

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DERİN ÖĞRENME YÖNTEMİ İLE HAVA KALİTESİ
PARAMETRELERİNİN TAHMİNİ: SAKARYA ÖRNEĞİ

YÜKSEK LİSANS TEZİ

Ayşegül ATALI

Çevre Mühendisliği Anabilim Dalı

ŞUBAT 2023

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DERİN ÖĞRENME YÖNTEMİ İLE HAVA KALİTESİ
PARAMETRELERİNİN TAHMİNİ: SAKARYA ÖRNEĞİ

YÜKSEK LİSANS TEZİ

Ayşegül ATALI

Çevre Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç.Dr. Beytullah EREN

ŞUBAT 2023

Ayşegül ATALI tarafından hazırlanan “Derin Öğrenme Yöntemi İle Hava Kalitesi Parametrelerinin Tahmini: Sakarya Örneği” adlı tez çalışması 17.02.2023 tarihinde aşağıdaki jüri tarafından oy birliği/oy çokluğu ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Çevre Mühendisliği Anabilim Dalı Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :	Doç.Dr. Ahmet ÇELEBİ Sakarya Üniversitesi	
Jüri Üyesi :	Doç.Dr. Beytullah EREN (Danışman) Sakarya Üniversitesi	
Jüri Üyesi :	Dr.Öğr.Üyesi Caner ERDEN Sakarya Uygulamalı Bilimler Üniversitesi	



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “DERİN ÖĞRENME YÖNTEMİ İLE HAVA KALİTESİ PARAMETRELERİNİN TAHMİNİ: SAKARYA ÖRNEĞİ” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(26/04/2023)

(imza)

Ayşegül ATALI





Eşime ve çocuklarıma



TEŞEKKÜR

Yüksek lisans eğitimim boyunca değerli bilgi ve deneyimlerinden yararlandığım, her konuda bilgi ve desteğini almaktan çekinmediğim, araştırmanın planlanmasından yazılmasına kadar tüm aşamalarında yardımlarını esirgemeyen, teşvik eden, aynı titizlikte beni yönlendiren değerli danışman hocam Doç.Dr.Beytullah EREN'e teşekkürlerimi sunarım. Bilgi ve deneyimleriyle yardımlarını esirgemeyen, sorularıma sabırla cevap veren sayın hocam Dr. Öğr. Üyesi Caner ERDEN'e teşekkür ederim.

Ayrıca bu çalışmanın ortaya çıkmasına vesile olan ve kişisel gelişimim ile ilgili her konuda bana destek olan değerli eşim Dr. Öğr. Üyesi Gökhan ATALI'ya, beni yetiştirip büyüten ve sözlerime kıymet verip eğitimim için emek veren değerli annem Seher ÜNAL'a ve değerli babam İbrahim ÜNAL'a, hayatımın neşesi canım kızlarım Elif ve Betül'e teşekkür ederim.

Çalışma sırasında iş yoğunluğumdan dolayı zorluğa düştüğümde, yükümü hafiletip işlerime yardımcı olan Esra ÇAKMAK ve Ayşegül YÜKSEL'e, çalışmalarına destek olan tüm kıymetli mesai arkadaşlarıma, ayrıca çalışmama imkân ve olanak sağlayarak destek olan, değerli müdürüm Tekin KAYA'ya minnet ve teşekkürlerimi sunarım.

Son olarak, doğduğu andan bugüne kadar hem en yakın arkadaşım, hem de en büyük destekçim olan, her konuda bilgi sahibi olmasıyla girişimlerimin hepsinde bana yardımcı olan biricik kardeşim Muhammet Fatih ÜNAL'a ve motive edici sözleri ile, asla yargılamadan her koşulda beni dinleyen ve anlayan, kardeş kadar yakın tüm güzel kalpli arkadaşlarıma teşekkür ederim.

Ayşegül ATALI



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SİMGELER	xv
TABLO LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxiii
1. GİRİŞ	1
1.1. Tezin Amacı	1
1.2. Kuramsal Temeller	2
1.3. Tezin Önemi	8
2. LİTERATÜR ÖZETİ	9
2.1. Hava Kirliliğinin Tahmininde İstatistiksel Yöntemlerin Kullanımı	9
2.2. Hava Kirliliğinin Tahmininde Yapay Zeka (AI) Yöntemlerinin Kullanımı ...	10
2.3. Hava Kirliliğinin Tahmininde Derin Öğrenme Yöntemlerinin Kullanımı.....	12
2.4. Hava Kirliliğinin Tahmininde LSTM Algoritmasının Kullanımı	13
3. MATERYAL VE YÖNTEM	15
3.1. Yazılım Araçları	15
3.1.1. Python	15
3.1.2. Kullanılan kütüphaneler	16
3.1.2.1. Matplotlib	16
3.1.2.2. Keras ve tensorflow	16
3.1.2.3. Pandas	16
3.1.2.4. Numpy	17
3.1.2.5. Keras-tuner	17
3.2. Veri Setleri	17
3.3. Veri Ön İşleme	19
3.4. Veri Tahmini Yöntemleri, Performans Ölçümleri ve Hiperparametreler	21
3.4.1. LSTM (long short term memory) modeli	21
3.4.2. Performans ölçümleri	24
3.4.3. Hiperparametreler	25
3.4.3.1. Random search	26
3.4.3.2. Bayesian search	26
3.4.3.3. Hyperband search	27
4. DENEYSEL ÇALIŞMA VE SONUÇLARI	29
4.1. Veri Ön İşleme Sonrasında Elde Edilen Sonuçlar	29
4.2. Hiperparametre Optimizasyonu	33
4.3. Veri Setine Ait Sonuçlar	38
4.3.1. NO ₂ parametresi için model sonuçlarının değerlendirilmesi	38

4.3.2. NO _x parametresi için model sonuçlarının değerlendirmesi.....	40
4.3.3. CO parametresi için model sonuçlarının değerlendirmesi	42
4.3.4. PM ₁₀ parametresi için model sonuçlarının değerlendirmesi	44
5. SONUÇ VE ÖNERİLER.....	47
KAYNAKLAR.....	49
ÖZGEÇMİŞ.....	55



KISALTMALAR

LSTM	: Uzun kısa süreli bellek
HKİ	: Hava Kalitesi İndeksi
EPA	: Çevre Koruma Ajansı
USEPA	: Amerika Birleşik Devletleri Çevre Koruma Ajansı
UHKİİ	: Ulusal Hava Kalitesi İzleme İstasyonları
BMİDÇS	: BM İklim Değişikliği Çerçeve Sözleşmesi
IPCC	: Hükümetlerarası İklim Değişikliği Paneli
UMSHAK	: Uzun Menzilli Sınırlar Ötesi Hava Kirliliği Sözleşmesi
AI	: Yapay zeka
ML	: Makine öğrenmesi
YSA	: Yapay sinir ağları
CNN	: Konvolüsyonel Sinir Ağları
RNN	: Tekrarlayan Sinir Ağları
DBN	: Derin İnanç Ağları
SOM	: Özdüzenleyici haritalar
GRU	: Kapılı Tekrarlayan Hücre
CTM	: Kimyasal-taşıma modeli
MSE	: Hata kareler ortalaması
MAE	: Hatanın mutlak ortalaması
RMSE	: Hata kareler ortalamasının karekökü
HPO	: Hiperparametre optimizasyonu
kNN	: En yakın komşu algoritması



SİMGELER

PM₁₀	: Partikül madde kirleticisi (Çapı < 10 µm)
NO₂	: Azot dioksit
NO_x	: Azot oksitler
CO	: Karbonmonoksit
PM_{2,5}	: Partikül madde kirleticisi (Çapı < 2,5 µm)
O₃	: Ozon
SO₂	: Kükürt dioksit
CH₄	: Metan
SO_x	: Kükürt oksitler
H₂SO₄	: Sülfirik Asit
HNO₃	: Nitrik Asit



