızı kirletici dağılımını belirleyen temel bir unsur olarak öne çıkarken, düşük hızlar $PM_{2.5}$ birikimini artırmıştır.

Çalışma, bazı sınırlılıklarla karşılaşmıştır. Veri seti, 2021-2023 dönemini kapsayan 3 yıllık bir zaman dilimiyle sınırlı kalmış; bu, uzun vadeli trendlerin tam olarak modellenmesini engellemiştir. GASKİ ve Meteoroloji istasyonlarında doğrudan $PM_{2.5}$ ölçümü olmaması, tahminlerin kesin doğruluğunu değerlendirme imkanını kısıtlamış; Fevzi Çakmak'taki başarı ($R^2 = 0.81$) bu tahminlerin güvenilirliğini desteklese de yerel farklılıklar belirsizlik yaratmıştır. AOD verileri, $PM_{2.5}$ ile negatif korelasyon (-0.58 'e kadar) göstererek güvenilir bir değişken olmaktan çıkmış; bu, uydu ölçümlerinin bölgesel meteorolojik koşullardan etkilendiğini düşündürmüştür. Derin öğrenme modelleri (LSTM ve GRU), veri miktarının yetersizliği nedeniyle LightGBM gibi makine öğrenmesi modellerinin gerisinde kalmış; Atapark'ta LSTM için $R^2 = 0.6184$ gibi düşük sonuçlar alınmıştır.

Bu çalışma, Gaziantep'te hava kirliliği yönetimine bilimsel bir katkı sunarken, veri odaklı yöntemlerin gücünü ortaya koymuştur. LightGBM'nin başarısı, zamansal ve mekânsal tahminlerde hızlı ve doğru sonuçlar üretebileceğini kanıtlamış; transfer öğrenme ise sensör eksikliği gibi pratik sorunlara yenilikçi bir çözüm getirmiştir. Bulgular, Beydilli gibi yüksek kirlilik bölgelerinde acil önlemlerin gerekliliğini vurgulamış; Nizip gibi daha temiz alanların korunmasının önemini göstermiştir. Çalışma, hava kalitesi izleme ağının genişletilmesi, uzun vadeli veri toplanması ve trafik-sanayi emisyonlarını azaltıcı politikalar gibi önerilerle tamamlanmıştır. Gelecekte, daha fazla istasyon verisiyle modellerin zenginleştirilmesi, diğer kirleticilerin analizi ve gerçek zamanlı tahmin sistemlerinin geliştirilmesi, bu çalışmanın etkisini artırabilir.



**SPATIO-TEMPORAL ESTIMATION OF PM_{2.5} CONCENTRATIONS
IN GAZIANTEP
USING TRANSFER LEARNING-BASED
HYBRID ARTIFICIAL INTELLIGENCE MODELS**

SUMMARY

Air pollution is now recognized as one of the most important global challenges for environmental sustainability and human health. Human activities such as emissions from industrial facilities, heavy vehicle traffic and energy consumption in households cause harmful gases and particles to accumulate in the atmosphere, disrupting the balance of ecosystems and triggering serious health problems. In this context, PM_{2.5} particles with a diameter of less than 2.5 micrometers have a special importance in air pollution research due to their easy access to the lungs through inhalation and their toxic effects. Gaziantep, located in the Southeastern Anatolia Region of Turkey, stands out as a city with a high risk of air pollution due to its leading position in industry and trade, rapid urbanization process and increasing population. This thesis aims to develop an innovative approach to predict the temporal and spatial distribution of PM_{2.5} concentrations in Gaziantep province, combining machine learning, deep learning and transfer learning techniques to provide a data-driven solution for air quality management.

The main objective of the study is to understand the changes in PM_{2.5} pollution in Gaziantep over time and its geographical distribution, to build reliable forecasting models based on these data, and to make pollution forecasts for regions where PM_{2.5} is not measured by solving the missing data problem. For this purpose, six air quality monitoring stations in Gaziantep (Atapark, Beydilli, Nizip, Fevzi Çakmak, GASKI and Meteorology) were taken as a basis and daily data between 01.01.2021 and 31.12.2023 were analyzed. However, the lack of PM_{2.5} measurements at GASKI and Meteorology stations necessitated an alternative forecasting method for these regions. The study was conducted in four main stages: data collection, pre-processing, modeling and evaluation. Data sources include PM_{2.5} measurements obtained from the National Air Quality Monitoring Network of the Ministry of Environment, Urbanization and Climate Change, meteorological data (temperature, relative humidity, wind speed, etc.) provided by Open-Meteo, Copernicus Atmosphere, Copernicus Atmosphere Monitoring Service (CAMS) reanalysis data, MODIS Aerosol Optical Depth (AOD) data, CORINE land use data, population density data from Global Human Settlement Layer and NDVI (vegetation cover index) data from Sentinel-2. Missing data were completed using Multiple Assignment by Chaining Equations (MICE), while outliers were identified and processed using the Interquartile Range (IQR) approach.

The modeling process followed a multi-layered strategy combining temporal and spatial analyses. For temporal forecasts, XGBoost, LightGBM, Random Forest, Decision Tree, LSTM and GRU models were tested using historical PM_{2.5} values, moving averages and meteorological parameters. Hyperparameter optimization was performed with Optuna tool to maximize the performance of the models. In the spatial analysis, variables such as geographical location of the stations, population density, industrial distance and land use were taken into account; LightGBM, XGBoost, Random Forest, Geographically Weighted Regression (GWR) and Graph Convolutional Network (GCN) models were applied. The most innovative aspect of the study is the integration of the predictions from the temporal model into the spatial model and the development of a hybrid system with transfer learning. This approach combines both temporal and spatial dimensions, which are usually considered separately in air pollution forecasting, and provides the opportunity to simultaneously model both changes over time and geographical distribution.

Another important component of this innovative framework is the development of a hybrid system with transfer learning. Transfer learning provided an effective solution to overcome practical challenges such as lack of data, particularly the lack of direct PM_{2.5} measurement data at GASKI and Meteorology stations. In this process, the rich dataset from Atapark, Beydilli, Nizip and Fevzi Çakmak stations was used as a source for transferring information to sensorless stations. Transfer learning enabled not only the completion of missing data but also the adaptability to different geographical conditions by increasing the generalization capacity of the models. The hybrid system combined the strengths of temporal and spatial models, while the ability of transfer learning to optimize this integration reinforced the methodological novelty of the study. This approach stands out as a practical and scalable method to overcome data constraints in air pollution forecasting.

Autoencoder-based Latent Domain Features (LDF) extraction, as one of the cornerstones of this hybrid system, played a critical role in completing missing data and transferring information to GASKI and Meteorological stations. Autoencoder generated compressed and meaningful representations (LDFs) from the raw data, revealing hidden patterns associated with PM_{2.5} concentrations. In this process, meteorological data, spatial variables and existing PM_{2.5} measurements were fed into Autoencoder, and the model extracted low-dimensional but information-rich feature vectors from this complex dataset. The LDFs were combined with Inverse Distance Weighting (IDW) to estimate PM_{2.5} levels, especially in areas without sensor data, such as GASKI and Meteorological stations, thus realizing a distance-based weighted information transfer from source stations to target stations. This method compensated for the lack of local measurements, while at the same time improving the reliability of the estimates by taking into account the spatial relationships between stations. The success of LDF extraction was also demonstrated by reducing noise in the dataset and facilitating the learning process of the models.

Performance evaluation was performed using the metrics of Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Coefficient of Determination (R^2). The results showed that the LightGBM model performed superiorly compared to the other methods. In temporal modeling, high accuracy rates such as $R^2 = 0.9224$ at Atapark

station, $R^2 = 0.9560$ at Beydilli and $R^2 = 0.9631$ at Nizip were obtained; this success was attributed to the effective use of variables such as moving average of the last 3 days (PM_{2.5}_RollingMean_3), PM_{2.5} value of the previous day (PM_{2.5}_Lag_1) and wind speed. In the spatial analysis, LightGBM gave the best results with $R^2 = 0.876$ in Atapark, $R^2 = 0.965$ in Beydilli and $R^2 = 0.879$ in Nizip; population density, percentage of built-up area and CAMS PM_{2.5} data were determined as the most important factors in pollution distribution. The integration of temporal estimates into the spatial model improved the model accuracy, especially in Nizip, increasing the R^2 value from 0.879 to 0.963; however, a slight decrease in R^2 was observed in Beydilli (from 0.965 to 0.951), which may be due to the different effects of local factors. The transfer learning supported hybrid approach proved its capability to forecast in sensorless areas with $R^2 = 0.82$ for GASKI, $R^2 = 0.84$ for Meteorology and $R^2 = 0.81$ for Fevzi Çakmak.

PM_{2.5} levels in Gaziantep showed striking differences between stations. Beydilli was found to have the highest pollution level at 35.4 $\mu\text{g}/\text{m}^3$, which is associated with dense urbanization, high population density (398.2 inhabitants/ km^2) and emissions from traffic. Nizip showed the lowest level at 22.9 $\mu\text{g}/\text{m}^3$, which is explained by its location away from urban areas and less industrial activity. Atapark showed a moderate level of pollution with 24.9 $\mu\text{g}/\text{m}^3$; 49.68% forest cover stabilized the pollution to some extent. The values estimated by transfer learning for GASKI and Meteorology stations were at medium-high levels due to proximity to urban dynamics. The impact of meteorological factors was also analyzed; wind speed was a key determinant of pollutant dispersion, with low wind speeds increasing PM_{2.5} deposition. Southerly winds affected pollution by increasing desert dust transport from Syria and Iraq, while relative humidity changed the suspension time of particles in the atmosphere.

The study faced some limitations. The dataset was limited to a 3-year timeframe covering the period 2021–2023, which prevented full modeling of long-term trends. The lack of direct PM_{2.5} measurements at GASKI and Meteorology stations limited the ability to assess the accuracy of the forecasts; local variations created uncertainty, although the success at Fevzi Çakmak ($R^2 = 0.81$) supports the reliability of these forecasts. AOD data were negatively correlated (up to -0.58) with PM_{2.5}, suggesting that satellite measurements are influenced by regional meteorological conditions. Deep learning models (LSTM and GRU) lagged behind machine learning models such as LightGBM due to insufficient amount of data; low results such as $R^2 = 0.6184$ for LSTM in Atapark were obtained.

This study provides a scientific contribution to air pollution management in Gaziantep and demonstrates the power of data-driven methods. The success of LightGBM proved that it can produce fast and accurate temporal and spatial forecasts, while transfer learning provided an innovative solution to practical problems such as lack of sensors. The findings emphasized the need for urgent measures in high pollution areas such as Beydilli and showed the importance of protecting cleaner areas such as Nizip. The study concludes with recommendations for expanding the air quality monitoring network, long-term data collection and policies to reduce traffic-industrial emissions. In the future, enriching the models with more station data, analyzing other pollutants and developing real-time forecasting systems can increase the impact of this study.

The combination of scientific and policy actions to improve Gaziantep's air quality will bring both environmental and societal benefits.



1. GİRİŞ

Hava kirliliği; endüstriyel faaliyetler, araç emisyonları, evsel yakıt kullanımı ve doğal süreçler gibi birçok kaynaktan atmosfere salınan zararlı gaz ve partiküllerin birikmesi sonucu oluşur. İnsan sağlığını, çevresel dengeyi ve ekosistemlerin sürdürülebilirliğini ciddi ölçüde etkileyen hava kirliliği, giderek büyüyen küresel bir sorundur. Hava kirliliğine neden olan temel kirleticiler arasında partiküler maddeler, karbon monoksit, kükürt oksitleri, nitrojen oksitleri ve hidrokarbonlar yer alır [1]. Bu kirleticiler, genellikle fabrika işlemleri gibi sabit kaynaklardan ya da taşıtlar gibi hareketli kaynaklardan ortaya çıkar ve atmosfere yayılır [2]. Atmosfere zararlı maddelerin salınmasıyla oluşan hava kirliliği, insan sağlığı ve çevre üzerinde önemli etkilere sahiptir. Hava kirliliği, farklı çevresel ortamlardaki ekosistemleri ve biyolojik çeşitliliği önemli ölçüde etkilemektedir. Kükürt dioksit, azot oksitler, ozon ve cıva gibi kirleticiler hem sucul hem de karasal ekosistemler üzerinde olumsuz etkiler yaratmaktadır [3]. Bu kirleticiler su kaynaklarının asitleşmesine, kıyı bölgelerinin ötrofikasyonuna ve besin zincirlerinde biyolojik birikime yol açmaktadır [3]. Karasal ekosistemlerde ise hava kirliliği, biyokimyasal döngüleri bozmakta, bitkilerin büyümesini ve üreme kapasitelerini olumsuz yönde etkilemektedir [4]. Bu etkiler, bireysel türler üzerindeki hafif düzeydeki bozulmalardan, ciddi vakalarda görülen hastalık ve ölümlere kadar geniş bir yelpazede gerçekleşebilir [4]. Ayrıca hava kirliliği, küresel ısınmayı ve ozon tabakasının incelmelerini şiddetlendirerek, ekosistemlerin dengesini bozacak düzeyde iklim değişikliklerine neden olmaktadır [5]. Etkiler, doğrudan zarar gören canlı türleriyle sınırlı kalmamakta, besin ağları ve ekosistem hizmetlerini de kapsamaktadır [6]. Bu konuda hâlen bilgi eksiklikleri olsa da, hava kirliliğinin biyolojik çeşitlilik ve ekosistem işleyişindeki yaygın etkileri, uzun vadeli koruma stratejilerinde dikkate alınmasını zorunlu kılmaktadır [3]. Hava kirliliğinin kontrolüne yönelik politikalar zaman içinde önemli değişiklikler göstermiş ve kirlilik kaynaklarıyla mücadele yöntemlerine göre çeşitlenmiştir. Hava kirliliği kaynakları; endüstri, ulaşım ve evsel enerji kullanımı

gibi farklı sektörlerden oluşmakta, bu sektörlerle yönelik geliştirilen politikalar ise teşvik edici, destekleyici ya da caydırıcı önlemler içermektedir [7]. Özellikle ekonomik teşvikler, destekleyici tedbirler ve cezai yaptırımlar sıklıkla kullanılan araçlar arasında yer almaktadır [7]. Ayrıca Kyoto Protokolü ve Montreal Protokolü gibi uluslararası anlaşmalar, hava kirliliğinin küresel ölçekte kontrol edilmesinde kritik bir rol oynamaktadır [8]. Türkiye özelinde gerçekleştirilen hava kalitesi araştırmaları, ülke genelindeki kirlilik eğilimlerini ortaya koymanın yanında mevcut yönetmeliklerin etkinliğini değerlendirmeyi amaçlamaktadır. Türkiye, 2019 yılı itibarıyla hava kalitesi standartlarını Avrupa Birliği direktiflerine uygun hale getirerek uyum sürecini hızlandırmıştır [9]. Son yıllarda partikül madde (PM₁₀) ve kükürt dioksit (SO₂) konsantrasyonlarında genel anlamda azalma eğilimi gözlenmekle birlikte bazı bölgelerde kirlilik düzeylerinde artışların sürdüğü belirlenmiştir. 2019 yılında Türkiye'nin 81 ilinden 49'unda PM₁₀ konsantrasyonları yüksek bulunmuş ve birçok ölçüm istasyonunda günlük sınır değerler aşılmıştır [10]. Bu durumun ortaya çıkmasında coğrafi konum, meteorolojik koşullar ve yerel emisyon kaynaklarının etkisi büyüktür [9]. Bazı şehirlerde hava kalitesinde belirgin iyileşmeler kaydedilirken, diğer bölgelerde kirlilik düzeyinin artış göstermesi, Türkiye'de hava kirliliğiyle mücadelede daha etkin ve kapsamlı politikalara duyulan ihtiyacı açıkça ortaya koymaktadır [11]. Hava kalitesinin değerlendirilmesinde sıklıkla ele alınan PM₁₀ ve SO₂ gibi kirleticilerin yanı sıra, özellikle insan sağlığı üzerindeki olumsuz etkileri nedeniyle son yıllarda önemi giderek artan partikül madde (PM_{2.5}), hava kirliliği çalışmalarının odak noktası hâline gelmiştir. PM_{2.5}, aerodinamik çapı 2.5 mikrometreden (μm) küçük olan ince partikül maddeleri ifade etmektedir [12] [13]. Bu ince partiküller uzun süre havada asılı kalabilmekte ve atmosferik koşullar nedeniyle uzun mesafelere taşınabilmektedir [13]. PM_{2.5}, hava kirliliğinin önemli bir göstergesi olup, özellikle Asya ülkelerindeki kentsel alanlarda sıklıkla Avrupa şehirlerine kıyasla daha yüksek konsantrasyonlarda bulunmaktadır [13]. PM_{2.5} maruziyeti, küresel çapta milyonlarca ölüme yol açan çeşitli solunum ve kardiyovasküler hastalıklara neden olmaktadır [12]. Başta taşıt emisyonları ve endüstriyel faaliyetler olmak üzere, büyük ölçüde insan kaynaklı faaliyetlerle ortaya çıkan PM_{2.5}, PM₁₀ gibi daha iri partiküllere kıyasla daha yüksek toksik etki potansiyeline sahiptir. Bu nedenle çevre otoriteleri tarafından

belirlenen $PM_{2.5}$ ölçüm standartları, insanların solunum sistemi üzerindeki etkileri daha iyi yansıtacak şekilde oluşturulmuştur [14]. $PM_{2.5}$ kirliliği, atmosferdeki yayılımı ve taşınım mekanizmaları açısından karmaşık bir yapıya sahiptir. Bölgesel taşınım süreçleri, $PM_{2.5}$ konsantrasyonlarını önemli ölçüde etkileyerek yerel kaynakların yanı sıra uzak mesafelerden taşınan kirleticilerin de değerlendirilmesini gerekli kılmaktadır [15] [16]. Ayrıca, $PM_{2.5}$ seviyeleri mevsimsel farklılıklar ve meteorolojik faktörlerden doğrudan etkilenmektedir [17]. Kış aylarında atmosferik sınır tabakasındaki değişimler ve inversiyon olayları, $PM_{2.5}$ 'in yoğunlaşmasına neden olabilmektedir. $PM_{2.5}$ 'in atmosferdeki hareketliliği ve meteorolojik değişkenlerle olan ilişkisi ilerleyen bölümlerde daha ayrıntılı bir şekilde ele alınacaktır.

1.1 Tezin Amacı

Hava kirliliği, günümüzde çevresel sürdürülebilirlik ve halk sağlığı açısından en önemli küresel sorunlardan biri olarak kabul edilmektedir. Özellikle $PM_{2.5}$ gibi ince partiküler maddelerin hava kalitesi üzerindeki etkileri, sanayileşme ve şehirleşme ile giderek artmaktadır. Gaziantep gibi sanayi ve trafik yoğunluğu yüksek şehirlerde, bu kirleticilerin zamansal ve mekânsal dağılımını anlamak, kirlilik yönetimi açısından kritik bir gereklilik hâline gelmiştir. Çalışmanın temel amacı, Gaziantep ilinde $PM_{2.5}$ kirliliğinin zamansal ve mekânsal tahminini gerçekleştirmek ve hava kirliliği tahmini için makine öğrenmesi (ML), derin öğrenme (DL) ve transfer öğrenme (TL) tekniklerinin etkinliğini incelemektir. Bu doğrultuda, hava kirliliğinin mevcut seviyelerini anlamak, hava kalitesini etkileyen faktörleri belirlemek ve tahmin modelleri geliştirerek $PM_{2.5}$ konsantrasyonlarının gelecekteki değişimlerini öngörmek hedeflenmektedir. Çalışmada hava kirliliği modelleme süreçlerinde geleneksel yöntemlerin yanı sıra modern yapay zeka tabanlı tekniklerin kullanımını incelenerek, özellikle transfer öğrenme yaklaşımlarının tahmin doğruluğunu nasıl artırabileceğini analiz edilmektedir. Gaziantep gibi sanayi ve trafik yoğunluğu yüksek bir şehirde, $PM_{2.5}$ kirliliğinin etkilerini en aza indirmek için uygulanabilecek veri odaklı stratejiler geliştirilmesi amaçlanmaktadır. Bu bağlamda, çalışma kapsamında hava kirliliği verileri toplanacak, eksik veriler doldurulacak, çeşitli makine öğrenmesi ve derin öğrenme modelleri eğitilecek ve performansları değerlendirilecektir. Buna ek olarak

çalışmada transfer öğrenme yaklaşımı sayesinde, hiç doğrudan PM 2.5 ölçümü yapılmayan istasyonlar için de tahmin yapılabilmektedir. Bu yöntem, özellikle GASKİ ve Meteoroloji istasyonları gibi verisi bulunmayan noktalar için hava kirliliği tahmini yapmayı mümkün kılmaktadır. Bu özellikler sayesinde PM_{2.5} seviyelerinin daha doğru izlenmesi ve modellenmesi mümkün olacak, geniş ölçekli hava kalitesi yönetimi için önemli bilimsel veriler sağlanacaktır. Böylece elde edilen bulgularla hava kirliliği kontrolüne yönelik politikaların ve yönetim stratejilerinin geliştirilmesine bilimsel katkıda bulunulması hedeflenmektedir.

1.2 Literatür Araştırması

Hava kalitesinin tahmin edilmesi, etkili çevre yönetimi ve insan sağlığının korunması açısından giderek önem kazanmaktadır. Son yıllarda geleneksel yöntemlerin yanı sıra makine öğrenimi, derin öğrenme ve transfer öğrenme gibi yöntemlerin kullanımı önemli ölçüde artmıştır. Bu yöntemler, hava kalitesi verilerindeki karmaşık ve doğrusal olmayan ilişkileri yakalayabilme potansiyelleri sayesinde, tahminlerin doğruluğunu artırmada önemli rol oynamaktadır.

1.2.1 PM_{2.5} tahminine yönelik makine öğrenmesi, derin öğrenme ve transfer öğrenme çalışmaları

Erken dönem hava kirliliği tahmin çalışmaları genellikle zaman serisi analizleri (ARIMA, regresyon modelleri) gibi istatistiksel yöntemlere dayanmıştır [18]. Ancak, bu yöntemler meteorolojik ve çevresel faktörlerin değişkenliğini yeterince modelleyemediği için günümüzde makine öğrenmesi ve derin öğrenme teknikleri daha sık tercih edilmektedir [19]. Makine öğrenmesi (ML) teknikleri, hava kirliliği tahminlerinde geniş ölçekli veri kümelerini işleyerek yüksek doğruluk oranları elde etmeyi mümkün kılmaktadır. Literatürde PM_{2.5} tahmini için kullanılan makine öğrenmesi yaklaşımlarını ele alan birçok çalışma bulunmaktadır. Mahalakshmi ve diğ. (2022), PM_{2.5} konsantrasyonlarını tahmin etmek amacıyla farklı makine öğrenimi algoritmalarını karşılaştırmış, Random Forest, Support Vector Regression (SVR), Lineer Regresyon ve Random Ferns yöntemlerini değerlendirmiştir. Araştırma sonuçları, dört algoritmanın bir arada (ensemble) kullanılması durumunda model

doğruluğunun %82'ye kadar yükseldiğini ortaya koymuştur. Deters ve diğ. (2017) ise, meteorolojik parametreleri kullanarak kentsel alanlarda $PM_{2.5}$ kirliliğinin modellenmesinde makine öğrenmesi yöntemlerinin etkinliğini incelemiştir. Araştırma kapsamında sıcaklık, rüzgâr hızı ve bağıl nem gibi meteorolojik parametrelerin model performansını artırdığı ve özellikle aşırı meteorolojik olaylarda tahmin doğruluğunu yükselttiği sonucuna ulaşılmıştır. Kumar ve diğ. (2024) tarafından Delhi şehrinde gerçekleştirilen çalışmada, $PM_{2.5}$ seviyelerinin tahmin edilmesinde çeşitli makine öğrenimi modelleri test edilmiştir. Çalışmada Random Forest, SVR ve Lineer Regresyon gibi yöntemlerle karşılaştırılan modeller arasında, Lineer Regresyonun en yüksek doğruluğa sahip olduğu ve günlük tahminlerde etkili sonuçlar verdiği ortaya konmuştur. Moursi ve diğ. (2019), $PM_{2.5}$ tahmininde farklı algoritmaların performanslarını karşılaştırmalı olarak analiz etmiştir. Çalışmada SVR, Random Forest, LSTM ve Extra Trees modelleri değerlendirilmiş olup, Extra Trees algoritmasının diğer yöntemlere göre daha yüksek bir tahmin performansı sunduğu belirtilmiştir. Brahmanandam (2021), atmosferik $PM_{2.5}$ seviyelerinin tahmini için derin öğrenme yöntemlerini incelemiş, özellikle uzun vadeli tahminlerde LSTM algoritmasının başarılı sonuçlar sağladığını vurgulamıştır. Derin öğrenme (DL) modelleri, $PM_{2.5}$ tahmini için gelişmiş zaman serisi ve mekânsal analiz yetenekleri sayesinde büyük başarılar elde etmiştir. Literatürde farklı DL yaklaşımlarının başarılarını inceleyen çeşitli çalışmalar bulunmaktadır. Ayturan ve diğ. (2020), kısa vadeli $PM_{2.5}$ tahmininde GRU (Gated Recurrent Unit) ve RNN (Recurrent Neural Network) gibi derin öğrenme modellerini birlikte kullanarak %82'lik yüksek bir doğruluk elde etmişlerdir. Benzer bir çalışmada Muthukumar ve diğ. (2022), $PM_{2.5}$ seviyelerinin tahmininde meteorolojik, çevresel ve uydu görüntülerinden elde edilen çok kaynaklı verileri kullanarak Graph Convolutional Network (GCN) ve Convolutional Long Short-Term Memory (ConvLSTM) yöntemlerini uygulamışlardır. Elde edilen sonuçlara göre, çok boyutlu verilerin ve gelişmiş derin öğrenme yöntemlerinin birlikte kullanılması tahmin performansını önemli ölçüde artırmaktadır. Brahmanandam (2021), Pekin'deki $PM_{2.5}$ konsantrasyonlarını tahmin etmek için farklı derin öğrenme modellerini karşılaştırmış ve bu modellerin özellikle uzun dönemli tahminlerde etkili sonuçlar verdiğini vurgulamıştır. Deters ve diğ. (2017),

kentsel alanlarda PM_{2.5} tahminlerinde derin öğrenme modellerinin meteorolojik parametrelerle desteklendiğinde daha iyi sonuçlar sağladığını ortaya koymuştur. Ayrıca Jeya ve Sankari (2020), akıllı şehir uygulamaları kapsamında iki yönlü Uzun Kısa Süreli Bellek (Bidirectional Long Short-Term Memory - Bi-LSTM) yöntemini kullanarak PM_{2.5} konsantrasyonlarını başarıyla tahmin etmiş ve geleneksel modellere kıyasla daha iyi sonuçlar aldıklarını belirtmişlerdir. Bu çalışmalara ek olarak kullanılan Transfer öğrenme (TL), özellikle veri eksikliği bulunan istasyonlarda PM_{2.5} tahminini iyileştirmek için geliştirilen bir yöntemdir. Son yıllarda PM_{2.5} tahmininde transfer öğrenme tabanlı derin öğrenme yöntemlerinin kullanımı önem kazanmıştır. Ni ve diğ. (2022), PM_{2.5} konsantrasyonlarının tahmininde hibrit transfer öğrenme tabanlı bir derin öğrenme modeli geliştirmiştir. Çalışmada, sınırlı veriye sahip istasyonların tahmin performansını iyileştirmek amacıyla, Maximum Mean Discrepancy (MMD) yöntemiyle en uygun kaynak istasyon belirlenmiş ve Domain Adversarial Neural Network (DANN) kullanılarak kaynak istasyonların verileri hedef istasyonlara aktarılmıştır. Sonuçlar, bu yöntemle veri eksikliği olan istasyonlarda tahmin doğruluğunun önemli ölçüde arttığını göstermiştir. Yang ve diğ. (2023), PM_{2.5} konsantrasyonlarının saatlik tahmini için transfer öğrenme ile desteklenen hibrit bir derin öğrenme modeli sunmuştur. Çalışmada, karmaşık ve yüksek boyutlu veri kümelerini modellemek amacıyla Uzun Kısa Süreli Bellek (LSTM) ve Çok Katmanlı Algılayıcı (MLP) modelleri kullanılmıştır. Transfer öğrenmenin, veri miktarı sınırlı bölgelerde tahmin performansını önemli ölçüde artırdığı gösterilmiştir. Dhole ve diğ. (2021), PM_{2.5} tahmini için çok kaynaklı transfer öğrenme yaklaşımlarını değerlendirmiştir. Çalışmada, birden fazla kaynak istasyondan alınan bilgilerin hedef istasyonda toplandığı bir ansamble yöntemi kullanılmış ve bu yaklaşımın, yeni kurulan ya da verisi sınırlı istasyonlarda tahmin modellerinin güvenilirliğini artırdığı sonucuna ulaşılmıştır. Gupta ve diğ. (2024), PM_{2.5} tahmini için transfer öğrenmenin etkinliğini artırmak amacıyla Latent Dependency Factor (LDF) adında yeni bir özellik sunmuştur. Çalışmada, iki aşamalı otoencoder kullanılarak kaynak ve hedef bölgeler arasındaki mekânsal ve anlamsal ilişkiler öğrenilmiş ve tahmin doğruluğunda %19.34 oranında bir iyileşme sağlanmıştır. Gupta ve diğ. (2024) ayrıca mekânsal transfer öğrenme yöntemleri kullanarak, veri yetersizliğinin olduğu

bölgelerde PM_{2.5} tahmin performansını geliştirmiştir. Çalışmada LDF kullanılarak veri açısından fakir bölgelerde PM_{2.5} konsantrasyonlarının tahmininde geleneksel transfer öğrenme yöntemlerine göre daha yüksek başarı elde edilmiştir.

1.2.2 Gaziantep ilini konu alan hava kirliliği çalışmaları

Gaziantep, Türkiye'nin sanayi ve ticaret merkezlerinden biri olup, hava kirliliği açısından riskli bölgeler arasında yer almaktadır. Sanayi faaliyetleri, trafik yoğunluğu ve meteorolojik koşullar, bölgedeki hava kirliliği seviyelerini önemli ölçüde etkilemektedir. Gaziantep özelinde yapılan bilimsel çalışmalar, hava kalitesinin mevcut durumunu belirlemek ve kirletici kaynakları analiz etmek amacıyla gerçekleştirilmiştir.

Gaziantep'te hava kirliliğine yönelik gerçekleştirilen çalışmalar, kirliliğin mekânsal ve zamansal değişimlerine yönelik önemli bulgular ortaya koymuştur. Öztürk ve Toros (2023), 2018-2023 yılları arasında Gaziantep'te hava kalitesi ölçümlerini inceleyerek SO₂ ve PM₁₀ değerlerinin mevsimlere göre değişimini analiz etmişlerdir. Araştırmada, SO₂ konsantrasyonlarının genel olarak 'iyi' düzeyde olduğu ancak PM₁₀ seviyelerinin kış aylarında 'hassas' hatta bazı dönemlerde 'sağlıksız' seviyelere çıktığı belirlenmiştir. Ayrıca kentin kuzey bölgelerinin güney bölgelerine kıyasla PM₁₀ açısından daha temiz olduğu tespit edilmiştir. Cuci ve Polat (2015), Gaziantep'teki trafik kaynaklı hava kirliliğinin mekânsal dağılımını belirlemek üzere detaylı bir kirlilik haritası oluşturmuştur. Çalışmanın sonuçları, trafik yoğunluğunun yüksek olduğu bölgelerde emisyonların belirgin şekilde arttığını ve özellikle Organize Sanayi Bölgesi (OSB) gibi sanayi alanlarında hava kirliliğinin yükseldiğini göstermiştir. Bu bulgular, şehirdeki hava kalitesini iyileştirmeye yönelik alınması gereken tedbirlerin belirlenmesinde önemli bir referans oluşturmaktadır. Yılmaz (2018), Güneydoğu Anadolu Bölgesi'nde 2011-2015 yılları arasındaki hava kalitesini değerlendirerek Gaziantep'in PM₁₀ seviyelerinin diğer illere kıyasla daha yüksek olduğunu vurgulamıştır. Araştırmada bölgedeki PM₁₀ ve SO₂ değerleri Dünya Sağlık Örgütü'nün (WHO) önerdiği sınır değerlerle karşılaştırılmış ve Gaziantep'in hava kalitesi detaylı olarak ele alınmıştır. Vural (2021), Güneydoğu Anadolu Bölgesi illerindeki hava kalitesinin 2007-2019 yılları arasındaki mekânsal değişimini Coğrafi

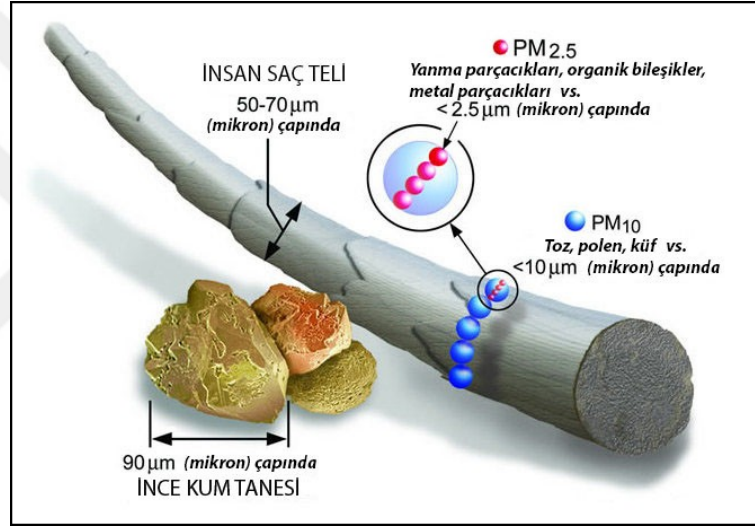
Bilgi Sistemleri (CBS) ile analiz etmiştir. Çalışma, Gaziantep'in hava kirliliğinin yıllar içinde artış gösterdiğini ve özellikle sanayi faaliyetlerinin yoğun olduğu alanlarda hava kalitesinin olumsuz etkilendiğini ortaya koymuştur.



2. PM_{2.5} KİRLİLİĞİ VE METEOROLOJİK FAKTÖRLER

2.1 PM_{2.5} Kaynakları ve Oluşumu

PM_{2.5} partikül maddeleri, çapı 2.5 mikrometreden (μm) daha küçük olan ince partiküllerdir ve havada asılı kalarak uzun mesafeler taşınabilirler. PM_{2.5} ve PM₁₀ partiküllerinin boyutları ve diğer maddelerle karşılaştırılması Şekil 2.1'de gösterilmektedir [20].



Şekil 2.1: Partikül Madde Boyutlarının Karşılaştırılması [20].

Bu partiküller çeşitli doğal ve antropojenik kaynaklardan atmosfere salınmaktadır. Bu partiküller, insan sağlığı ve ekosistemler üzerinde ciddi etkilere sahip olup, hava kalitesinin belirlenmesinde kritik bir role sahiptir. PM_{2.5}'in kaynaklarını anlamak, hava kirliliği kontrol stratejilerinin geliştirilmesi açısından önem taşımaktadır. PM_{2.5}'in başlıca kaynakları doğal ve antropojenik olmak üzere iki ana gruba ayrılmaktadır. Doğal kaynaklar arasında volkanik patlamalar, orman yangınları, deniz spreyi, toprak ve biyolojik süreçler sonucu atmosfere yayılan organik ve inorganik partiküller bulunmaktadır [21]. Antropojenik kaynaklar ise motorlu taşıt emisyonları, sanayi faaliyetleri, enerji üretimi, evsel ısınma ve tarımsal uygulamalar sonucunda

atmosfere yayılan partiküllerden oluşmaktadır [21]. Kentsel ve kırsal alanlardaki $PM_{2.5}$ kaynak dağılımı farklılık göstermektedir. Kentsel alanlarda özellikle trafik emisyonları, sanayi tesislerinden salınan gazlar ve biyokütle yakımı gibi faktörler ön plandayken, kırsal bölgelerde toprak kaynaklı partiküller ve biyolojik süreçler daha yaygın görülmektedir [22] [23]. Yapılan çalışmalar, $PM_{2.5}$ 'in bileşiminin genellikle sülfat, nitrat ve amonyum gibi inorganik bileşiklerden oluştuğunu ve bu bileşenlerin toplam $PM_{2.5}$ kütlesinin %50'sinden fazlasını oluşturduğunu göstermektedir [24]. $PM_{2.5}$ 'in atmosferde oluşumu birincil ve ikincil süreçler olmak üzere iki ana mekanizmaya dayanmaktadır. Birincil $PM_{2.5}$, doğrudan emisyon kaynaklarından atmosfere salınan partiküllerden oluşurken, bu emisyonlar genellikle fosil yakıtların yanması, sanayi tesisleri ve taşıtlardan kaynaklanmaktadır [25]. İkincil $PM_{2.5}$ ise atmosferde gazların kimyasal reaksiyonları sonucunda oluşan partiküllerden meydana gelmektedir. Özellikle SO_2 ve NO_x emisyonlarının atmosferde oksitlenerek sülfat ve nitrat partiküllerine dönüşmesi, $PM_{2.5}$ kütlesine önemli katkı sağlamaktadır [26]. Bu süreçler, fotokimyasal oksidasyon, heterojen reaksiyonlar ve gazdan partiküle dönüşüm mekanizmalarını içermektedir [23] [27].

Ayrıca, meteorolojik faktörler de $PM_{2.5}$ seviyeleri üzerinde doğrudan etkilidir; bağıl nem, rüzgar hızı ve sıcaklık değişimleri, partikül oluşum sürecini belirleyen temel unsurlar arasında yer almaktadır [28]. Sanayi süreçlerinde ise kalsiyum bazlı ince malzemeler, potasyum buharları ve SO_2 arasındaki etkileşimler sonucunda $PM_{2.5}$ oluşumu gerçekleşebilmektedir [29]. $PM_{2.5}$ seviyelerinin belirlenmesi ve kaynaklarının anlaşılması için Pozitif Matris Faktörizasyonu (PMF) gibi teknikler yaygın olarak kullanılmaktadır [24]. Bu yöntemler, kirliliğin kaynağını tespit etmek ve hava kalitesini iyileştirmeye yönelik etkin stratejiler geliştirmek açısından kritik önem taşımaktadır [30].

Bu bölümde ele alınan çalışmalar, $PM_{2.5}$ 'in kaynakları ve oluşum süreçlerinin anlaşılmasının, hava kalitesi yönetimi için kritik olduğunu göstermektedir. Özellikle ikincil $PM_{2.5}$ oluşum mekanizmalarının incelenmesi, hava kirliliğini azaltmaya yönelik daha etkili politikaların belirlenmesini sağlayacaktır.

2.2 PM_{2.5} ve Sağlık Üzerine Etkileri

Modern dünyanın hızla gelişen kentleşme ve sanayileşme süreçleri, hava kalitesinin bozulmasına yol açarak insan sağlığı üzerinde önemli riskler oluşturmaktadır. Bu risklerin başında, atmosferde asılı kalan ve çapı 2.5 mikrometreden küçük olan PM_{2.5} partikülleri gelmektedir. Gözle görülmeyecek kadar küçük olmalarına rağmen, solunum yoluyla doğrudan akciğerlere ulaşabilen bu partiküller, uzun vadede ciddi sağlık sorunlarına neden olabilmektedir. PM_{2.5}'in başlıca kaynakları arasında fosil yakıtların yanması, sanayi tesisleri, araç emisyonları ve doğal süreçler yer almakta olup, bu kirleticiler hava kalitesini olumsuz etkileyerek toplum sağlığı açısından önemli bir tehdit oluşturmaktadır.

Solunum sistemi, hava kirleticilerine karşı ilk savunma hattı olduğundan PM_{2.5} maruziyetine karşı özellikle savunmasızdır. Bu ince partiküller, solunduğunda akciğerlerin derinliklerine nüfuz ederek alveol düzeyine ulaşabilmekte ve burada iltihaplanmaya, oksidatif strese ve alveolar duvarlarda hasara yol açabilmektedir [31] [32]. Uzun süreli PM_{2.5} maruziyeti, kronik obstrüktif akciğer hastalığı (KOAH), astım ve akciğer kanseri gibi ciddi solunum yolu hastalıklarının gelişme riskini artırmaktadır [33] [34] [35]. Ayrıca, hava kirliliği solunum yolu enfeksiyonlarını tetikleyerek toplum sağlığı üzerindeki olumsuz etkilerini artırmaktadır.

Özellikle çocuklar, yaşlılar ve solunum yolu hastalıkları olan bireyler, PM_{2.5}'in zararlı etkilerine karşı daha hassas kabul edilmektedir. Partiküllerin akciğer fonksiyonlarını olumsuz yönde etkilediği, solunum yolu iltihaplarını artırdığı ve bağışıklık sistemini zayıflatarak enfeksiyonlara yatkınlığı artırdığı bilimsel çalışmalarla gösterilmiştir [32] [33]. Bu nedenle, hava kirliliğinin azaltılması, solunum yolu hastalıklarının önlenmesi ve yönetimi açısından büyük önem taşımaktadır [35]. PM_{2.5} seviyelerini kontrol altına almak için çevresel politikalar geliştirmek, hava kalitesini iyileştirmek ve bireylerin maruziyetini azaltmaya yönelik önlemler almak, toplum sağlığını koruma açısından kritik bir adımdır.

PM_{2.5} maruziyeti yalnızca solunum yolları ile sınırlı kalmayıp kardiyovasküler sağlık üzerinde de önemli olumsuz etkilere yol açmaktadır. Yapılan araştırmalar,

hem kısa hem de uzun süreli PM_{2.5} maruziyetinin ateroskleroz, hipertansiyon, aritmi ve miyokard enfarktüsü gibi kardiyovasküler hastalıklarla ilişkili olduğunu göstermektedir [36] [37]. PM_{2.5} partikülleri solunum yoluyla vücuda girdikten sonra sistemik dolaşıma geçerek damar endotelinde iltihaplanma, oksidatif stres ve damar sertliğine neden olmakta, bu da kalp hastalıkları riskini artırmaktadır. Ayrıca, hava kirliliği nedeniyle kan basıncında dalgalanmalar meydana gelmekte ve kardiyovasküler sistem üzerindeki yük artmaktadır. Bu nedenle, PM_{2.5} maruziyetinin azaltılmasına yönelik bireysel ve hükümet düzeyinde alınacak önlemler, hava kirliliğine bağlı kardiyovasküler hastalıkların önlenmesi açısından hayati önem taşımaktadır [38] [39].

PM_{2.5} kirliliğinin insan sağlığı üzerindeki olumsuz etkileri, yalnızca solunum yolları ve kalp-damar sistemiyle sınırlı kalmayıp; gebelik sürecinden erken çocukluk dönemine kadar uzanarak yeni nesillerin sağlığını da riske atıyor. Gebelik sırasında PM_{2.5}'e maruz kalmak, düşük doğum ağırlığı, erken doğum ve ölü doğum riskini artırırken [40], doğum sonrası maruziyet çocuklarda akciğer fonksiyonlarının zayıflamasına ve astım gelişme riskine neden olabilmektedir [41] [42]. Çocukların gelişim süreçlerinde daha hassas olmaları nedeniyle, PM_{2.5}'in uzun vadeli sağlık etkileri bu yaş grubunda daha belirgin şekilde ortaya çıkmaktadır [43].

Tüm bu sağlık riskleri göz önüne alındığında, hava kirliliğini kontrol altına almak ve toplum sağlığını korumak amacıyla ulusal ve uluslararası düzeyde sınır değerler belirlenmiştir. Dünya Sağlık Örgütü (WHO), hava kalitesi kılavuzlarını düzenli olarak güncelleyerek PM_{2.5}'in insan sağlığı üzerindeki olumsuz etkilerini en aza indirecek limitler önermektedir. Benzer şekilde, Amerikan Çevre Koruma Ajansı (EPA) ve Avrupa Birliği (EU) gibi kurumlar da halk sağlığını korumak amacıyla farklı hava kalitesi standartları oluşturmuştur. Türkiye'de ise, Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, hava kalitesi yönetmelikleri çerçevesinde PM_{2.5} seviyelerini kontrol altına almayı hedeflemektedir.

Dünya Sağlık Örgütü (WHO), bu konuda en önemli rehberlerden biri olup, PM_{2.5} maruziyetinin sağlık üzerindeki ciddi etkilerini göz önünde bulundurarak 2021 yılı hava kalitesi kılavuzlarını güncellemiştir. WHO, PM_{2.5}'in yıllık ortalama

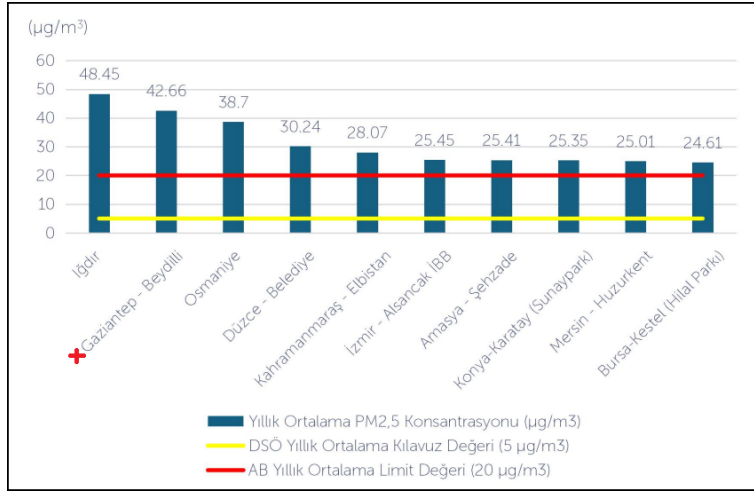
konsantrasyonunun $5 \mu\text{g}/\text{m}^3$ 'ü, 24 saatlik ortalamasının ise $15 \mu\text{g}/\text{m}^3$ 'ü aşmaması gerektiğini önermektedir. Bu limitler, özellikle hassas grupların korunmasını hedefleyerek, hava kirliliğine bağlı erken ölümleri ve hastalık yükünü azaltmayı amaçlamaktadır. Bununla birlikte, Avrupa Birliği (AB) ve Amerikan Çevre Koruma Ajansı (EPA) gibi diğer düzenleyici kurumlar da kendi hava kalitesi standartlarını belirlemiştir. AB limit değerleri, WHO tarafından önerilen sınırların üzerinde olup, yıllık $\text{PM}_{2.5}$ limitini $20 \mu\text{g}/\text{m}^3$ olarak belirlemiştir. EPA ise yıllık ortalama için $9 \mu\text{g}/\text{m}^3$, 24 saatlik ortalama için ise $35 \mu\text{g}/\text{m}^3$ sınırını kabul etmektedir [44] [45] [46]. Tüm bu limitler, Çizelge 2.1'de verilmiştir.

Çizelge 2.1: $\text{PM}_{2.5}$ için DSÖ, AB, Ulusal ve EPA limit değerleri ($\mu\text{g}/\text{m}^3$ cinsinden).

Kirletici	Ortalama Süre	DSÖ 2021 Kılavuz ($\mu\text{g}/\text{m}^3$)	AB Limit	Ulusal Limit	EPA Limit
$\text{PM}_{2.5}$	Yıllık	5	20	–	9
	24 Saatlik	15 ^a	–	–	35

Bu tablodaki a değeri, $\text{PM}_{2.5}$ ölçümünün limit değerlerde verilen konsantrasyon değerini yani $15 \mu\text{g}/\text{m}^3$ 'ü bir yılda 3-4 günden fazla aşamayacağını ifade etmektedir. Limitlerdeki bu farklılıklar, ülkelerin hava kirliliği ile mücadeleye yönelik politikalarının ve yerel koşullarının değişkenliğini yansıtmaktadır. Hava kirliliği ile mücadelede, WHO ve diğer uluslararası kuruluşların belirlediği katı sınır değerlerine uyum sağlamak, halk sağlığını korumanın en etkili yollarından biridir. Ancak, birçok ülkede bu seviyelere ulaşmak zor olduğundan, kademeli geçiş süreçleri ve hava kirliliğini azaltmaya yönelik stratejik eylem planları uygulanmaktadır. Türkiye'nin de bu süreçte yenilenebilir enerjiye geçiş, ulaşımda düşük emisyon politikaları ve sanayi kaynaklı emisyonların sıkı kontrolü gibi önlemleri güçlendirmesi gerekmektedir.

Bu bağlamda, Türkiye'de özellikle sanayi yoğun bölgeler, büyük şehirler ve ulaşımın yoğun olduğu alanlarda $\text{PM}_{2.5}$ seviyelerinin yüksek olduğu görülmektedir. Gaziantep, sanayi faaliyetleri ve artan kentleşme ile birlikte hava kirliliği açısından kritik illerden biri olarak öne çıkmaktadır. 2023 verilerine göre Türkiye'nin en yüksek $\text{PM}_{2.5}$ seviyelerine sahip 10 ili Şekil 2.2 de gösterilmektedir [46]. Bu durum, Gaziantep'te $\text{PM}_{2.5}$ kirliliğinin Dünya Sağlık Örgütü (WHO) ve Avrupa Birliği (AB) tarafından belirlenen sınır değerleri önemli ölçüde aştığını göstermektedir.



Şekil 2.2: Türkiye'nin en yüksek PM_{2.5} seviyelerine sahip 10 ili (2023 verilerine göre) [46].

Özellikle Gaziantep-Beydilli istasyonunda yıllık ortalama PM_{2.5} seviyesi 42.66 µg/m³ olarak ölçülmüş olup, bu değer WHO'nun önerdiği 5 µg/m³ sınırının yaklaşık 8.5 katı, AB'nin belirlediği 20 µg/m³ limitinin ise iki katından fazladır. Yüksek seviyelerdeki bu kirlilik, uzun süreli maruziyetin sağlık üzerindeki olumsuz etkilerini artırmakta ve bölge halkı için ciddi bir risk oluşturmaktadır.

Ayrıca, Türkiye'de 2022 yılında hava kirliliğine atfedilen ölümlerin oransal olarak en fazla olduğu ilk 10 il, Şekil 2.3'de hava kirliliğinin insan sağlığı üzerindeki ölümcül etkilerini ortaya koymaktadır [46].

Gaziantep'te 2022 yılında PM_{2.5} kaynaklı hava kirliliğine bağlı ölümlerin sayısının 1.714 kişi olduğu tahmin edilmektedir. PM_{2.5} maruziyeti yalnızca solunum yolu hastalıklarını artırmakla kalmamakta, aynı zamanda kardiyovasküler hastalıklar, inme ve akciğer kanseri gibi ciddi sağlık sorunlarına da yol açmaktadır.

Bölgesel ölçekte Gaziantep gibi sanayi ve ulaşım kaynaklı emisyonların yüksek olduğu şehirlerde, hava kirliliğini azaltmaya yönelik politikaların uygulanması büyük önem taşımaktadır. Bu noktada, Gaziantep'in PM_{2.5} seviyelerini belirleyen meteorolojik faktörleri ele almak, bölgedeki hava kirliliğinin nedenleri ve dinamikleri hakkında daha derinlemesine bir anlayış sağlayacaktır.

İlin adı	İlin yıllık PM2,5 ortalaması (µg/m³)	PM2,5 için sınır değer 5 µg/m³			PM2,5 için sınır değer 10 µg/m³		
		Atfedilen ölüm sayısı	Atfedilen ölüm oranı (%)	Atfedilen ölüm hızı	Atfedilen ölüm sayısı	Atfedilen ölüm oranı (%)	Atfedilen ölüm hızı
Hakkâri	73,81	230	41,11	215,44	217	38,8	203,33
Batman	59,39	525	34,2	210	486	31,62	194,16
Şirnak	51,79	283	30,24	154,83	258	27,5	140,81
Muş	51,78	335	30,23	216,17	305	27,5	196,6
Malatya	50,26	1.294	29,41	286,48	1.172	26,64	259,51
İğdir	41,40	180	24,43	197,66	158	21,47	173,67
Şanlıurfa	41,21	1.192	24,32	157,85	1.047	21,35	138,58
Ağrı	40,42	346	23,86	179,67	303	20,87	157,17
Osmaniye	39,70	649	23,44	221,48	566	20,43	193,09
Gaziantep	39,06	1.714	23,06	181,01	1.490	20,04	157,32

Şekil 2.3: Türkiye’de 2022 Yılında Hava Kirliliğine Atfedilen Ölümlerin Oransal Olarak En Fazla Olduğu İlk 10 İli [46].

2.3 Meteorolojik Faktörlerin PM_{2.5} Üzerindeki Etkisi

Hava kirliliğinin dağılımı ve yoğunluğu yalnızca emisyon kaynaklarına bağlı olmayıp, atmosferdeki meteorolojik faktörler tarafından da büyük ölçüde belirlenmektedir. Sıcaklık, bağıl nem, rüzgar hızı, yağış ve atmosferik karışım yüksekliği gibi değişkenler, PM_{2.5}’in havadaki konsantrasyonunu ve taşınımını etkileyerek solunabilir partikül maddelerin birikmesine veya dağılmasına neden olabilir. Özellikle sanayi ve ulaşım gibi yoğun insan faaliyetlerinin bulunduğu bölgelerde, bu faktörlerin etkisi daha belirgin hale gelmekte ve hava kalitesini doğrudan etkilemektedir.

PM_{2.5} konsantrasyonlarının değişimini anlamak için meteorolojik süreçlerin hava kirliliği üzerindeki etkileri detaylı bir şekilde incelenmelidir. Bu süreçler, atmosferde meydana gelen fiziksel ve kimyasal mekanizmalar aracılığıyla partiküllerin taşınmasını, çökmesini ve kimyasal dönüşümünü belirler. Bu bağlamda, bölgesel meteorolojik koşulların anlaşılması, hava kirliliği yönetimi açısından kritik bir rol oynamaktadır. PM_{2.5} konsantrasyonları ile sıcaklık arasındaki ilişki mevsimsel ve bölgesel farklılıklar göstermektedir. Genellikle yüksek sıcaklıklar PM_{2.5} seviyelerinde artışa neden olabilmektedir [47] [48], özellikle yaz aylarında bu durum belirginleşmektedir [49]. Bununla birlikte, bazı çalışmalar belirli mevsimlerde veya bölgelerde sıcaklık ile PM_{2.5} konsantrasyonları arasında negatif korelasyon bulunduğunu ortaya koymaktadır [50] [51].

PM_{2.5}’in sağlık üzerindeki olumsuz etkileri, yüksek sıcaklık koşullarında daha da şiddetlenebilmektedir [52] [53]. Sıcak hava dalgaları sırasında PM_{2.5} maruziyetinin

solunum yolu hastalıkları ve kardiyovasküler sorunları artırdığı gözlemlenmiştir. Ayrıca, sıcaklığın partikül madde oluşumu üzerindeki etkisi, atmosferik reaksiyonlar ve nem oranlarıyla doğrudan ilişkilidir. Bu nedenle, $PM_{2.5}$ seviyelerinin yönetimi açısından sıcaklık değişimlerinin dikkate alınması, hava kalitesi yönetim politikalarının etkinliğini artırmak için kritik bir faktördür.

$PM_{2.5}$ konsantrasyonları, bağıl nemin yüksek olduğu dönemlerde havadaki nemin, partikülleri absorbe etmesiyle artabilir, bu da ince partiküllerin atmosferde daha uzun süre asılı kalmasına yol açabilir. $PM_{2.5}$ konsantrasyonları ile bağıl nem (RH) arasındaki ilişki, kirliliğin kaynağı ve bölgesel koşullara bağlı olarak farklılık göstermektedir. Trafik yoğunluğunun yüksek olduğu kentsel alanlarda, bağıl nem arttıkça $PM_{2.5}$ seviyelerinde genel bir yükselme eğilimi görülmektedir. Ancak, sanayi bölgelerinde bu ilişki tam tersi bir eğilim gösterebilir ve bağıl nemin artışı ile $PM_{2.5}$ konsantrasyonlarında azalma gözlemlenebilir [54].

Bağıl nem, atmosferdeki su buharı miktarını etkileyerek $PM_{2.5}$ 'in kimyasal bileşimini değiştirmekte ve suya çözünebilir bileşenlerin oluşumunu teşvik etmektedir. Özellikle, bağıl nemin yüksek olduğu ortamlarda ikincil organik aerosollerin ve suda çözünebilir partiküllerin oluşumu hızlanmaktadır [55]. Hava kirliliği yönetimi açısından bağıl nem değişimlerinin dikkatle izlenmesi, $PM_{2.5}$ seviyelerinin kontrol altına alınmasına yönelik önlemler geliştirilmesi için önemli bir unsurdur.

Rüzgar hızı genellikle $PM_{2.5}$ seviyeleri ile negatif bir korelasyon göstermektedir [48] [56] [57] [58]. Yüksek rüzgar hızları, atmosferdeki kirleticilerin seyrelmesini sağlayarak $PM_{2.5}$ konsantrasyonlarının azalmasına yardımcı olmaktadır. Rüzgar, özellikle açık alanlarda ve kıyı bölgelerinde hava kirliliğini dağıtarak hava kalitesini iyileştirebilir. Ancak, düşük rüzgar hızlarında, atmosferik durgunluk nedeniyle kirletici maddelerin birikmesi artmakta ve hava kirliliği seviyeleri yükselmektedir. Rüzgar yönü de $PM_{2.5}$ seviyeleri üzerinde önemli bir etkiye sahiptir. Kirlilik kaynaklarına bağlı olarak belirli rüzgar yönleri, kirleticilerin taşınmasına ve bölgesel hava kirliliği dağılımının değişmesine neden olabilir. Örneğin, sanayi bölgelerinden esen rüzgarlar, $PM_{2.5}$ yoğunluğunu çevre yerleşim alanlarına taşıyabilirken, kırsal bölgelerden gelen rüzgarlar daha temiz hava getirerek partikül seviyelerinin düşmesine

katkı sağlayabilir. Gaziantep gibi Türkiye'nin güney sınırlarında yer alan şehirlerde, güneyden esen rüzgarların etkisiyle PM_{2.5} konsantrasyonlarında artış beklenir; çünkü bu rüzgarlar, genellikle Suriye ve Irak gibi bölgelerdeki çöl alanlarından kaynaklanan toz taşınımını beraberinde getirir ve bu çöl tozları PM_{2.5} partiküllerinin yoğunluğunu artırır. Bu nedenle, hava kalitesi yönetimi açısından rüzgar hızı ve yönü gibi meteorolojik değişkenlerin dikkate alınması, PM_{2.5} maruziyetini azaltmaya yönelik stratejilerin geliştirilmesinde önemli bir rol oynamaktadır.

Yağışın hava kalitesi üzerindeki etkisi, atmosferdeki partikül maddelerin giderilmesi açısından önemli bir mekanizma olarak kabul edilmektedir. Yağmur, atmosferde asılı kalan PM_{2.5} partiküllerini yıkayarak hava kirliliğini azaltan bir ıslak çökeltme etkisi oluşturur. Yapılan çalışmalar, 5 mm'lik bir yağışın PM_{2.5} konsantrasyonlarını 10-30 µg/m³ oranında düşürebildiğini göstermektedir [59]. Türkiye'de yapılan araştırmalarda da benzer şekilde, yağışlı dönemlerde PM_{2.5} seviyelerinin düştüğü ve bunun en önemli nedenlerinden birinin ıslak çökeltme olduğu belirlenmiştir [17] [60]. Bu süreç, partiküllerin su damlacıkları ile birleşerek atmosfere geri dönmesini sağlarken, özellikle endüstriyel bölgelerde ve yoğun trafik alanlarında hava kalitesinin iyileştirilmesine katkıda bulunmaktadır.

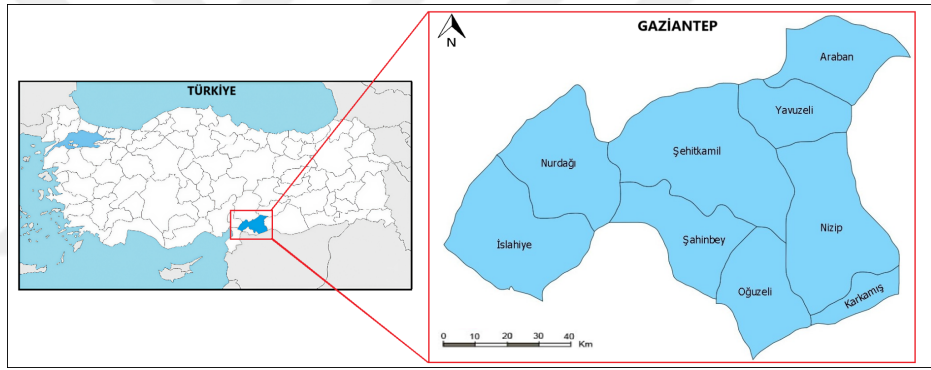
Atmosferik karışım yüksekliği ile PM_{2.5} konsantrasyonları arasındaki ilişki oldukça karmaşık olup, mevsimsel değişkenlikler ve meteorolojik faktörlere bağlı olarak farklılık göstermektedir. Yapılan araştırmalar, sınır tabaka yüksekliği ile PM_{2.5} seviyeleri arasında genellikle negatif bir korelasyon olduğunu ortaya koymaktadır. Sınır tabaka yüksekliğinin arttığı durumlarda, partikül maddelerin atmosferde daha geniş bir hacme yayılması nedeniyle PM_{2.5} konsantrasyonları azalabilmektedir. Ancak bu ilişkinin mevsimsel koşullara ve bölgesel hava dinamiklerine bağlı olarak değişkenlik gösterebileceği belirtilmektedir [61] [62]. PM_{2.5}'in dikey dağılımı genellikle yükseklik arttıkça azalma eğilimi gösterse de, bazı atmosferik koşullar bu dağılımı etkileyerek belirli yüksekliklerde partikül birikimine neden olabilmektedir. Özellikle kararlı atmosfer koşulları ve sıcaklık terselmesi gibi durumlar, partikül maddelerin alt atmosferde birikmesine yol açarak PM_{2.5} seviyelerinin yüksek kalmasına neden olabilir [58]. Bu nedenle, atmosferik karışım yüksekliğinin izlenmesi

ve $PM_{2.5}$ 'in dikey dağılımının anlaşılması, hava kirliliği kontrolü ve yönetimi açısından büyük önem taşımaktadır.



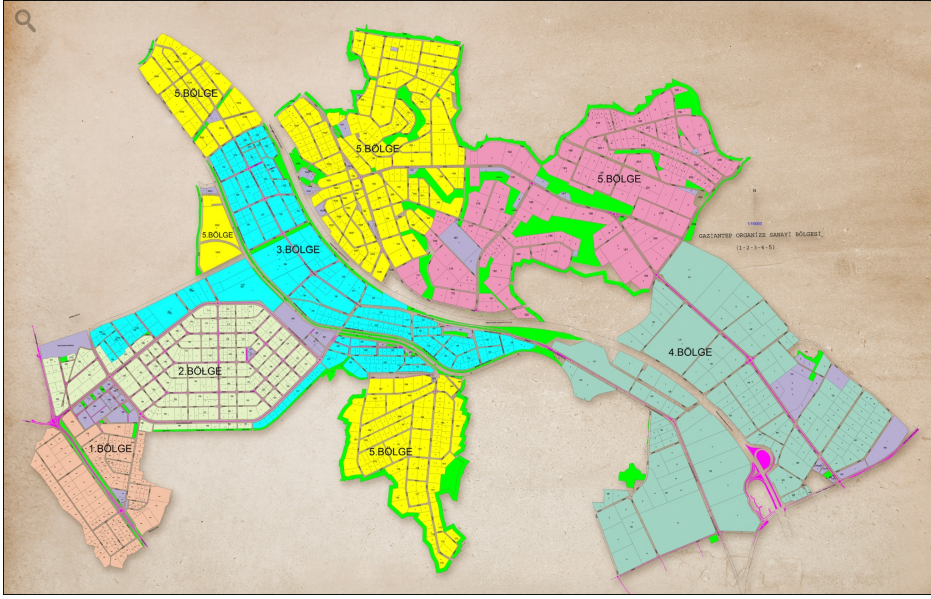
3. GAZİANTEP İLİ'NİN COĞRAFİK KOŞULLARI VE İKLİMİ

Gaziantep, Türkiye'nin güneydoğusunda yer alan ve sanayi, ticaret ve kültürel mirasıyla öne çıkan önemli bir kenttir. Coğrafi olarak 36° 56' - 38° 04' kuzey enlemleri ile 36° 28' - 38° 01' doğu boylamları arasında konumlanmaktadır. Gaziantep, 6222 km²'lik alanıyla Türkiye topraklarının yaklaşık olarak %1'lik bölümünü kapsamaktadır. Gaziantep'in doğusunda Şanlıurfa, batısında Osmaniye ve Hatay, kuzeyinde Kahramanmaraş, güneyinde Suriye, kuzeydoğusunda Adıyaman ve güneybatısında Kilis illeri bulunmaktadır [63]. Gaziantep'in Türkiye içindeki konumu ve ilçelerinin haritası Şekil 3.1'de görülmektedir.



Şekil 3.1: Gaziantep ilinin Türkiye haritası üzerindeki konumu ve ilçe sınırları.

Gaziantep, 2 milyonun üzerindeki nüfusuyla Türkiye'nin en kalabalık illerinden biri olup, hızlı nüfus artışı ve sanayileşme süreci ile birlikte hava kirliliği sorunlarıyla da karşı karşıya kalmaktadır. Şehir, özellikle Gaziantep Organize Sanayi Bölgesi (OSB) başta olmak üzere yoğun sanayi faaliyetlerine sahiptir. Gaziantep OSB, Türkiye'nin en büyük sanayi bölgelerinden biri olup, 3.2'de görüldüğü üzere beş ayrı bölgeden oluşmaktadır ve tekstil, gıda, makine, kimya ve plastik sanayi başta olmak üzere birçok sektörde faaliyet gösteren işletmeleri barındırmaktadır [64]. Güneydoğu Anadolu Bölgesi'nin en büyük şehirlerinden biri olan Gaziantep, Türkiye'nin sanayi ve ticaret merkezlerinden biri olarak önemli bir ekonomik potansiyele sahiptir.



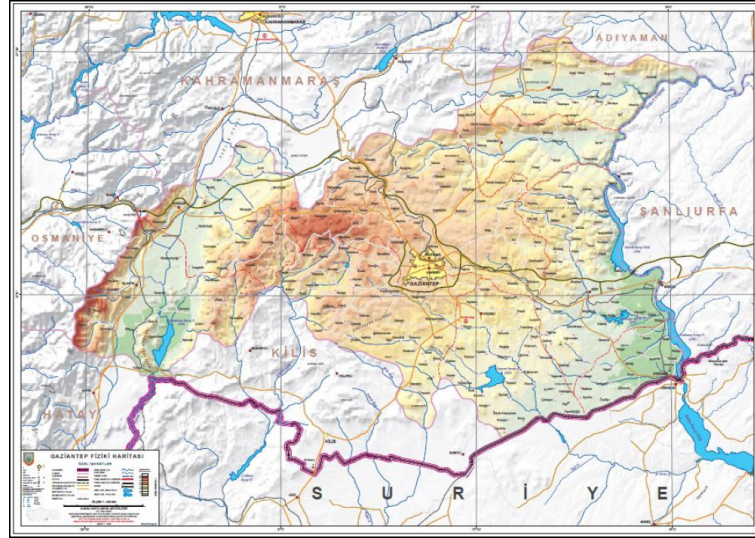
Şekil 3.2: Gaziantep organize sanayi bölgesi haritası [64].

3.1 Gaziantep'in Coğrafi Koşulları

Gaziantep, Akdeniz ile Güneydoğu Anadolu bölgeleri arasında bir geçiş noktası oluşturarak, hem iklimsel hem de topoğrafik açıdan çeşitlilik gösteren bir yapıya sahiptir. Deniz seviyesinden ortalama 850 metre yükseklikte bulunan Gaziantep'in kuzeyi ve kuzey batısı genellikle dağlık ve engebeli bir yapıya sahipken, güney ve güney doğu kesimleri daha düz ve tarıma elverişli alanlardan oluşmaktadır. Gaziantep'in fiziki haritası Şekil 3.3'de gösterilmektedir [65].

İlin en önemli yükseltileri Sof Dağı (1496 m) ve Dülükbaba Dağı (1250 m)'dir. Ayrıca, Gaziantep'in güneyinde yer alan Fırat Nehri'ne yakın bölgeler, topografik yapının daha alçak olduğu alanlar olarak dikkat çekmektedir.

Gaziantep'in yer altı su kaynakları bakımından sınırlı olduğu, ancak tarımsal faaliyetler açısından verimli topraklara sahip olduğu bilinmektedir. Şehrin kuzey kesimlerinde daha sert ve taşlık araziler yaygınken, güney ve batı bölgeleri tarımsal üretime daha elverişlidir. Topoğrafyanın bu farklılıkları, kentsel gelişimi, sanayi bölgelerinin konumlandırılmasını ve dolaylı olarak hava kirliliğinin dağılımını etkilemektedir.



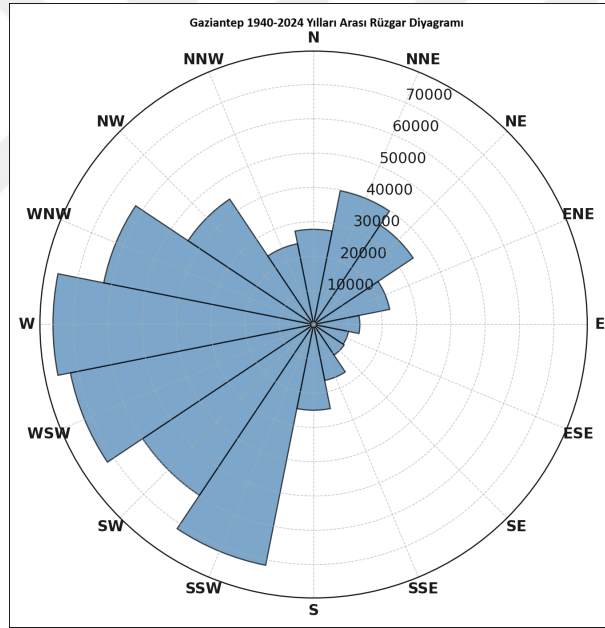
Şekil 3.3: Gaziantep ilinin fiziki haritası [65].

3.2 Gaziantep'in İklimi

Gaziantep'in iklimi, farklı iklim sınıflandırmalarına göre değişkenlik göstermekle birlikte, genel olarak Akdeniz ve karasal iklimin geçiş özelliklerini taşıyan yarı kurak bir iklim olarak tanımlanmaktadır. Şehir, yaz aylarında yüksek sıcaklıkların etkisinde kalırken, kış aylarında soğuk hava dalgalarına maruz kalabilir. Yıllık yağış miktarı Türkiye ortalamasının altında olup, kurak dönemler belirgin şekilde hissedilmektedir. Gaziantep'in iklimi çeşitli sınıflandırmalara göre değerlendirildiğinde, genellikle kurak ve yarı nemli özellikler taşıdığı görülmektedir. Aydeniz İklim Sınıflandırmasına göre şehir kurak iklim tipinde yer alırken, Erinç İklim Sınıflandırmasına göre yarı nemli olarak tanımlanmakta ve yağış etkinlik indisi 26,15 olarak belirlenmektedir. DeMartonne İklim Sınıflandırmasına göre ise yarı kurak ile nemli arasında bir geçiş iklimine sahip olup, kuraklık indisi 11,72'dir. Bununla birlikte, Trewartha İklim Sınıflandırmasına göre Gaziantep, kışları serin (ortalama 3,91°C), yazları ise çok sıcak (ortalama 28,72°C) bir yapıya sahiptir. Thornthwaite İklim Sınıflandırmasına göre ise yarı kurak ve az nemli bir iklime sahip olup, su fazlasının özellikle kış mevsiminde yoğun olduğu, yaz aylarında ise buharlaşma oranının %57,3'e ulaştığı belirtilmektedir. Gaziantep'in sıcaklık ve yağış özellikleri, uzun yıllara dayanan ölçümler temelinde incelendiğinde karasal ve yarı kurak iklim özellikleri gösterdiği görülmektedir.

Meteoroloji Genel Müdürlüğü (MGM) verilerine göre, Gaziantep'te yıllık ortalama sıcaklık 15,3°C olarak kaydedilirken, en düşük sıcaklık -17,5°C (Şubat 1940), en yüksek sıcaklık ise 44,0°C (Temmuz ve Ağustos 2021) olarak ölçülmüştür [66]. En sıcak aylar genellikle Temmuz ve Ağustos olurken, en soğuk dönem Ocak ve Şubat aylarında yaşanmaktadır. Yağış açısından incelendiğinde, Gaziantep'in yıllık ortalama yağış miktarı 564,1 mm olarak belirlenmiş olup, en yüksek yağış Aralık ve Ocak aylarında gözlemlenmektedir. En az yağış ise Temmuz ve Ağustos aylarında görülmektedir. Yıllık toplam yağışlı gün sayısı 83,2 gün olup, kış aylarında yağışlı gün sayısı daha fazladır.

Gaziantep'te hakim rüzgar yönü ve şiddeti hava kalitesini doğrudan etkileyen önemli meteorolojik faktörler arasındadır. Gaziantep iline ait 1940-2024 yılları arasında hakim rüzgar yönü ve esme frekanslarını gösteren rüzgar diyagramı, uzun yıllara dayanan gözlem istatistiklerinden yararlanılarak hesaplanmış ve şekil 3.4'de verilmiştir.



Şekil 3.4: Gaziantep ilinde 1940–2024 yılları arasında gözlemlenen rüzgar yönlerinin dağılımını gösteren rüzgar diyagramı.

Diyagram, batı-kuzeybatı (WNW) yönünden esen rüzgarların en baskın olduğunu göstermekte olup, bu yönün bölgedeki hava kirliliğinin taşınması ve yayılımı üzerindeki etkisini vurgulamaktadır. Aynı zamanda, güneybatı (SW) yönlü rüzgarların da yüksek frekansta esmesi, bölgeye güneyden, özellikle çöl kaynaklı toz taşınımının

gerçekleşebileceğini göstermektedir. Bu durum, Gaziantep'in hava kalitesinin sadece yerel emisyonlardan değil, aynı zamanda atmosferik taşınım süreçlerinden de etkilendiğine işaret etmektedir.



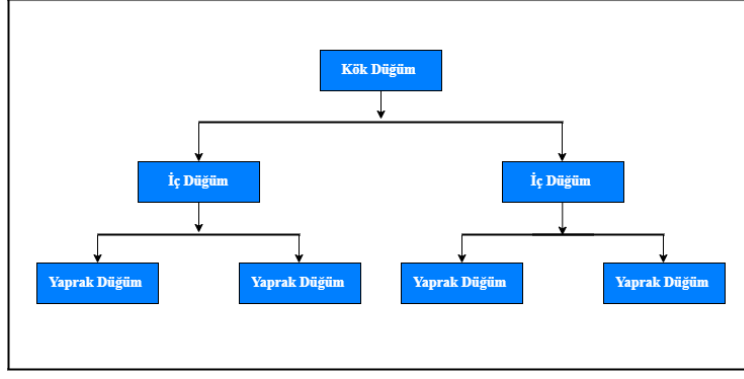


4. HAVA KİRLİLİĞİ TAHMİNİNDE KULLANILAN YÖNTEMLER

Son yıllarda hava kirliliği tahmini üzerine yapılan araştırmalar, makine öğrenimi ve derin öğrenme tekniklerine odaklanmaktadır. Geleneksel istatistiksel yöntemlerin yerini giderek daha fazla veri odaklı yaklaşımlar almakta, hava kirliliği tahmin modellerinde doğruluğun artırılması hedeflenmektedir. Bu bağlamda, transfer öğrenme, hibrit derin öğrenme yaklaşımları, rastgele ormanlar (random forests) ve gradyan artırma (gradient boosting) gibi yöntemler yaygın olarak kullanılmaktadır [67] [68] [69]. Son yıllarda yapılan çalışmalar, farklı modellerin hava kirliliği seviyelerini öngörme başarısını karşılaştırarak hangi yöntemlerin daha etkili olduğunu belirlemeye odaklanmaktadır [70]. Bunun yanı sıra, hava kirliliği tahmininde kullanılan bağımsız değişkenlerin (predictor variables) seçimi, modellerin değerlendirme metrikleri ve coğrafi dağılımları da incelenerek alanın genel durumu hakkında kapsamlı bir çerçeve sunulmuştur [71]. Bu alanda yapılan araştırmalar, gelecekte daha hassas tahminlerin elde edilmesi ve şehirlerde hava kalitesini iyileştirmeye yönelik politika önerilerinin geliştirilmesi açısından büyük önem taşımaktadır. Gelişen teknoloji ile birlikte, yapay zeka ve büyük veri analitiğinin bu alana daha fazla entegre edilmesi, tahmin modellerinin doğruluğunu artırmada kritik bir rol oynayacaktır.