TABLO LİSTESİ

Sayfa

Tablo 1.1. Ulusal HKİ kesme noktaları	3
Tablo 1.2. İndeks hesaplanan parametrelerin sınır değerleri	3
Tablo 3.1. Sakarya Hava Kalitesi İstasyon konumları ve ölçülen parametreler	18
Tablo 3.2. Sakarya Merkez İstasyon PM ₁₀ , CO, NO _x , NO ₂ ölçüm değerleri.....	20
Tablo 3.3. Sakarya Merkez İstasyon PM ₁₀ , CO, NO _x , NO ₂ için minimum ve maximum değerleri	20
Tablo 4.1. 2020-2021 yılları arasında hava kirleticilerinin saatlik değişimlerin ortalama değerleri.....	32
Tablo 4.2. 2020-2022 yılları kirletici konsantrasyonu ortalamaları	33
Tablo 4.3. Hiperparametre kıyaslanması.....	34
Tablo 4.4. Ortak parametreler	36
Tablo 4.5. Parametre arama uzayı	36
Tablo 4.6. Arama algoritmalarının elde ettiği en iyi parametreler	36
Tablo 4.7. Arama algoritması çalışma süreleri.....	38
Tablo 4.8. NO ₂ parametresi için en iyi arama algortması performans sonuçları.....	40
Tablo 4.9. NO _x parametresi için en iyi arama algortması performans sonuçları.....	42
Tablo 4.10. CO parametresi için en iyi arama algortması performans sonuçları.....	44
Tablo 4.11. PM ₁₀ parametresi için en iyi arama algortması performans sonuçları ..	46



ŞEKİL LİSTESİ

Sayfa

Şekil 1.1. Makine öğrenmesi algoritması ve derin öğrenme algoritması kıyaslanması	6
Şekil 3.1. Sakarya merkez ve diğer istasyonların harita üzerinde konumu	19
Şekil 3.2. Veri sirkülasyonu	19
Şekil 4.1. PM ₁₀ 'nun zamana bağlı değişimi.....	30
Şekil 4.2. CO'nun zamana bağlı değişimi	30
Şekil 4.3. NO ₂ 'nin zamana bağlı değişimi.....	31
Şekil 4.4. NO _x 'in zamana bağlı değişimi	31
Şekil 4.5. Kirleticilerin violin grafiği.....	33
Şekil 4.6. Hiperparametre algoritmalarının kıyaslanması (Test seti)	35
Şekil 4.7. Hiperparametre algortimalarının kıyaslanması (Eğitim seti)	35
Şekil 4.8. Hiperparametre algoritmaları için R ² kıyaslaması (Test seti).....	37
Şekil 4.9. Hiperparametre algoritmaları için R ² kıyaslaması (Eğitim seti).....	37
Şekil 4.10. NO ₂ konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (eğitim seti)	39
Şekil 4.11. NO ₂ konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (test seti).....	39
Şekil 4.12. NO _x konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (eğitim seti)	41
Şekil 4.13. NO _x konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (test seti).....	41
Şekil 4.14. CO konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (eğitim seti)	43
Şekil 4.15. CO konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (test seti).....	43
Şekil 4.16. PM ₁₀ konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (eğitim seti)	45
Şekil 4.17. PM ₁₀ konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (test seti).....	45



DERİN ÖĞRENME YÖNTEMİ İLE HAVA KALİTESİ PARAMETRELERİNİN TAHMİNİ: SAKARYA ÖRNEĞİ

ÖZET

Hava kirliliği, son yıllarda küresel bir kriz olarak değerlendirilmektedir. Araştırmacılar hava kirliliği konusu üzerinde yoğun bir şekilde durmaktadır. Bunun sebebi, hava kirliliğinin insan sağlığı, bitki örtüsü ve su kaynakları için olumsuz etkilere sebep olmasıdır. Canlı yaşamı için önemli bir tehlike unsuru olan bu problem ile baş edilmesi için önlemler alınması gerekmektedir.

Bu çalışmada hava kirliliği sorunu için alternatif bir öneri olarak, hava kalitesi modellemesi ile hava kalitesinin tahmini ve analizi üzerinde durulmuştur. Hava kirliliği, LSTM (Long Short-Term Memory) algortiması kullanılarak modellenmiştir. Bu model ile geçmiş veriler kullanılarak gelecekteki hava kirliliği seviyelerinin tahmin edilmesi amaçlanmıştır.

Çalışmada, hava kalitesi ölçüm parametrelerinin saatlik ölçümüne dayalı veriler toplanmıştır ve kirliletiçi parametre olarak; PM_{10} , NO_2 , NO_x , CO kullanılmıştır. Ölçüm değeri UHKİA (Ulusal Hava Kalitesi İzleme Ağı) tarafından kurulan, Sakarya Endüstri Meslek Lisesinin bahçesinde konumlanan, hava kirliliği ölçüm istasyonundan kaydedilen verilerden oluşmaktadır. Modellemede 2020 Ocak ayından, 2022 Eylül ayına kadar kaydedilmiş veriler kullanılmıştır.

Hava kirliliği verilerinin toplanmasından sonra, veri ön işleme yöntemleri ile temizleme çalışması yapılmıştır. 2020 Ocak ve 2022 Eylül ayları arasında her bir parametre için, 24.282 adet veri olmak üzere toplam 97.128 adet veri elde edilmiştir. Veri ön işleme için kullanılan farklı iki yöntem ile iki ayrı veri seti elde edilmiştir. Bu iki adet veri seti arasında mukayeseler yapılmıştır.

Elde edilen 97.128 veri ile LSTM algortiması kullanılarak bir model oluşturulmuş, eğitim verileri kullanılarak model eğitilmiştir. Elde edilen doğrulama verileri kullanılarak modelin performansı, hiperparametre optimizasyonu ile değerlendirilmiştir. Kullanılan model, daha sonra gelecekteki hava kirliliği seviyelerini tahmin etmek için kullanılabileceği görülmüştür. Önerilen modelin hava kirliliği tahmini için yüksek performansa sahip olduğu görülmüştür.



FORECASTING AIR QUALITY PARAMETERS BY DEEP LEARNING METHOD: A CASE STUDY OF SAKARYA

SUMMARY

In recent years, air pollution has been recognized as a global crisis. Researchers are mainly focused on the issue of air pollution, as the problems caused by air pollution pose a vital danger. When studies related to air pollution are examined, their results generally indicate that the components in the air that are the source of vital functions are being disrupted, and understanding the consequences of this, as well as understanding that it is not possible to completely eliminate pollutant sources but that they can be reduced by stopping them through deep and detailed interdisciplinary studies, is crucial for human health. Air pollution poses a significant threat to human health, causing many diseases such as asthma, bronchitis, upper respiratory tract infections, and in some cases, life-threatening diseases.

In addition, acid rain caused by air pollution has deep and disturbing effects on the ecosystem. Air pollution, in the broadest sense, means the chemical, physical, and biological changes in the quality of air due to substances that pollute the air. Air pollutants can arise as a result of natural cycles or as a result of activities necessary for life. Common air pollutants include particulate matter (PM_{10} , $PM_{2.5}$), carbon monoxide (CO), ozone (O_3), nitrogen oxides (NO_2 , NO_x), and sulfur dioxide (SO_2).