4.1 Karar Ağacı (Decision Tree)

Karar ağaçları, veri setindeki örnekleri belirli kurallar çerçevesinde dallara ayırarak tahmin yapmayı sağlayan, makine öğrenmesi alanında hem sınıflandırma hem de regresyon problemlerini çözmek için yaygın olarak kullanılan hiyerarşik bir makine öğrenmesi algoritmasıdır. Ağaç yapısı, kök düğüm (root node) ile başlar ve veri belirli özelliklere göre bölündükçe iç düğümler (internal nodes) aracılığıyla alt gruplara ayrılır. Nihayetinde, yaprak düğümler (leaf nodes) karar verilen sınıfları veya tahmin edilen değerleri temsil eder. Şekil 4.1'de karar ağacının temel yapısı gösterilmektedir.



Şekil 4.1: Bir karar ağacının yapısı: kök düğüm, iç düğümler ve yaprak düğümler.

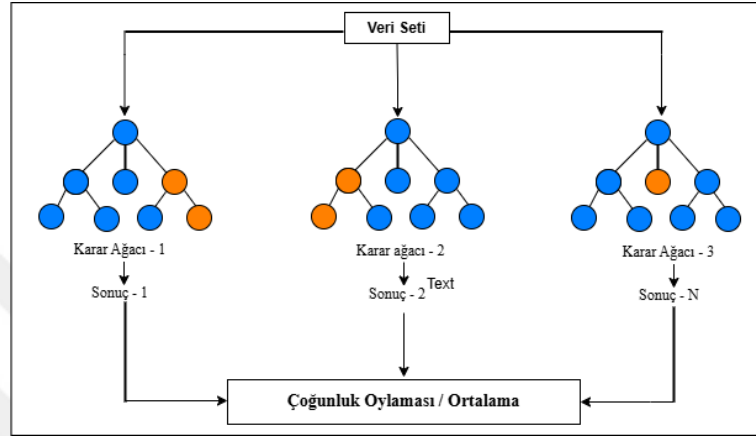
Modelin temel amacı, veriyi mümkün olduğunca homojen alt gruplara ayırarak, yeni veriler için doğru tahminler yapmaktır. Bölünme noktalarının belirlenmesinde bilgi kazancı (information gain) veya Gini katsayısı (Gini index) gibi ölçütler kullanılır. Karar ağaçları, sezgisel ve yorumlanabilir bir yapı sunduğu için görselleştirilebilir ve insan tarafından kolayca anlaşılabilir. Ancak, aşırı büyüyen ağaçlar modelin aşırı öğrenmesine sebep olabilir. Bu durumu önlemek için budama (pruning) teknikleri veya Rastgele Orman (Random Forest) gibi topluluk öğrenme yöntemleri kullanılır. Karar ağaçları, finans, sağlık, pazarlama ve eğitim gibi birçok alanda etkili karar destek sistemleri oluşturmak için tercih edilmektedir.

4.2 Rastgele Orman (Random Forest)

Rastgele Orman, birden fazla karar ağacının (decision tree) birleşiminden oluşan, topluluk öğrenme (ensemble learning) yöntemine dayalı bir makine öğrenmesi algoritmasıdır. Model, her bir karar ağacını farklı veri örnekleri ve değişken alt kümeleriyle eğiterek, genelleme hatalarını minimize etmeyi amaçlar. Leo Breiman tarafından geliştirilen bu yöntem, hem sınıflandırma hem de regresyon problemlerinde yaygın olarak kullanılmaktadır. Modelin temel çalışma prensibi, bootstrap örnekleme (bootstrap sampling) ile her karar ağacının eğitim verisinin farklı bir alt kümesiyle oluşturulmasıdır. Bunun yanı sıra, her düğümde karar verirken tüm özellikler yerine yalnızca rastgele seçilmiş bir alt özellik kümesi dikkate alınır. Böylece, bireysel ağaçların birbirinden bağımsız olması sağlanarak modelin aşırı öğrenme (overfitting) eğilimi azaltılır.

Sınıflandırma problemlerinde, Rastgele Orman her bir ağacın verdiği tahmin sonucunu çoğunluk oylaması (majority voting) ile belirler.

Regresyon problemlerinde ise farklı ağaçlardan gelen tahminlerin ortalaması alınarak nihai sonuç elde edilir. Bu yapı, modelin bireysel karar ağaçlarındaki hataları ortalama etkisiyle dengeleyerek daha güvenilir tahminler yapmasını sağlar. Şekil 4.2'de Rastgele Orman algoritmasının temel çalışma mantığı gösterilmektedir.



Şekil 4.2: Rastgele orman (Random Forest) algoritmasının temel çalışma prensibi.

4.3 Gradyan Artırma (Gradient Boosting)

XGBoost (eXtreme Gradient Boosting), gradyan artırma karar ağaçlarını (Gradient Boosted Decision Trees - GBDT) temel alan, ölçeklenebilir ve yüksek performanslı bir makine öğrenmesi algoritmasıdır. Tianqi Chen tarafından geliştirilen bu yöntem, karar ağaçlarını kullanarak tahmin hatalarını iteratif bir şekilde azaltmayı amaçlar. Bu süreçte her bir yeni ağaç, önceki modelin hatalarını analiz ederek modelin doğruluğunu artıran güncellemeler yapar. XGBoost, karar ağaçları, topluluk öğrenme (ensemble learning) ve gradyan artırma (gradient boosting) gibi temel makine öğrenmesi yöntemlerinin birleşimiyle oluşturulmuştur.

XGBoost'un temel çalışma mekanizması, veri seti üzerinde ilk olarak bir karar ağacı eğiterek başlar. Bu modelin ürettiği tahminler ile gerçek değerler arasındaki farklar, yani rezidüel hatalar (residuals) hesaplanır. Daha sonra, bir sonraki ağaç bu hatalara odaklanarak eğitilir ve model, hata oranını azaltacak şekilde güncellenir. Bu işlem belirlenen iterasyon sayısına kadar devam eder ve model her adımda daha

dođru hale gelir. XGBoost'un başarısının temel nedeni, gradyan iniři (gradient descent) algoritmasıyla modeli sürekli olarak iyileřtirmesi ve hata oranlarını minimize etmesidir.

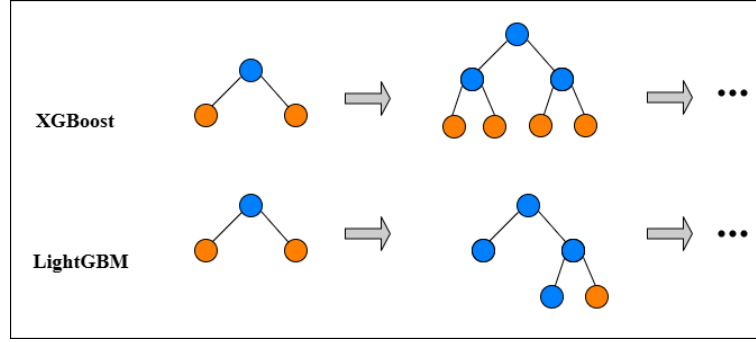
Rastgele Orman (Random Forest) modeliyle benzer řekilde birden fazla karar ađacı kullanmasına rađmen, XGBoost'un farkı ađaçların birbirinden bađımsız deđil, sıralı olarak eđitilmesidir. Bu nedenle, Rastgele Orman modeli ařırı öğrenmeyi (overfitting) önlemek için bagging (bootstrap aggregating) yöntemini kullanırken, XGBoost hata oranlarını azaltmak için boosting mekanizmasını kullanır. XGBoost, hızlı çalıřma süresi, paralel işlem yapabilmesi ve büyük veri kümelerinde verimli çalıřmasıyla makine öğrenmesi yarışmalarında ve endüstriyel uygulamalarda en çok tercih edilen yöntemlerden biri haline gelmiřtir.

4.4 Hafif Gradyan Artırma Makinesi (LightGBM - Light Gradient Boosting Machine) Algoritması

LightGBM, Microsoft tarafından geliştirilen, karar ađaçlarını temel alan bir gradyan artırma (gradient boosting) algoritmasıdır. Ölçeklenebilirliđi, verimliliđi ve hız avantajları sayesinde özellikle büyük veri kümelerinde yaygın olarak kullanılmaktadır. Gradyan artırma mekanizmasını kullanarak zayıf karar ađaçlarını sıralı olarak eđitir ve her adımda hataları minimize eder. Ancak, LightGBM, geleneksel gradyan artırma yöntemlerinden bazı temel farklılıklara sahiptir. Bu algoritma, hızlı eđitim süresi ve düşük bellek kullanımı sađlayan bazı yenilikçi teknikler içerir. Bu algortimada kullanılan Gradient-Based One-Side Sampling (GOSS) yöntemi, büyük gradyan deđerlerine sahip örnekleri koruyarak, küçük gradyanlara sahip örnekleri kısmen dıřlayarak hesaplama yükünü azaltır. Bu sayede, modelin dođruluđu düřmeden işlem süresi hızlanır. Exclusive Feature Bundling (EFB) tekniđi, birbirini dıřlayan özellikleri birleřtirerek modelin daha verimli çalıřmasını sađlar.

LightGBM'nin en belirgin farklarından biri yaprak bazlı (leaf-wise) ađaç büyütme stratejisidir. XGBoost gibi geleneksel gradyan artırma algoritmaları seviye bazlı (level-wise) büyüme stratejisini kullanırken, LightGBM her iterasyonda en fazla hata düşüřü sađlayan yaprađı genişletir. Bu, modelin daha karmařık karar sınırları oluřturmasını sađlar ve büyük veri setlerinde daha yüksek dođruluk elde edilmesine

yardımcı olur. 4.3'de XGBoost ve LightGBM algoritmasının temel işleyişi gösterilmektedir.



Şekil 4.3: XGBoost ve LightGBM algoritmasının temel işleyişi.

4.5 Uzun Kısa Süreli Bellek (LSTM)

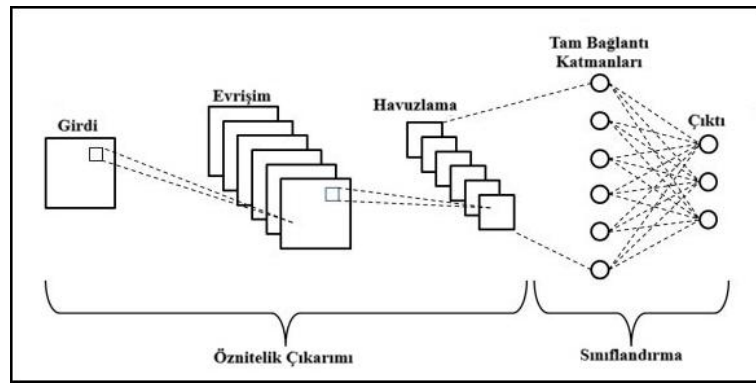
Derin öğrenmede zamansal verilerle çalışırken, modelin geçmiş bilgileri hatırlayabilmesi büyük önem taşır. Bu noktada, ardışık veriler üzerinde öğrenme yapabilen Özyinelemeli Sinir Ağları (RNN) devreye girer. Özyinelemeli Sinir Ağları (RNN - Recurrent Neural Networks), zamansal veriler üzerinde tahmin yapmak için geliştirilmiş yapay sinir ağlarıdır. Bu model, uzun dizilerdeki bilgileri hatırlamakta yetersiz kalır. Bunun başlıca nedeni, gradyan kaybolması (vanishing gradient) problemidir. RNN'ler geçmiş bilgileri işleyerek bir sonraki adıma aktarsa da uzun vadeli bağımlılıkları öğrenmekte zorlanır.

Bu sorunu çözmek için geliştirilen LSTM (Long Short-Term Memory) ağları, uzun süreli bağımlılıkları daha iyi öğrenmek amacıyla özel bir bellek mekanizması kullanır. LSTM, klasik RNN'lerden farklı olarak hücre durumu (cell state) adı verilen bir yapıya sahiptir. Bu yapı, bilgiyi zaman adımları boyunca taşıyarak, modelin önemli verileri uzun süre boyunca saklamasını sağlar. LSTM'nin başarısı, unutma kapısı (forget gate), giriş kapısı (input gate) ve çıkış kapısı (output gate) olmak üzere üç temel kapı mekanizmasına dayanır. Unutma kapısı, önceki bilgilerin ne kadarının silineceğine karar verirken, giriş kapısı, yeni bilgilerin ne kadarının belleğe eklenmesi gerektiğini belirler. Çıkış kapısı ise hücrede saklanan bilginin ne kadarının bir sonraki zaman adımına aktarılacağını kontrol eder. Bu mekanizmalar sayesinde, LSTM modelleri, uzun diziler üzerinde daha stabil öğrenme yapabilir. 4.4'de RNN'nin genel

GRU'nun uzun vadeli bağımlılıkları korurken gereksiz hesaplamaları azaltmasını sağlar. GRU'nun çalışma prensibi, önceki zaman adımındaki gizli durumu (hidden state) ve mevcut giriş verisini birleştirerek yeni bir çıktı üretmesi esasına dayanır. Model, güncelleme kapısını kullanarak eski bilginin ne kadar korunacağına, sıfırlama kapısını kullanarak da hangi bilgilerin unutulacağına karar verir. Böylece GRU, daha az hesaplama gücü gerektirerek hızlı ve verimli öğrenme süreci sağlar.

4.7 Evrişimli Sinir Ağları (CNN - Convolutional Neural Networks)

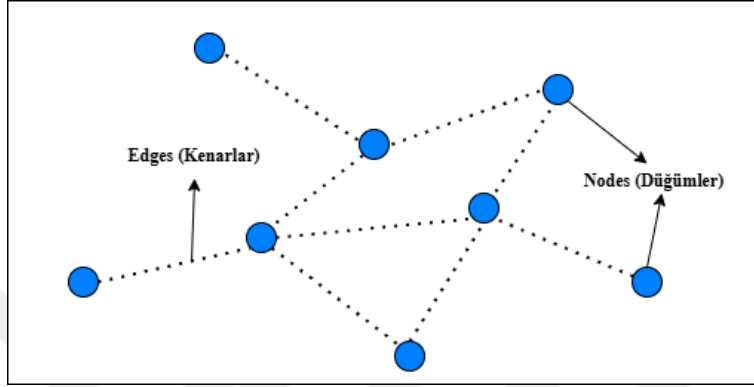
Evrişimli Sinir Ağları (CNN - Convolutional Neural Networks), görüntü işleme ve derin öğrenme alanında en yaygın kullanılan yapay sinir ağı modellerinden biridir. CNN, özellik çıkarımı (feature extraction) ve sınıflandırma (classification) olmak üzere iki temel aşamadan oluşur. Özellik çıkarımı aşamasında, giriş görüntüsünden önemli özellikler tespit edilir ve filtreler kullanılarak evrişim (convolution) işlemi gerçekleştirilir. Bu aşamada görüntü, katmanlar boyunca ilerledikçe daha yüksek seviyede özelliklere dönüştürülür. Daha sonra havuzlama (pooling) katmanları, evrişim katmanları tarafından çıkarılan özellik haritalarını küçülterek modelin genelleştirme yeteneğini artırır. İkinci aşama olan sınıflandırma aşamasında, çıkarılan özellikler tam bağlantılı (fully connected) katmanlara aktarılır. Bu katmanda, verinin hangi sınıfa ait olduğu belirlenir. Tam bağlantılı katmanlar, geleneksel yapay sinir ağlarında olduğu gibi çalışarak, modelin nihai tahminini yapmasını sağlar. Şekil 4.6'de, temel bir CNN mimarisinin bileşenleri ve bilgi akışı gösterilmektedir.



Şekil 4.6: Bir Konvolüsyonel Sinir Ağı (CNN) mimarisinin temel yapısı: öznitelik çıkarımı (evrişim ve havuzlama katmanları) ve sınıflandırma (tam bağlantılı katmanlar).

4.8 Grafik Sinir Ağları (GNN - Graph Neural Networks)

Grafik Sinir Ağları (GNN - Graph Neural Networks), düzensiz ve birbiriyle bağlantılı verileri işleyebilen derin öğrenme modelleridir. Geleneksel sinir ağları, genellikle satır ve sütunlar halinde düzenlenmiş verilerle çalışırken, GNN modelleri düğümler (nodes) ve kenarlar (edges) ile temsil edilen grafik veri yapıları üzerinde işlem yapar. Basit bir GNN yapısı ve modeldeki düğümler ve kenarların gösterimi 4.7'de gösterilmektedir.



Şekil 4.7: GNN yapısının basit gösterimi.

GNN'ler, grafik üzerindeki her bir düğümün kendi komşularından bilgi toplamasına dayalı bir öğrenme süreci kullanır. Bu süreç, genellikle bilgi yayılımı (message passing) adı verilen teknikle gerçekleştirilir. Model, her düğümün kendisiyle bağlantılı diğer düğümlerden özellikler almasını ve zaman içinde bu bilgiyi güncellemesini sağlar. Bu sayede, düğümler yalnızca kendi verilerine değil, aynı zamanda ağdaki ilişkilerine de dayalı olarak öğrenme yapabilir. Sonuç olarak, GNN'ler grafik yapılarındaki veri ilişkilerini öğrenme konusunda güçlü bir yöntem sunarak, geleneksel yöntemlerin yetersiz kaldığı durumlarda yüksek doğrulukta sonuçlar elde edilmesini sağlar. Özellikle karmaşık ilişkilerin bulunduğu veri setlerinde, düğümler arasındaki bağıntıları anlamlandırmada büyük bir avantaj sunar.

4.9 Gizli Bağımlılık Faktörü (LDF - Latent Dependency Factor)

Latent Dependency Factor (LDF), veri setlerinde eksik verileri tamamlamak ve mekânsal bağıntıları modellemek için kullanılan bir özellik çıkarım yöntemidir. Özellikle transfer öğrenme (transfer learning) süreçlerinde, farklı veri kaynaklarından elde edilen bilgiler kullanılarak eksik veri noktalarının tahmin edilmesine yardımcı

olur. LDF, autoencoder tabanlı bir yaklaşım ile veri kümesindeki bağımlılıkları öğrenerek, gizli temsiller (latent representations) oluşturur. LDF'nin temel çalışma prensibi, girdi verilerini daha düşük boyutlu ancak anlamlı bir vektör uzayına sıkıştırarak mekânsal ve semantik ilişkileri öğrenmektir. Bu süreçte, coğrafi, meteorolojik ve çevresel değişkenler dikkate alınarak sensörler arasındaki bağıntılar yakalanır. Autoencoder mimarisinin encoder aşaması, ham verileri özetleyici özelliklere dönüştürürken, decoder aşaması bu bilgileri geri açarak modelin doğruluğunu artırmayı amaçlar.

LDF, makine öğrenmesi ve derin öğrenme modellerinde bir ön işleme adımı olarak kullanılabilir. Özellikle mekânsal tahmin, çevresel modelleme ve eksik veri tamamlama gibi alanlarda etkili bir özellik çıkarım tekniği olarak öne çıkmaktadır. Bu sayede, transfer öğrenme yöntemleriyle farklı konumlar arasındaki ilişkiler modellenilebilir ve eksik veri içeren istasyonlarda daha güvenilir tahminler yapılabilir.

4.10 Performans Değerlendirme Metrikleri

Bu çalışmada, önerilen modelin tahmin performansını değerlendirmek amacıyla üç temel metrik kullanılmıştır: Ortalama Mutlak Hata (Mean Absolute Error - MAE), Kök Ortalama Kare Hata (Root Mean Squared Error - RMSE) ve Belirlilik Katsayısı (Coefficient of Determination - R²). Bu metrikler, modelin doğruluğunu, hata büyüklüğünü ve gerçek verilere ne kadar uyum sağladığını ölçmek için seçilmiştir. Aşağıda, her bir metriğin tanımı ve matematiksel formülleri sunulmuştur. MAE, tahmin edilen değerler ile gerçek değerler arasındaki mutlak farkların ortalamasını hesaplayarak modelin hata miktarını basit ve anlaşılır bir şekilde ifade eder. Matematiksel olarak MAE, (4.1)'de görüldüğü şekildedir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.1)$$

Burada, n veri noktası sayısını, y_i gerçek değeri, \hat{y}_i ise tahmin edilen değeri temsil eder. Bu metrik, hataların büyüklüğünü doğrudan yansıttığı için yorumlaması kolaydır ve çalışmada modelin genel tahmin sapmasını değerlendirmede etkili bir araç olarak kullanılmıştır. RMSE, tahmin hatalarının karelerinin ortalamasının karekökü alınarak

hesaplanır ve büyük hatalara daha fazla ağırlık verir. RMSE'nin formülü (4.2)'de görüldüğü şekildedir.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.2)$$

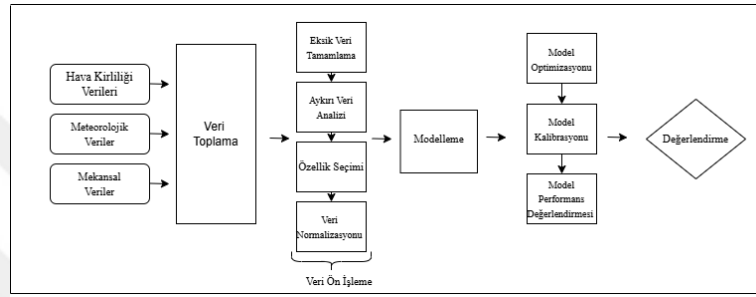
Bu metrik, uç değerlerdeki hataları daha belirgin hale getirerek modelin performansını hassas bir şekilde analiz etme imkânı sunar. Çalışmada, RMSE tahminlerin tutarlılığını ve hata dağılımının büyüklüğünü incelemek amacıyla tercih edilmiştir. R^2 metriği, modelin bağımlı değişkendeki varyansın ne kadarını açıkladığını gösterir ve 0 ile 1 arasında bir değer alır. R^2 'nin formülü (4.3)'de görüldüğü şekildedir.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (4.3)$$

Burada \bar{y} , gerçek değerlerin ortalamasını ifade eder. 1'e yakın bir R^2 değeri, modelin veriye yüksek derecede uyum sağladığını gösterir. Bu çalışmada, R^2 modeli genelleme yeteneği ve açıklayıcılığı açısından değerlendirmek için kullanılmıştır. Bu üç metrik bir arada, modelin performansını çok yönlü bir şekilde analiz etme imkânı sunmuştur.

5. VERİ VE YÖNTEM

Çalışmada, Gaziantep ili için $PM_{2.5}$ konsantrasyonlarının tahmin edilmesi amacıyla zamansal ve mekânsal veri analizi gerçekleştirilmiştir. Çalışma süreci veri toplama, veri ön işleme, modelleme ve değerlendirme olmak üzere dört ana aşamadan oluşmaktadır. Çalışmanın genel akış süreci 5.1’de özetlenmiştir.



Şekil 5.1: Çalışmada izlenen veri analizi ve modelleme süreci.

5.1 Veriler Hakkında Bilgiler

Gaziantep ili için zamansal ve mekânsal $PM_{2.5}$ tahmini amacıyla geliştirilen modellerde, $PM_{2.5}$ sensör verileri, meteorolojik veriler ve mekânsal veriler kullanılmıştır. Çalışmada yer alan tüm veriler, 01.01.2021 - 31.12.2023 tarihleri arasındaki günlük değerlerden oluşmaktadır. Her istasyon için kullanılan verilerin istatistiksel özetleri aşağıda sunulmuştur: Çizelge 5.1’de Atapark istasyonunun, Çizelge 5.2’de Beydilli istasyonunun, Çizelge 5.3’te Nizip istasyonunun, Çizelge 5.4’te Fevzi Çakmak istasyonunun, Çizelge 5.6’da GASKİ istasyonunun ve son olarak Çizelge 5.5’de Meteoroloji istasyonunun verilerinin istatistiksel özetleri verilmektedir.

5.1.1 $PM_{2.5}$ sensör verisi

$PM_{2.5}$ konsantrasyonu verileri, T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı tarafından sağlanan Ulusal Hava Kalitesi İzleme Ağı (UHKİA) aracılığıyla elde edilmiştir. Bu veriler, <https://sim.csb.gov.tr/> uzantısından kamuya açık olarak sağlan-

Çizelge 5.1: Atapark istasyonuna ait değişkenlerin istatistiksel özeti
(01.01.2021–31.12.2023).

Parametre	Birim	Ortalama Değer	Eksik Veri Sayısı	Eksik Veri Yüzdesi
Sensör PM _{2.5}	µg/m ³	25.10	255	23.29
Sıcaklık	°C	16.89	0	0.00
Bağıl Nem	%	54.33	0	0.00
Çiy Noktası Sıcaklığı	°C	5.17	0	0.00
Yağış	mm	0.04	0	0.00
Deniz Seviyesine İndirilmiş Basınç	hPa	1013.44	0	0.00
Yüzey Basıncı	hPa	921.20	0	0.00
Rüzgar Hızı (10m)	km/h	14.29	0	0.00
Rüzgar Yönü (10m)	°	212.203	0	0.00
CAMS PM _{2.5}	µg/m ³	15.06	0	0.00
NDVI	-	0.03	694	63.38

Çizelge 5.2: Beydilli istasyonuna ait değişkenlerin istatistiksel özeti
(01.01.2021–31.12.2023).

Parametre	Birim	Ortalama Değer	Eksik Veri Sayısı	Eksik Veri Yüzdesi
Sensör PM _{2.5}	µg/m ³	46.20	191	17.44
Sıcaklık	°C	16.60	0	0.00
Bağıl Nem	%	53.02	0	0.00
Çiy Noktası Sıcaklığı	°C	4.71	0	0.00
Yağış	mm	0.04	0	0.00
Deniz Seviyesine İndirilmiş Basınç	hPa	1013.32	0	0.00
Yüzey Basıncı	hPa	915.7	0	0.00
Rüzgar Hızı (10m)	km/h	13.50	0	0.00
Rüzgar Yönü (10m)	°	205.475	0	0.00
CAMS PM _{2.5}	µg/m ³	16.98	0	0.00
NDVI	-	0.10	666	60.82

maktadır [72]. Çalışmada, Gaziantep ilinde hava kirliliği ölçümleri gerçekleştiren Atapark, Fevzi Çakmak, GASKİ, Beydilli, Nizip ve Meteoroloji istasyonlarına ait veriler analiz edilmiştir. Ancak, GASKİ ve Meteoroloji istasyonlarında PM_{2.5} ölçümleri bulunmadığı için bu istasyonlar için doğrudan veri kullanılmamış, bunun yerine çalışma kapsamında bu istasyonlar için PM_{2.5} tahmini yapılmıştır. Şekil 5.2’de Gaziantep’te bulunan hava kirliliği ölçüm istasyonlarının dağılımını göstermektedir. Görüldüğü üzere, beş istasyon Gaziantep’in merkez ilçelerinde konumlanmışken, Nizip istasyonu şehrin doğusundaki Nizip ilçesinde yer almaktadır.

Çizelge 5.3: Nizip istasyonuna ait değişkenlerin istatistiksel özeti
(01.01.2021–31.12.2023).

Parametre	Birim	Ortalama Değer	Eksik Veri Sayısı	Eksik Veri Yüzdesi
Sensör PM _{2.5}	µg/m ³	21.91	548	50.05
Sıcaklık	°C	19.06	0	0.00
Bağıl Nem	%	49.97	0	0.00
Çiy Noktası Sıcaklığı	°C	5.92	0	0.00
Yağış	mm	0.05	0	0.00
Deniz Seviyesine İndirilmiş Basınç	hPa	1012.74	0	0.00
Yüzey Basıncı	hPa	951.13	0	0.00
Rüzgar Hızı (10m)	km/h	13.79	0	0.00
Rüzgar Yönü (10m)	°	223.49	0	0.00
CAMS PM _{2.5}	µg/m ³	12.97	0	0.00
NDVI	-	0.06	665	60.73

Çizelge 5.4: Fevzi Çakmak istasyonuna ait değişkenlerin istatistiksel özeti
(01.01.2021–31.12.2023).

Parametre	Birim	Ortalama Değer	Eksik Veri Sayısı	Eksik Veri Yüzdesi
Sensör PM _{2.5}	µg/m ³	25.40	385	35.16
CAMS PM _{2.5}	µg/m ³	16.98	0	0.00
Deniz Seviyesine İndirilmiş Basınç	hPa	1013.58	0	0.00
Rüzgar Hızı (10m)	km/h	10.33	0	0.00
Rüzgar Yönü (10m)	°	205.95	0	0.00
NDVI	-	0.08	880	80.37

Çizelge 5.5: Meteoroloji istasyonuna ait değişkenlerin istatistiksel özeti
(01.01.2021–31.12.2023).

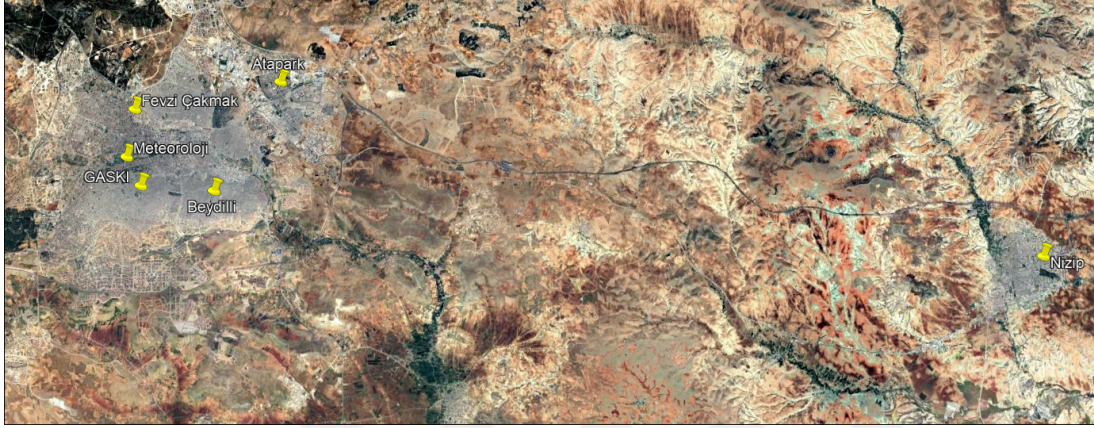
Parametre	Birim	Ortalama Değer	Eksik Veri Sayısı	Eksik Veri Yüzdesi
Sensör PM _{2.5}	µg/m ³	-	1095	100.00
CAMS PM _{2.5}	µg/m ³	16.98	0	0.00
Deniz Seviyesine İndirilmiş Basınç	hPa	1013.32	0	0.00
Rüzgar Hızı (10m)	km/h	13.52	0	0.00
Rüzgar Yönü (10m)	°	205.55	0	0.00
NDVI	-	0.07	880	80.37

Çalışmada kullanılan hava kalitesi izleme istasyonları, Gaziantep ilinin farklı bölgelerine yerleştirilmiş olup, konum, enlem ve boylam bilgileri Çizelge 5.7’de sunulmaktadır.

İstasyonların bulunduğu bölgelerin fiziksel ve çevresel koşulları, hava kirliliği ölçümlerini etkileyebilecek faktörler arasında yer almaktadır. Atapark istasyonu bir park alanında yer alırken, Fevzi Çakmak istasyonu Çevre, Şehircilik ve İklim

Çizelge 5.6: GASKİ istasyonuna ait değişkenlerin istatistiksel özeti
(01.01.2021–31.12.2023).

Parametre	Birim	Ortalama Değer	Eksik Veri Sayısı	Eksik Veri Yüzdesi
Sensör PM _{2.5}	µg/m ³	-	1095	100.00
CAMS PM _{2.5}	µg/m ³	16.98	0	0.00
Deniz Seviyesine İndirilmiş Basınç	hPa	1013.32	0	0.00
Rüzgar Hızı (10m)	km/h	13.52	0	0.00
Rüzgar Yönü (10m)	°	205.51	0	0.00
NDVI	-	0.11	872	79.63



Şekil 5.2: Hava kalitesi ölçüm istasyonlarının konumları.

Çizelge 5.7: Gaziantep'te yer alan UHKİA istasyonlarına ait konum bilgileri.

İstasyon Adı	İlçe (Konum)	Enlem (Latitude)	Boylam (Longitude)
Atapark	Gaziantep (Şahinbey)	37° 5' 16.01" N	37° 25' 38.86" E
Beydilli	Gaziantep (Şahinbey)	37° 2' 40.02" N	37° 23' 43.26" E
Fevzi Çakmak	Gaziantep (Şehitkamil)	37° 4' 41.38" N	37° 21' 16.49" E
Nizip	Gaziantep (Nizip)	37° 0' 35.24" N	37° 48' 13.97" E
GASKİ	Gaziantep (Şahinbey)	37° 3' 33.01" N	37° 22' 57.00" E
Meteoroloji	Gaziantep (Şehitkamil)	37° 3' 54.00" N	37° 22' 40.80" E

Değişikliği İl Müdürlüğü bahçesinde konumlandırılmıştır. Meteoroloji istasyonu, Gaziantep Meteoroloji Müdürlüğü yerleşkesinde bulunurken, GASKİ istasyonu Gaziantep Su ve Kanalizasyon İdaresi'nin D-6 nolu su deposunun olduğu alanda yer almaktadır. Beydilli istasyonu, su deposu yakınında, Nizip istasyonu ise Nizip İlçe Tarım Müdürlüğü'nün bahçesinde konumlanmıştır.

5.1.2 Meteorolojik veriler

Bu çalışmada kullanılan meteorolojik veriler, Open-Meteo'nun Best Match algoritması aracılığıyla ECMWF IFS ve ERA5 reanalysis veri setlerinden elde edilmiştir. IFS (Integrated Forecasting System) modeli, 9 km uzaysal çözünürlüğe sahip olup saatlik tahminler sunarak kısa vadeli hava analizlerinde yüksek doğruluk sağlamaktadır [73]. ERA5 veri seti ise 25 km çözünürlüğe sahip olup, geniş ölçekli iklim analizleri için önerilmektedir ve 1940 yılından günümüze kadar veri sağlayabilmektedir [73]. Çalışmada, bölgenin coğrafi yapısı ve analiz ihtiyacına en uygun çözünürlüğü sunan model Best Match algoritması tarafından otomatik olarak belirlenmiş ve bu doğrultuda veriler işlenmiştir.

Meteorolojik değişkenler, 01.01.2021 - 31.12.2023 tarihleri arasında saatlik ölçümlerden elde edilen günlük ortalama değerler olarak hesaplanmış ve zamansal modelleme sürecinde kullanılmıştır. Çalışmada kullanılan değişkenler 2 metre sıcaklığı (°C), 2 metre bağıl nem (%), 2 metre çiy noktası sıcaklığı (°C), Deniz seviyesi basıncı (hPa), Yüzey basıncı (hPa), 10 metre rüzgar hızı (km/h), 10 metre rüzgar yönü (°) şeklindedir. Bu meteorolojik parametreler, hava kirliliği tahmini açısından kritik olup, PM_{2.5} seviyeleri ile atmosferik değişkenler arasındaki ilişkileri anlamak için modele dahil edilmiştir. Sıcaklık ve bağıl nem, partikül maddelerin atmosferde kalma süresi ve dağılımını etkileyen temel faktörlerdir. Çiy noktası sıcaklığı, yoğunlaşma süreçleriyle bağlantılı olarak partiküllerin kimyasal dönüşümünü etkileyebilirken, basınç değişimleri atmosferik dolaşım sistemlerini belirleyerek hava kütlelerinin hareketini ve PM_{2.5} konsantrasyonlarını etkilemektedir. Ayrıca, rüzgar hızı kirleticilerin taşınması, birikmesi veya seyrelmesi açısından önemli bir değişkendir [73] [74].

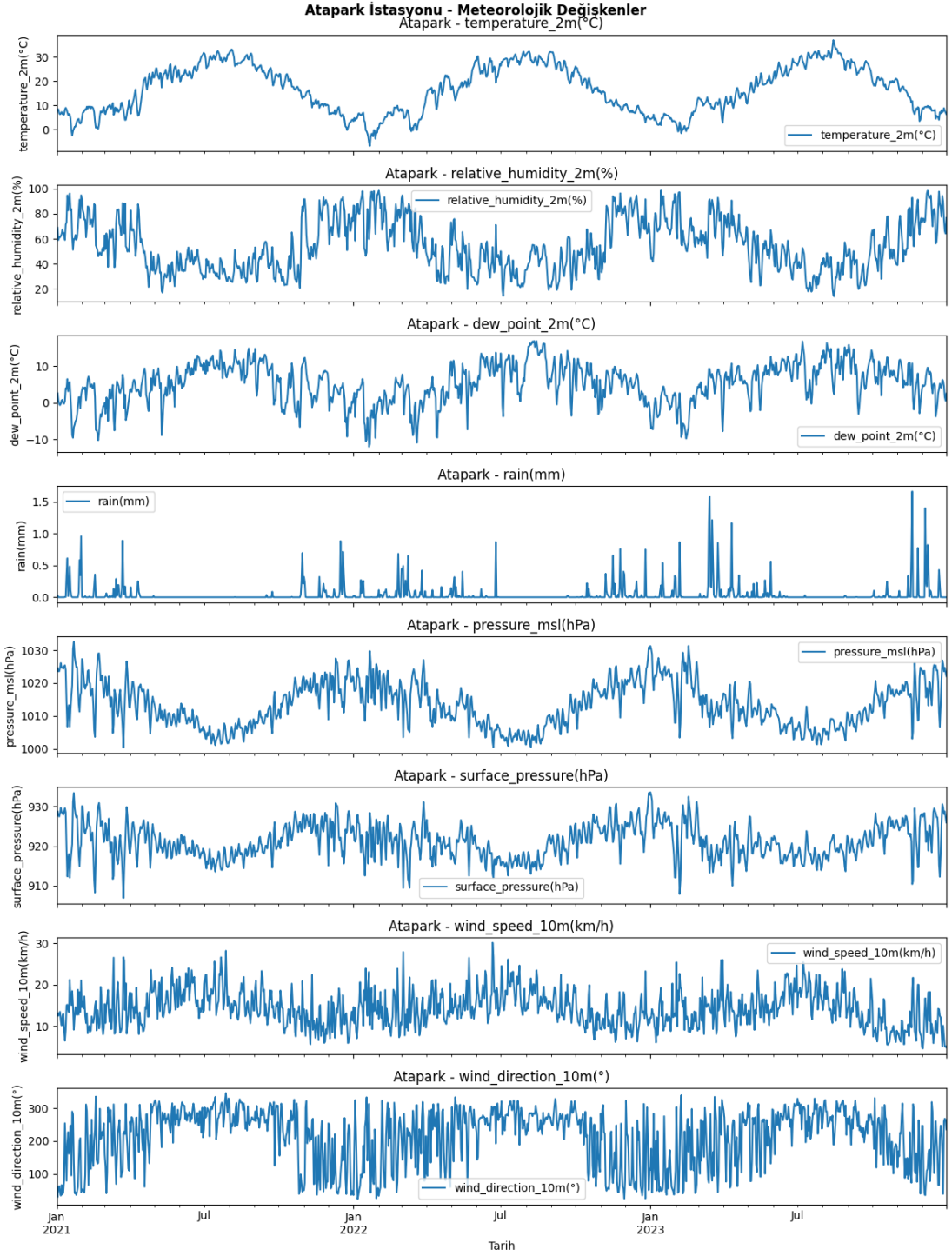
Meteorolojik veriler, Atapark, Beydilli ve Nizip istasyonları için sağlanarak zamansal modelleme sürecinde kullanılmıştır. Fevzi Çakmak, GASKİ ve Meteoroloji istasyonları mekânsal modelleme kapsamında değerlendirilmiş ve ilk üç istasyondan elde edilen bilgilerle bu istasyonlar için eksik veriler tahmin edilmiştir.

Best Match algoritması, konuma ve analiz ihtiyacına göre en uygun reanalysis veri setini seçmektedir. IFS modeli (9 km çözünürlük) saatlik tahmin sağlarken, ERA5 modeli (25 km çözünürlük) geniş ölçekli analizler için günlük olarak güncellenmektedir [73] [75]. Bu çalışmada, yerel ölçekli tahmin doğruluğunu artırmak amacıyla Best Match algoritması kullanılmış ve en uygun çözünürlüğe sahip meteorolojik model otomatik olarak belirlenmiştir. Best Match algoritmasının seçilme nedeni, farklı reanalysis veri setlerini karşılaştırarak bölgenin coğrafi ve iklimsel özelliklerine en uygun meteorolojik modeli otomatik olarak belirlemesidir. IFS modeli, yüksek uzaysal çözünürlüğü (9 km) ve saatlik güncellenen tahminleri ile kısa vadeli hava tahminleri için üstün doğruluk sağladığı için tercih edilmiştir. Alternatif olarak kullanılacak ERA5 ve ERA5-Land veri setleri, geniş ölçekli iklim analizleri için uygundur, ancak bu çalışmada yerel düzeyde hassas modelleme yapılması gerektiği için Best Match algoritmasının sunduğu veriler daha uygun bulunmuştur. Bunun yanı sıra, Best Match algoritması, kullanıcılarına açık kaynaklı, erişilebilir ve sürekli güncellenen veriler sunarak, çalışmanın tekrarlanabilirliğini ve veri sürekliliğini sağlamaktadır.

Çalışmada kullanılan meteorolojik verilerin zaman içindeki değişimi, her bir istasyon için ayrı ayrı incelenmiştir. Atapark, Beydilli ve Nizip istasyonlarında ölçülen sıcaklık, bağıl nem, çiy noktası sıcaklığı, yağış miktarı, basınç, yüzey basıncı, rüzgar hızı ve rüzgar yönü değişkenleri için zaman serisi grafikleri oluşturulmuştur. 5.3'te, 5.4'te ve 5.5'te sırasıyla Atapark, Beydilli ve Nizip istasyonlarındaki meteorolojik değişkenlerin zaman içindeki değişimleri verilmiştir. Bu görseller, istasyonlar arasındaki meteorolojik benzerlikleri ve farklılıkları analiz etmeye yardımcı olmaktadır.

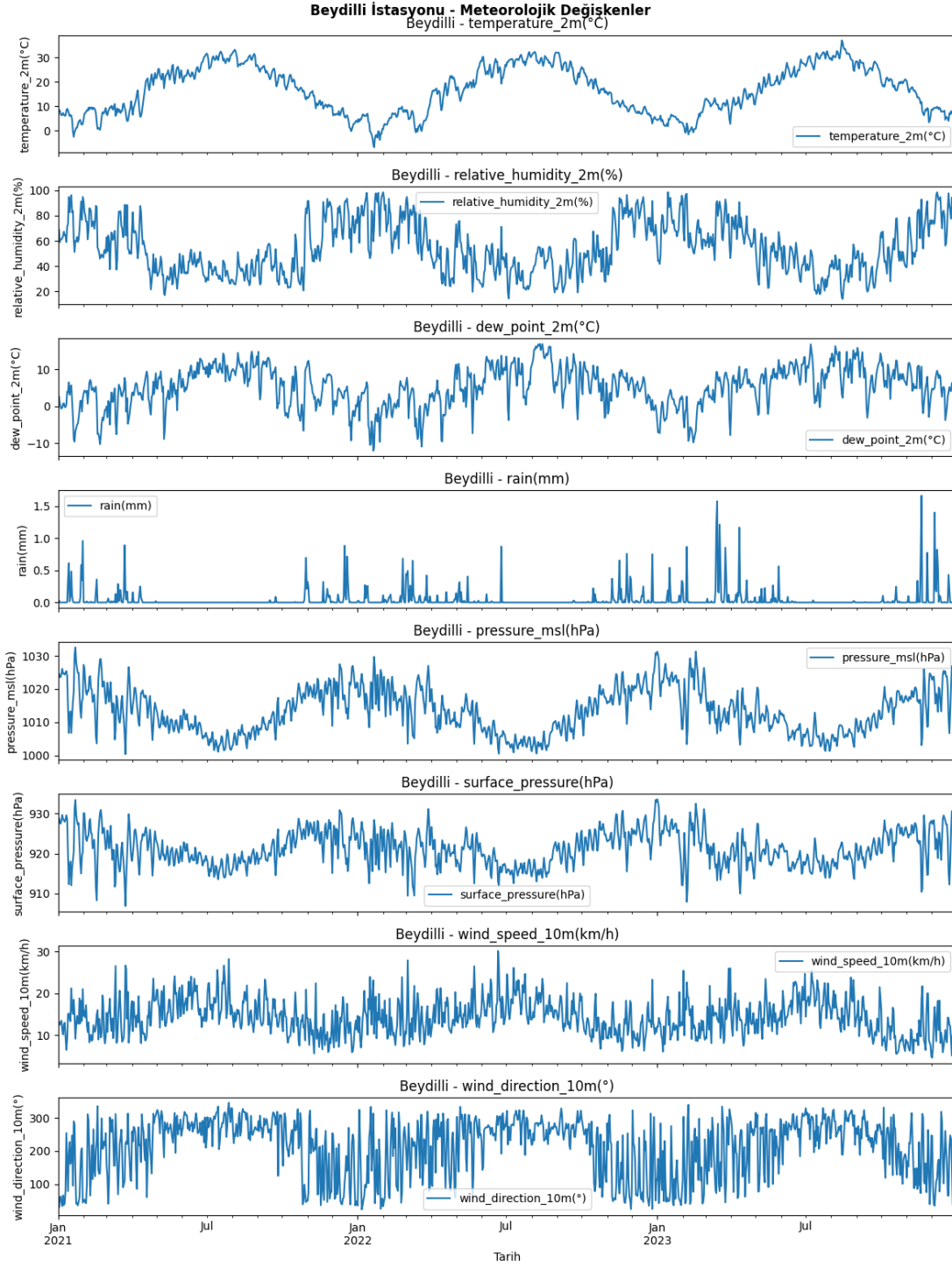
5.1.3 Copernicus Atmosphere Monitoring Service (CAMS) yeniden analiz verileri

Çalışmada eksik PM_{2.5} verilerini tamamlamak ve modelleme sürecine ek destek sağlamak amacıyla Copernicus Atmosphere Monitoring Service (CAMS) hava kalitesi yeniden analiz (reanalysis) verileri kullanılmıştır. CAMS, Avrupa'daki hava kalitesini izlemek ve tahmin etmek amacıyla geliştirilen, yüksek çözünürlüklü küresel ve



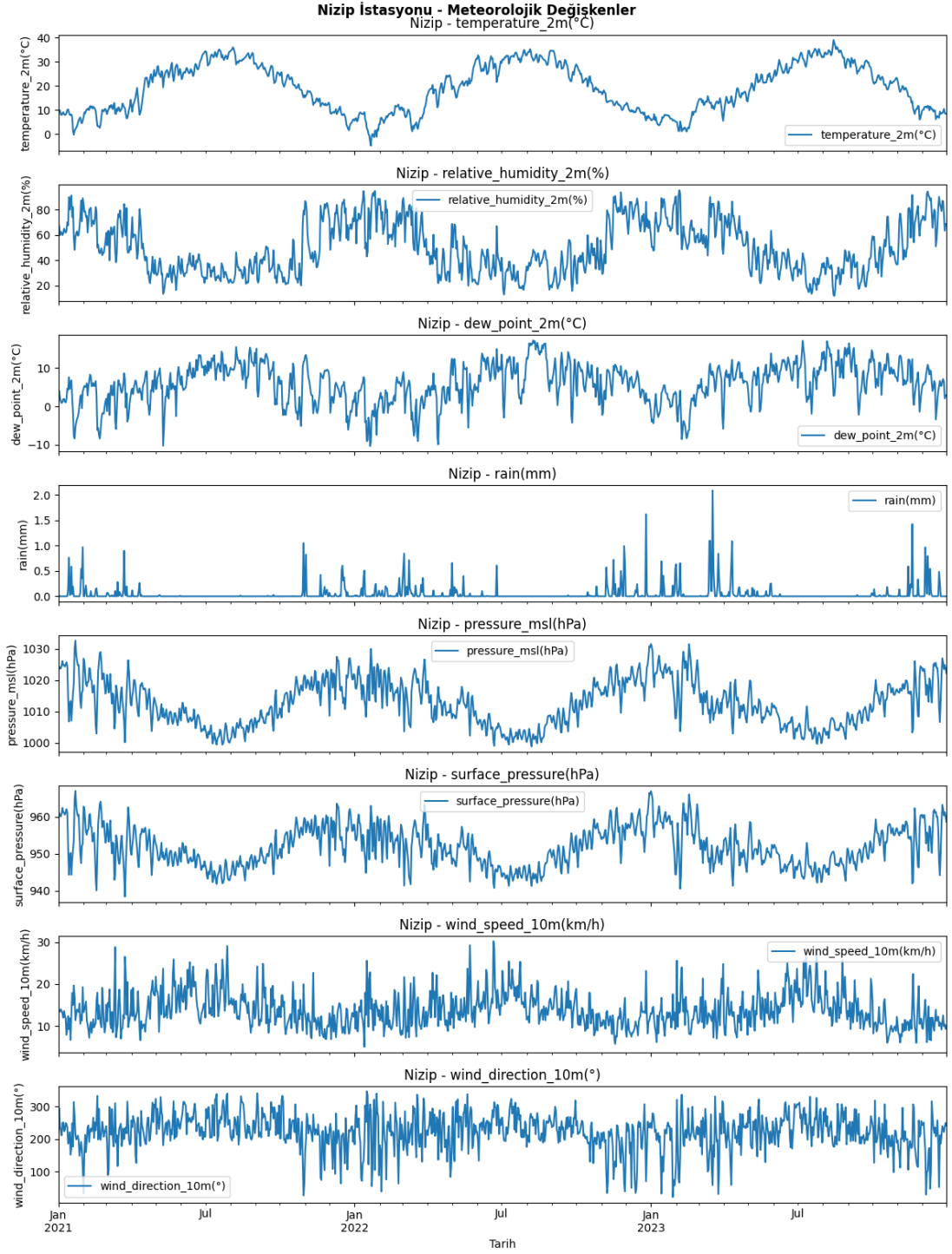
Şekil 5.3: 2021–2023 yılları arasında Atapark istasyonuna ait ortalama meteorolojik değişkenlerin zamansal değişimi.

bölgesel atmosfer modeli tabanlı bir veri kümesidir. Bu veri seti, uydular, yer istasyonları ve atmosferik modellemeye dayalı veri entegrasyonu sağlayarak, yerel



Şekil 5.4: 2021–2023 yılları arasında Beydilli istasyonuna ait ortalama meteorolojik değişkenlerin zamansal değişimi.

ölçümlerin eksik olduğu durumlarda güvenilir bir kaynak olarak değerlendirilmiştir [76]. Kullanılan CAMS verileri 0.1° (~ 10 km x 10 km) uzaysal çözünürlüğe sahip



Şekil 5.5: 2021–2023 yılları arasında Nizip istasyonuna ait ortalama meteorolojik değişkenlerin zamansal değişimi.

olup, her istasyona en yakın grid noktasından alınarak modelleme sürecine dahil edilmiştir. Veriler saatlik çözünürlüğe sahip olmakla birlikte, analiz sürecinde günlük

ortalamalara dönüştürülerek kullanılmıştır. $PM_{2.5}$ konsantrasyonları için referans olarak değerlendirilen CAMS verileri, hava kirliliğinin zaman içindeki değişimini anlamada ve eksik veri noktalarının tamamlanmasında önemli bir rol oynamıştır.

5.1.4 MODIS MAIAC Aerosol Optik Derinlik (AOD) verisi

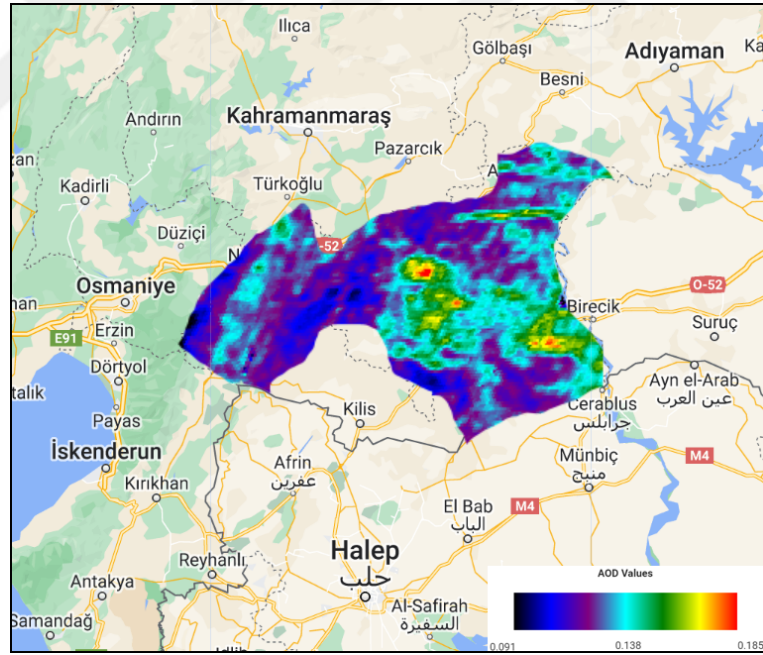
$PM_{2.5}$ tahmin modellerinde önemli bir girdi olan Aerosol Optical Depth (AOD), atmosferdeki partikül yoğunluğunu belirleyen ve uydu tabanlı gözlemlerden elde edilen bir parametredir. Bu çalışmada, AOD verileri MODIS (Moderate Resolution Imaging Spectroradiometer) sensöründen alınmış olup, MODIS/061/MCD19A2_GRANULES veri seti kullanılmıştır. Veriler, Google Earth Engine (GEE) platformu aracılığıyla işlenmiş ve istasyonlara yakın grid hücrelerinden elde edilmiştir. Kullanılan veri seti Optical_Depth_055 bandına dayanmakta olup, 550 nm dalga boyunda AOD ölçümleri sağlamaktadır. MODIS verilerinin uzamsal çözünürlüğü 1 km olup, veriler günlük ölçümlerden hareketle aylık ortalamalar alınarak modellenmiştir. Uydu tabanlı ölçümlerden kaynaklanan hataları minimize etmek için MODIS verileri 0.001 faktörü ile ölçeklendirilmiştir.

Verilerin işleme sürecinde, Gaziantep'te yer alan Atapark, Beydilli ve Nizip istasyonları belirlenmiş ve her istasyon için 5 km yarıçapında tampon alanlar oluşturulmuştur. Bu tampon alanlar, istasyon çevresindeki ortalama AOD değerlerini hesaplamak için kullanılmıştır. Çalışmada kullanılan AOD verileri, bölgedeki hava kalitesini anlamak ve $PM_{2.5}$ tahminlerinde ek bir değişken olarak değerlendirmek amacıyla incelenmiştir. Ancak yapılan korelasyon analizleri, AOD ile $PM_{2.5}$ konsantrasyonları arasında negatif bir ilişki olduğunu göstermiştir. Bu durum, bu çalışma için AOD verilerinin $PM_{2.5}$ tahminlerinde güvenilir bir değişken olarak kullanılamayabileceğini ortaya koymuştur.

Atapark istasyonu için AOD ile $PM_{2.5}$ arasında -0.23, Beydilli için -0.26 ve Nizip için -0.18 değerleri elde edilmiştir. Beydilli AOD ile $PM_{2.5}$ arasında -0.58, Nizip AOD ile $PM_{2.5}$ arasında ise -0.20 gibi belirgin negatif korelasyon değerleri tespit edilmiştir. Özellikle Beydilli istasyonunda AOD ile $PM_{2.5}$ arasında en güçlü negatif ilişki (-0.58) görülmüş olup, bu durum AOD'nin bu bölgede $PM_{2.5}$ konsantrasyonlarının tahmini için güvenilir bir gösterge olmayabileceğini ortaya koymaktadır.

Negatif korelasyonun ortaya çıkmasında çeşitli meteorolojik ve çevresel faktörlerin etkili olabileceği değerlendirilmektedir. Özellikle şiddetli rüzgarların partiküllerin taşınmasına neden olması, AOD ile $PM_{2.5}$ arasındaki doğrudan ilişkinin zayıflamasına yol açabilir. Ayrıca farklı nem oranları, atmosferdeki aerosollerin optik özelliklerini değiştirerek uydu ölçümlerinin doğruluğunu etkileyebilir. Bunun yanı sıra, optik ölçüm hataları ve bölgesel aerosol bileşimi farklılıkları, AOD değerlerinin $PM_{2.5}$ seviyelerini tam olarak yansıtmasını engelleyebilir. Tüm bu etkenler göz önünde bulundurularak, AOD verileri modelleme sürecinde kullanılmamış, ancak atmosferik analizlerin desteklenmesi açısından incelenmiştir.

Çalışmada AOD verilerinin Gaziantep ili için mekansal dağılımı şekil 5.6'da sunulmaktadır. Görseller, Gaziantep ve çevresinde elde edilen AOD değerlerini ve bunların bölgesel dağılımını göstermektedir. Ancak, yukarıda belirtilen nedenlerden dolayı AOD verileri modelleme sürecinden çıkarılmıştır ve tahmin modellerine dahil edilmemiştir.



Şekil 5.6: MODIS uydu verileri kullanılarak oluşturulan Gaziantep ve çevresine ait ortalama AOD (Aerosol Optical Depth) dağılım haritası.

5.1.5 Arazi kullanımı verisi

Arazi kullanımı verileri, hava kirliliğinin mekânsal dağılımını anlamada kritik bir rol oynamaktadır. Bu çalışmada, Gaziantep'te bulunan PM_{2,5} izleme istasyonlarının çevresindeki arazi kullanım özelliklerini belirlemek amacıyla CORINE Land Cover 2018 veri seti kullanılmıştır. Avrupa Çevre Ajansı (European Environment Agency - EEA) tarafından sağlanan bu veri seti, kentsel alanlar, tarım arazileri, ormanlık bölgeler ve su kütleleri gibi farklı arazi sınıflarının mekânsal dağılımını sunmaktadır [77]. Arazi kullanım analizi, her istasyonun çevresindeki yapısal alanlar, tarım alanları, ormanlar ve su kütleleri gibi farklı arazi sınıflarının yüzdesel dağılımını ortaya koymuştur. Çalışmada analiz edilen altı istasyonun çevresindeki arazi kullanım yüzdeleri Çizelge 5.8 'de sunulmuştur. Bu tablo, her istasyonun bulunduğu çevresel yapıyı detaylandırarak, bölgesel farklılıkları daha net bir şekilde göstermektedir.

Çizelge 5.8: İstasyonlara ait yapısal alan, tarım alanı, orman ve su kütlesi yüzdeleri.

İstasyon	Yapısal Alan (%)	Tarım Alanları (%)	Ormanlar (%)	Su Kütleleri (%)
Fevzi Çakmak	70,32	7,68	18,00	4,00
Meteoroloji	52,52	20,00	11,48	16,00
GASKİ	68,16	4,00	25,00	2,84
Atapark	41,08	0,96	49,68	8,28
Beydilli	98,08	0,00	0,00	1,92
Nizip	50,00	40,65	1,61	7,74

Tablodan da görüldüğü üzere, Beydilli istasyonu %98,08 oranında yapısal alanlarla çevrili olup, tamamen kentleşmiş bir bölgede yer almaktadır. Fevzi Çakmak istasyonu, %70,32 oranında yapısal alan içermesine rağmen, %18 oranında ormanlık alan barındırarak daha dengeli bir arazi yapısına sahiptir. Atapark istasyonu, %41,08 yapısal alana sahip olmasına karşın, %49,68 oranında ormanlık alan içerdiğinden, yeşil alanların yoğun olduğu bir konumda yer almaktadır. Nizip istasyonu, %50 oranında yapısal alan ve %40,65 oranında tarım arazisi ile kırsal ve kentsel alanların bir arada bulunduğu bir bölgededir. GASKİ ve Meteoroloji istasyonları, farklı arazi türlerinin karışımına sahiptir. Özellikle Meteoroloji istasyonu, %52,52 yapısal alan ve %20 tarım alanı ile daha dengeli bir dağılıma sahipken, GASKİ istasyonu %25 oranında ormanlık alan barındırarak diğer istasyonlardan farklı bir mekânsal yapıya sahiptir.

Bu dağılım, hava kalitesini etkileyen mekânsal faktörlerin belirlenmesi açısından önemli bir veri kaynağı sunmaktadır. Özellikle yoğun yapısal alan içeren istasyonlarda hava kirliliği seviyelerinin daha yüksek olması beklenirken, ormanlık veya tarım alanlarının bulunduğu bölgelerde atmosferik dağılım süreçleri nedeniyle daha düşük PM_{2.5} seviyeleri gözlemlenebilir. Arazi kullanımı verileri, hava kirliliği tahmin modellerine mekânsal değişkenler eklenerek, istasyon bazlı hava kirliliği dağılımının daha doğru bir şekilde modellenmesine olanak sağlamaktadır.

5.1.6 Nüfus yoğunluğu verisi

Hava kirliliğinin mekânsal dağılımını ve insan sağlığı üzerindeki etkilerini değerlendirmek amacıyla, PM_{2.5} ölçüm istasyonlarının çevresindeki nüfus yoğunluğu verileri analiz edilmiştir. Nüfus yoğunluğu, özellikle kentleşmiş alanlarda hava kirliliği seviyelerini anlamak açısından önemli bir değişken olup, yoğun insan faaliyetlerinin bulunduğu bölgelerde emisyonların daha fazla olmasına neden olmaktadır. Bu nedenle, hava kalitesinin nüfus yoğunluğu ile ilişkisini belirlemek, kirliliğin etkilerini daha iyi anlamaya yardımcı olmaktadır.

Bu çalışmada, Avrupa Komisyonu Ortak Araştırma Merkezi (European Commission, Joint Research Centre – JRC) tarafından sağlanan Global Human Settlement Layer (GHSL) 2020 veri seti kullanılmıştır. GHSL, yüksek çözünürlüklü uydu görüntüleri ve nüfus sayımı verilerini birleştirerek, dünya genelinde detaylı mekânsal nüfus dağılımı sunan kapsamlı bir veri kaynağıdır [78]. Çalışmada, istasyonların bulunduğu 90 metre çözünürlüğe sahip nüfus yoğunluğu verileri kullanılmış ve her istasyon için 1 km²'lik alandaki kişi sayısı belirlenmiştir. İstasyon bazında nüfus yoğunluğu değerleri Çizelge 5.9 'da sunulmuştur.

Çizelge 5.9: PM_{2.5} istasyonlarına ait 1 km² alandaki nüfus yoğunluğu değerleri (kişi/km²).

İstasyon	Nüfus Yoğunluğu (kişi/km ²)
Fevzi Çakmak	87,6
Beydilli	398,2
Atapark	80,3
Nizip	128,5
Meteoroloji	86,03
GASKİ	142,49

Tabloda görüldüğü üzere, Beydilli istasyonu yoğun kentleşmenin bulunduğu bir bölgede yer almakta ve en yüksek nüfus yoğunluğuna sahiptir. Buna karşılık, Atapark ve Meteoroloji istasyonları daha düşük nüfus yoğunluğuna sahip alanlarda konumlanmıştır. Nizip istasyonu, hem kırsal hem de kentleşmiş alanların bulunduğu bir bölgede yer aldığından, orta seviyede bir nüfus yoğunluğu değeri göstermektedir. Bu nüfus yoğunluğu verileri, PM_{2.5} tahmin modellerine mekânsal değişkenler eklenerek, hava kirliliği ve demografik faktörler arasındaki ilişkinin daha iyi anlaşılmasına olanak sağlamaktadır.

5.1.7 NDVI (Bitki Örtüsü İndeksi) verisi

Bitki örtüsü yoğunluğu, atmosferdeki PM_{2.5} seviyelerinin mekânsal dağılımını etkileyen önemli çevresel faktörlerden biridir. Yeşil alanlar, bitki örtüsü aracılığıyla hava kirliliğinin azaltılmasına katkı sağlayarak, partikül maddelerin tutulmasına ve hava kalitesinin iyileştirilmesine yardımcı olmaktadır. Bu nedenle, bitki örtüsünün yoğunluğu ve dağılımını gösteren Normalized Difference Vegetation Index (NDVI), PM_{2.5} tahmin modellerine eklenen mekânsal değişkenlerden biri olarak kullanılmıştır. NDVI, bitki örtüsünün sağlığını ve yoğunluğunu değerlendirmek için kullanılan bir uydu tabanlı indeks olup, farklı dalga boylarındaki yansıma özelliklerinden türetilmektedir. Bu çalışmada NDVI verileri, Sentinel-2 uydusundan elde edilmiş ve Google Earth Engine (GEE) platformu kullanılarak işlenmiştir. Sentinel-2, Avrupa Uzay Ajansı (European Space Agency - ESA) tarafından sağlanan yüksek çözünürlüklü bir uydu görüntüleme sistemi olup, bitki örtüsü dinamiklerini izlemek için yaygın olarak kullanılmaktadır [79].

NDVI değerleri, Beydilli ve Fevzi Çakmak gibi yoğun kentleşmiş alanlarda düşük; Nizip ve Atapark gibi daha yeşil alanlara sahip bölgelerde ise yüksek olduğunu göstermektedir. Bu durum, bitki örtüsü oranı arttıkça PM_{2.5} seviyelerinde potansiyel bir düşüş olabileceğini göstermekte olup, yeşil alanların hava kirliliğini azaltmadaki rolünü vurgulamaktadır.

5.1.8 Rakım verisi

Rakım, hava kirliliği seviyelerinin mekânsal dağılımını etkileyebilecek önemli bir coğrafi faktördür. Yüksek rakımlı bölgelerde hava akımları süreçleri farklılık gösterebilir, bu da $PM_{2.5}$ konsantrasyonlarının dağılımını etkileyebilir. Bu çalışmada, $PM_{2.5}$ ölçüm istasyonlarının rakım verileri modele dahil edilmiş, ancak model performansına olumsuz etkisi olduğu tespit edilmiştir.

Modelleme sürecinde yapılan değerlendirmelerde, rakım değişkeninin tahmin doğruluğunu düşürdüğü görülmüş ve bu nedenle analizlerden çıkarılmıştır. Bu durum, mevcut model yapısının rakım değişkeni ile uyumlu olmadığını göstermektedir. Sonuç olarak, rakım verisi modelleme aşamalarında kullanılmamıştır.

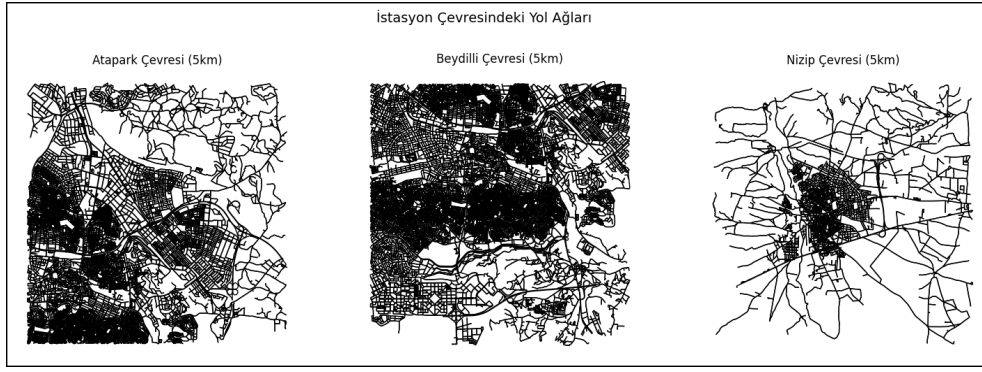
5.1.9 Yol yoğunluğu verisi

Hava kirliliği tahmini çalışmalarında yol yoğunluğu, trafik kaynaklı emisyonların belirlenmesi açısından kritik bir değişkendir. Özellikle yoğun trafik akışına sahip bölgelerde, egzoz gazları ve yol yüzeylerinden kaynaklanan partiküller $PM_{2.5}$ konsantrasyonlarını artırabilmektedir. Bu nedenle, istasyonların çevresindeki yol yoğunluğu analizi yapılarak, trafikle ilişkili hava kirliliğinin mekânsal dağılımı değerlendirilmiştir.

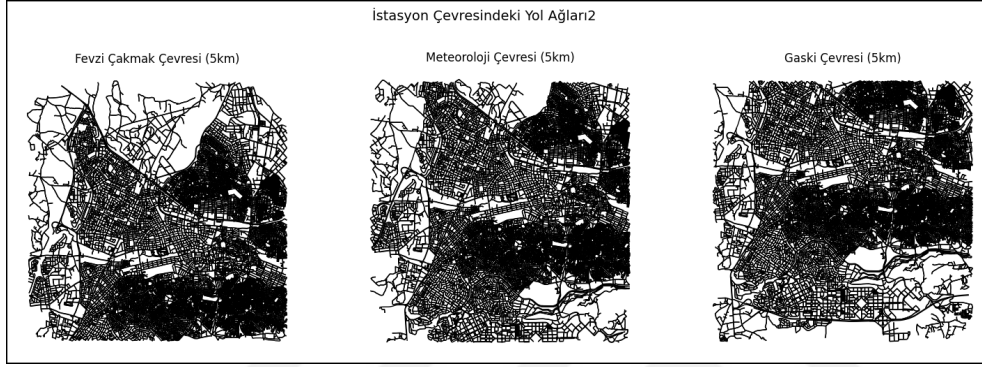
Çalışmada yol yoğunluğu verileri OpenStreetMap (OSM) veritabanından elde edilmiş ve osmnx Python kütüphanesi kullanılarak hesaplanmıştır [80]. OSM, açık kaynaklı bir harita veri sağlayıcısı olup, küresel ölçekte yol ağı bilgilerini içermektedir. Yol ağları, istasyonların çevresindeki 5 km çapındaki bir alan için çıkarılmış ve bu ağların toplam uzunluğu analiz edilmiştir.

Şekil 5.7 ve Şekil 5.8'de istasyonların çevresindeki yol ağı yoğunluğu mekânsal olarak gösterilmiştir. Haritalarda, kentsel alanlar ve kırsal bölgeler arasındaki yol yoğunluğu farkları açıkça gözlemlenebilmektedir.

Yol yoğunluğu, her istasyon için 5 km yarıçapındaki bölgedeki toplam yol uzunluğunun alan büyüklüğüne oranlanmasıyla hesaplanmıştır. İlk olarak, her



Şekil 5.7: Atapark, Beydilli ve Nizip istasyonlarının 5 km çevresindeki yol ağlarının mekânsal dağılımı.



Şekil 5.8: Fevzi Çakmak, Meteoroloji ve GASKİ istasyonlarının 5 km çevresindeki yol ağlarının mekânsal dağılımı.

İstasyonun koordinatları kullanılarak OpenStreetMap'ten (OSM) ilgili bölgenin yol ağı verileri çekilmiş ve ağ bağlantıları (edges) çıkarılmıştır. Daha sonra, 5 km yarıçapındaki tüm yolların toplam uzunluğu hesaplanmış ve bu değer, $\pi \times r^2$ formülü ile hesaplanan alan büyüklüğüne bölünerek, km/km^2 cinsinden yol yoğunluğu elde edilmiştir. OSMnx kütüphanesi kullanılarak gerçekleştirilen hesaplamalar sonucunda, istasyonlara ait yol yoğunluğu değerleri Çizelge 5.10'da verilmiştir.

Çizelge 5.10: İstasyonlara ait 5 km yarıçaplı bölgelerde hesaplanan yol yoğunluğu değerleri.

İstasyon	Yol Yoğunluğu (km/km^2)
Atapark	37,571308
Beydilli	48,978923
Nizip	12,995968
Fevzi Çakmak	48,268640
Meteoroloji	50,479027
GASKİ	51,107559

Elde edilen yol yoğunluğu verilerine göre, GASKİ istasyonu en yüksek yol yoğunluğuna sahipken, Nizip istasyonu en düşük yol yoğunluğuna sahiptir. Bu durum, GASKİ'nin kentsel bir bölgede, yoğun yol ağına sahip bir alanda konumlanmasıyla açıklanabilirken, Nizip'in daha düşük nüfus yoğunluğuna ve kırsal bir bölgede yer almasına bağlıdır.

5.1.10 Sanayi ağırlıklı mesafe verisi

Sanayi tesisleri, hava kirliliğinin en önemli kaynaklarından biri olup, özellikle PM_{2.5} emisyonlarının yoğun olarak salındığı bölgeler arasındadır. Bu nedenle, Gaziantep'teki hava kalitesi ölçüm istasyonlarının en yakın organize sanayi bölgelerine (OSB) olan mesafeleri hesaplanarak, sanayi kaynaklı emisyonların PM_{2.5} konsantrasyonlarına etkisi değerlendirilmiştir.

Her istasyonun en yakın organize sanayi bölgesine (OSB) olan mesafesi, coğrafi koordinat sistemi kullanılarak hesaplanmıştır. OSB'lerin ve hava kalitesi istasyonlarının enlem ve boylam koordinatları belirlenmiş ve bu noktalar arasındaki mesafeler, Geopy kütüphanesinin geodesi fonksiyonu ile hesaplanmıştır. Bu yöntem, iki nokta arasındaki küresel yüzey üzerindeki en kısa mesafeyi (büyük çember mesafesi) belirlemek için kullanılmaktadır.

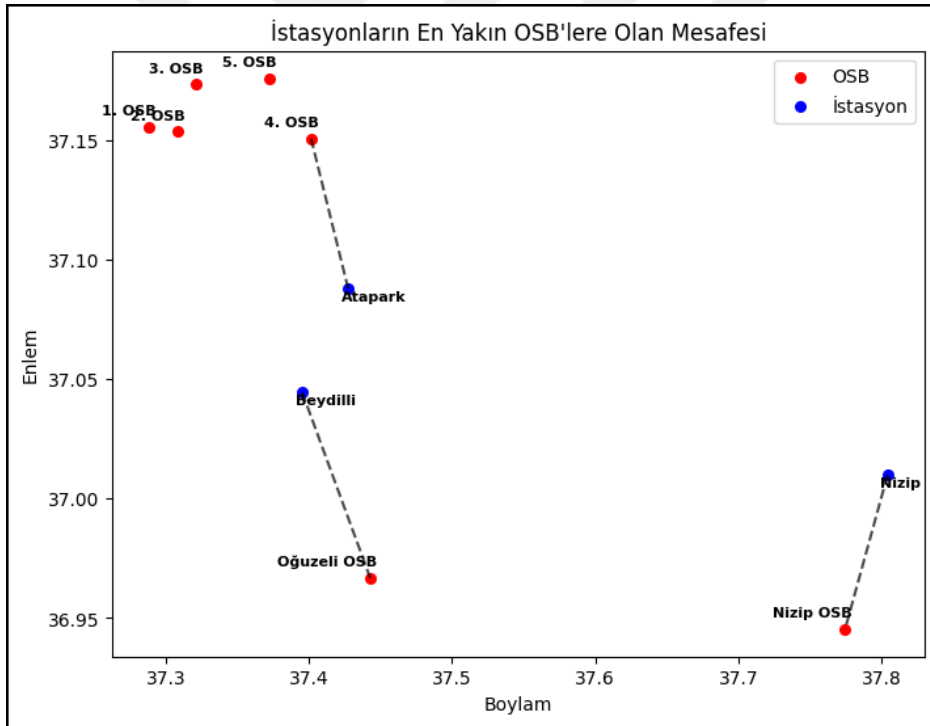
İlk olarak, Gaziantep il sınırları içindeki beş büyük OSB ile Nizip ve Oğuzeli ilçelerindeki OSB'ler referans alınarak konumları belirlenmiştir. Daha sonra, her hava kalitesi ölçüm istasyonu için tüm OSB'lere olan mesafeler hesaplanmış ve en kısa mesafeye sahip OSB, ilgili istasyona "en yakın OSB" olarak atanmıştır.

Ancak, sanayi tesislerinin hava kalitesi üzerindeki etkisi yalnızca en yakın OSB'ye olan mesafe ile sınırlı değildir. Sanayi kaynaklı emisyonlar, farklı yönlerdeki birden fazla OSB'nin etkisi altında olabilir. Bu nedenle, sanayi bölgelerinin etkisini daha bütüncül değerlendirebilmek adına, ağırlıklı mesafe yöntemi kullanılmıştır. Ağırlıklı mesafe hesaplanırken, her istasyonun birden fazla OSB'ye olan mesafeleri dikkate alınmış ve bu mesafeler ters ağırlıklandırma yöntemi ile işlenmiştir.