Each of these pollutants must be restricted from separate sources and evaluated separately. For this reason, the density of each air quality measurement station is measured at regular intervals, and threshold values are determined to pose the least threat to health. Additionally, air quality is determined based on measurement results. In recent years, air pollution modeling has been performed based on obtained data, and efforts have been made to facilitate the prediction of future data using various traditional methods, as well as to take measures and precautions against possible air pollution events. The predictions obtained as a result of these studies make it easier to take future precautions, while also allowing measurement errors that may arise due to any technical failure in daily data to be easily detected, and error analyses can be more accurately evaluated to reach effective results. It is necessary to pay attention to regular interval measurements of air quality data, as in the example of Sakarya. Furthermore, air pollution modeling based on measurement results and artificial intelligence, as well as the prediction of future data, will facilitate taking measures and precautions against possible air pollution events. The prediction data obtained as a result of these studies will facilitate future measurements. In addition, measurement errors that may arise due to any technical problem in daily data will be easily detected. Error analyses will be evaluated more accurately to achieve effective results.

In this study, air quality prediction was made for Sakarya province. The air quality data was obtained from the official website of the Ministry of Environment and Urbanization, www.havaizleme.gov.tr. The study covered the period from January 2020 to September 2022.

Air quality data was collected by measuring stations established under the National Air Quality Monitoring Network. In the study, one of the four air quality measurement stations in Sakarya was used. The station used in the study is Sakarya Central station, which is located on Sakarya Street near the Vocational High School. It is located 3.5 meters from the nearest residence and 0.5 meters from the road. This station was established to measure air pollution caused by traffic and has been continuously measuring since March 2013. Four parameters were used for the prospective estimation of air quality in the study, which include PM₁₀, NO₂, NO_x, and CO. Hourly measurement results between January 2020 and September 2022 were used for each of the four parameters of the central station. There should be 97,128 data in total, 24,282 for each parameter, within the specified date range.

However, partial data deficiencies have been observed due to sensor errors and similar situations. Data cleaning was performed using Python in the dataset. Two separate datasets were created by completing the missing values in the dataset with mean and nearest neighbor (kNN) methods. After data preprocessing, missing data in the dataset were detected. Missing data were filled using two pure methods. Then, the values in the dataset were observed, and the periods when the values of the pollutant parameters were the most intense and the periods when they were the lowest were evaluated separately. A violin graph was created for the datasets filled with the mean and kNN methods considering the MSE results. Based on this graph, a comparison was made between the mean and kNN methods.

According to the obtained MSE results, it was observed that the distribution and average of the neighbor search algorithm were better than NO_x. It has been observed that the NO_x parameter gives better results when filled with the mean. Thus, the datasets to be used in the second stage were decided. The dataset filled with the mean method was used for NO_x. For NO₂, PM₁₀, and CO, the dataset filled with the kNN algorithm was used.

The dataset used in this study includes air quality data collected between January 2020 and September 2022, comprising a total of 97,128 samples, with 24,282 samples for each pollutant parameter. Mean and kNN methods were used to fill the missing values in the datasets, and a violin graph was created to compare the Mean and kNN methods based on the MSE results. According to the obtained MSE results, it was observed that the distribution and average of the kNN algorithm were better than that of NO_x. On the other hand, the NO_x parameter gave better results when filled with the mean method. Therefore, it was decided to use the mean method for NO_x and the kNN algorithm for filling the missing values in NO₂, PM₁₀, and CO datasets in the second stage.

After the data preprocessing, a total of 24,282 datasets were obtained for each pollutant parameter, and 80% (19,423 samples) of the train-split dataset were allocated for training, while the remaining 20% (4,859 samples) were reserved for testing. Upon examining the dataset, a strong relationship was observed between PM₁₀ and the pollutants NO₂, NO_x, and CO.

In the study, LSTM (Long Short Term Memory), one of the deep learning methods, was preferred for predictive analysis. Hyperparameter optimization was applied to increase the prediction performance of the LSTM used in the study.

Correct hyperparameter selection for each parameter of air quality is provided by hyperparameter optimization. Random search, bayesian search, and hyperband search algorithms were adapted from hyperparameter search algorithms. For hyperparameter

optimization for each pollutant, the RMSE, and MSE value was calculated. By examining the results obtained, hyperparameter selection was made. Hyperband for NO_x , Bayesian for NO_2 , PM_{10} , and hyperparameter for CO were found suitable.

The hyperparameter values determined by the best search algorithm were used for retraining. Estimates created after this stage were evaluated separately. The LSTM method used for forecasting is a preferred method in the forward forecasting of air quality.

As a result, it is seen that the model used is a method with high estimation accuracy. In this context, it shows that the LSTM Algorithm can be widely used to predict air quality data. In addition, in order to draw attention to air pollution, it is thought that the study is the right way, and it is thought that if the estimation steps are expanded and developed, it can produce more efficient results.





1. GİRİŞ

1.1. Tezin Amacı

Bu çalışma, yaşamsal fonksiyonların kaynağını oluşturan havadaki bileşenlerin değişimiyle ortaya çıkan hava kirliliğinin, olumsuz sonuçlarına farklı bir çözüm üretebilme amacı ile başlatılmıştır. Kirliliğe neden olan kirletici kaynakların tamamen ortadan kaldırılması mümkün değildir ancak derin ve ayrıntılı disiplinler arası çalışmalar ile durdurularak azaltılabileceği düşünülmüştür.

Bu çalışmada, yapay zeka teknikleri ile hava kirliliği modellenmesi planlanmıştır. Yapay zeka teknikleri, hava kirliliği analizi için kullanılabilir, elde edilen karmaşık verileri işleyerek tahminlerde bulunabilir. Ayrıca tahmin sonuçlarını değerlendirebilir ve buna bağlı olarak önlemler alınmasına da olanak sağlayabilir. Yapay zeka teknikleri ile yapılan tahminlerin, farklı istatistiksel ve fiziksel yöntemler ile yapılan tahminlere göre, daha gerçeğe yakın olacağı düşünülmüştür. Özetle, hava kirliliği modellemesinde yapay zeka kullanımı, hava kirliliğinin kontrolü için önemli bir adımdır.

Çalışmada kullanılması planlanan yöntem, yapay zeka alanının bir alt dalı olan LSTM (Long Short-Term Memory) isimli bir derin öğrenme algoritmasıdır. Bu algoritma, çok katmanlı yapılardan oluşan sinir ağı kullanılarak geliştirilen, bir makine öğrenmesi türüdür. LSTM tabanlı hava kirliliği tahmin modelinin oluşturulabilmesi için, Çevre Şehircilik ve İklim Değişikliği Bakanlığı Ulusal Hava Kalitesi İzleme Ağından temin edilen, Sakarya merkez istasyonuna ait veriler kullanılacaktır. Bu istasyonda ölçülen, azot dioksit (NO_2), azotoksitler (NO_x), partikül maddeler (PM_{10}) ve karbon monoksit (CO) konsantrasyonlarına ait bir veri seti oluşturularak, veri ön işleme ve veri temizleme çalışması yapılacaktır. Veri ön işleme sonrasında elde edilen veriler, LSTM algoritması kullanılarak bir model oluşturacaktır. Modelin performansı hiperparametre optimizasyonu ile değerlendirilecektir.

Kullanılan modelin, hava kirliliği tahmini için düşük maliyetli ve gerçeğe yakın olacağı ayrıca bir çok kurum ve kuruluş tarafından kullanılabileceği düşünülmektedir.

1.2. Kuramsal Temeller

Hava kirliliğinin kontrolü ve yönetimi için dünya genelinde hava kalitesi standartları oluşturulmuştur. Ayrıca kirlетici kaynakların kontrolü için çeşitli yasal düzenlemeler yapılarak, dünya genelinde birtakım sınırlandırmalar getirilmiş ve bazı kimyasal materyallerinin kullanımı yasaklanmıştır.