Yakın OSB'lerin etkisi daha yüksek olacak şekilde, uzak OSB'lerin etkisi daha düşük olacak şekilde ağırlıklandırma yapılmıştır. Ağırlıklı mesafe hesaplamasında aşağıdaki formül kullanılmıştır:

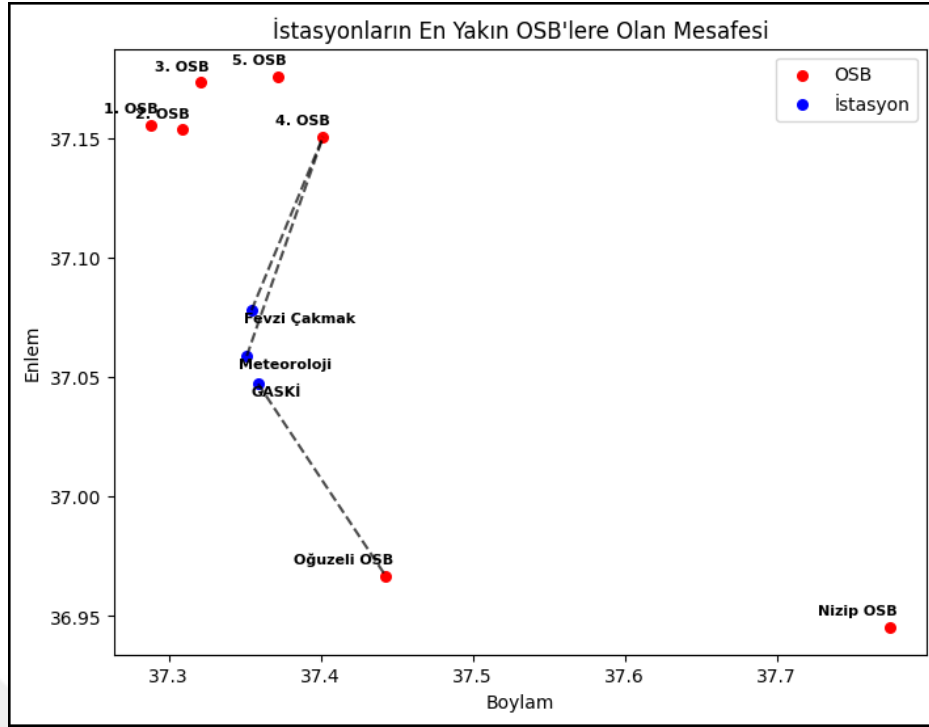
$$D_w = \frac{\sum_{i=1}^n \frac{1}{d_i}}{\sum_{i=1}^n \frac{1}{d_i^2}} \quad (5.1)$$

Burada D_w ağırlıklı mesafeyi, d_i her bir OSB'ye olan mesafeyi, n ise hesaplamada dikkate alınan OSB sayısını ifade etmektedir. Bu yöntem sayesinde, istasyonların sanayi kaynaklı hava kirliliğine maruziyeti daha hassas bir şekilde modellenmiştir. Böylece, yalnızca en yakın OSB değil, tüm OSB'lerin istasyonlar üzerindeki etkisi mekânsal olarak değerlendirilmiştir. Şekil 5.9 ve Şekil 5.10'da, hava kalitesi ölçüm istasyonları ile organize sanayi bölgeleri (OSB) arasındaki mesafeler mekânsal olarak gösterilmektedir.



Şekil 5.9: Atapark, Beydilli ve Nizip istasyonlarının en yakın OSB'ler arasındaki mekânsal mesafelerin görselleştirilmesi.

Görselleştirmede, istasyonlar mavi, OSB'ler ise kırmızı renkte gösterilmiş olup, her istasyonun en yakın OSB'sine olan mesafesi kesikli çizgiler ile bağlanarak haritalandırılmıştır. Bu görseller aracılığıyla, istasyonların sanayi bölgelerine olan



Şekil 5.10: Fevzi Çakmak, Meteoroloji ve GASKİ istasyonlarının en yakın OSB'ler arasındaki mekânsal mesafelerin görselleştirilmesi.

yakınlığı mekânsal olarak değerlendirilmiş ve bu durumun hava kirliliği üzerindeki etkileri analiz edilmiştir.

5.2 Veri Ön İşleme

5.2.1 Eksik veri tamamlanması

5.2.1.1 Eksik sensör verilerinin tamamlanması

Hava kalitesi ölçüm istasyonlarından elde edilen $PM_{2.5}$ verileri, sensör arızaları, bakım çalışmaları veya veri iletim problemleri gibi çeşitli nedenlerle eksik gözlemler içermektedir. Eksik verilerin doğrudan çıkarılması, bilgi kaybına yol açabileceğinden ve model performansını olumsuz etkileyebileceğinden, bu çalışmada eksik veri tamamlama süreci uygulanmıştır.

Bu kapsamda, Zincirleme Denklemlerle Çoklu Atama Multiple Imputation by Chained Equations- MICE) yöntemi kullanılarak eksik $PM_{2.5}$ değerleri tamamlanmıştır. MICE

yöntemi, eksik verileri iteratif bir regresyon süreciyle tahmin ederek tamamlamakta ve verinin istatistiksel yapısını koruyarak bilgi kaybını en aza indirmektedir.

Çalışmada kullanılan istasyonlardaki eksik PM_{2.5} oranları incelenmiş olup, eksik veri sayısı ve oranları Çizelge 5.11’de verilmiştir.

Çizelge 5.11: İstasyonlara ait PM_{2.5} verilerindeki eksik gözlem sayısı ve oranları.

İstasyon	Eksik Veri Sayısı	Eksik Veri Yüzdesi (%)
Atapark	255	23,29
Beydilli	191	17,44
Nizip	548	50,05
Fevzi Çakmak	278	28,14
Meteoroloji	Veri Yok	100
GASKİ	Veri Yok	100

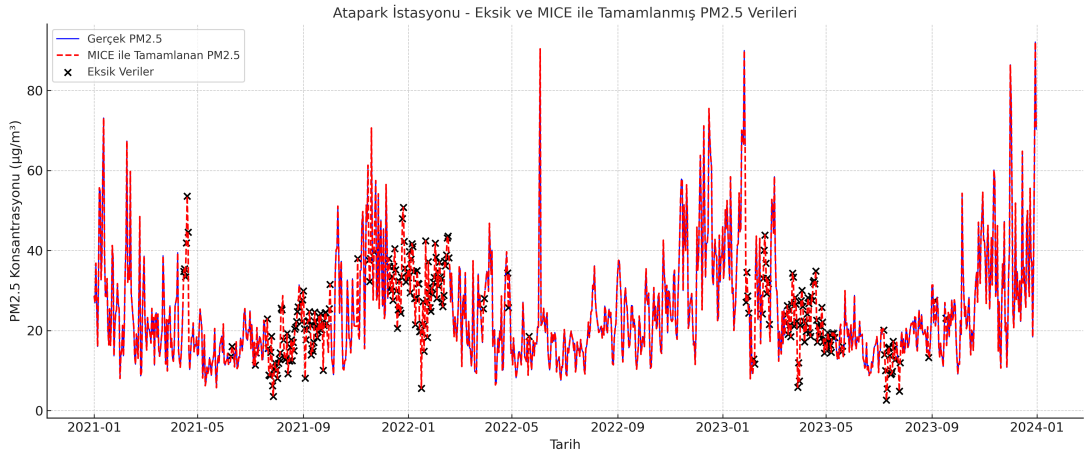
Buna göre, Nizip (%50,05) ve Fevzi Çakmak (%28,14) istasyonlarında eksik veri oranlarının daha yüksek olduğu belirlenmiştir. Bu durum, özellikle yüksek eksik veri oranına sahip istasyonlarda uygun bir tamamlama yönteminin kullanılmasını gerektirmiştir. MICE yöntemi, eksik PM_{2.5} değerlerini tahmin etmek için meteorolojik değişkenler ve uydu tabanlı hava kalitesi verileriyle desteklenmiştir. Eksik verilerin tamamlanmasında aşağıdaki değişkenler kullanılmıştır: CAMS PM_{2.5} konsantrasyonu, sıcaklık, bağıl nem, çiy noktası sıcaklığı, yağış, deniz seviyesi basıncı, yüzey basıncı, rüzgar hızı ve rüzgar yönü. Bu değişkenler, PM_{2.5} seviyeleri ile doğrudan ilişkili olup eksik gözlemlerin tahmin edilmesinde modelin doğruluğunu artırmak amacıyla modele dahil edilmiştir.

MICE yöntemi iteratif bir şekilde uygulanmış olup, ilk aşamada eksik değerler geçici bir yöntemle doldurulmuş, ardından her eksik gözlem diğer değişkenlere bağlı olarak tahmin edilerek güncellenmiştir. Bu işlem, belirli bir eşik değere ulaşıncaya kadar devam ettirilmiş ve sonrasında tamamlanan veri setinin analizi gerçekleştirilmiştir. Tamamlanan PM_{2.5} değerlerinin zaman serisi analizleri yapılmış ve eksik verilerin tamamlanan verilerle istatistiksel olarak uyumlu olup olmadığı incelenmiştir. MICE yöntemi ile tamamlanan verilerin, orijinal veri dağılımına büyük ölçüde benzer bir yapı gösterdiği gözlemlenmiştir.

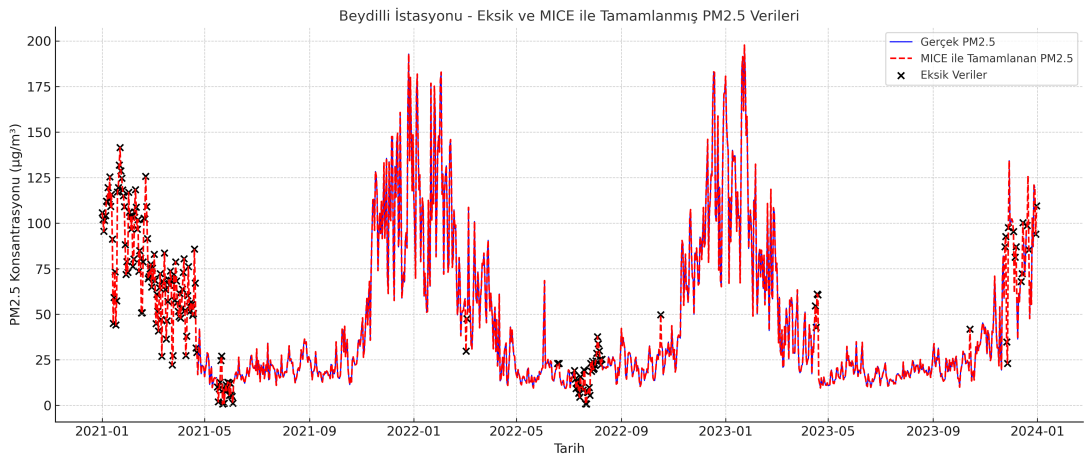
Eksik verilerin tamamlanmasının ardından istasyon bazında elde edilen model başarısı R² değerleri ile değerlendirilmiş olup, Atapark (%63,58), Beydilli (%87,34), Nizip

(%71,83) ve Fevzi Çakmak (%77,20) istasyonları için tamamlanan verilerin modelle uyumlu olduğu belirlenmiştir. Bu sonuçlar, MICE yönteminin özellikle yüksek eksik veri oranına sahip istasyonlarda başarılı bir şekilde uygulanabildiğini ve eksik verilerin güvenilir bir şekilde tamamlanabildiğini göstermektedir.

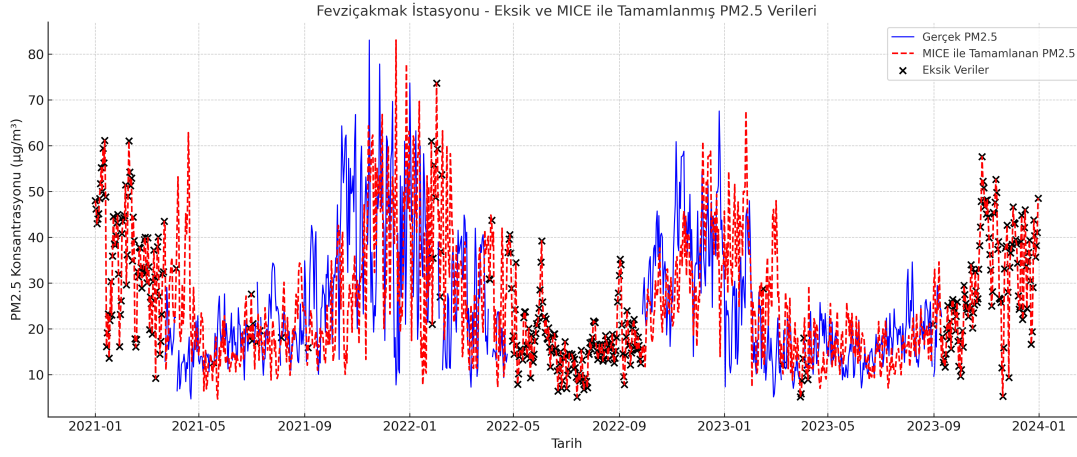
Sonuçların görselleştirilmesi amacıyla eksik PM_{2.5} verilerinin tamamlanma süreci ve elde edilen sonuçların değerlendirilmesi amacıyla çeşitli görselleştirmeler yapılmıştır. Şekil 5.11, Şekil 5.12, Şekil 5.13 ve Şekil 5.14'de eksik PM_{2.5} verilerinin zaman serisi grafikleri sunulmuştur. Bu grafiklerde mavi çizgiler gerçek PM_{2.5} değerlerini, kırmızı kesikli çizgiler tamamlanan verileri ve siyah çarpılar (X) eksik gözlemleri temsil etmektedir.



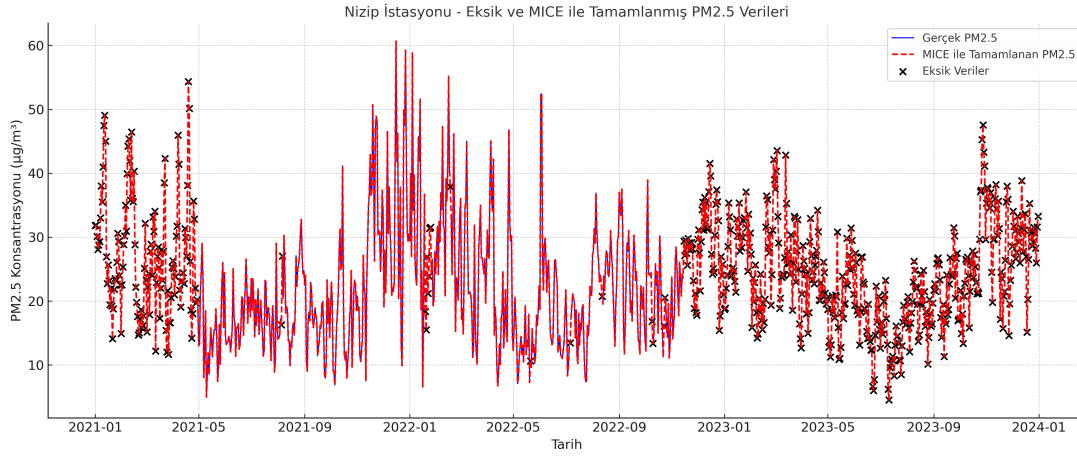
Şekil 5.11: Atapark istasyonu için eksik ve tamamlanmış PM_{2.5} verilerinin zaman serisi gösterimi.



Şekil 5.12: Beydilli istasyonu için eksik ve tamamlanmış PM_{2.5} verilerinin zaman serisi gösterimi.



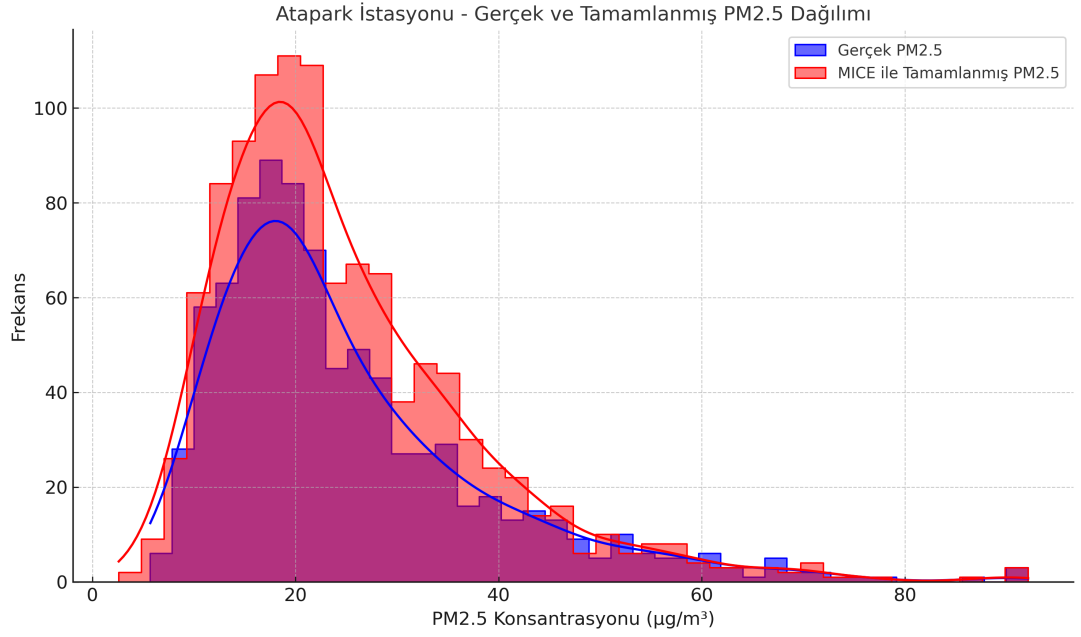
Şekil 5.13: Fevzi Çakmak istasyonu için eksik ve tamamlanmış PM2.5 verilerinin zaman serisi gösterimi.



Şekil 5.14: Nizip istasyonu için eksik ve tamamlanmış PM2.5 verilerinin zaman serisi gösterimi.

Tamamlanan PM_{2.5} verilerinin dağılımı da histogram analizleriyle değerlendirilmiştir. Şekil 5.15, Şekil 5.16, Şekil 5.17 ve Şekil 5.18'de istasyonlara ait histogram grafikleri sunulmuştur. Mavi histogramlar gerçek verileri, kırmızı histogramlar ise MICE yöntemi ile tamamlanan verileri göstermektedir.

Sonuç olarak, MICE yöntemi ile eksik PM_{2.5} verileri başarılı bir şekilde tamamlanarak modelleme için eksiksiz bir veri seti oluşturulmuştur. Böylece, eksik veriler nedeniyle oluşabilecek bilgi kaybı önlenmiş ve hava kirliliği tahmin modelinin daha sağlıklı sonuçlar üretebilmesi sağlanmıştır. Tüm istasyonların eksik verilerinin tamamlanmış olduğu PM_{2.5} veri setinin zamansal dağılım grafiği şekil 5.19'da verilmiştir.



Şekil 5.15: Atapark istasyonunda tamamlanan PM_{2.5} verilerinin histogram analizi.

5.2.1.2 Eksik NDVI (Bitki Örtüsü İndeksi) verilerinin tamamlanması

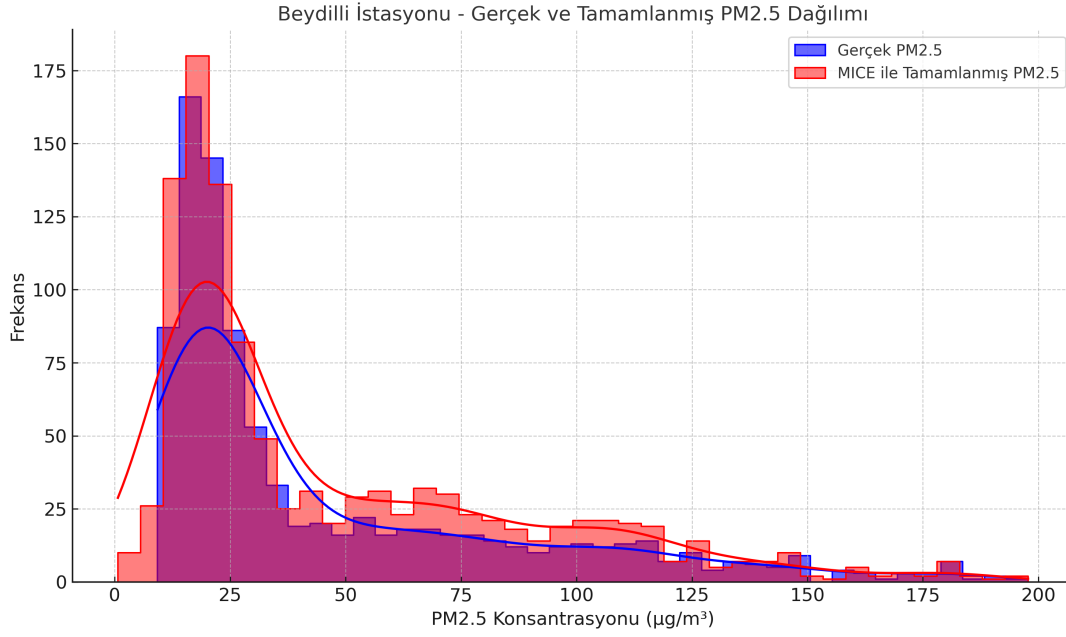
Bitki örtüsü yoğunluğunu ifade eden Normalized Difference Vegetation Index (NDVI) verileri, uydu gözlemlerine dayandığından bazı zaman dilimlerinde eksiklik gösterebilmektedir.

Eksik verilerin oluşmasının temel nedenleri arasında, yoğun bulut örtüsü, sensör arızaları ve atmosferik koşulların etkisi yer almaktadır. Bu çalışmada, NDVI verilerinde bulunan eksiklikleri gidermek amacıyla interpolasyon yöntemi kullanılmıştır.

Eksik verileri tamamlamak için zaman serisi interpolasyonu uygulanmıştır. Bu yöntem, eksik değerleri zaman içindeki eğilimleri takip ederek tahmin etmeye olanak tanır. Özellikle lineer interpolasyon kullanılarak, eksik olan günlerdeki NDVI değerleri, önceki ve sonraki gözlem noktalarına dayanarak tahmin edilmiştir. Interpolasyon işlemi sonrasında, eksik NDVI değerlerinin tamamlandığı görülmüş ve böylece tüm istasyonlar için eksiksiz bir veri seti elde edilmiştir.

5.2.2 Aykırı veri tespiti

Hava kalitesi verilerinde sensör hataları, ani meteorolojik değişimler veya aşırı hava olayları nedeniyle olağan dışı gözlemler görülebilmektedir. Bu tür veriler, modelleme



Şekil 5.16: Beydilli istasyonunda tamamlanan PM_{2.5} verilerinin histogram analizi.

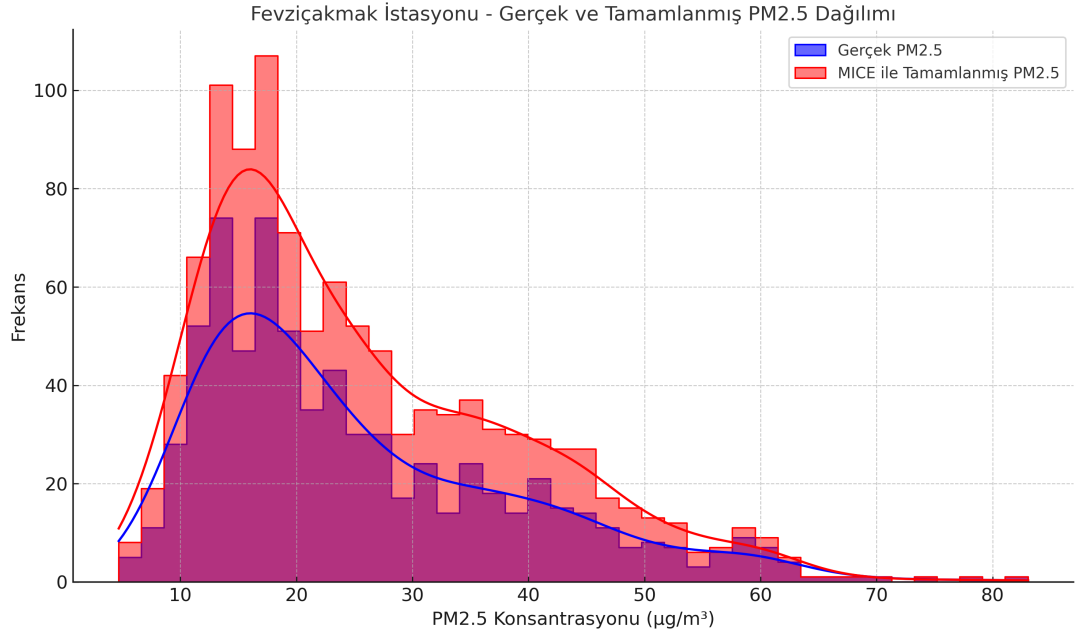
sürecinin doğruluğunu olumsuz etkileyebileceğinden, aykırı değerlerin belirlenmesi ve uygun şekilde işlenmesi gerekmektedir.

Bu süreçte, öncelikle verilerin normal dağılıma uygun olup olmadığı Shapiro–Wilk Testi ile incelenmiştir. Shapiro–Wilk Testi, özellikle küçük ve orta ölçekli veri setlerinde normallik analizinde yaygın olarak kullanılan bir yöntem olup, $p < 0.05$ olduğu durumlarda ilgili değişkenin normal dağılıma uymadığı kabul edilmektedir.

Yapılan testler sonucunda, çoğu değişken için $p < 0.05$ olduğu belirlenmiş ve verilerin normal dağılım göstermediği tespit edilmiştir. Yalnızca nüfus yoğunluğu, yapısal alan, tarım alanları, ormanlar, su kütleleri, enlem, boylam ve istasyon mesafeleri gibi sabit değişkenlerin normal dağılıma uygun olduğu gözlemlenmiştir.

Verilerin normal dağılım göstermemesi nedeniyle, aykırı değerlerin tespitinde Çeyrekler Arası Açıklık (Interquartile Range - IQR) yöntemi tercih edilmiştir. IQR yöntemi, verilerin çarpık (skewed) veya normal olmayan bir dağılıma sahip olduğu durumlarda güvenilir sonuçlar sağlaması nedeniyle tercih edilmiştir.

IQR yöntemi, veri setindeki birinci çeyrek (Q_1) ve üçüncü çeyrek (Q_3) değerlerini esas alarak değişkenlerin normal sınırlarını belirler. Aykırı değerler aşağıdaki formüller kullanılarak hesaplanmıştır:



Şekil 5.17: Fevzi Çakmak istasyonunda tamamlanan PM_{2.5} verilerinin histogram analizi.

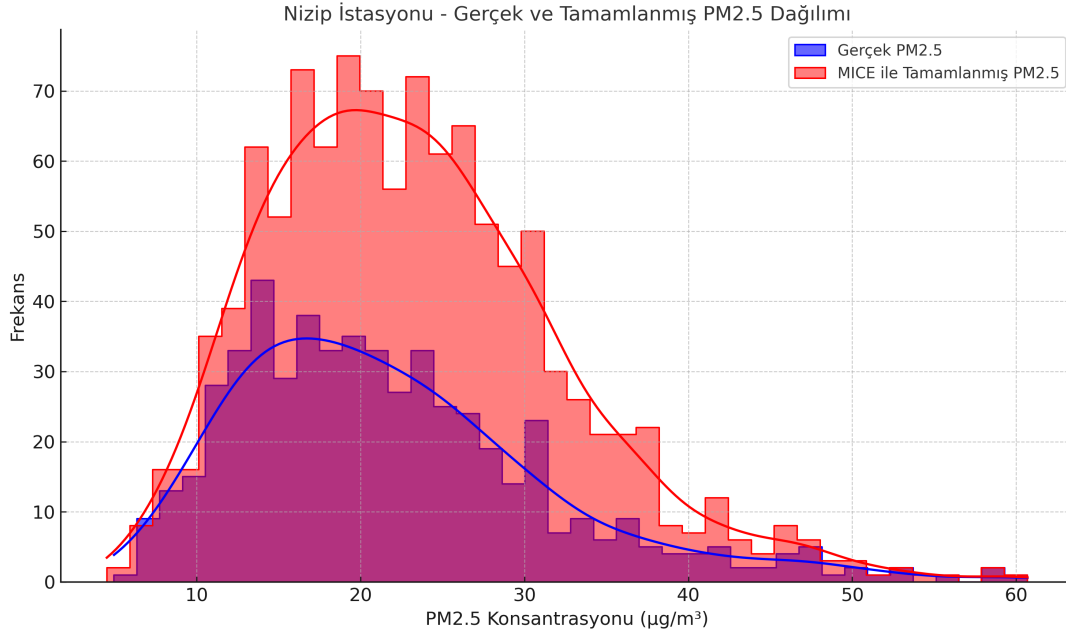
$$IQR = Q_3 - Q_1 \quad (5.2)$$

$$\text{Alt Sınır} = Q_1 - 1.5 \times IQR \quad (5.3)$$

$$\text{Üst Sınır} = Q_3 + 1.5 \times IQR \quad (5.4)$$

Bu sınırların dışında kalan veriler aykırı değer olarak değerlendirilmiştir. IQR yöntemi ile yapılan analiz sonucunda, PM_{2.5} ve meteorolojik değişkenlerde belirli ölçüde aykırı değerler tespit edilmiştir. Tespit edilen aykırı değerler, her istasyon için ayrı ayrı hesaplanarak Çizelge 5.12’de sunulmuştur.

Analiz sonucunda, özellikle yağış değişkeninde aşırı aykırı değerler olduğu tespit edilmiştir. Verilerin büyük bir kısmı “0” değerinden oluşurken, bazı zaman dilimlerinde olağanüstü yüksek yağış miktarları gözlemlenmiştir. Bu durum, istasyonlar arasında veri dengesizliği yaratabileceği için, yağış değişkeni analizden çıkarılmıştır.



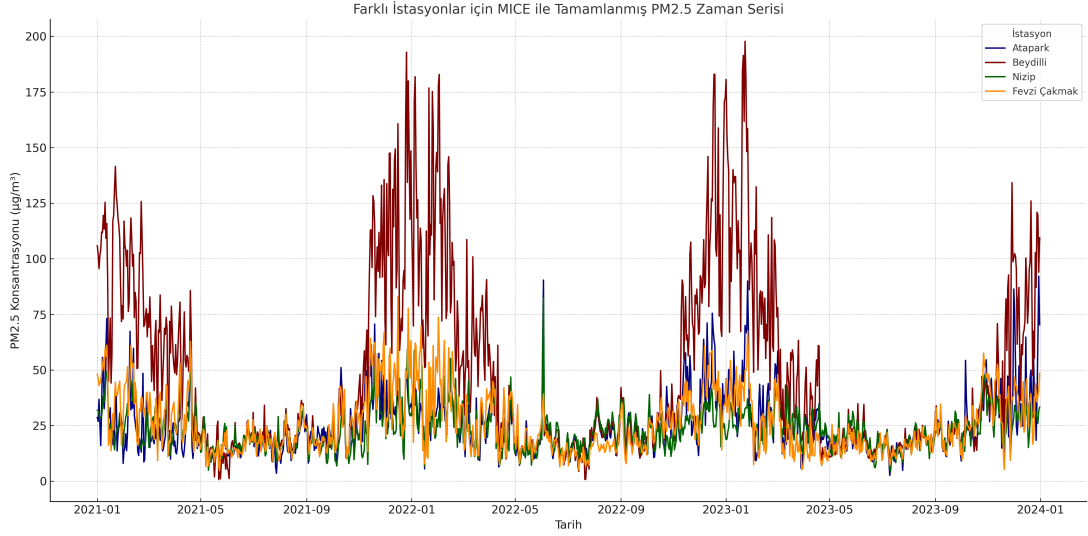
Şekil 5.18: Nizip istasyonunda tamamlanan PM_{2.5} verilerinin histogram analizi.

Çizelge 5.12: İstasyonlara göre aykırı değerlerin sayısı.

İstasyon	Değişken	Aykırı Değer Sayısı
Atapark	Sensör PM _{2.5}	43
	CAMS PM _{2.5}	25
	Yağış	237
Beydilli	Sensör PM _{2.5}	26
	CAMS PM _{2.5}	32
	Yağış	245
Nizip	Sensör PM _{2.5}	22
	CAMS PM _{2.5}	23
	Yağış	236
	Rüzgar Yönü	53

Buna ek olarak, PM_{2.5} (hem sensör hem de CAMS verileri) ve rüzgar yönü değişkenlerinde de aykırı değerler belirlenmiştir. Ancak, bu değişkenlerdeki aykırı değerlerin modelleme sürecindeki önemini koruduğu ve istasyonların farklı coğrafi koşullara sahip olmasından kaynaklanabileceği değerlendirilerek, bu değişkenlerde herhangi bir veri çıkarımı yapılmamış, yalnızca ölçeklendirme uygulanmıştır.

Bu süreç sonucunda, modelleme sürecinde kullanılan veri seti daha homojen bir yapıya kavuşturulmuş; uç değerlerin neden olabileceği çarpıklıklar ve sapmalar minimize edilmiştir. Böylece, istatistiksel modelleme ve makine öğrenimi algoritmalarının daha sağlıklı öğrenme yapabilmesi sağlanmış, tahminleme performansı ve genellenebilirlik



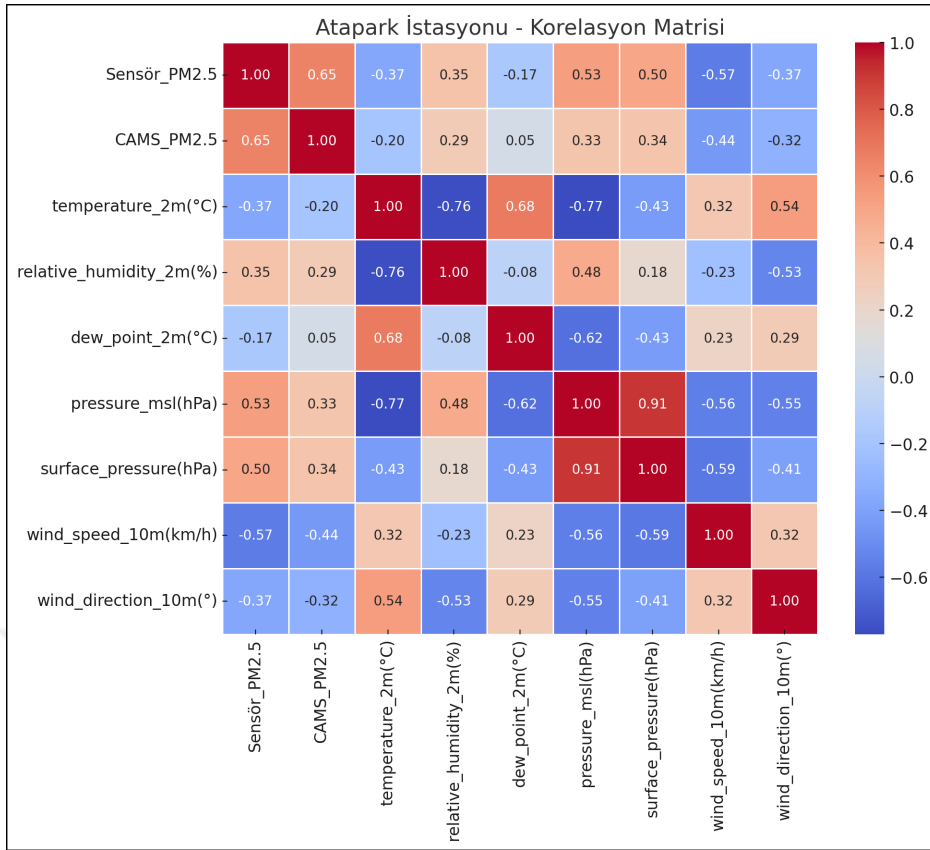
Şekil 5.19: Farklı istasyonlar için PM_{2.5} zaman serisi (2021–2023).

düzeyi anlamlı ölçüde iyileştirilmiştir. Aykırı verilerin etkisinin azaltılması, modelin hem eğitilme sürecindeki kararlılığını hem de farklı veri kümeleri üzerindeki doğruluğunu olumlu yönde etkilemiştir.

5.2.3 Özellik seçimi

Özellik seçimi sürecinde, model doğruluğunu artırmak ve değişkenler arasındaki ilişkileri analiz edebilmek amacıyla korelasyon analizi uygulanmıştır. Korelasyon analizi, değişkenler arasındaki doğrusal ilişkilerin değerlendirilmesinde yaygın olarak kullanılan istatistiksel bir yöntem olup, bu çalışmada Pearson korelasyon katsayısı esas alınmıştır. Her bir hava kalitesi izleme istasyonu için ayrı ayrı oluşturulan korelasyon matrisleri aracılığıyla, hem meteorolojik değişkenlerin kendi aralarındaki ilişkileri hem de bu değişkenlerin PM_{2.5} konsantrasyonu ile olan korelasyonları analiz edilmiştir. Korelasyon matrisleri Şekil 5.20, Şekil 5.21 ve Şekil 5.22’de sunulmuştur.

Pearson korelasyon katsayısı, iki değişken arasındaki doğrusal ilişkinin gücünü -1 ile +1 arasında bir değerle ifade eder. Bu katsayı, +1’e yaklaştıkça değişkenler arasında güçlü bir pozitif ilişki olduğunu, -1’e yaklaştıkça ise güçlü bir negatif ilişki olduğunu gösterir. Katsayının 0’a yakın olması, değişkenler arasında anlamlı bir doğrusal ilişki bulunmadığını ifade eder.

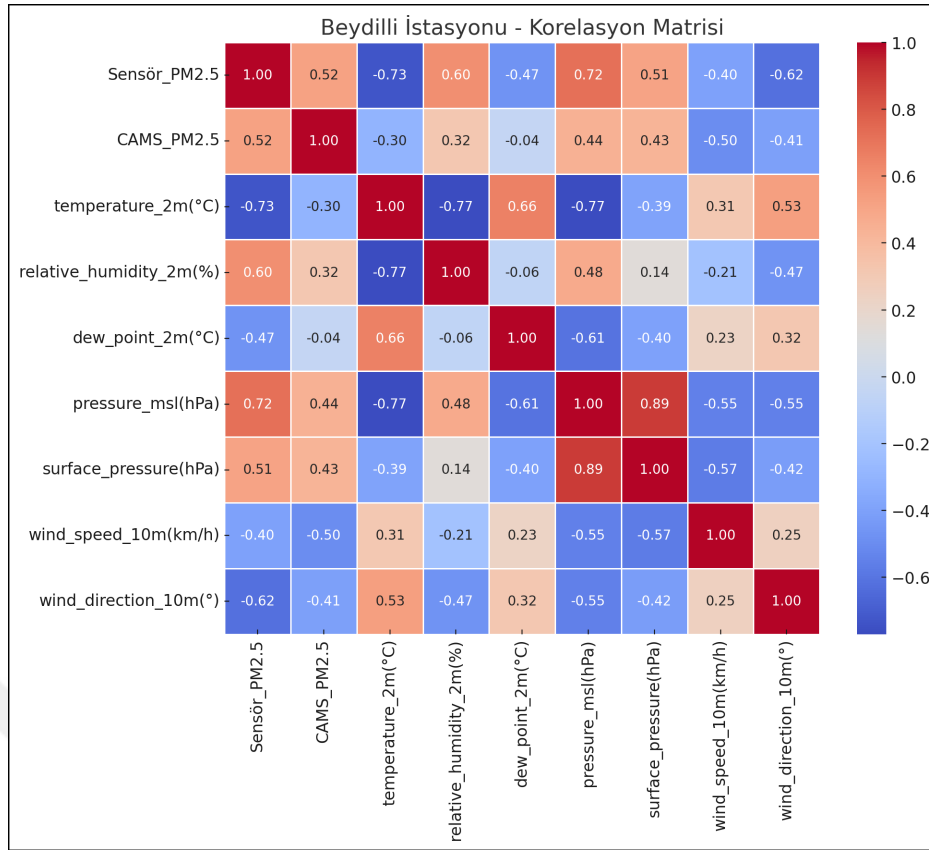


Şekil 5.20: Atapark İstasyonu için değişkenler arası Pearson korelasyon matrisi.