Hava kirliliğinin seviyesini halkın daha kolay anlaması için Hava Kalitesi İndeksi (HKİ) olarak tanımlanan bir sınıflandırma sistemi geliştirilmiştir. Mevcut hava kalitesi bu indeksler yardımı ile iyi, orta, kötü ve tehlikeli şeklinde derecelendirilmektedir. Ülkeler indeks hesaplamasında kullanılan yöntem ve kriterleri kendi hava kalitesi standartlarına uygun bir şekilde oluşturmaktadır [1].

Hava Kalitesi İndeksi 0-500 aralığındadır. HKİ değerleri yükseldikçe hava kirliliğinin arttığı düşünülmektedir. Her bir HKİ kategorisi bir renkle sembolize edilmiştir, bu şekilde toplumun hava kirliliğinin sağlıklı ya da sağlıklı olmadığı konusunda fikir sahibi olması amaçlanmıştır [2]. Tablo 1.1.'de EPA "Hava Kalitesi İndeksi kategorileri", sağlık için uygun olan seviyeleri, renkleri ve anlamları görülmektedir [2]. Hava kalitesi indeksinde hesaplanan 5 temel kirlетici bulunmaktadır. Bunlar; partikül maddeler (PM₁₀), karbon monoksit (CO), kükürt dioksit (SO₂), azot dioksit (NO₂) ve ozon (O₃)'dur.

UHKİ, Amerika Birleşik Devletleri Çevre Koruma Ajansı (USEPA) tarafından geliştirilen hava kalitesi indeksinin, Ulusal Mevzuat ve sınır değerlere uyarlanmış halidir. Ülkemizde, hava kalitesi sınır değerleri için, Avrupa Birliği limit değerleri hedeflenmekte ve bu hedefe kademeli olarak ulaşmak planlanmaktadır. Planlanan bu hedefe göre; 2024 tarihinden itibaren tüm kirlетici parametreler için AB mevzuatı uygulanacaktır [3]. Ülkemizde ve AB ülkelerinde İndeks Hesaplanan Parametrelerin Sınır Değerleri Tablo 1.2'de gösterilmiştir. Bu tabloda her kirlетici için istenilen sınır değerler, ülkemiz ve AB ülkeleri için hava kalitesinin mukayese edilmesine olanak sağlamaktadır.

Tablo 1.1. Ulusal HKİ kesme noktaları [2]

İndeks	HKİ	SO ₂ (µg/m ³)	NO ₂ (µg/m ³)	CO (µg/m ³)	O ₃ (µg/m ³)	PM ₁₀ (µg/m ³)	Renk Anlamı
		1 Sa. Ort.	1 Sa. Ort.	8 Sa. Ort.	8 Sa. Ort.	24 Sa. Ort.	
İyi	0-50	0-100	0-100	0-5500	0-120	0-50	Hava kalitesi memnun edici ve hava kirliliği az riskli veya hiç risk teşkil etmiyor.
Orta	51 - 100	101- 250	101- 200	5501- 10000	121- 160	51-100	Hava kalitesi uygun fakat alışılmadık şekilde hava kirliliğine hassas olan çok az sayıdaki insanlar için bazı kirleticiler açısından orta düzeyde sağlık endişesi oluşabilir.
Hassas	101- 150	251- 500	201- 500	10001- 16000	161- 180	101- 260	Hassas gruplar için sağlık etkileri oluşabilir. Genel olarak kamunun etkilenmesi olası değildir.
Sağlıksız	151- 200	501- 850	501- 1000	16001- 24000	181- 240	261- 400	Herkes sağlık etkileri yaşamaya başlayabilir, hassas gruplar için ciddi sağlık etkileri söz konusu olabilir.
Kötü	201- 300	851- 1100	1001- 2000	24001- 32000	241- 700	401- 520	Sağlık açısından acil durum oluşturabilir. Nüfusun tamamının etkilenme olasılığı yüksektir.
Tehlikeli	301- 500	>1101	>2001	>32001	>701	>521	Sağlık alarmı: Herkes daha ciddi sağlık etkileri ile karşılaşabilir.

Tablo 1.2. İndeks hesaplanan parametrelerin sınır değerleri [4]

Parametre	SO ₂ (µg/m ³) 1 Sa. Ort.	NO ₂ (µg/m ³) 1 Sa. Ort.	CO (µg/m ³) 8 Sa. Ort.	O ₃ (µg/m ³) 8 Sa. Ort.	PM ₁₀ (µg/m ³) 24 Sa. Ort.
Ulusal Sınır Değer	410	270	10.000	120	70
AB Üye Ülkelerin Sınır Değeri	350	200	10.000	120	50

Kirletici parametreler hakkında kamuoyu tarafından oluşan farkındalık sanayi devrimine kadar uzanmaktadır. Tarihsel olarak, her birinin zararlarının keşfi ve sebep olduğu sonuçların açığa çıkması ayrı ayrı değerlendirilir. Ancak 20. Yüzyılın başlarından itibaren bu konu ile ilgili ayrıntılı bilimsel çalışmalar yapılmaya başlandığı söylenebilir. Örneğin, karbonmonoksitin tarihsel olarak yaygın bir kirletici ve bir tehlike unsuru olarak değerlendirilmesi aşamalarını incelediğimizde, Recihle ve arkadaşları tarafından yapılan bir çalışma göze çarpmaktadır. Bu çalışmada 1981

yılında, araştırmacılar uydu üzerinden hava kirliliği parametrelerinin ölçümünü yapmıştır. Yapılan ölçüm sonrası, CO konsantrasyonunun yüksek seviyelere ulaştığı görülmüştür ve bu durumun dünya çapında bir tehlike teşkil ettiği savunulmuştur. [5]. Bir diğer parametrede O_3 'dür. O_3 renksiz ve kokusuz bir gaz olmakla birlikte oldukça karmaşık reaksiyonlardan kaynaklanır ve doğrudan havaya yayılmaz. Bu özellikleri onun diğer kirleticiler arasında benzersiz bir yapıya sahip olduğunu göstermektedir. O_3 yüksek reaktiviteye sahiptir ve atmosferin iki bölümünde (stratosfer ve troposfer) eser miktarda oluşmaktadır. Dünyayı, güneşin ultraviyole ışınlarından korur. Ancak, doğal olmayan antropojenik ve doğal biyogenik kaynaklardan oluşan O_3 kirliliği yer seviyesinde olduğunda dünyadaki canlı yaşamı için zararlı olabilmektedir [6]. Başka bir kirleticisi ise PM_{10} 'dur. Endüstriyel kaynaklı olabilir, motorlu taşıtlardan ve fosil yakıt kullanımından da kaynaklanabilir [7]. Havadaki partikül maddelerin, iklim değişikliklerine sebep olduğu, insan sağlığına yönelik olumsuz sonuçlarına ve ekosistemlerdeki biyojeokimyasal döngülere neden olduğu yapılan çalışmalar sonucunda ortaya çıkmıştır [8].

Yukarıda bahsedilen üç ana kirleticisi parametre ve diğerleri, çoğunlukla antropojenik kaynaklardan yayılır ve uzun mesafelere kadar taşınabilir. Bazı hava kirleticileri atmosferde son derece kararludur ve çevrede, besin zincirinde birikerek insanları, hayvanları ve doğal biyoçeşitliliği etkileyebilir [9]. Bu sebeple kaynakları net bir şekilde araştırılmalı ve sınırlandırılmalıdır.

Kirleticisi parametrelerin kaynakları farklılık göstermekle birlikte, kendi aralarında üç ana başlığa ayrılmaktadır. Bunlar ısınmadan, motorlu taşıtlardan ve sanayiden kaynaklı hava kirlilikleridir. Isınmadan kaynaklı hava kirliliği; düşük kalorili, kalitesiz kömür veya yakma ürünü kullanılmasından kaynaklanır. Motorlu taşıtlardan kaynaklanan hava kirliliği, nüfus artışı, gelir düzeyi artışı ile motorlu taşıt sayısının artmasından kaynaklanır ayrıca bunun en önemli bileşeni ise dizel araçlardır [10]. Dizel motorlar, ulaşım, demir yolu, denizcilik, madencilik, tarım ve çeşitli üretim alanları, endüstride kullanılan araçlar olmak üzere birçok alanda araçlara güç sağlamak için kullanılmaktadır. Dizel motorların egzozu 40'tan fazla toksik hava kirleticisi içermektedir. Saniyeden kaynaklanan hava kirliliği ise sanayi tesisinin yanlış seçilmesi, çevre koruma tedbirlerinin alınmaması, uygun teknolojik gelişmelerin takibinin yapılamamasından kaynaklanır.

Yukarıda saydığımız kirleticiler ve kaynakları göz önüne alınarak uygun hava kalitesi izleme istasyonları konumlandırılmakta bu şekilde veri toplanmaktadır. Ülkemizde, Çevre ve Şehircilik Bakanlığının kontrolü altında, 81 ilde belirlenen noktalarda kurulan istasyonlar aracılığıyla sürekli olarak ölçümler yapılmaktadır. Ulusal Hava Kalitesi İzleme İstasyonları (UHKİİ) olarak isimlendirilen sistem aracılığıyla ortamdaki kirletici miktarı ölçülebilmekte ve buna bağlı olarak hava kalitesi belirlenmektedir. Tüm ölçüm istasyonlarının hava kalitesi indeksleri hesaplanarak www.havaizleme.gov.tr web sitesinde yayımlanmaktadır [4]. Yayımlanan bu veriler araştırmacılar tarafından çeşitli çalışmalar için kullanılmaktadır.

Hava kalitesi verileri kullanılarak; hava kirliliğinin önlenmesi ve kontrolü için hava kirliliği modellemesi yapılması son yıllarda tercih edilen bir yöntemdir. Hava kalitesini önceden tahmin ederek, canlıların yaşamını tehdit eden etkenleri en aza indirmek için zamanında uyarı ve önlemler alınabilir [11]. Çevre bilimi ve çeşitli bilim dallarının bileşimi ile etkili çözümlere ulaşılmaktadır. Hava kirliliğinin geleceğe dönük tahmini için kullanılan birçok yöntem olmasına rağmen, çoğunlukla sığ modeller kullanılmakta ve bu modellerin ulaştırdığı sonuçlar yeterince gerçeğe yakın olmamaktadır [12]. Ayrıca nesnel cihaz ölçümleri, yanıltıcı olabilir. Sensör hatalarından kaynaklı oluşan değerlendirmeler, sonuçların analizini ve geleceğe yönelik tahminleri güçleştirmektedir. Sonuç olarak, hem olası sensör hatalarının oluşturduğu problemleri çözmek hem de geleceğe yönelik tahminleri kolaylaştırmak için farklı yöntemlere başvurmak gerekmektedir.

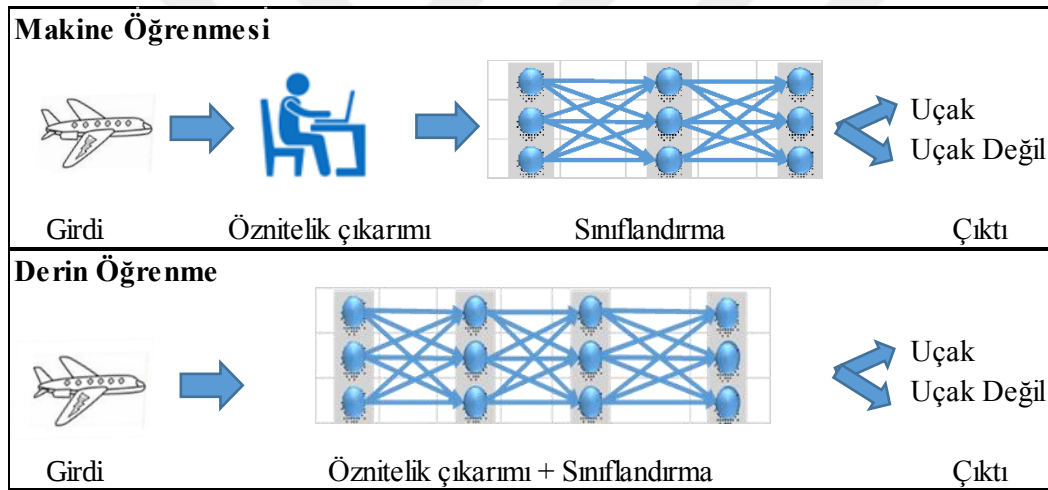
Hava kalitesi süreci doğası gereği karmaşık olduğundan, zamansal eğilimleri ve mekânsal dağılımı hava kirletici emisyonları ve birikimi, hava koşulları, trafik akışı, insan faaliyetleri vb. gibi çeşitli faktörlerden etkilenmektedir. Bu sebeple tahminlerde kullanılacak yöntemlerin bu unsurları da barındırması gerekmektedir. Hava kalitesinin özelliklerinin uygun bir şekilde değerlendirilebilmesi için geleneksel modellerin kullanılması zorluk teşkil etmektedir.

Derin öğrenme algoritmaları, veri katmanının doğal özelliklerini en alttan en üst düzeye çıkarmak için çok katmanlı mimariler kullanır ve verilerdeki temsili yapıyı tanımlayabilir. Bu özelliğiyle günümüz koşullarında hava kalitesi verilerinin tahmini için kullanılacak, uygun ve gerçeğe yakın bir yöntem olarak düşünülebilir. Derin öğrenme algoritmaları, önceden bilgi sahibi olmadan temsili hava kalitesi özelliklerini

çıkarabilir ve hava kalitesi tahminleri için iyi bir performans sağlayabilir. Derin öğrenme yöntemi, makine öğrenmesi yöntemi olup bir çok alanda diğer tahmin modellerinden ve geleneksel veri analizi yöntemlerinden daha iyi sonuçlar ortaya çıkarmaktadır [13].

Derin öğrenme, sinir ağları algoritmaları içerisinde diğerlerine göre daha yeni ve işlevsel bir tür olması ile birlikte makine öğrenmesinin alt dallarından biridir. Bir ya da daha fazla girdisi vardır. Çok fazla sayıda katman barındırmaktadır ve son çıktısı bir ya da daha fazla olan bir algoritmadır. Her katmanda kendisinden önceki verileri birleştirerek farklı türden ve anlamlı sonuçları olan değerler ortaya çıkarmaktadır.

Derin öğrenme Şekil 1.1.'de görüldüğü üzere makine öğrenmesinden farklı olarak kaynaklarındaki verilerin özelliklerini göz önüne alarak gizli katmanlarında öğrenmektedir.



Şekil 1.1. Makine öğrenmesi algoritması ve derin öğrenme algoritması kıyaslanması [14]

Hava kirliliğini azaltabilmek için doğru hava kirliliği tahmin metotları kullanmak gerekmektedir. Tahmin için kullanılan istatistiksel yöntemler oldukça belirsiz olmakla birlikte, eksik ve sınırlı bir yaklaşım oluşturmaktadır. Son yıllarda tercih edilen derin öğrenme yöntemi, büyük miktarda veri ve büyük nöron ağları kullanarak karmaşık bilgileri çözüp yönetebilmek için ideal bir yaklaşım olmuştur. Hava kalitesi modellemesi ve tahmini çalışmalarında, popüler hale gelmiştir.