Yürütülen korelasyon analizi doğrultusunda, özellik seçimi sürecinde birbiriyle yüksek düzeyde korelasyon gösteren değişkenlerden yalnızca biri modele dahil edilmiştir. Örneğin, deniz seviyesine indirgenmiş basınç ile yüzey basıncı değişkenleri arasında oldukça yüksek bir korelasyon gözlemlenmiş, bu sebeple yalnızca deniz seviyesi basıncı değişkeni modele dahil edilerek yüzey basıncı çıkarılmıştır. Benzer şekilde, sıcaklık ile çiy noktası sıcaklığı değişkenleri arasındaki yüksek korelasyon nedeniyle, çiy noktası sıcaklığı modelden çıkarılmıştır. Bu tercih, sıcaklık değişkeninin çiy noktası sıcaklığını yeterli ölçüde temsil ettiği varsayımına dayanmaktadır.

PM_{2.5} konsantrasyonu ile yüksek korelasyon gösteren değişkenler ise modelde tutulmuştur. Özellikle sıcaklık, bağıl nem, rüzgar hızı ve basınç değişkenleri, hava kirliliğinin atmosferik koşullara bağlı değişimini açıklayabilme potansiyelleri nedeniyle modelleme sürecinde yer almıştır.

Bu sistematik değişken seçimi ile çoklu doğrusal bağlantı (multicollinearity) sorunlarının önüne geçilmesi ve modelin tahmin doğruluğunun artırılması hedeflenmiştir.



Şekil 5.21: Beydilli İstasyonu için değişkenler arası Pearson korelasyon matrisi.

Seçilen bu değişkenler, sonraki aşamada modelleme sürecine hazırlanmak üzere normalizasyon işlemine tabi tutulmuştur.

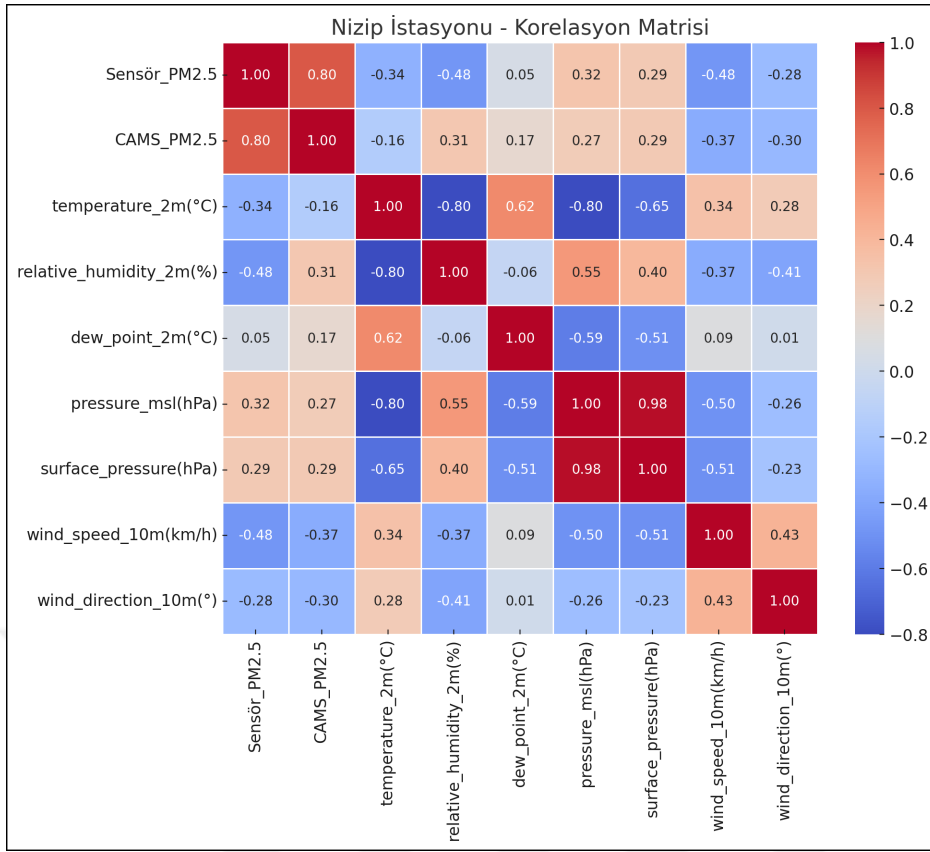
5.2.4 Veri normalizasyonu

Modelleme sürecinde, değişkenlerin farklı ölçüm birimlerine sahip olması model performansını olumsuz etkileyebileceğinden, tüm değişkenler Min-Max ölçeklendirme yöntemi kullanılarak $[0, 1]$ aralığına dönüştürülmüştür.

Min-Max ölçeklendirme yöntemi, her bir değişkenin minimum ve maksimum değerleri arasındaki oranı koruyarak verileri normalize eder. Bu yöntem aşağıdaki formül ile uygulanmıştır:

$$X_{\text{normalized}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (5.5)$$

Bu işlem, özellikle farklı büyüklüklere sahip değişkenlerin aynı ölçek üzerinde değerlendirilmesini sağlamakta ve modelin dengesiz ağırlıklardan etkilenmesini



Şekil 5.22: Nizip İstasyonu için değişkenler arası Pearson korelasyon matrisi.

önmektedir. Normalizasyon işlemi, zamansal ve mekânsal veri setleri için ayrı ayrı gerçekleştirilmiştir. Sabit değerler içeren değişkenler (örneğin enlem, boylam ve istasyon mesafeleri) normalizasyon işlemine tabi tutulmamıştır. Normalizasyonun ardından, tüm değişkenlerin ölçeklenmiş değerleri tekrar kontrol edilerek, modelin giriş değişkenleri için uygun bir yapı oluşturulmuştur.

5.3 Yöntem

Gaziantep ili için $PM_{2.5}$ tahmini yapmak amacıyla gerçekleştirilen bu çalışmada, zamansal ve mekânsal modelleme yaklaşımları birleştirilerek hava kalitesi tahminlerinin doğruluğu artırılmaya çalışılmıştır.

İlk olarak, zamansal modelleme kapsamında geçmiş $PM_{2.5}$ değerleri, hareketli ortalamalar ve çevresel değişkenler kullanılarak tahminler yapılmış; bu süreçte makine öğrenmesi modellerinden XGBoost, LightGBM, Random Forest ve Decision Tree

ile derin öğrenme modellerinden LSTM ve GRU denenmiştir. Hiperparametre ayarlamaları ise Optuna optimizasyon kütüphanesi ile gerçekleştirilmiştir.

Daha sonra, mekânsal modelleme aşamasında istasyonların coğrafi konumları, nüfus yoğunluğu ve sanayi mesafesi gibi mekânsal değişkenler dikkate alınarak istasyonlar arası ilişkiler değerlendirilmiştir. Bu bağlamda, yine XGBoost, LightGBM, Random Forest gibi makine öğrenmesi algoritmalarının yanı sıra mekânsal regresyon tekniklerinden Geographically Weighted Regression (GWR) ve grafik tabanlı öğrenme için Graph Convolutional Network (GCN) modeli uygulanmıştır.

Son olarak, hibrit modelleme yaklaşımı ile zamansal ve mekânsal modellerin çıktıları birleştirilerek genel tahmin performansı iyileştirilmiştir. Bu süreçte zamansal model tahminleri mekânsal modele entegre edilmiş; Autoencoder modeli ile Latent Dependency Factor (LDF) üretilerek eksik verilerin tamamlanması amacıyla hedef istasyonlara transfer edilmiş ve nihai PM_{2.5} tahminleri LightGBM modeli ile elde edilmiştir. Uygulanan bu bütüncül yöntemler, çalışmanın hava kalitesi tahminlerini daha güvenilir ve anlamlı hale getirme amacını desteklemiştir.

5.3.1 Zamansal modelleme

5.3.1.1 Makine öğrenmesi modelleri

Çalışmada PM_{2.5} tahmini yapmak amacıyla XGBoost (XGB), Random Forest (RF), Decision Tree (DT) ve LightGBM (LGBM) olmak üzere dört farklı makine öğrenmesi modeli kullanılmıştır. Bu modeller, geçmiş hava kirliliği seviyeleri ve meteorolojik değişkenler kullanılarak tahminleme amacıyla eğitilmiştir. Aşağıda, her bir modelin istasyon bazlı performans sonuçları ve karşılaştırmalı analizleri sunulmaktadır.

XGBoost, gradient boosting algoritmasına dayalı olarak çalışan ve genellikle yüksek doğruluk oranları elde eden bir modeldir. PM_{2.5} tahmini için Nizip istasyonunda en yüksek başarı oranını yakalamış, ancak diğer istasyonlarda LightGBM modelinin gerisinde kalmıştır. Modelin aşırı öğrenme (overfitting) riskini azaltmak amacıyla Optuna ile optimum hiperparametre ayarları gerçekleştirilmiştir. XGB modelinin istasyon bazlı performans sonuçları Çizelge 5.13'de verilmiştir.

Çizelge 5.13: XGBoost modelinin istasyon bazlı performans metrikleri.

İstasyon	MAE	RMSE	R ²
Atapark	0.0102	0.0221	0.9395
Beydilli	0.0153	0.0263	0.9547
Nizip	0.0050	0.0064	0.9706

XGBoost modeli, özellikle Nizip istasyonunda $R^2 = 0.9706$ değeriyle en başarılı tahmini gerçekleştirmiştir.

Random Forest modeli, birden fazla karar ağacının birleştirilmesiyle çalışan ve genellikle aşırı öğrenmeyi önleme konusunda etkili bir ansambl öğrenme yöntemidir. Bu modelde de aşırı öğrenme riskini azaltmak için Optuna ile hiperparametre optimizasyonu yapılmıştır. RF modelinin istasyonlara göre performans metrikleri Çizelge 5.14’te sunulmuştur.

Çizelge 5.14: Random Forest modelinin istasyon bazlı performans metrikleri.

İstasyon	MAE	RMSE	R ²
Atapark	0.0145	0.0221	0.8832
Beydilli	0.0174	0.0308	0.9393
Nizip	0.0070	0.0088	0.9434

Random Forest modeli, XGBoost’a yakın bir performans sergilemiş ancak genellikle geride kalmıştır. Özellikle Atapark ve Nizip’te hata oranları görece yüksek çıkmıştır.

Decision Tree modeli, veri setindeki değişkenlere dayalı olarak kurallar oluşturarak tahmin yapan temel bir modeldir. Modelin aşırı öğrenme riskini azaltmak amacıyla Optuna ile hiperparametre optimizasyonu gerçekleştirilmiştir. DT modelinin performans metrikleri Çizelge 5.15’de gösterilmiştir.

Çizelge 5.15: Decision Tree modelinin istasyon bazlı performans metrikleri.

İstasyon	MAE	RMSE	R ²
Atapark	0.0183	0.0269	0.8273
Beydilli	0.0217	0.0387	0.9048
Nizip	0.0098	0.0125	0.8862

Decision Tree modeli diğer modellere göre daha basit bir yapıya sahip olduğu için doğruluk oranı düşüktür. Ancak modelin yorumlanabilirliği ve kolay uygulanabilirliği nedeniyle karşılaştırmalar açısından referans olarak kullanılmıştır.

Decision Tree modeli, en düşük performansı sergileyen model olmuştur. Tüm istasyonlarda en düşük R^2 değerlerine sahip olduğu için diğer modellere kıyasla başarısız bulunmuştur.

LightGBM, özellikle büyük veri setlerinde ve zaman serisi tahminlerinde yüksek performans gösteren bir gradient boosting modelidir. Optuna ile yapılan hiperparametre optimizasyonu sayesinde tüm istasyonlarda en iyi performansı göstermiştir. LightGBM modelinin istasyon bazlı performans metrikleri Çizelge 5.16'da verilmiştir.

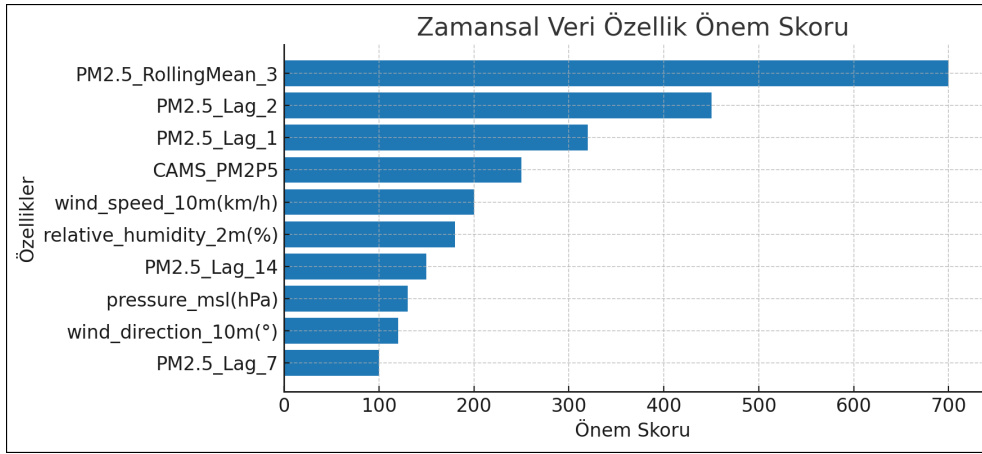
Çizelge 5.16: LightGBM modelinin istasyon bazlı performans metrikleri.

İstasyon	MAE	RMSE	R^2
Atapark	0.0109	0.0181	0.9224
Beydilli	0.0160	0.0263	0.9560
Nizip	0.0053	0.0071	0.9631

LightGBM modeli, yüksek doğruluk ve düşük hata oranları sunarak $PM_{2.5}$ tahmini için en başarılı model olarak belirlenmiştir. Bu başarının temel nedenleri arasında, modelin veri içindeki ilişkileri iyi öğrenebilmesi, aşırı öğrenmeyi (overfitting) minimize etmesi ve Optuna optimizasyonu ile en iyi hiperparametrelerin belirlenmesi yer almaktadır.

Bu başarıyı daha iyi anlamak ve modelin karar mekanizmasını analiz etmek için Özellik Önem Skoru (Feature Importance) analizi gerçekleştirilmiştir. Bu analiz, modelin $PM_{2.5}$ tahmini yaparken hangi değişkenlere ağırlık verdiğini açıklamaktadır. LightGBM modeli için yapılan Özellik Önem Skoru analizi, modelin hangi değişkenleri daha fazla kullandığını ve tahmin sürecinde hangi değişkenlerin kritik olduğunu ortaya koymuştur. Üç istasyonun (Atapark, Beydilli ve Nizip) özellik önem skorlarının ortalaması alınarak genel bir değerlendirme yapılmıştır. Bu doğrultuda, zamansal modelleme için elde edilen özellik önem sıralaması Şekil 5.23'de verilmiştir.

Analiz sonucunda en önemli değişkenin, son 3 günün hareketli ortalaması olan $PM_{2.5_RollingMean_3}$ olduğu belirlenmiştir. Bu değişken, modelin tahmin doğruluğu üzerinde en yüksek etkiye sahip olmuştur. İkinci en önemli değişken



Şekil 5.23: Zamansal modelleme için LightGBM ile elde edilen özellik önem skoru.

PM2.5_Lag_2 (2 gün önceki PM_{2.5} değeri) olup, mevcut hava kirliliği seviyelerini tahmin etmede kritik bir rol oynamaktadır. PM2.5_Lag_1 ise (1 gün önceki PM_{2.5} değeri), kısa vadeli PM_{2.5} değişimlerini anlamada önemli bir değişken olarak öne çıkmıştır. Rüzgar hızı, hava sirkülasyonunun PM_{2.5} seviyeleri üzerindeki etkisi nedeniyle önemli bir faktör olarak değerlendirilmiştir. CAMS_PM2P5 değişkeni, Avrupa hava tahmin sistemi tarafından sağlanan verilerle modelin doğruluğunu artırmıştır. Bağıl nem (relative humidity), PM_{2.5}'in atmosferdeki su buharı ile etkileşimini açıklamada önemli bir değişken olarak değerlendirilmiştir.

Genel olarak zamansal modelleme sürecinde LightGBM modeli, diğer modellere kıyasla en yüksek doğruluğu sağlamış ve tüm istasyonlar için en iyi performansı göstermiştir. Bu nedenle, mekânsal modelleme sürecine entegre edilerek nihai tahmin modeli olarak kullanılmıştır.

5.3.1.2 Derin öğrenme modelleri

Zamansal PM_{2.5} tahmini kapsamında, makine öğrenmesi modellerine ek olarak Uzun Kısa Süreli Bellek (LSTM - Long Short-Term Memory) ve Kapılı Tekrarlayan Birim (GRU - Gated Recurrent Unit) modelleri de uygulanmıştır. Bu modeller, zamansal bağımlılığı öğrenme konusunda güçlü yapay sinir ağlarıdır ve özellikle zaman serisi tahmin problemlerinde sıklıkla kullanılmaktadır.

LSTM ve GRU, geçmiş zamandaki değerlerin gelecekteki tahminlere etkisini modellemek için tasarlanmıştır ve PM_{2.5} tahmini gibi zaman serisi analizlerinde

sıklıkla tercih edilmektedir. Bu modeller, bellek hücreleri ve kapılar kullanarak geçmiş bilgileri koruma yeteneğine sahiptir ve bu sayede uzun vadeli bağımlılıkları öğrenebilirler. Optuna ile hiperparametre optimizasyonu uygulanarak LSTM ve GRU modelleri için en iyi parametreler belirlenmiş ve bu modeller Atapark, Beydilli ve Nizip istasyonları için eğitilmiştir.

LSTM ve GRU modellerinin test veri seti üzerindeki MAE, RMSE ve R^2 performansları Çizelge 5.17’te sunulmuştur.

Çizelge 5.17: LSTM ve GRU modellerinin performans karşılaştırması.

Model	İstasyon	MAE	RMSE	R^2
LSTM	Atapark	0.0636	0.0971	0.6184
LSTM	Beydilli	0.0338	0.0546	0.8115
LSTM	Nizip	0.0416	0.0542	0.7250
GRU	Atapark	0.0621	0.0954	0.6317
GRU	Beydilli	0.0349	0.0582	0.6050
GRU	Nizip	0.0430	0.0558	0.6180

Çizelgeye göre, LSTM modeli, GRU modeline kıyasla daha iyi performans göstermiştir, ancak fark çok büyük olmamış, iki model de benzer sonuçlar vermiştir. Makine öğrenmesi modellerine kıyasla R^2 değerleri önemli ölçüde düşük kalmıştır. Özellikle Atapark ve Beydilli istasyonlarında LightGBM modeli çok daha başarılı sonuçlar vermiştir. GRU modeli, hesaplama açısından daha verimli olmasına rağmen, tahmin doğruluğu açısından LSTM kadar iyi sonuç verememiştir.

Derin öğrenme ve makine öğrenmesi modelleri karşılaştırıldığında; her iki model grubu da aynı veri seti ile eğitilmiş, ancak LSTM ve GRU’nun tahmin doğruluğu, LightGBM, XGBoost ve Random Forest gibi makine öğrenmesi modellerine kıyasla daha düşük kalmıştır. LightGBM modeli, test setinde en düşük RMSE ve en yüksek R^2 değerlerini sağlayarak en başarılı model olmuştur. Derin öğrenme modellerinin düşük performansı, veri miktarının bu tür modeller için yeterli olmamasından kaynaklanmış olabilir. LSTM ve GRU’nun yüksek doğrulukla çalışabilmesi için daha büyük veri setleri ve daha uzun eğitim süreleri gerekmektedir. Bu nedenlerle, LightGBM modeli zamansal tahmin süreci için en başarılı model olarak belirlenmiş ve mekânsal modelleme sürecine entegre edilmiştir.

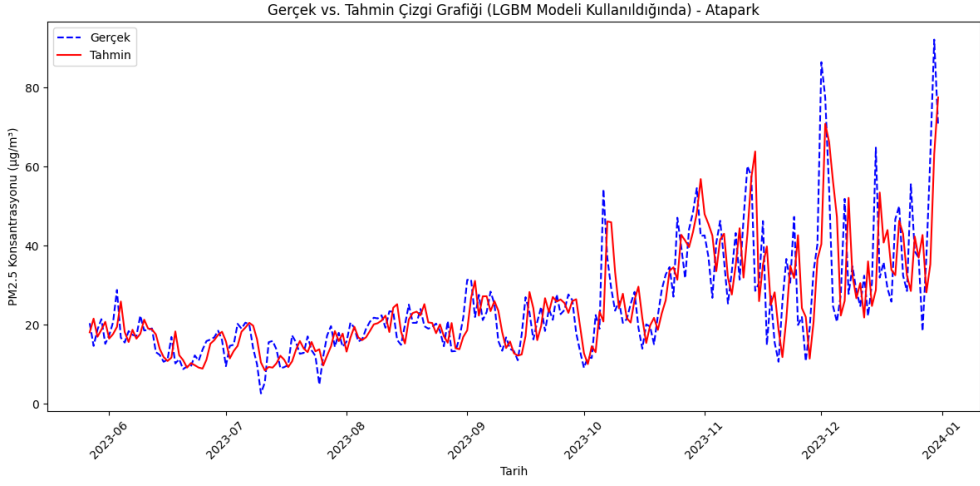
Zamansal modelleme aşamasında kullanılan tüm modellerin istasyon bazlı performans metrikleri Çizelge 5.18’de sunulmuştur.

Çizelge 5.18: Zamansal modelleme sonuçları: tüm modellerin performans karşılaştırması.

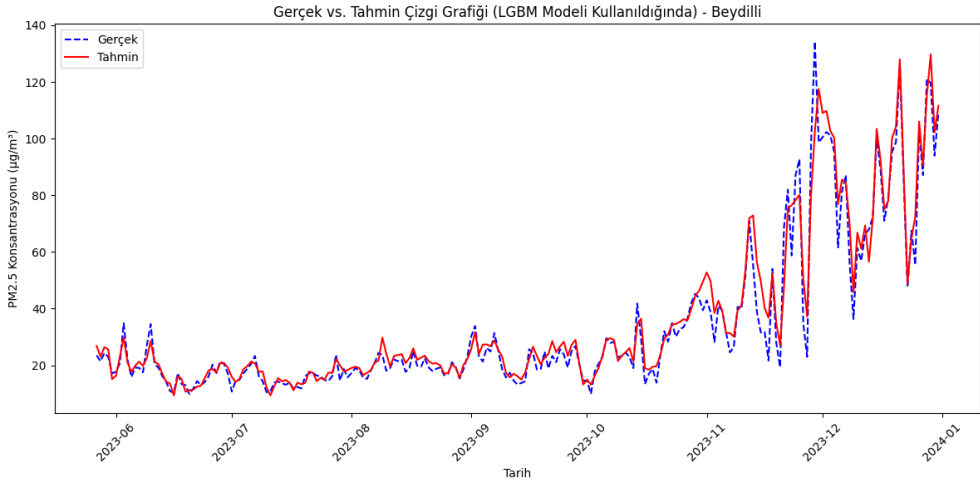
Model	İstasyon	MAE	RMSE	R ²
XGBoost	Atapark	0.0102	0.0221	0.9395
XGBoost	Beydilli	0.0153	0.0263	0.9547
XGBoost	Nizip	0.0050	0.0064	0.9706
Random Forest	Atapark	0.0145	0.0221	0.8832
Random Forest	Beydilli	0.0174	0.0308	0.9393
Random Forest	Nizip	0.0070	0.0088	0.9434
Decision Tree	Atapark	0.0183	0.0269	0.8273
Decision Tree	Beydilli	0.0217	0.0387	0.9048
Decision Tree	Nizip	0.0098	0.0125	0.8862
LightGBM	Atapark	0.0109	0.0181	0.9224
LightGBM	Beydilli	0.0160	0.0263	0.9560
LightGBM	Nizip	0.0053	0.0071	0.9631
LSTM	Atapark	0.0636	0.0971	0.6184
LSTM	Beydilli	0.0338	0.0546	0.8115
LSTM	Nizip	0.0416	0.0542	0.7250
GRU	Atapark	0.0621	0.0954	0.6317
GRU	Beydilli	0.0349	0.0582	0.6050
GRU	Nizip	0.0430	0.0558	0.6180

Zamansal modelleme sürecinde, tahmin edilen PM_{2.5} değerleri ile gerçek PM_{2.5} ölçümleri arasındaki ilişkileri değerlendirmek amacıyla çeşitli görseller oluşturulmuştur. Bu görseller, tahmin modellerinin performansını değerlendirmek ve modelleme sürecinin doğruluğunu analiz etmek açısından kritik öneme sahiptir. Şekil 5.24, Şekil 5.25 ve Şekil 5.26’da, Atapark, Beydilli ve Nizip istasyonları için gerçek PM_{2.5} ölçümleri ile LightGBM modeli tarafından tahmin edilen değerler zaman serisi grafikleriyle sunulmuştur.

Grafikte, gerçek değerler mavi kesik çizgiyle, tahmin edilen değerler ise kırmızı düz çizgiyle gösterilmiştir. Atapark istasyonu için grafik incelendiğinde; tahmin edilen değerlerin genel olarak gerçek değerleri yakından takip ettiği görülmektedir. Ancak, bazı ani artış ve düşüşlerde modelin gerçek değerlerden hafif sapmalar gösterdiği gözlemlenmiştir. Beydilli istasyonu için grafik incelendiğinde; özellikle PM_{2.5} seviyelerinin yüksek olduğu dönemlerde model tahminleri ile gerçek değerler arasında benzer bir trend izlenmiştir. Nizip istasyonu için grafik incelendiğinde;



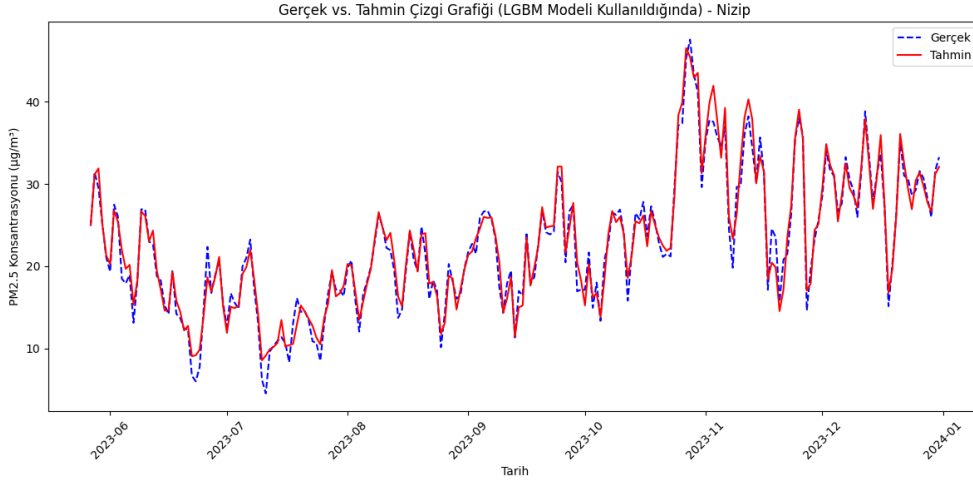
Şekil 5.24: Atapark istasyonu için gerçek ve LightGBM modeli ile tahmin edilen $PM_{2.5}$ zaman serisi.



Şekil 5.25: Beydilli istasyonu için gerçek ve LightGBM modeli ile tahmin edilen $PM_{2.5}$ zaman serisi.

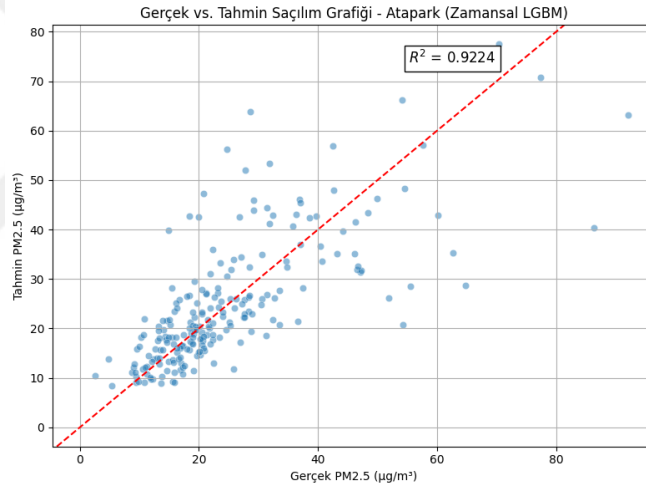
model, genellikle gerçek değerleri başarıyla tahmin etmiş, ancak bazı dönemlerde tahmin edilen değerlerin hafif bir kayma gösterdiği gözlemlenmiştir. Bu durum, zamansal tahmin modellerinin daha uzun süreli geçmiş verilerle beslenmesi gerektiğini göstermektedir.

Bu sonuçlar, LightGBM modelinin zamansal $PM_{2.5}$ tahmininde yüksek doğruluk sunduğunu, ancak ani dalgalanmaları yakalamada zaman zaman sapmalar yaşanabileceğini göstermektedir. İstasyonlar için gerçek $PM_{2.5}$ ölçümleri ile tahmin edilen değerler arasındaki ilişkinin saçılım grafiği Şekil 5.27, Şekil 5.28 ve Şekil 5.29'da



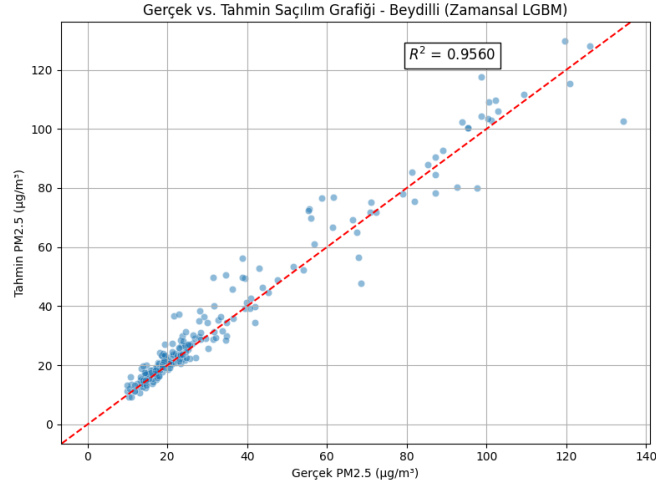
Şekil 5.26: Nizip istasyonu için gerçek ve LightGBM modeli ile tahmin edilen PM_{2.5} zaman serisi.

görselleştirilmiştir. Kırmızı kesik çizgi, mükemmel tahmin hattını ($y = x$ doğrusu) temsil etmektedir.

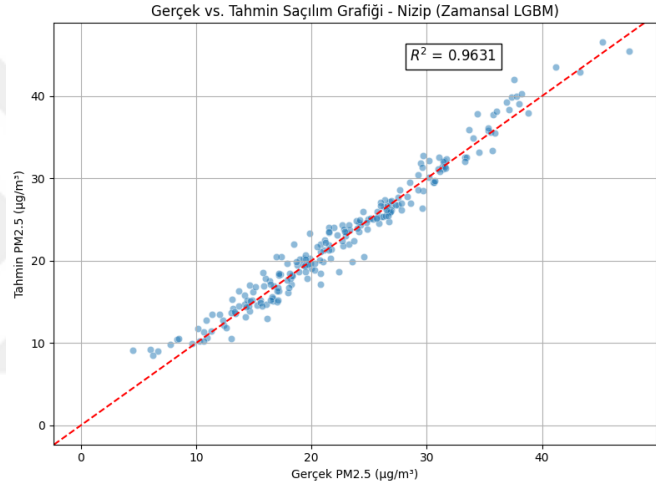


Şekil 5.27: Atapark istasyonu için gerçek ve tahmin edilen PM_{2.5} değerlerinin saçılım grafiği.

Bu saçılım grafiklerine göre, Atapark istasyonu incelendiğinde, tahmin edilen ve gerçek değerler büyük ölçüde doğrusal bir ilişki sergilemektedir. Ancak, bazı yüksek değerlerde modelin hafif düşük tahminler yaptığı gözlemlenmiştir. Beydilli istasyonu incelendiğinde, model, geniş bir aralıktaki PM_{2.5} değerlerini başarılı şekilde tahmin etmiş ve R^2 değerinin yüksek olmasıyla doğrulanmıştır. Nizip istasyonu incelendiğinde, modelin tahminleri, düşük ve orta seviyeli PM_{2.5} değerlerinde oldukça



Şekil 5.28: Beydilli istasyonu için gerçek ve tahmin edilen $PM_{2.5}$ değerlerinin saçılım grafiği.

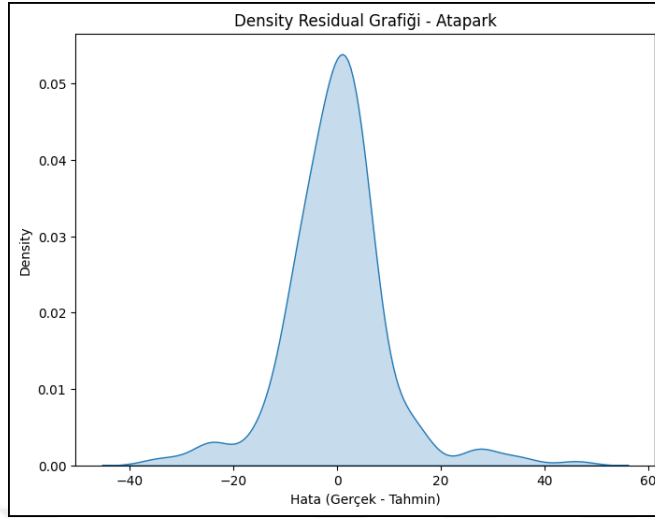


Şekil 5.29: Nizip istasyonu için gerçek ve tahmin edilen $PM_{2.5}$ değerlerinin saçılım grafiği.

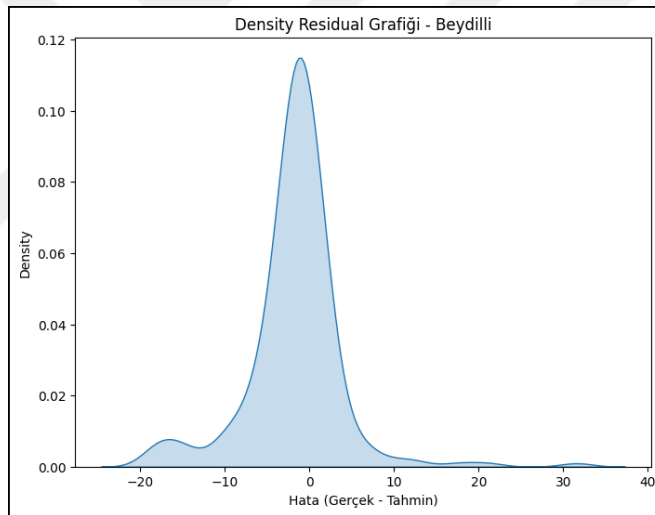
tutarlı gözükmemektedir, ancak yüksek $PM_{2.5}$ seviyelerinde tahmin edilen değerler bazen gerçeğin biraz altında kalabilmektedir.

Son olarak model tahminlerinin başarısını daha iyi anlamak için, hata dağılımı analizleri gerçekleştirilmiştir. Şekil 5.30, Şekil 5.31 ve Şekil 5.32 numaralı şekillerde gösterilen Density Residual (Hata Dağılımı) Grafikleri, tahmin edilen $PM_{2.5}$ değerleri ile gerçek ölçümler arasındaki hata farklarının yoğunluk dağılımını görselleştirmektedir. Grafiklerdeki yoğunluk eğrileri, model hatalarının genellikle 0'a yakın ve simetrik olarak dağıldığını göstermektedir. Bu durum, modelin sistematik bir

sapma üretmeden tahmin gerçekleştirdiğini ve hata dağılımının büyük ölçüde normal dağılıma yakın olduğunu göstermektedir.

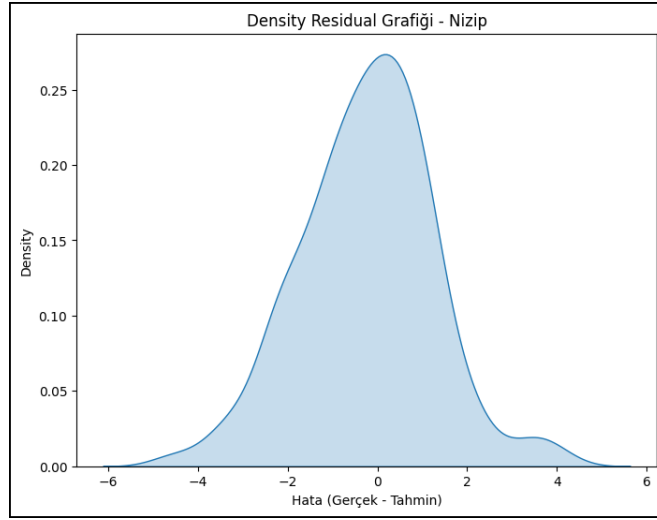


Şekil 5.30: Atapark istasyonu için LightGBM modeline ait hata dağılımı (residual) yoğunluk grafiği.



Şekil 5.31: Beydilli istasyonu için LightGBM modeline ait hata dağılımı (residual) yoğunluk grafiği.