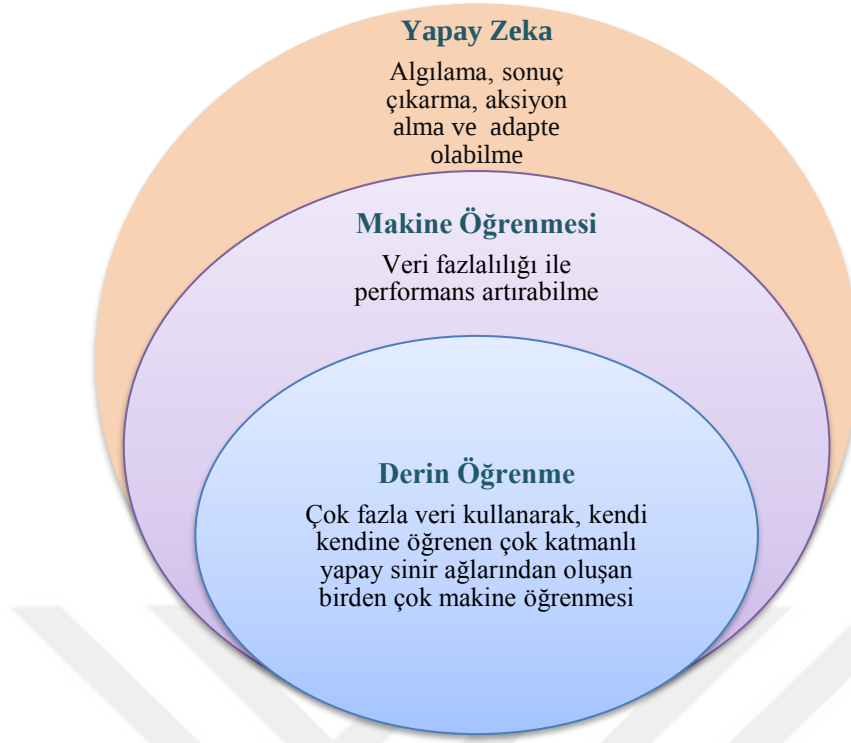
Derin öğrenme yapay sinir ağları yaklaşımlarının arasında ileri düzey bir yöntemdir. Tek katmanlı yapay sinir ağları gelişerek, çok katmanlı sinir ağlarına ve daha da gelişerek derin öğrenme algoritmalarına doğru bir yol kat etmiştir.

Yapay sinir ağlarının gelişiminin temeli, insan beynindeki hücrelerin matematiksel işleyişinin keşfi ile başlamış ve bilgisayarın bu matematiksel modeller kullanılarak, veri sınıflandırılması, tanı, teşhis vb. gibi birçok alanda kullanılmasına örnek teşkil etmiştir [15].

Derin öğrenme sayesinde, verilerin aynı türden ve birimden tek bir özelliğini öğrenmekten ziyade, alt düzeyde verilerin birçok özelliğinin bileşiminden oluşan hiyerarşi ile üst düzey verilerin hiyerarşisini öğrenebilmek mümkün olmaktadır. Üst düzey veri setinin özellikleri, alt düzey veri setinin özelliklerinden türetilmektedir. Bu şekilde türetilen üst düzey veriler birden çok veri özelliğini temsil etmektedir. Bu bağlamda, veri miktarı arttıkça ve makine öğrenimi yöntemleri geliştikçe, derin öğrenme yöntemlerinin güçlü öğrenme yeteneklerinin, gittikçe daha da önemli hale geleceği düşünülmektedir [16].

Derin öğrenme, makine öğrenmesi ve yapay zekâ arasındaki ilişkiyi anlatmak gerekirse; birbirlerini içeren alt küme şeklinde değerlendirmek mümkün olabilmektedir. Bu kümelenme modeli Şekil 1.2.'de gösterilerek detaylandırılmıştır.

Yapay zekâ, makinelerin insan zekâsını taklit ederek oluşturduğu zekâ türü, makine öğrenmesi yapay zekâyâ ait bir araç, derin öğrenme de makine öğrenimin bir yöntemidir [17].



Şekil 1.2. Makine öğrenmesi, derin öğrenme ve yapa zekâ algoritmaları kümelenmesi [18]

1.3. Tezin Önemi

Sakarya’da hava kalitesinin geleceğe yönelik tahminini içeren bu çalışma, hava kirliliği seviyesinin anlaşılmasına, yönetilmesine ve iyileştirilmesine yardımcı olacaktır.

Halihazırda, Sakarya Büyükşehir Belediyesi tarafından hava kalitesi tahmini yapmak için birtakım çalışmalar yapılmaktadır. Hava kalitesi ölçümlerinin yapıldığı istasyonlardan alınan veriler ile hava kirliliği kaynaklarının tespiti ve izlenmesi, hava kirliliği verilerinin toplanması ve analiz edilmesi gibi faaliyetler gerçekleştirilmektedir. Ancak hava kalitesi modellemesi için yeni ve güncel yöntemlere ihtiyaç duyulmaktadır. Bu açıdan çalışmanın yeni bir bakış açısı getirebileceği düşünülmektedir.

Ayrıca çalışmada kullanılan LSTM yönteminin etkili ve doğru bir tercih olduğu düşünülmektedir. Bu çalışma, Sakarya’da hava kirliliği tahmini yapmak için LSTM algoritmasının kullanılabileceğini göstermektedir. LSTM algoritmasının yanı sıra, diğer makine öğrenimi algoritmaları kullanılarak daha fazla araştırma yapılması, hava kirliliği tahminlerinin doğruluğunu artırmaya yardımcı olabilir.

2. LİTERATÜR ÖZETİ

Hava kirliliği modellemesi, matematiksel bir modelleme türüdür. Bu modelleme türü, hava kirliliğine sebep olan kaynakları ve meteorolojik unsurları birlikte değerlendirerek hava kirliliğini ileriye dönük tahmin etmeyi amaçlar. Günümüzde hava kalitesi tahmini için istatistiksel modeller, yapay sinir ağları ve makine öğrenimi algoritmaları gibi çeşitli teknikler kullanılmaktadır. İstatistiksel modeller, güncel hava kalitesi verilerini analiz ederek gelecekteki hava kalitesini tahmin eder. Yapay sinir ağları, hava kalitesi tahmini için öğrenme ve özyinelemeli işlemler kullanır. Farklı parametrelere ait karmaşık verileri işler. Makine öğrenimi algoritmaları ise hava kalitesi tahmini için daha büyük miktarda veriyi işleyebilir. Bu bölümde başlıklara ayrılarak bu çalışmalar hakkında bilgiler sunulacaktır.

2.1. Hava Kirliliğinin Tahmininde İstatistiksel Yöntemlerin Kullanımı

Hava Kirliliğinin tahmininde istatistiksel yöntemlerin kullanımı oldukça yaygındır. Örneğin, Lei ve arkadaşlarının, 2019 yılında yapmış oldukları bir çalışmada; Çin, Büyük Körfez Bölgesi'nde yer alan kentlerdeki hava kirliliği modellemesi ve tahmini için istatistiksel bir yöntem önerilmektedir. Bu bölgedeki hava kalitesi seviyesinin çoğu zaman Dünya Sağlık Örgütü tarafından önerilen seviyelerin üstüne çıktığı görülmüş, önlem alması için bir hava kalitesi tahmini geliştirilmesi amaçlanmıştır. Çalışmada, kirlletici parametrelerden PM_{10} ve $PM_{2.5}$ konsantrasyonlarının bir sonraki güne ilişkin tahmini üzerinde durulmaktadır. Tahmin için çoklu regresyon analizine dayalı istatistiksel modeller kullanılmıştır, %95 güven aralığına sahip istatistiksel olarak anlamlı bir tahmin geliştirilmiştir [19].