Buna göre, Atapark istasyonunda hataların büyük bir kısmı 0 civarında yoğunlaşmış, ancak bazı uç değerlerde pozitif ve negatif hatalar gözlemlenmiştir. Beydilli istasyonunda hatalar nispeten simetrik dağılmış olup, modelin genellikle dengeli tahminler sunduğu görülmektedir. Nizip istasyonu için modelin tahmin hataları, diğer istasyonlara kıyasla daha küçük aralıkta dağılmış olup, Nizip'te modelin daha istikrarlı tahmin yaptığı gözlemlenmiştir. Genel olarak, hata dağılımı analizleri modelin ortalama olarak gerçek değerleri iyi tahmin ettiğini, ancak bazı dönemlerde



Şekil 5.32: Nizip istasyonu için LightGBM modeline ait hata dağılımı (residual) yoğunluk grafiği.

küçük sapmaların meydana geldiğini göstermektedir. Sonuç olarak, zaman serisi grafiklerinde, modelin tahminlerinin gerçek değerlerle büyük ölçüde örtüştüğü ancak bazı dönemlerde ani artış ve düşüşleri tam olarak yakalayamadığı belirlenmiştir. Saçılım grafikleri, tahmin edilen ve gerçek değerler arasındaki ilişkinin güçlü olduğunu, ancak yüksek $PM_{2.5}$ seviyelerinde hafif tahmin sapmalarının meydana geldiğini göstermektedir. Density residual (hata yoğunluğu) grafikleri ise modelin hata dağılımının dengeli olduğunu ve büyük hata değerlerinin nadir görüldüğünü ortaya koymuştur.

5.3.2 Mekânsal modelleme

Hava kirliliğinin mekânsal dağılımını daha iyi anlamak ve $PM_{2.5}$ tahminlerini iyileştirmek amacıyla farklı mekânsal modelleme yaklaşımları kullanılmıştır. Bu kapsamda, makine öğrenmesi modelleri (XGBoost, LightGBM, Random Forest) ve mekânsal regresyon ve grafik tabanlı modeller (Geographically Weighted Regression - GWR ve Graph Convolutional Network - GCN) uygulanmıştır. Bu modellerin performansı üç farklı istasyon (Atapark, Beydilli, Nizip) için değerlendirilmiş ve sonuçlar karşılaştırılmıştır.

5.3.2.1 Makine öğrenmesi modelleri

Bu aşamada mekânsal tahminleme için XGBoost (XGB), Random Forest (RF) ve LightGBM (LGBM) modelleri uygulanmıştır. Makine öğrenmesi modelleri ile istasyon bazında hava kirliliği tahmini gerçekleştirilmiş ve model performansları MAE, RMSE ve R^2 değerleri üzerinden karşılaştırılmıştır. Model parametreleri Optuna ile optimize edilerek en iyi performans elde edilmiştir. XGBoost modeli, mekânsal PM2.5 tahmini için kullanılmış olup, özellikle Beydilli istasyonunda en yüksek doğruluk oranını sağlamıştır. Modelin aşırı öğrenme (overfitting) riskini azaltmak için Optuna ile hiperparametre optimizasyonu gerçekleştirilmiştir. XGBoost modeline ait istasyon bazlı performans sonuçları Çizelge 5.19'da sunulmuştur.

Çizelge 5.19: XGBoost mekânsal model performans sonuçları.

İstasyon	MAE	RMSE	R^2
Atapark	0.0144	0.0261	0.8333
Beydilli	0.0241	0.0366	0.9602
Nizip	0.0101	0.0191	0.8195

Bu sonuçlara göre, XGBoost modeli özellikle Beydilli istasyonunda en yüksek doğruluk oranını sağlamıştır ($R^2 = 0.9602$). Ancak Atapark ve Nizip istasyonlarında modelin tahmin doğruluğu görece daha düşük kalmıştır.

Mekansal modelleme kullanılan Random Forest modeli, birden fazla karar ağacının birleşiminden oluşan ansambl öğrenme modeli olup, özellikle Nizip istasyonunda en yüksek doğruluk oranını sağlamıştır. Modelin aşırı öğrenme (overfitting) riskini azaltmak için Optuna ile hiperparametre optimizasyonu gerçekleştirilmiştir. Random Forest modelinin performans metrikleri Çizelge 5.20'de sunulmuştur.

Çizelge 5.20: Random forest mekânsal model performans sonuçları.

İstasyon	MAE	RMSE	R^2
Atapark	0.0159	0.0268	0.8246
Beydilli	0.0248	0.0370	0.9591
Nizip	0.0105	0.0178	0.8433

Bu sonuçlara göre, Random Forest modeli Beydilli ve Nizip istasyonlarında iyi performans göstermiş, ancak Atapark istasyonunda hata oranı daha yüksek kalmıştır.

Özellikle Nizip'te XGBoost modeline kıyasla daha yüksek bir R^2 değeri elde edilmiştir ($R^2 = 0.8433$).

Mekansal modellemede, zamansal modelleme için kullanılan Decision Tree modeli mekânsal bağımlılığı veya konular arası ilişkileri doğrudan modelleyemediği için kullanılmamıştır.

Son olarak mekansal verilerle çalıştırılan LightGBM modeli, özellikle büyük veri setlerinde ve mekânsal tahminleme problemlerinde yüksek başarı gösteren bir gradient boosting modelidir. Optuna ile yapılan hiperparametre optimizasyonu sayesinde tüm istasyonlarda en iyi performansı göstermiştir.

LightGBM modelinin performans metrikleri Çizelge 5.21'de sunulmuştur.

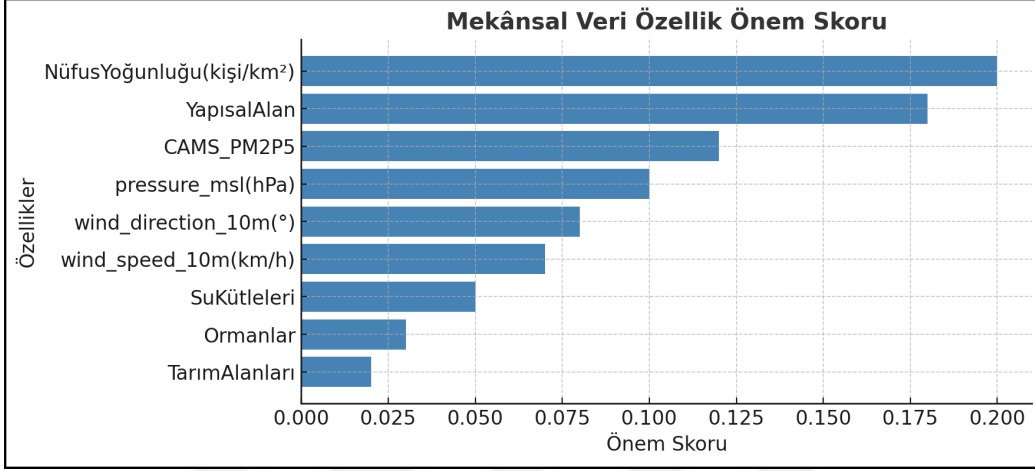
Çizelge 5.21: LightGBM mekânsal model performans sonuçları.

İstasyon	MAE	RMSE	R^2
Atapark	0.0131	0.0221	0.876
Beydilli	0.0223	0.0340	0.965
Nizip	0.0095	0.0158	0.879

LightGBM modeli, yüksek doğruluk ve düşük hata oranları sunarak mekânsal $PM_{2.5}$ tahmini için en başarılı model olarak belirlenmiştir. Modelin başarısı özellikle Atapark ve Beydilli istasyonlarında açıkça gözlemlenmiş olup, Nizip istasyonunda XGBoost modeli ile benzer doğruluk seviyelerine ulaşmıştır.

LightGBM modeli, makine öğrenmesi yöntemleri arasında en başarılı model olarak belirlenmiştir. Beydilli istasyonunda, XGBoost ve Random Forest modellerine kıyasla en yüksek doğruluk oranı sağlanmıştır. Atapark ve Nizip istasyonlarında LightGBM, XGBoost ile benzer performans göstermiş ancak hata oranları daha düşük kalmıştır. XGBoost modeli, Beydilli istasyonunda yüksek performans göstermesine rağmen, Atapark ve Nizip istasyonlarında hata oranları artmıştır. Random Forest modeli, genel olarak XGBoost'tan biraz daha düşük performans göstermiş, özellikle Atapark istasyonunda hata oranı en yüksek model olmuştur. Bu doğrultuda, LightGBM modeli mekânsal modelleme sürecinde en başarılı model olarak belirlenmiş ve mekânsal regresyon ve grafik tabanlı modeller ile karşılaştırılmak üzere analiz sürecine devam edilmiştir.

Mekansal modelleme en iyi model olan LightGBM modeli için yapılan Feature Importance analizi ile elde edilen özellik önem sıralaması Şekil 5.33'te verilmiştir. Analiz sonucunda mekânsal modelleme sürecinde en etkili değişkenler Nüfus Yoğunluğu, Yapısal Alan Yüzdesi, CAMS verisi ve deniz seviyesine indirgenmiş basınç olmuştur.



Şekil 5.33: Mekânsal modelleme için LightGBM ile elde edilen özellik önem skoru.

LightGBM modeline göre, en önemli değişken olan nüfus yoğunluğu, PM_{2.5} seviyeleri üzerinde en büyük etkiye sahip değişken olarak belirlenmiştir. Özellikle sanayi bölgelerine ve kent merkezlerine yakın alanlarda hava kirliliği seviyelerinin arttığı gözlemlenmiştir.

İkinci en büyük etkiye sahip olan yapısal alan, şehirleşmenin yoğun olduğu bölgelerde hava kirliliğinin daha yüksek olduğunu göstermektedir. CAMS_PM_{2.5} verisi, Avrupa Uzay Ajansı tarafından sağlanan PM_{2.5} tahmin verileri olup, modelin genel doğruluğunu artırmada önemli bir rol oynamıştır. Son olarak, basınç (hPa), rüzgar yönü (°) ve rüzgar hızı (km/h) gibi meteorolojik değişkenlerin de PM_{2.5} seviyelerinin belirlenmesinde etkili olduğu tespit edilmiştir.

5.3.2.2 Mekânsal regresyon ve grafik tabanlı modeller

Mekânsal modelleme sürecinde coğrafi değişkenlerin etkisini daha iyi anlamak ve mekânsal bağımlılığı modellemek amacıyla geleneksel regresyon tekniklerinden farklı olarak Geographically Weighted Regression (GWR) ve Graph Convolutional Network (GCN) modelleri kullanılmıştır. Bu yöntemler, hava kirliliğinin mekânsal dağılımını

ve farklı konumlardaki PM_{2.5} seviyelerinin birbirleriyle olan ilişkisini daha doğru bir şekilde yakalamayı hedeflemektedir.

Geographically Weighted Regression (GWR), mekânsal bağımlılığı göz önünde bulundurarak farklı coğrafi konumlarda değişkenlerin etkisini modellemek için kullanılan bir regresyon yöntemidir. Bu yöntem, klasik doğrusal regresyonun aksine, model katsayılarını her lokasyon için farklı olarak hesaplamaktadır. Böylece, hava kirliliği seviyelerini etkileyen faktörlerin mekâna bağlı olarak nasıl değiştiği analiz edilebilmektedir. GWR modeli Atapark, Beydilli ve Nizip istasyonları için çalıştırılmış ve sonuçlar Çizelge 5.22’de verilmiştir.

Çizelge 5.22: GWR modeli ile hesaplanan istasyon bazlı R² değerleri.

İstasyon	R ²
Atapark	0.6044
Beydilli	0.6853
Nizip	0.6818

Atapark istasyonunda modelin tahmin performansı R² = 0.6044 olarak hesaplanmıştır. Beydilli istasyonunda hata oranı diğer istasyonlara göre bir miktar düşük olsa da R² = 0.6853 ile güçlü bir doğruluk elde edilmiştir. Nizip istasyonunda, GWR modeli R² = 0.6818 ile mekânsal bağımlılığı modellemiş ancak geleneksel makine öğrenmesi modellerine kıyasla daha düşük performans göstermiştir.

Bu sonuçlar, GWR modelinin mekânsal faktörleri hesaba katarak PM_{2.5} tahmininde belirli bir başarı gösterdiğini ancak genel doğruluk seviyesinin daha düşük kaldığını ortaya koymaktadır. Modelin katsayılarının her konum için değişmesi, yorumlamayı zorlaştırmakta ve genelleştirme kapasitesini sınırlamaktadır.

Kullanılan diğer model olan Graph Convolutional Network (GCN), mekânsal bağımlılığı modellemek için kullanılan bir derin öğrenme modelidir. Coğrafi ağ yapısını kullanarak mekânsal ilişkileri öğrenen GCN modeli, hava kirliliği seviyelerinin mekânsal yayılımını tahmin etme açısından oldukça etkili olabilmektedir.

GCN modelinde istasyonlar birer düğüm (node) olarak kabul edilmiştir. İstasyonlar arasındaki mekânsal ilişkiyi belirlemek için mesafeye dayalı bir ağırlıklı bağlantı matrisi kullanılmıştır.

Son olarak, model hava kirliliği seviyelerini çevre istasyonlardan gelen bilgileri kullanarak tahmin etmiştir. GCN modeli üç istasyon için ayrı ayrı çalıştırılmış ve model performansı Çizelge 5.23'te verilmiştir.

Çizelge 5.23: GCN modeli performans sonuçları (mekânsal modelleme).

İstasyon	MAE	RMSE	R ²
Atapark	0.2684	0.3621	0.5285
Beydilli	0.3163	0.4189	0.7155
Nizip	0.1945	0.3074	0.4672

Buna göre, Atapark istasyonunda model, $R^2 = 0.5285$ ile nispeten düşük bir doğruluk göstermiştir. Beydilli istasyonunda GCN modeli, $R^2 = 0.7155$ ile orta seviyede bir doğruluk sağlamıştır. Nizip istasyonunda GCN modeli, $R^2 = 0.4672$ ile en düşük doğruluğa sahip olmuştur. Bu sonuçlar, GCN modelinin mekânsal verileri öğrenme kapasitesinin beklenenden düşük kaldığını ve mekânsal bağımlılığı klasik makine öğrenmesi modellerine kıyasla daha düşük doğrulukla temsil ettiğini göstermektedir.

GCN modelinin başarısının istasyonlar arasındaki coğrafi ilişkileri doğrudan modellemesinden kaynaklandığı düşünülmektedir. Ancak, GCN modelinin karmaşıklığı ve yüksek hesaplama maliyeti nedeniyle, modelin daha büyük veri setleriyle eğitilmesi gerektiği sonucuna varılmıştır. Mekânsal modelleme aşamasında kullanılan tüm modellerin sonuçları Çizelge 5.24'te gösterilmiştir.

Çizelge 5.24: Mekânsal modelleme sonuçlarının karşılaştırılması.

Model	İstasyon	MAE	RMSE	R ²
XGBoost	Atapark	0.0144	0.0261	0.8333
XGBoost	Beydilli	0.0241	0.0366	0.9602
XGBoost	Nizip	0.0101	0.0191	0.8195
Random Forest	Atapark	0.0159	0.0268	0.8246
Random Forest	Beydilli	0.0248	0.0370	0.9591
Random Forest	Nizip	0.0105	0.0178	0.8433
LGBM	Atapark	0.0131	0.0221	0.8760
LGBM	Beydilli	0.0223	0.0340	0.9650
LGBM	Nizip	0.0095	0.0158	0.8790
GCN	Atapark	0.2684	0.3621	0.5285
GCN	Beydilli	0.3163	0.4189	0.7155
GCN	Nizip	0.1945	0.3074	0.4672
GWR	Atapark	-	-	0.6044
GWR	Beydilli	-	-	0.6853
GWR	Nizip	-	-	0.6818

5.3.3 Zamansal modelden elde edilen verinin mekânsal modelde kullanımı

Mekânsal modelleme sürecinde, zamansal modelleme aşamasında elde edilen PM_{2.5} tahminleri mekânsal modele eklenerek modelin genel tahmin doğruluğunun artırılması hedeflenmiştir. Bu yaklaşım, önceden öğrenilmiş bilgilerin yeni bir modele entegre edilmesiyle tahmin performansını iyileştirmeyi amaçlayan çok aşamalı modelleme (multi-stage modeling) veya model-tabanlı özellik üretimi (model-based feature engineering) olarak literatürde bilinmektedir.

Bu yöntemin temel mantığı, zamansal modelin geçmiş verilerden öğrendiği hava kirliliği desenlerini kullanarak, mekânsal modelin tahmin sürecine ek bağlamsal bilgi sağlamaktır. Özellikle, zamansal modelleme ile elde edilen PM_{2.5} tahminlerinin, mekânsal modelleme sürecinde ek bir özellik (feature) olarak kullanılması, mekânsal modelin yalnızca coğrafi ve çevresel değişkenlere dayanarak tahmin yapmasını değil, aynı zamanda zamansal eğilimleri de dikkate almasını sağlamaktadır.

Bu bağlamda, zamansal modelden elde edilen PM_{2.5} tahminleri, mekânsal modelin girdilerine doğrudan eklenmiş ve istasyon bazlı olarak 27.05.2023 sonrası için gerçek PM_{2.5} verilerinin yerine kullanılmıştır. Böylece, mekânsal model hem mekânsal bağımlılığı hem de zamansal değişkenlikleri göz önünde bulundurarak daha dengeli ve doğru tahminler yapabirmiştir. Model performansını optimize etmek amacıyla, Optuna ile hiperparametre ayarlamaları yeniden yapılmış ve hem genel model doğruluğu hem de istasyon bazlı performans analiz edilmiştir.

Bu yaklaşım, özellikle hava kirliliği gibi hem mekânsal hem de zamansal bağımlılık gösteren değişkenlerin tahmini için oldukça mantıklı bir strateji olup, veri kümesindeki ek bilgi kullanımını maksimize eden bir modelleme tekniğidir. Böylece, sadece mekânsal değişkenlere dayalı tahmin yerine, geçmiş zamansal tahminlerin de mekânsal bağlamda değerlendirilmesi sağlanarak modelin genel başarımı artırılmıştır. Özellik entegrasyonu öncesi ve sonrası R² performans sonuçları Çizelge 5.25'te gösterilmiştir.

Bu sonuçlar, zamansal modelden elde edilen PM_{2.5} tahminlerinin mekânsal modele entegre edilmesinin model performansı üzerindeki etkisini açıkça göstermektedir. Özellikle Atapark istasyonu için R² değeri 0.876'dan 0.928'e yükselmiş, dolayısıyla

Çizelge 5.25: Özellik entegrasyonu öncesi ve sonrası R^2 performans sonuçları.

İstasyon	Entegrasyon Öncesi R^2	Entegrasyon Sonrası R^2
Atapark	0.876	0.928
Beydilli	0.965	0.951
Nizip	0.879	0.963

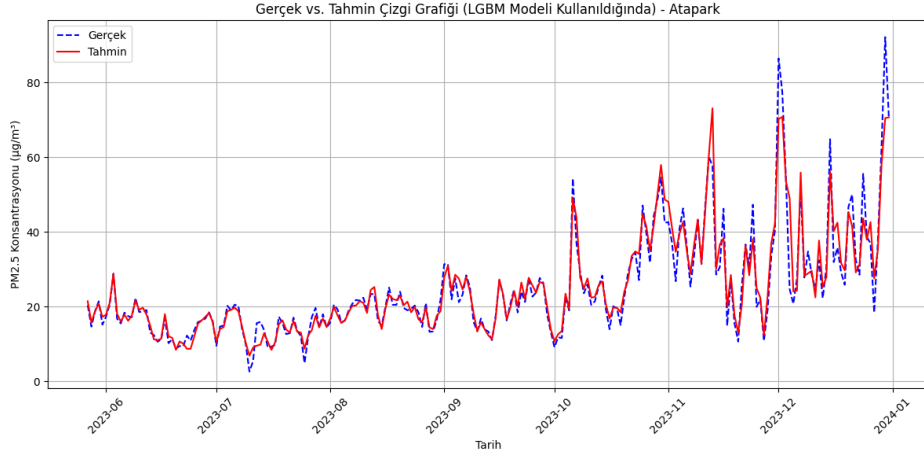
tahmin doğruluğunda önemli bir iyileşme sağlanmıştır. Benzer şekilde, Nizip istasyonunda da R^2 değeri 0.879'dan 0.963'e çıkarak zamansal verinin mekânsal modele eklenmesinin tahmin başarısını artırdığı görülmüştür.

Ancak, Beydilli istasyonunda entegrasyon sonrası R^2 değerinde küçük bir düşüş gözlemlenmiştir (0.965'den 0.951'e). Bu durum, zamansal model tahminlerinin bazı mekânsal bölgelerde modelin doğruluğunu artırırken, bazı bölgelerde beklenen iyileşmeyi tam olarak sağlamadığına işaret etmektedir. Bunun başlıca sebepleri arasında istasyonlar arasındaki mekânsal bağımlılığın değişkenliği, yerel çevresel faktörlerin etkisinin farklı olması ve mekânsal modelin, zamansal model tahminleriyle bazı durumlarda çakışan veya farklılaşan ilişkileri öğrenmesi yer almaktadır.

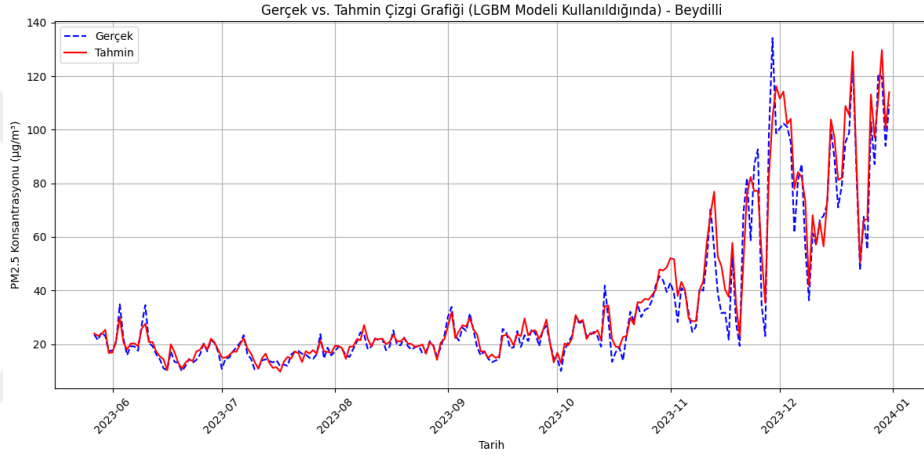
Genel olarak, özellik entegrasyonu sonrası RMSE ve MAE değerlerinde artış gözlemlense de, bu artış modelin daha gerçekçi tahminler yapmasını sağlayarak genelleme yeteneğini geliştirmiştir. Bu durum, multi-stage modeling gibi yaklaşımların model başarısını iyileştirme potansiyeli taşıdığını ancak her istasyon için aynı düzeyde iyileşme sağlamayabileceğini göstermektedir.

Mekânsal modelleme sürecinde, tahmin edilen $PM_{2.5}$ değerleri ile gerçek $PM_{2.5}$ ölçümleri arasındaki ilişkileri değerlendirmek amacıyla çeşitli görseller oluşturulmuştur. Bu görseller, mekânsal tahmin modellerinin performansını analiz etmek ve tahmin hatalarını detaylandırmak açısından kritik öneme sahiptir. Şekil 5.34, Şekil 5.35, ve Şekil 5.36'da Atapark, Beydilli ve Nizip istasyonları için gerçek $PM_{2.5}$ ölçümleri ile LightGBM modeli tarafından tahmin edilen değerleri içeren zaman serisi grafiklerini göstermektedir.

Grafikte, gerçek değerler mavi kesik çizgiyle, tahmin edilen değerler ise kırmızı düz çizgiyle temsil edilmiştir. Atapark istasyonunda modelin tahminleri genel olarak gerçek değerleri takip etmiş, ancak bazı yüksek seviyelerde hafif sapmalar gözlemlenmiştir; buna rağmen genel eğilimler başarılı bir şekilde yansıtılmıştır.

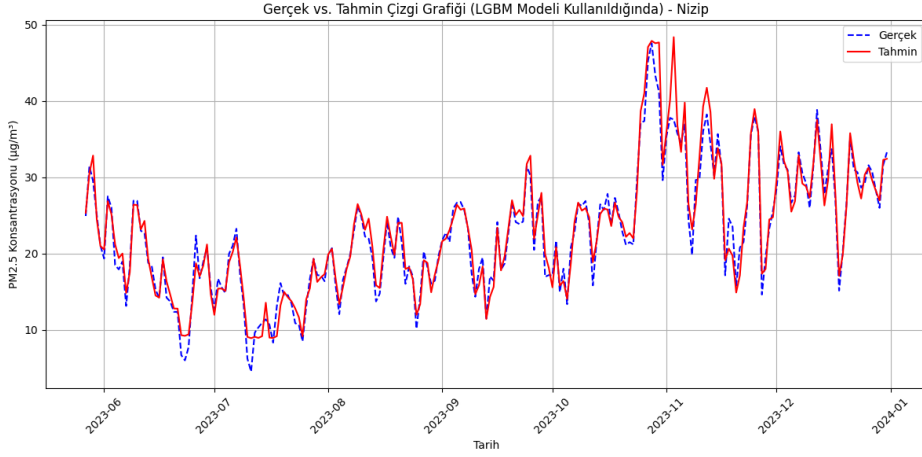


Şekil 5.34: Atapark istasyonu için gerçek ve tahmin $PM_{2.5}$ değerlerinin karşılaştırılması (Mekânsal Model).



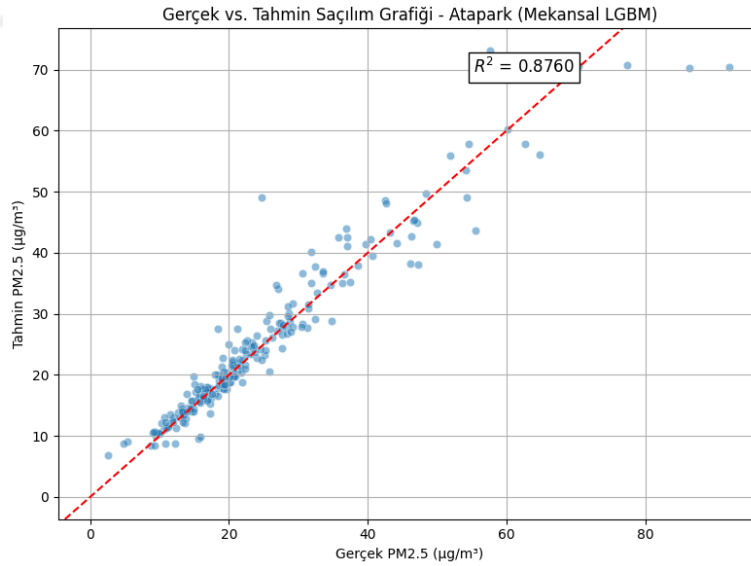
Şekil 5.35: Beydilli istasyonu için gerçek ve tahmin $PM_{2.5}$ değerlerinin karşılaştırılması (Mekânsal Model).

Beydilli istasyonunda $PM_{2.5}$ seviyelerinin yüksek olduğu dönemlerde modelin tahmin doğruluğu görece daha yüksek olmuş, fakat aşırı hava olayları gibi ani değişimlerin yaşandığı noktalarda tahminler ile gerçek değerler arasında küçük farklar tespit edilmiştir. Nizip istasyonunda ise model genel eğilimleri başarıyla tahmin edebilmiş, ancak düşük $PM_{2.5}$ seviyelerinde tahmin edilen değerlerde zaman zaman hafif kaymalar gözlenmiştir; bu durum, mekânsal tahmin modellerinin daha fazla istasyon verisiyle desteklenmesi gerektiğini ortaya koymaktadır. Bu sonuçlar, LightGBM modelinin mekânsal $PM_{2.5}$ tahminlerinde genel olarak yüksek doğruluk sunduğunu, ancak belirli istasyonlarda kısa süreli ani değişimleri tam olarak yakalayamayabileceğini göstermektedir.



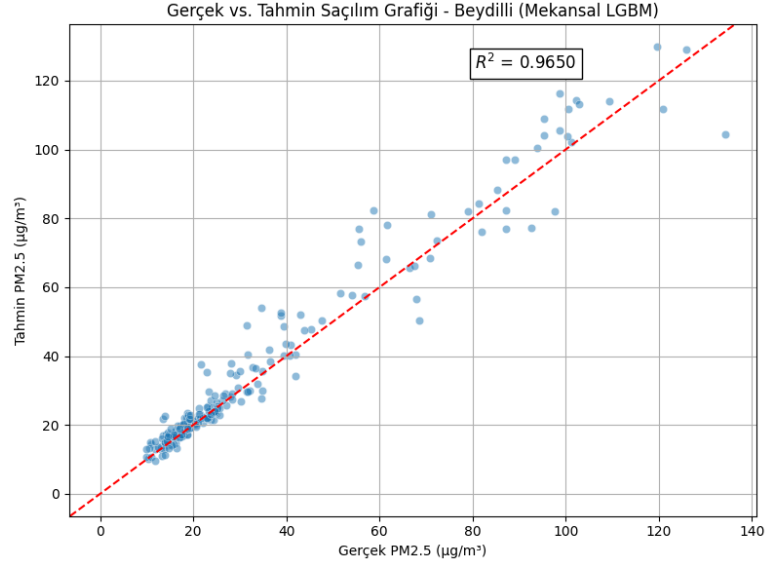
Şekil 5.36: Nizip istasyonu için gerçek ve tahmin $PM_{2.5}$ değerlerinin karşılaştırılması (Mekânsal Model).

Mekânsal modelleme sürecinde, tahmin edilen $PM_{2.5}$ değerleri ile gerçek ölçüm değerleri arasındaki ilişkiyi analiz edebilmek amacıyla saçılım grafikleri oluşturulmuştur. Şekil 5.37, Şekil 5.38 ve Şekil 5.39, sırasıyla Atapark, Beydilli ve Nizip istasyonları için model tarafından tahmin edilen $PM_{2.5}$ değerleri ile gerçek ölçümler arasındaki ilişkiyi göstermektedir. Grafikte yer alan kırmızı kesikli çizgi, mükemmel tahmin hattını ($y = x$ doğrusu) temsil etmektedir.



Şekil 5.37: Gerçek vs. tahmin saçılım grafiği - Atapark (Mekânsal).

Grafiklerde kırmızı kesik çizgi, mükemmel tahmin hattını ($y = x$ doğrusu) temsil etmekte olup, tahmin edilen değerlerin bu çizgiye yakınlığı modelin başarısını ortaya koymaktadır. Atapark istasyonunda tahmin edilen ve gerçek değerler büyük ölçüde

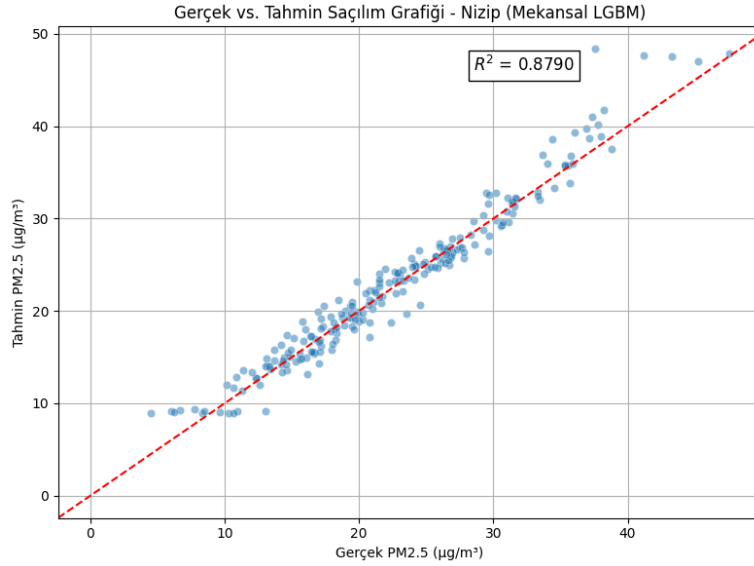


Şekil 5.38: Gerçek vs. tahmin saçılım grafiği - Beydilli (Mekânsal).

doğrusal bir ilişki sergilemiş, ancak bazı yüksek değerlerde modelin hafif düşük tahminler yaptığı gözlemlenmiştir. Beydilli istasyonunda model, geniş bir $PM_{2.5}$ aralığını başarılı bir şekilde tahmin etmiş ve yüksek R^2 değeriyle doğruluğunu kanıtlamıştır. Nizip istasyonunda ise düşük ve orta seviyeli $PM_{2.5}$ değerlerinde tahminler oldukça tutarlı olmuş, fakat yüksek $PM_{2.5}$ seviyelerinde modelin tahmin değerlerinin genellikle gerçek değerlerin altında kaldığı tespit edilmiştir. Bu saçılım analizleri, mekânsal modelin genellikle istasyonlar arasındaki mekânsal ilişkiyi başarılı bir şekilde öğrendiğini, ancak bazı aşırı durumlarda modelin tahminlerinin gerçek değerlere kıyasla düşük seviyelerde kalabileceğini göstermektedir.

Model tahminlerinin başarısını daha iyi anlayabilmek için hata dağılımı analizleri gerçekleştirilmiştir. Şekil 5.40, Şekil 5.41 ve Şekil 5.42’de sunulan Density Residual (Hata Dağılımı) grafiklerinde, modelin tahmin hatalarının dağılımı görselleştirilmiştir. Bu grafikler, her istasyon için tahmin ile gerçek değerler arasındaki farkların yoğunluğunu göstermekte olup, modelin sapma ve hata desenlerini ortaya koymaktadır.

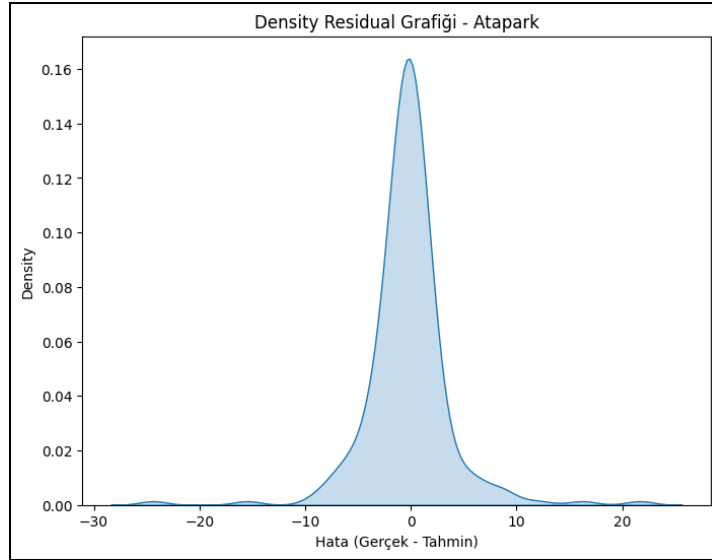
Bu grafikler, modelin tahmin ettiği $PM_{2.5}$ değerleri ile gerçek değerler arasındaki farkların istatistiksel dağılımını incelemek amacıyla kullanılmıştır. Atapark istasyonunda hataların büyük bir kısmı 0 civarında yoğunlaşmış, ancak bazı uç değerlerde pozitif ve negatif hatalar gözlemlenmiştir; bu durum, modelin genel olarak istikrarlı tahminler ürettiğini, fakat zaman zaman sapmalar yaşadığını ortaya



Şekil 5.39: Gerçek vs. tahmin saçılım grafiği - Nizip (Mekânsal).

koymaktadır. Beydilli istasyonunda hata dağılımı nispeten simetrik bir yapı sergilemiş ve modelin dengeli tahminler sunduğunu göstermiştir, ancak uç değerlerde belirli seviyelerde hata yoğunluğu dikkat çekmiştir. Nizip istasyonunda ise modelin tahmin hataları genellikle dar bir aralıkta dağılmış, bu da Nizip'te modelin daha stabil tahminler ürettiğini ifade etmektedir. Genel olarak, hata dağılımı analizleri mekânsal modelin ortalama olarak doğru tahminler yaptığını, ancak belirli dönemlerde küçük sapmaların meydana geldiğini ortaya koymaktadır; özellikle Atapark ve Beydilli istasyonlarında bazı uç değerlerde hatalar gözlenirken, Nizip istasyonunda modelin tahmin hatalarının daha dengeli olduğu görülmüştür.

Zaman serisi grafiklerinde, modelin tahminlerinin genel olarak gerçek değerlerle uyumlu olduğu, ancak bazı ani dalgalanmaları tam olarak yakalayamadığı belirlenmiştir. Saçılım grafikleri, tahmin edilen ve gerçek değerler arasındaki ilişkinin güçlü olduğunu ortaya koymuş, fakat yüksek PM_{2.5} seviyelerinde modelin tahminlerinin genellikle gerçek değerlere kıyasla daha düşük kaldığını göstermiştir. Density residual (hata yoğunluğu) grafikleri ise modelin hata dağılımının nispeten dengeli olduğunu ve büyük hata değerlerinin nadir görüldüğünü ortaya koymuştur. Bu analizler, mekânsal modelin tahmin performansının başarılı olduğunu kanıtlamakla birlikte, model doğruluğunu artırmak için ek mekânsal ve zamansal faktörlerin modele dahil edilmesinin faydalı olabileceğini işaret etmektedir. Ayrıca, mekânsal modelin daha



Şekil 5.40: Atapark istasyonu için density residual (hata dağılımı) Grafiği

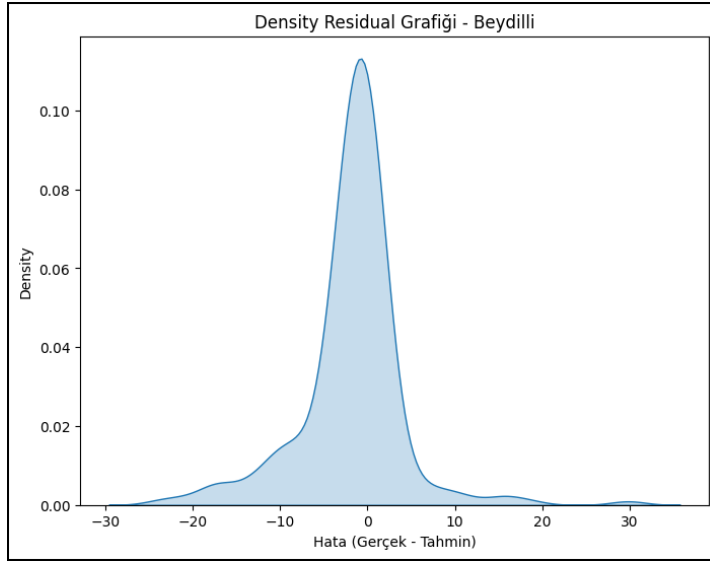
geniş bir coğrafi alanda daha fazla istasyon verisi ile beslenmesi, modelin genelleme yeteneğini güçlendirebilir ve tahmin hatalarını daha da azaltabilir.

5.3.4 Transfer learning destekli hibrit modelleme

Çalışmada, eksik verilerin tamamlanması, $PM_{2.5}$ tahmin doğruluğunun artırılması ve sensör verisi olmayan istasyonlarda tahmin yapılabilmesi amacıyla Autoencoder tabanlı Gizli Bağımlılık Faktörü (LDF) ile LightGBM modelini birleştiren bir hibrit modelleme süreci önerilmiştir. Bu süreç, üç temel aşamadan oluşur: İlk olarak, Autoencoder kullanılarak Latent Dependency Factor (LDF) çıkarımı gerçekleştirilir; ardından Transfer Learning yöntemiyle eksik veriler tamamlanır ve son olarak LightGBM ile $PM_{2.5}$ tahminleri yapılır. Bu yaklaşımda, Atapark, Beydilli, Nizip ve Fevzi Çakmak istasyonlarından elde edilen özellikler kullanılarak, sensör verisi bulunmayan GASKİ ve Meteoroloji istasyonları için $PM_{2.5}$ tahminleri başarıyla üretilmiştir.

5.3.4.1 Autoencoder modeli ile gizli bağımlılık faktörü çıkarımı (LDF)

Transfer Learning sürecinin ilk aşaması, $PM_{2.5}$ tahmini için derin özelliklerin (Latent Dependency Factor - LDF) çıkarılmasıdır ve bu işlem Autoencoder modeli kullanılarak gerçekleştirilmiştir. LDF, Autoencoder tarafından öğrenilen sıkıştırılmış özellikler olup, hava kirliliği tahmininde önemli olan ancak doğrudan ölçülemeyen gizli ilişkileri



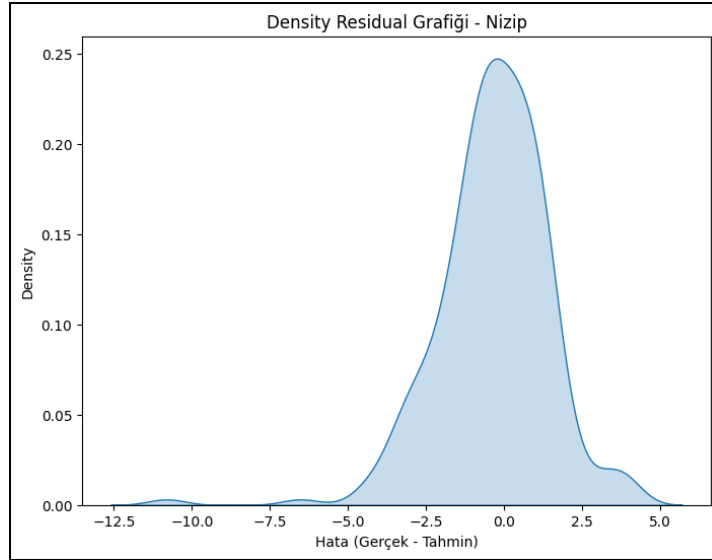
Şekil 5.41: Beydilli istasyonu için density residual (hata dağılımı) Grafiği

temsil eder. Hava kalitesi verileri genellikle eksiklikler içerdiğinden, Autoencoder bu eksiklikleri gidermek, istasyonlar arasında genelleme sağlamak ve sensör bulunmayan istasyonlara tahmin transferi yapmak amacıyla tercih edilmiştir. Ayrıca, ham verilerin doğrudan kullanılması model karmaşıklığını artırırken, LDF kullanımı veriyi sıkıştırarak modelin öğrenme sürecini kolaylaştırmıştır.

5.3.4.2 Transfer learning ile veri tamamlama

Transfer Learning ile veri tamamlama aşamasında, sensör verisi bulunmayan GASKİ ve Meteoroloji istasyonları için $PM_{2.5}$ tahminleri yapabilmek amacıyla Autoencoder ve IDW (Inverse Distance Weighting) yöntemlerinden faydalanılmıştır. Bu süreçte, transfer edilecek istasyonlar ve kullanılan yöntemler belirli bir düzen içinde ele alınmıştır. Kaynak istasyonlar olarak Atapark, Beydilli, Nizip ve Fevzi Çakmak belirlenmiş; hedef istasyonlar ise GASKİ ve Meteoroloji olarak seçilmiştir. Bu kapsamda, Autoencoder tabanlı LDF transferi ve IDW yöntemi ile mekânsal veri tahmini olmak üzere iki temel yöntem kullanılmıştır.

IDW ile LDF transferi sürecinde, GASKİ ve Meteoroloji istasyonlarında $PM_{2.5}$ ölçümü yapılmadığı için bu istasyonlara ait LDF değerlerinin tahmini gerçekleştirilmiştir. Bu amaçla, en yakın üç istasyondan alınan LDF değerleri esas alınarak mesafeye bağlı ağırlıklı ortalama yöntemiyle hesaplamalar yapılmıştır. Böylece, GASKİ ve



Şekil 5.42: Nizip istasyonu için density residual (hata dağılımı) Grafiği

Meteoroloji istasyonları için LDF vektörleri başarılı bir şekilde elde edilmiştir. Bu yaklaşım, eksik verilerin tamamlanmasında etkili bir çözüm sunarak tahminlerin doğruluğunu artırmayı hedeflemiştir.

5.3.4.3 Transfer learning ile PM_{2.5} tahmini

Transfer Learning sürecinin son aşamasında, elde edilen Latent Dependency Factor (LDF) ve diğer mekânsal ile meteorolojik değişkenler kullanılarak PM_{2.5} konsantrasyonlarının tahmini gerçekleştirilmiştir. Bu aşamada, Fevzi Çakmak, GASKİ ve Meteoroloji istasyonları için LightGBM modeli kullanılarak PM_{2.5} tahminleri yapılmıştır. Modelde bağımsız değişkenler olarak meteorolojik değişkenler (rüzgar hızı, rüzgar yönü, bağıl nem, basınç vb.), mekânsal değişkenler (nüfus yoğunluğu, sanayi mesafesi, yol yoğunluğu vb.) ve Autoencoder ile üretilen LDF vektörleri (LDF_1, LDF_2, LDF_3, LDF_4) kullanılmıştır. Bağımlı değişken ise PM_{2.5} konsantrasyonu olarak belirlenmiştir. Modelin eğitimi ve test edilmesi sürecinde, transfer edilen LDF'lerin modelin tahmin doğruluğuna etkisi analiz edilmiştir.

Fevzi Çakmak istasyonu, sensör verisi bulunan bir nokta olması nedeniyle modelin doğruluğunu test etmek için kullanılmış ve başarılı bir performans göstermiştir ($R^2 = 0.81$, MAE = 4,47, RMSE = 5,88). Gerçek ölçüm verileri ile karşılaştırma yapılarak modelin güvenilirliği değerlendirilmiştir. GASKİ ve Meteoroloji istasyonlarında

doğrudan PM_{2.5} ölçüm verisi bulunmadığı için bu istasyonlarda model yalnızca tahmin amaçlı uygulanmış; dolayısıyla bu noktalarda performans metrikleri hesaplanmamıştır. Ancak Fevzi Çakmak istasyonunda doğrulanan model performansı, diğer istasyonlardaki tahminlerin görece güvenilirliğine işaret etmektedir.

Model Performans Sonuçları aşağıdaki Çizelge 5.26'da sunulmuştur.

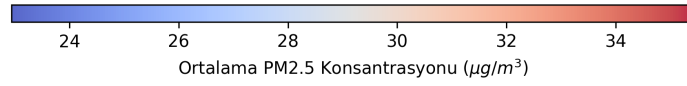
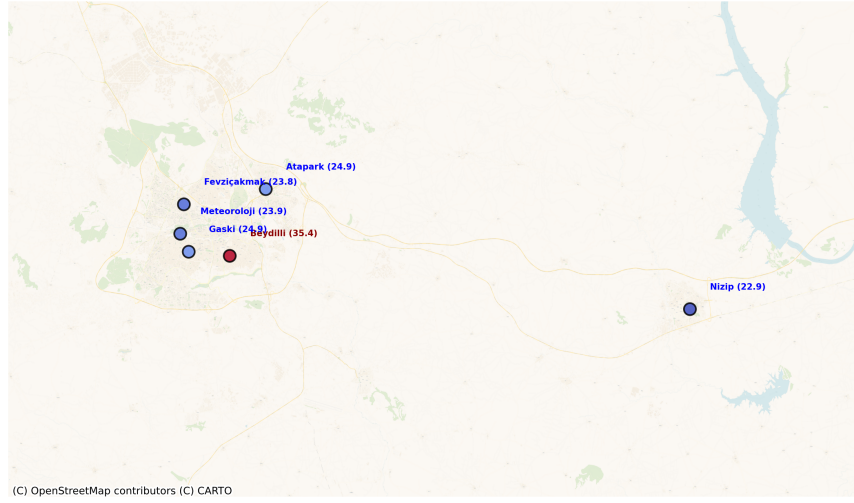
Çizelge 5.26: Transfer learning modeli performans sonuçları

İstasyon	Model	MAE	RMSE	R²
Fevzi Çakmak	LightGBM	4,47	5,88	0,81

Genel olarak, Transfer Learning destekli LightGBM modeli, eksik verilerin tamamlanması ve PM_{2.5} tahminlerinin iyileştirilmesi açısından başarılı sonuçlar sunmuştur. Özellikle Fevzi Çakmak istasyonunda elde edilen düşük hata değerleri, modelin transfer edilen LDF'leri etkin bir şekilde öğrenebildiğini göstermektedir. Bu yöntem, sensör verisi bulunmayan istasyonlarda da güvenilir PM_{2.5} tahminleri yapmak için güçlü bir alternatif sunmaktadır.

Şekil 5.43, nihai tahmin edilen PM_{2.5} ortalamalarının mekânsal dağılımını göstermektedir. Modelleme sürecinde elde edilen tahmin değerleri kullanılarak, her istasyon için ortalama PM_{2.5} seviyeleri hesaplanmış ve coğrafi olarak görselleştirilmiştir. Haritada kırmızı tonlar hava kirliliğinin daha yüksek olduğu bölgeleri, mavi tonlar ise daha düşük seviyedeki PM_{2.5} konsantrasyonlarını ifade etmektedir. Beydilli istasyonu, 35.4 µg/m³ ile en yüksek PM_{2.5} değerine sahip bölge olarak belirlenirken, Nizip istasyonu 22.9 µg/m³ ile en düşük PM_{2.5} seviyesine sahip bölge olmuştur. GASKİ ve Meteoroloji istasyonlarında sensör verisi bulunmadığından, bu noktaların PM_{2.5} değerleri transfer öğrenme modeli ile tahmin edilmiştir. Bu mekânsal tahmin haritası, modelin farklı istasyonlardaki PM_{2.5} seviyelerini başarılı bir şekilde yansıttığını göstermektedir.

Tahmin Edilen PM2.5 Değerlerinin Mekansal Olarak Ortalama Dağılımı



Şekil 5.43: Tahmin edilen nihai PM_{2.5} ortalama dağılım haritası.



6. SONUÇ VE ÖNERİLER

Bu çalışma, Gaziantep ilinde $PM_{2.5}$ kirliliğinin zamansal ve mekânsal düzeyde tahmin edilmesini amaçlamış; makine öğrenmesi, derin öğrenme ve transfer öğrenme yaklaşımlarının etkinliğini kapsamlı biçimde değerlendirmiştir. Atapark, Beydilli, Nizip, Fevzi Çakmak, GASKİ ve Meteoroloji istasyonlarına ait hava kalitesi, meteorolojik ve mekânsal veriler analiz edilmiş; eksik veri tamamlama, aykırı değer kontrolü, öznitelik mühendisliği ve modelleme süreçleri yürütülerek $PM_{2.5}$ tahminleri gerçekleştirilmiştir.

Elde edilen bulgular, hem bilimsel modelleme başarısını hem de politika geliştirme açısından önemli çıkarımları ortaya koymaktadır. LightGBM algoritması gerek zamansal gerekse mekânsal tahminlerde diğer modellere kıyasla üstün performans sergilemiştir. Özellikle gecikmeli değişkenler ve hareketli ortalamaların modele dahil edilmesi, tahmin doğruluğunu artırmıştır. $PM_{2.5}$ 'in bir önceki gün değeri ve son üç günlük ortalaması, tahmin sürecinde belirleyici öznitelikler olmuştur. Mekânsal modellemede ise nüfus yoğunluğu, yapılaşma oranı, yol yoğunluğu ve sanayi mesafesi gibi değişkenler önemli açıklayıcı faktörler arasında yer almıştır. Zamansal çıktılar mekânsal modele entegre edildiğinde, özellikle Nizip gibi istasyonlarda model performansında anlamlı bir artış gözlenmiştir.

Sensör verisi bulunmayan GASKİ ve Meteoroloji istasyonlarında transfer öğrenme destekli hibrit modelleme kullanılmış, Autoencoder ile elde edilen Latent Dependency Factor (LDF) temsilleri LightGBM modeline entegre edilerek tahminler gerçekleştirilmiştir. Fevzi Çakmak istasyonunda yapılan testler, modelin genellenebilirliğini ve güvenilirliğini desteklemiştir.

İstasyonlar arasında gözlenen $PM_{2.5}$ farklılıkları, yerel emisyon kaynaklarının ve fiziksel çevre koşullarının etkisini açıkça ortaya koymuştur. Beydilli gibi kentsel bölgelerde yüksek konsantrasyonlar gözlenirken, Nizip gibi kırsal bölgelerde daha

düşük seviyeler kaydedilmiştir. Ayrıca kış aylarında $PM_{2.5}$ seviyelerindeki artış, evsel ısınma kaynaklı emisyonların hava kalitesi üzerindeki etkisini vurgulamıştır.

Meteorolojik değişkenlerin $PM_{2.5}$ seviyeleri üzerindeki etkisi kapsamlı şekilde değerlendirilmiştir. Güneyden esen rüzgârların Gaziantep'e çöl kökenli toz taşınımı getirdiği gözlenmiş; bağıl nemin ise partiküllerin atmosferde kalma süresini artırarak $PM_{2.5}$ konsantrasyonlarını yükselttiği tespit edilmiştir. Bu durum, Gaziantep'in yalnızca yerel değil, aynı zamanda bölgesel hava kirliliği kaynaklarından da etkilendiğini ortaya koymaktadır. Bu çerçevede, şehrin güneyine ve kuzeyine kurulacak arka plan hava kalitesi izleme istasyonları, güneyden sınır ötesi toz taşınımı ve kuzeyde konumlanan organize sanayi bölgelerinden kaynaklı sanayi etkilerini izlemek açısından önem taşımaktadır. Çalışmada kullanılan uydu temelli Aerosol Optik Derinlik (AOD) verileri, MODIS-MAIAC ürününden elde edilmiştir. Ancak AOD verilerinin yer seviyesindeki $PM_{2.5}$ konsantrasyonları ile doğrudan birebir ilişkili olmaması, atmosferik partiküllerin dikey dağılımı, optik özellikleri ve yerel meteorolojik koşullar gibi faktörlere bağlıdır. Bu nedenle AOD verilerinin $PM_{2.5}$ tahmininde daha etkili kullanılabilmesi için lokal kalibrasyon yöntemlerine ve dikey profil verilerine ihtiyaç duyulmaktadır.

Çalışmanın sınırlılıkları arasında, yalnızca 2021–2023 yıllarını kapsayan bir zaman dilimi kullanılması nedeniyle uzun vadeli eğilim analizlerinin yapılamamış olması yer almaktadır. Ayrıca sensör verisi bulunmayan istasyonlardaki tahminlerin doğruluğu doğrudan ölçümle teyit edilemediği için belirli düzeyde belirsizlik içermektedir. Derin öğrenme modelleri (LSTM, GRU) ise veri yetersizliği ve yüksek hesaplama maliyetleri nedeniyle beklenen performansı gösterememiştir.

Tüm bu bulgular ışığında Gaziantep'te hava kalitesi yönetiminin geliştirilmesi için bazı öneriler sunulabilir: Mevcut hava kalitesi izleme ağı sanayi, yoğun trafik ve sınır bölgelerini kapsayacak şekilde genişletilmeli; özellikle şehrin güney ve kuzey bölgelerine arka plan istasyonları kurulmalıdır. Kış aylarında $PM_{2.5}$ artışlarını azaltmak adına düşük emisyonlu yakıt kullanımı ve filtreleme sistemleri desteklenmelidir. MODIS MAIAC AOD verileri yerel koşullara göre kalibre edilerek

modelleme sürecine entegre edilmeli, sanayi ve trafik kaynaklı emisyon verileri de düzenli olarak toplanarak açıklayıcı modeller oluşturulmalıdır.

Sonuç olarak, bu tez çalışması PM_{2.5} hava kirliliğinin zamansal ve mekânsal yönlerini anlamaya yönelik çağdaş modelleme tekniklerini etkili biçimde kullanarak hem bilimsel hem de uygulamaya dönük katkılar sunmuştur. LightGBM ve transfer öğrenme yaklaşımları, veri eksikliği olan bölgelerde dahi yüksek doğrulukla tahmin yapılabileceğini göstermiştir. Geliştirilen modelleme stratejileri, benzer kirlilik profiline sahip diğer şehirler için de uygulanabilir bir çerçeve oluşturarak çevresel sürdürülebilirlik ve halk sağlığının korunmasına katkı sağlamaktadır.





KAYNAKLAR

- [1] **Wark, K. ve Warner, C.F.** (1981). *Air pollution: its origin and control*, Harper and Row Publishers, New York, NY.
- [2] **Heinsohn, R.J. ve Kabel, R.L.** (1998). *Sources and control of air pollution*, Prentice Hall.
- [3] **Lovett, G.M., Tear, T.H., Evers, D.C., Findlay, S.E., Cosby, B.J., Dunscomb, J.K., Driscoll, C.T. ve Weathers, K.C.** (2009). Effects of air pollution on ecosystems and biological diversity in the eastern United States, *Annals of the New York Academy of Sciences*, 1162(1), 99–135.
- [4] **Smith, W.H.** (1974). Air pollution—effects on the structure and function of the temperate forest ecosystem, *Environmental Pollution (1970)*, 6(2), 111–129.
- [5] **Sharma, S.B., Jain, S., Khirwadkar, P. ve Kulkarni, S.** (2013). The effects of air pollution on the environment and human health, *Indian Journal of Research in Pharmacy and Biotechnology*, 1(3), 391.
- [6] **Sundar, S.I.K.** (2018). A review of studies on ecological impacts of air pollution, *Journal of Ecology and Natural Resources*, 2(6), 1–4.
- [7] **Jonidi Jafari, A., Charkhloo, E. ve Pasalari, H.** (2021). Urban air pollution control policies and strategies: a systematic review, *Journal of Environmental Health Science and Engineering*, 19, 1911–1940.
- [8] **Ruchi Singh, R.S. ve Amit Kumar, A.K.**, (2019). Air pollution control: policies and legislations., Air pollution: sources, impacts and controls, CAB International Wallingford UK, s.198–215.
- [9] **Tepe, A.M. ve Doğan, G.** (2019). Türkiye'nin güney sahilinde yer alan dört şehrin hava kalitelerinin incelenmesi, *Mühendislik Bilimleri ve Tasarım Dergisi*, 7(3), 585–595.
- [10] **Zeydan, Ö.** (2021). Assessment of Particulate Matter (PM10) Pollution in Turkey in 2019, *Journal of the Institute of Science and Technology*, 11(1), 106–118, <http://dx.doi.org/10.21597/jist.745539>.
- [11] **Şişman, E.** (2019). Türkiye'de seçilen hava kalitesi izleme istasyonları için eğilim (trend) değerlendirmeleri, *Doğal Afetler ve Çevre Dergisi*, 5(1), 134–152, gönderim Tarihi: 17/07/2018, Kabul Tarihi: 02/10/2018, ORCID: 0000-0003-3696-9967.

- [12] **Thangavel, P., Park, D. ve Lee, Y.C.** (2022). Recent insights into particulate matter (PM_{2.5})-mediated toxicity in humans: an overview, *International journal of environmental research and public health*, 19(12), 7511.
- [13] **Gautam, S., Yadav, A., Tsai, C.J. ve Kumar, P.** (2016). A review on recent progress in observations, sources, classification and regulations of PM_{2.5} in Asian environments, *Environmental Science and Pollution Research*, 23, 21165–21175.
- [14] **Fromme, H.** (2012). Particles in the indoor environment, *Air Quality–Monitoring and Modeling*, 117–144.
- [15] **Wang, J., Zhang, M., Bai, X., Tan, H., Li, S., Liu, J., Zhang, R., Wolters, M.A., Qin, X., Zhang, M. ve diğerleri** (2017). Large-scale transport of PM_{2.5} in the lower troposphere during winter cold surges in China, *Scientific reports*, 7(1), 13238.
- [16] **Liu, C., Huang, J., Wang, Y., Tao, X., Hu, C., Deng, L., Xu, J., Xiao, H.W., Luo, L., Xiao, H.Y. ve diğerleri** (2020). Vertical distribution of PM_{2.5} and interactions with the atmospheric boundary layer during the development stage of a heavy haze pollution event, *Science of the Total Environment*, 704, 135329.
- [17] **Bozkurt, Z.** (2018). PM₁₀ ve PM_{2.5} boyutundaki atmosferik partiküllerin bölgesel, mevsimsel değişimlerinin ve meteorolojik parametrelerle ilişkilerinin incelenmesi, *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 6(4), 1305–1320.
- [18] **Anchan, A., Shedthi, B.S. ve Manasa, G.** (2022). Models Predicting PM_{2.5} Concentrations—A Review, *Recent Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE 2020*, 65–83.
- [19] **Narkhede, G., Hiwale, A.S., Pawar, M. ve Dixit, S.** (2023). Comparative Analysis of Prediction Models for Particulate Matter (PM_{2.5}) Prediction, *2023 First International Conference on Advances in Electrical, Electronics and Computational Intelligence (ICAECCI)*, IEEE, s.1–6.
- [20] **Vesilind, P.A., Morgan, S.M. ve Heine, L.G.** (2010). *Introduction to Environmental Engineering-SI Version*, Cengage Learning.
- [21] **Yerramilli, A., Dodla, V.B.R. ve Yerramilli, S.** (2011). Air pollution, modeling and GIS based decision support systems for air quality risk assessment, *Advance air pollution*, 295–324.
- [22] **Liu, H., Tian, H., Zhang, K., Liu, S., Cheng, K., Yin, S., Liu, Y., Liu, X., Wu, Y., Liu, W., Bai, X., Wang, Y., Shao, P., Luo, L., Lin, S., Chen, J. ve Liu, X.** (2019). Seasonal variation, formation mechanisms and potential sources of PM_{2.5} in two typical cities in the Central Plains Urban Agglomeration, China, *Science of The Total Environment*, 657, 657–670, <https://www.sciencedirect.com/science/article/pii/S0048969718349088>.

- [23] **Kuang, X., Wang, X., Wang, W. ve Hu, M.** (2022). Formation mechanisms and sources of PM_{2.5}: Current knowledge and future perspectives, *Environmental Pollution*, 311, 119845, <https://doi.org/10.1016/j.envpol.2022.119845>.
- [24] **Singh, N., Murari, V., Kumar, M., Barman, S.C. ve Banerjee, T.** (2017). Fine particulates over South Asia: Review and meta-analysis of PM_{2.5} source apportionment through receptor model, *Environmental Pollution*, 223, 121–136.
- [25] **Shi, Z., Li, J., Huang, L., Wang, P., Wu, L., Ying, Q., Zhang, H., Lu, L., Liu, X., Liao, H. ve Hu, J.** (2017). Source apportionment of fine particulate matter in China in 2013 using a source-oriented chemical transport model, *Science of The Total Environment*, 601-602, 1476–1487, <https://www.sciencedirect.com/science/article/pii/S0048969717314237>.
- [26] **Wang-Li, L.** (2015). Insights to the formation of secondary inorganic PM_{2.5}: Current knowledge and future needs, *International Journal of Agricultural and Biological Engineering*, 8(2), 1–13.
- [27] **Zhang, R., Wang, G., Guo, S., Zamora, M.L., Ying, Q., Lin, Y., Wang, W., Hu, M. ve Wang, Y.** (2015). Formation of urban fine particulate matter, *Chemical reviews*, 115(10), 3803–3855.
- [28] **Wan, X., Kang, S., Xin, J., Liu, B., Wen, T., Wang, P., Wang, Y. ve Cong, Z.** (2016). Chemical composition of size-segregated aerosols in Lhasa city, Tibetan Plateau, *Atmospheric Research*, 174, 142–150.
- [29] **Ji, Z., Fan, X., Gan, M., Chen, X., Lv, W., Li, Q., Zhou, Y., Tian, Y. ve Jiang, T.** (2017). Influence of sulfur dioxide-related interactions on PM_{2.5} formation in iron ore sintering, *Journal of the Air & Waste Management Association*, 67(4), 488–497, <https://doi.org/10.1080/10962247.2016.1259692>, published online: 23 Feb 2017.
- [30] **Bhargava, A.** (2020). Air pollution-sources and classification, *International Journal Rural and Regional Planning Development*, 6(1), 32–38.
- [31] **Xing, Y.F., Xu, Y.H., Shi, M.H. ve Lian, Y.X.** (2016). The impact of PM_{2.5} on the human respiratory system, *Journal of thoracic disease*, 8(1), E69.
- [32] **Hou, T., Zhu, L., Wang, Y. ve Peng, L.** (2024). Oxidative stress is the pivot for PM_{2.5}-induced lung injury, *Food and Chemical Toxicology*, 184, 114362.
- [33] **Yang, L., Li, C. ve Tang, X.** (2020). The impact of PM_{2.5} on the host defense of respiratory system, *Frontiers in cell and developmental biology*, 8, 91.
- [34] **Cao, L., Zhou, Y., Zhang, Z., Sun, W., Mu, G. ve Chen, W.** (2016). Impacts of airborne particulate matter and its components on respiratory system health, *Zhonghua yu Fang yi xue za zhi [Chinese Journal of Preventive Medicine]*, 50(12), 1114–1118.

- [35] **Li, X. ve Liu, X.** (2021). Effects of PM_{2.5} on chronic airway diseases: A review of research progress, *Atmosphere*, 12(8), 1068.
- [36] **Yang, L., Zhang, Y., Qi, W., Zhao, T., Zhang, L., Zhou, L. ve Ye, L.** (2022). Adverse effects of PM_{2.5} on cardiovascular diseases, *Reviews on Environmental Health*, 37(1), 71–80.
- [37] **Hamanaka, R.B. ve Mutlu, G.M.** (2018). Particulate matter air pollution: effects on the cardiovascular system, *Frontiers in endocrinology*, 9, 680.
- [38] **Basith, S., Manavalan, B., Shin, T.H., Park, C.B., Lee, W.S., Kim, J. ve Lee, G.** (2022). The impact of fine particulate matter 2.5 on the cardiovascular system: a review of the invisible killer, *Nanomaterials*, 12(15), 2656.
- [39] **Krittanawong, C., Qadeer, Y.K., Hayes, R.B., Wang, Z., Thurston, G.D., Virani, S. ve Lavie, C.J.** (2023). PM_{2.5} and cardiovascular diseases: State-of-the-Art review, *International Journal of Cardiology Cardiovascular Risk and Prevention*, 19, 200217.
- [40] **Li, Z., Tang, Y., Song, X., Lazar, L., Li, Z. ve Zhao, J.** (2019). Impact of ambient PM_{2.5} on adverse birth outcome and potential molecular mechanism, *Ecotoxicology and environmental safety*, 169, 248–254.
- [41] **Jedrychowski, W.A., Perera, F.P., Maugeri, U., Mroz, E., Klimaszewska-Rembiasz, M., Flak, E., Edwards, S. ve Spengler, J.D.** (2010). Effect of prenatal exposure to fine particulate matter on ventilatory lung function of preschool children of non-smoking mothers, *Paediatric and perinatal epidemiology*, 24(5), 492–501.
- [42] **Jung, C.R., Chen, W.T., Tang, Y.H. ve Hwang, B.F.** (2019). Fine particulate matter exposure during pregnancy and infancy and incident asthma, *Journal of allergy and clinical immunology*, 143(6), 2254–2262.
- [43] **Amnuaylojaroen, T. ve Parasin, N.** (2024). Pathogenesis of PM_{2.5}-related disorders in different age groups: children, adults, and the elderly, *Epigenomes*, 8(2), 13.
- [44] **EPA** (2024). *Final Rule to Strengthen the National Ambient Air Quality Standards for Particulate Matter*, <https://www.epa.gov/naaqs>.
- [45] **DSÖ** (2021). *Küresel Hava Kalitesi Rehberi*, Dünya Sağlık Örgütü, Cenevre.
- [46] **Temiz Hava Hakkı Platformu** (2024). *Kara Rapor 2024: Hava Kirliliği ve Sağlık Etkileri*, https://www.temizhavahakki.org/wp-content/uploads/2024/10/KaraRapor-2024_w_26.09.2024.pdf.
- [47] **Shapovalov, A.V., Riabova, S.A., Krashennnikov, A.V., Loktev, D.N. ve Soloviev, S.P.** (2023). Effect of summer air temperature on the concentration of microparticles according to the data from the Mikhnevo observatory, *29th International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*, cilt12780, SPIE, s.1117–1120.

- [48] **Yang, Q., Yuan, Q., Li, T., Shen, H. ve Zhang, L.** (2017). The relationships between PM_{2.5} and meteorological factors in China: seasonal and regional variations, *International journal of environmental research and public health*, 14(12), 1510.
- [49] **Vannucci, P.F., Foley, K., Murphy, B.N., Hogrefe, C., Cohen, R.C. ve Pye, H.O.** (2024). Temperature-dependent composition of summertime PM_{2.5} in observations and model predictions across the Eastern US, *ACS Earth and Space Chemistry*, 8(2), 381–392.
- [50] **Vaishali, Verma, G. ve Das, R.M.** (2023). Influence of temperature and relative humidity on PM_{2.5} concentration over Delhi, *Mapan*, 38(3), 759–769.
- [51] **Sarkar, S., Srivastava, R. ve Sagar, K.** (2015). Diurnal monitoring of surface ozone and PM_{2.5} concentration and its correlation with temperature, *International Journal of Technology Enhancements and Emerging Engineering Research*, 3, 121–129.
- [52] **Steenland, K., Vu, B. ve Scovronick, N.** (2022). Effect modification by maximum temperature of the association between PM_{2.5} and short-term cardiorespiratory mortality and emergency room visits in Lima, Peru, 2010–2016, *Journal of exposure science & environmental epidemiology*, 32(4), 590–595.
- [53] **Mirabelli, M.C., Vaidyanathan, A., Flanders, W.D., Qin, X. ve Garbe, P.** (2016). Outdoor PM_{2.5}, ambient air temperature, and asthma symptoms in the past 14 days among adults with active asthma, *Environmental health perspectives*, 124(12), 1882–1890.
- [54] **Zalakeviciute, R., López-Villada, J. ve Rybarczyk, Y.** (2018). Contrasted effects of relative humidity and precipitation on urban PM_{2.5} pollution in high elevation urban areas, *Sustainability*, 10(6), 2064.
- [55] **Cheng, Y., He, K.b., Du, Z.y., Zheng, M., Duan, F.k. ve Ma, Y.l.** (2015). Humidity plays an important role in the PM_{2.5} pollution in Beijing, *Environmental pollution*, 197, 68–75.
- [56] **Arslan, H. ve Toltar, A.** (2023). The influence of meteorological parameters on PM_{2.5} and PM₁₀ values in Ümraniye and Silivri districts of İstanbul, *Environmental Research and Technology*, 6(4), 288–301.
- [57] **ÜNAL AKDENİZ, C. ve ÖZEL KADILAR, G.** (2023). Ankara İli Meteoroloji Parametrelerinin Hava Kirliliği Üzerindeki Etkilerinin Regresyon Analizi ile İncelenmesi, *Nicel bilimler dergisi (Online)*, 5(2).
- [58] **Babaan, J., Ballori, J., Tamondong, A., Ramos, R. ve Ostrea, P.** (2018). Estimation of PM_{2.5} vertical distribution using customized UAV and mobile sensors in Brgy. UP Campus, Diliman, Quezon City, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 89–103.

- [59] **Ouyang, W., Guo, B., Cai, G., Li, Q., Han, S., Liu, B. ve Liu, X.** (2015). The washing effect of precipitation on particulate matter and the pollution dynamics of rainwater in downtown Beijing, *Science of the Total Environment*, 505, 306–314.
- [60] **Öztürk, S., Gerçek, D., Güven, İ.T., Gaga, E., Üzmez, Ö.Ö. ve Civan, M.** (2021). Kocaeli İzmit İlçesi'nde Partikül Madde (PM_{2.5}) Konsantrasyon Seviyeleri, Mekânsal ve Mevsimsel Değerlendirilmesi, *Mühendislik Bilimleri ve Tasarım Dergisi*, 9(3), 809–821.
- [61] **Wang, C., Jia, M., Xia, H., Wu, Y., Wei, T., Shang, X., Yang, C., Xue, X. ve Dou, X.** (2019). Relationship analysis of PM_{2.5} and boundary layer height using an aerosol and turbulence detection lidar, *Atmospheric Measurement Techniques*, 12(6), 3303–3315.
- [62] **Gupta, P. ve Christopher, S.A.** (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach, *Journal of Geophysical Research: Atmospheres*, 114(D14).
- [63] **Gaziantep Valiliği** (2024). *İlimiz Gaziantep*, <http://www.gaziantep.gov.tr/ilimiz-gaziantep>.
- [64] **Gaziantep Organize Sanayi Bölgesi (GAOSB)** (2024). *Bölge krokisi*, <https://www.gaosb.org/tr/genel-sayfa/bolgeler/bolge-krokisi-17.html>, gaziantep Organize Sanayi Bölgesi Resmi Web Sitesi. Erişim tarihi: 23 Mart 2025.
- [65] **Milli Savunma Bakanlığı Harita Genel Müdürlüğü (MSB HGM)** (2024). *Gaziantep Fiziki İl Haritası*, <https://www.harita.gov.tr/urun/gaziantep-fiziki-il-haritasi/406>, harita Genel Müdürlüğü Resmi Web Sitesi. Erişim tarihi: 23 Mart 2025.
- [66] **Meteoroloji Genel Müdürlüğü (MGM)** (2024). *İl ve ilçeler istatistikleri: Gaziantep*, <https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx?k=A&m=GAZIANTEP>, meteoroloji Genel Müdürlüğü Resmi Web Sitesi. Erişim tarihi: 23 Mart 2025.
- [67] **Yang, J. ve Ismail, A.W.** (2022). Air quality forecasting using deep learning and transfer learning: a survey, *2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT)*, IEEE, s.1–6.
- [68] **Mane, M.V., Kumar, D. ve Agarwal, K.** (2022). Detection and prediction of air pollution using machine learning and deep learning techniques, *2022 international conference on computing, communication, and intelligent systems (ICCCIS)*, IEEE, s.145–150.
- [69] **Geetha, D. ve diğerleri** (2022). Prediction Of Concentration of Air Pollution Using Deep and Machine Learning, *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, cilt 1, IEEE, s.1369–1375.

- [70] **Sinnott, R.O. ve Guan, Z.** (2018). Prediction of air pollution through machine learning approaches on the cloud, *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, IEEE, s.51–60.
- [71] **Méndez, M., Merayo, M.G. ve Núñez, M.** (2023). Machine learning algorithms to forecast air quality: a survey, *Artificial Intelligence Review*, 56(9), 10031–10066.
- [72] **T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı** (2024). *Ulusal Hava Kalitesi İzleme Ağı (UHİKA) Veri Portalı*, https://sim.csb.gov.tr/STN/STN_Report/StationDataDownloadNew, erişim tarihi: 25 Mart 2025.
- [73] **Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D. ve Thépaut, J.N.** (2023). *1940'tan günümüze kadar olan tek seviyelerdeki ERA5 saatlik verileri [Veri seti]*, <https://doi.org/10.24381/cds.adbb2d47>, veri seti.
- [74] **Zippenfenig, P.** (2023). *Open-Meteo.com Hava Durumu API [Bilgisayar yazılımı]*, <https://doi.org/10.5281/ZENODO.7970649>, bilgisayar yazılımı.
- [75] **Muñoz Sabater, J.** (2019). *ERA5-Land hourly data from 2001 to present*, <https://doi.org/10.24381/CDS.E2161BAC>, dataset, Copernicus Climate Change Service (C3S) via Climate Data Store (CDS), ECMWF.
- [76] **Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.M., Dominguez, J.J., Engelen, R., Eskes, H., Flemming, J. ve diğerleri** (2019). The CAMS reanalysis of atmospheric composition, *Atmospheric Chemistry and Physics*, 19(6), 3515–3556.
- [77] **European Environment Agency (EEA)** (2018). *CORINE Land Cover 2018*, <https://doi.org/10.2909/960998c1-1870-4e82-8051-6485205ebbac>, veri seti, Avrupa Çevre Ajansı (EEA) tarafından yayımlanmıştır.
- [78] **Schiavina, M., Freire, S., Carioli, A. ve MacManus, K.** (2023). *GHS-POP R2023A - GHS population grid dataset (1975–2030)*, <https://doi.org/10.2905/2ff68a52-5b5b-4a22-8f40-c41da8332cfe>, dataset published by the European Commission, Joint Research Centre (JRC).
- [79] **Agency, E.S.** (n.d.). *Sentinel-2 Mission Overview*, https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2, accessed: March 2025.
- [80] **Boeing, G.** (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks, *Computers, Environment and Urban Systems*, 65, 126–139.



ÖZGEÇMİŞ

Ad SOYADI: Türkan ZENGİN GÖMLEKSİZ

ÖĞRENİM DURUMU:

- **Lisans:** 2020, İstanbul Teknik Üniversitesi, Uçak ve Uzay Bilimleri Fakültesi, Meteoroloji Mühendisliği

MESLEKİ DENEYİMLER VE ÖDÜLLER:

- 2021-2022 yılları arasında Meteoroloji Genel Müdürlüğü, Batman Meydan Meteoroloji Müdürlüğü'nde Meteoroloji Mühendisi olarak çalıştı.
- 2022 yılından beri Meteoroloji Genel Müdürlüğü, Kilis Meteoroloji Müdürlüğü'nde Meteoroloji Mühendisi olarak çalışmaktadır.

YÜKSEK LİSANS TEZİNDEN TÜRETİLEN YAYINLAR, SUNUMLAR VE PATENTLER:

- **ZENGİN GÖMLEKSİZ, Türkan.** (2024). *Enhancing PM_{2.5} Forecasting through Integrated Data Sets in the Gaziantep Region.* 11th International Symposium on Atmospheric Sciences, 23–25 October 2024, İstanbul, Türkiye.

DİĞER YAYINLAR, SUNUMLAR VE PATENTLER:

- Zengin Gömleksiz, T. (2023). *The Relationship Between Planetary Boundary Layer Height and PM₁₀ Concentration in Adana.* **Research in Atmospheric Science**, 2023(2), 23–28. DOI: <http://dx.doi.org/10.29228/resatmsci.74323>