2020 yılında Jeong ve arkadaşlarının yaptıkları bir çalışmada ise, Güney Kore'deki kış dönemi (Aralık-Şubat) boyunca partikül madde (PM_{10}) konsantrasyonlarındaki mevsimsel değişimleri tahmin etmek için istatistiksel-dinamik bir model önerilmektedir. Doğu Asya'daki hava kirliliği emisyonlarının artışı ve buna bağlı olarak artan endişelerden dolayı partikül madde konsantrasyonlarının doğru şekilde tahmininin, hava kirliliğinin önlenmesi için faydalı olacağı düşünülmüştür. Çalışmada, son üç yılda ölçümü yapılan PM_{10} konsantrasyonları ve meteorolojik değişkenler

kullanılarak, Güney Kore'deki mevsimsel PM₁₀ konsantrasyonlarının tahmini yapılmıştır. Bu entegre modelin yüksek tahmin yeteneğine ulaştığı ancak geliştirilebilir olduğu vurgulanmıştır [20].

Marinov ve arkadaşları da hava kirliliği tahmini için istatistiksel yöntemleri kullanmışlardır. 2022 yılında yaptıkları bir çalışmada, Sofya'daki hava kalitesinin ileriye dönük tahmini için zaman serisi analizi yöntemlerine başvurmuş, 5 adet istasyondan 2015-2019 yılları arası temin etmiş oldukları saatlik CO, NO₂, O₃ ve PM_{2.5} verilerini toplamışlardır. Eksik verileri ön işleme yöntemleri ile düzenlemiş, AutoRegressive Integrated Moving Average (ARIMA) yöntemi ile analiz etmişlerdir. Bu çalışmada Dünya Sağlık Örgütü'nün kirletici parametre konsantrasyonları için belirlediği sınırların aşılp aşılmadığı gözlemlenmiş ve farklı istasyonlardan alınan veriler kıyaslanmıştır. Çalışma, hava kirliliği modellemesi ve tahmini için zaman serisi yöntemlerinin etkili sonuç verdiğini göstermektedir [21].

Doğu Çin'de yapılan bir başka çalışmada Zhao ve arkadaşları, Doğu Çin'in hava kalitesi yönünden düşük seviyeli olan bir bölgesindeki 34 adet şehrin hava kalitesi verilerini kullanarak, hava kalitesi modellemesi ve tahmini için mekânsal-zamansal bir yöntem ile istatistiksel bir öğrenme yöntemi önermektedir. Korelasyon analizi ve zaman serisi analizi ile mekânsal ve zamansal özellikler kullanılmıştır. Kullanılan modelin deney sonuçları ile başka bir takım temel algoritmalar ile karşılaştırılmış ve bu kıyaslama sonucunda kullanılan modelin daha iyi sonuç verdiği gözlemlenmiştir. Çalışma sonucunda hava kalitesi tahmininin ekolojik çevre sorunlarına bir çözüm getirebileceği düşünülmüştür [22].

Çalışmaların tümünde varılan ortak kanı, hava kirliliğinin ileriye dönük tahmininin, oluşabilecek çevre sorunlarına önleme niteliğinde olduğudur. Kullanılan istatistiksel yöntemlerin hava kalitesi tahmininde iyi bir performans gösterdiği görülmüştür. Ancak bu çerçevenin geliştirilmesi mümkündür, buna bağlı olarak farklı yöntemler ile hava kalitesi tahmini yapılması yönünde çalışmalar devam etmektedir.

2.2. Hava Kirliliğinin Tahmininde Yapay Zeka (AI) Yöntemlerinin Kullanımı

Hava kalitesinin yapay zekâ ile tahminine dayalı birçok çalışma yapılmıştır. Bu bölümde bu çalışmalardan bazılarına yer verilecektir. Örneğin, 2016 yılında Oprea ve arkadaşları yapay zekâ ile hava kalitesi tahmini için bir çalışma yapmıştır. Bu çalışmada yapay sinir ağı (YSA) modellerinin, PM_{2.5} tahmininde kullanımı

incelenmiştir. $PM_{2,5}$ tahmini için kullanılan veri seti Romanya'da bulunan Ploiesti kentinde yer alan bir hava kalitesi izleme istasyonundan temin edilmiş olup sıcaklık ve $PM_{2,5}$ konsantrasyonlarını içermektedir. Entegre, YSA modelleri kullanılarak kıyaslamalarda bulunulmuş, kıyaslaması yapılan her iki YSA modelinin de iyi bir performansa sahip olduğu gözlemlenmiştir [23].

Montalvo ve arkadaşları, Lima şehrindeki hava kirliliği tahmini üzerine farklı bir bakış açısı getirmektedir. Hava kalitesi izleme istasyonlarının kurulumu, tamiri ve bakımı oldukça maliyetli olduğu için bazı bölgelerde hava kalitesi izleme aşamaları zorluk teşkil edebilmektedir. Bu çalışmada; halkın, şu an soluduğu havanın kalitesini ve yakın gelecekteki (6 saat sonra) durumunu izleyebilmesine olanak sağlayan bir web sitesi tasarımı üzerinde durulmaktadır. Bu web sitesindeki veriler, düşük maliyetli IoT (nesnelerin interneti) istasyonlarından temin edilen $PM_{2,5}$ konsantrasyon verilerini içermektedir. Elde edilen veriler AI modelleri ile analiz edilmiş, bölgesel kirlilik haritaları çıkarılmıştır. Ayrıca yakın geleceğe yönelik tahminlerde bulunulmuştur. Bu çalışma referans ağlarının mümkün olmadığı bölgeler için bir çözüm sunmakta ve AI modellerinin tahmin yeteneğinin tatmin edici olduğunu göstermektedir [24].

Atina'da ise saatlik benzen konsantrasyonunun tahmini için AI yöntemlerine başvurulmuştur. Karatzas ve arkadaşları bu çalışmada, hava kirliliğinin önlenmesi için hava kirliliği tahmininin önemini vurgulamış, bu amaçla yapay zekâ yöntemlerinin kullanımının doğruluğunu araştırmışlardır. Çalışmada öncelikle benzenin temel kaynağının trafik olduğu bulunmuştur. Bu bulgu ile benzenin azaltılmasının trafikteki araçlara yapılacak müdahale ile mümkün olduğu belirtilmiştir. Daha sonra oluşturan iki ayrı YSA modeliyle ileriye dönük tahmin yapılmış ve sonuçlar kıyaslanmıştır. Sonuç olarak iki YSA modelinin de başarılı olduğu ancak kullanılan Model 1'in daha güçlü bir performans sergilediği görülmüştür [25].

2015 yılında, Agra'da, hava kirliliği parametrelerinden biri olan NO_2 için bir AI'ya bağlı tahmin modeli oluşturulmuştur. Son on yıllık hava kirliliği verileri toplanarak bir analiz yapılmış, Kasım ayına ait verilerin tüm senelerde yüksek olduğu görülmüştür. Mishra ve Goyal tarafından yapılan bu çalışmada öncelikle çoklu doğrusal regresyon yöntemi kullanılmıştır. Daha sonra farklı parametreler ve meteorolojik veriler kullanılarak bir yapay sinir ağı (ANN) modeli geliştirilmiştir. Sonuç olarak geliştirilen bu modelin, Agra'da hava kalitesi parametrelerinden NO_2 için güvenilir olduğuna karar verilmiştir [26].

Verilen bu örnek çalışmalarda istatistiksel yöntemler gibi kullanılan YSA modellerinin entegre edilerek veya edilmeyerek kullanılmasıyla iyi sonuçlar verdiği gözlemlenmiştir.

2.3. Hava Kirliliğinin Tahmininde Derin Öğrenme Yöntemlerinin Kullanımı

Hava kalitesinin derin öğrenme ile tahmini, yapay zekâ ve makine öğrenmesi tekniklerinde son zamanlarda tercih edilen bir yöntemdir. Karmaşık hava kirliliği verilerini kullanarak, çok katmanlı sinir ağlarıyla, kirlletici konsantrasyonunu tahmin için yapılan çalışmalarda derin öğrenme yöntemleri kullanılabilir. Bu konu ile ilgili yapılan çalışmalardan bu bölümde bahsedilecektir.

Farklı derin öğrenme yöntemleri ile karşılaştırmalı olarak yapılan çalışmalara, literatürde oldukça sık rastlanmaktadır. Murad ve arkadaşları, bayes ağları, derin ensemble, Monte Carlo (MC) dropout ve stokastik ağırlık ortalaması-Gaussian (SWAG) gibi farklı derin öğrenme yöntemlerini kullanarak Norveç'in havanın kalitesini ölçmeyi amaçlamıştır. Çalışmada bu yöntemler ile ilgili deneysel değerlendirmelerde bulunulmuştur. Sonuç olarak, önerilen yöntemlerin hava kalitesi tahmininde daha önce yapılan çalışmalardan daha güçlü bir performans gösterdiğini görülmüştür. Kullanılan yöntemler arasında en güvenilir sonuçları veren Bayes sinir ağları olarak tespit edilmiş, ancak pratikte uygulama zorlukları olduğu görülmüştür. Diğer yöntemlerin ise daha az güvenilir olduğu, ancak kullanımının daha kolay olduğu görülmektedir [27].

Derin öğrenme yöntemleri ile makine öğrenmesi yöntemlerinin kıyaslanması için yapılan çalışmalar da mevcuttur. Lin ve arkadaşları 2021 yılında yapmış oldukları bir çalışmada, hava kalitesi tahmininde derin öğrenmenin geleneksel makine öğrenmesine göre daha iyi tahmin performansı gösterdiğini doğrulamıştır. Bu çalışmada, hava kalitesi tahmininin doğruluğunu artırmak için öncelikle farklı tahmin modelleri (GRU13d, GRUAW13d, GRUAW14d, GRUSS13d ve GRUST14d gibi) tasarlamak için bir GRU (Gated Recurrent Unit) derin öğrenme ağı mimarisi kullanılarak entegre bir sistem tasarlanmıştır. 2013-2019 yılları arası, Tayvan'da yer alan 67 adet hava kalitesi izleme istasyonundan temin edilen veriler kullanılarak analizler yapılmıştır. Çalışmanın sonucunda önerilen entegre derin öğrenme yönteminin diğer yöntemlerden daha verimli olduğu görülmüştür [28].

Yapılan çalışmalarda farklı derin öğrenme yöntemleri kullanılarak hava kalitesi tahmini yüksek performans göstermiştir.

2.4. Hava Kirliliğinin Tahmininde LSTM Algoritmasının Kullanımı

LSTM yöntemi literatürde özellikle zaman serisi verilerinin tahmini için kullanılır. Hava kirlleticilerinin konsantrasyonu içeren verileri kullanarak ileriye dönük tahminlerde bulunabilir. Bu yöntem ile yapılan birçok çalışma mevcuttur ve günümüzde de yaygınlaşmaktadır.

Örneğin, Alghieth ve arkadaşları Suudi Arabistan'daki partikül madde konsantrasyonunun tahmini için LSTM algoritması kullanmıştır. LSTM algoritması söz konusu ülkede, yüksek toz kirliliğinin önüne geçmek için önlemler alınması ve ileriye partikül madde konsantrasyonunun tahmininin yapılabilmesi için ideal bir yöntem olarak düşünülmüştür. Bu çalışmada LSTM modeli eğitilmiş, test verileri kullanılarak modelin doğruluğu ölçülmüştür. Veriler, modelin performansını değerlendirmek için kullanılmıştır. Sonuçlar, hata metrikleri kullanılarak değerlendirilmiştir. Çalışma sonucunda, LSTM tekniğinin uygulanmasıyla düşük hata oranları elde edilmiştir. MAE hata oranı 0,98, RMSE hata oranı 8,68, R-Kare ise 0,999'dur. Modelin mükemmel yakın doğruluk elde ettiği görülmüştür [29].

LSTM yöntemi çalışmalarda, bazen başlı başına bir yöntem olarak kullanılmakta, bazen de iki veya daha fazla yöntemle hibrit bir model elde edilmektedir. Du ve arkadaşları tarafından yazılan bir makalede, hava kalitesi tahmini için hibrid bir derin öğrenme yöntemi önerilmektedir. Çalışmada, DAQFF olarak adlandırılan, konvolüsyonel sinir ağları ve LSTM yöntemini içeren bir model kullanılmıştır. İki ayrı veri seti kullanarak kapsamlı deneysel değerlendirmeler yapılmış, PM_{2,5} konsantrasyonunun tahmini üzerinde durulmuştur. Sonuç olarak önerilen modelin hem sığ istatistiksel modellerden hem de derin öğrenme modellerinden daha iyi performans gösterdiği keşfedilmiştir. Gelecekteki hava kirliliğinin ani değişimi için uygulanabilir bir yöntem olarak düşünülmüştür [30].

Bazı çalışmalarda, hava kirliliği verileri kullanılarak hava kalitesi tahmininin yanında kirliliğe neden olan faktörlerde tahmin edilmiştir. Awan ve arkadaşları, trafik tahmininde hava kirliliği verilerinin kullanılmasını önermiştir. Bunun için İspanya Madrid'deki açık veri portalından temin edilen; CO, NO, NO₂, NO_x ve O₃ gibi kirleticiler parametreler, basınç, sıcaklık, rüzgâr yönü ve hızı gibi meteorolojik parametreler

kullanılmıştır. Çalışma da bu veriler kullanılarak LSTM modeli eğitilmiş, trafik tahmininde bulunulmuştur. Sonuç olarak kirletici ve meteorolojik parametreler kullanılarak trafik tahmininin doğruluğunun arttığı görülmüştür. İspanya'nın Madrid şehri için yapılan çalışmanın genişletilebileceği yönünde kanaat getirilmiştir [31].

Mani ve Volety; sensör birleştirme yöntemiyle, hava kirliliği seviyelerinin tahmini için LSTM ve ARIMA yöntemlerini karşılaştırmışlardır. Düşük maliyetli IoT kullanarak temin edilen hava kirliliği verileri ve istasyon verileri sensör birleştirme yöntemi ile birleştirilmiştir. LSTM ve ARIMA yöntemleri ile gelecekteki değerleri tahmin edilmiştir. Bu çalışmada iki ayrı yöntemde sensör birleştirme ile daha doğru sonuçlar verdiğini ortaya çıkarmıştır [32].

LSTM, RNN temelli bir yaklaşımdır. İkisinin birlikte kullanıldığı çalışmalara literatürde rastlanmaktadır. Tsai ve arkadaşları, Tayvan Çevre Koruma İdaresi'nden alınan, 2012-2016 yılları arasında ölçümü yapılan 20 boyutlu verileri kullanarak, PM_{2.5} konsantrasyonunu önümüzdeki dört saatte tahmin etmeyi amaçlamışlardır. Bunun için RNN (Tekrarlayan Sinir Ağı) ile LSTM kullanmışlardır. Python'da yazılan Keras kullanılarak bir sinir ağı oluşturulmuştur. Tensorflow üzerinden RNN ile LSTM çalıştırılmıştır. Sonuçlar, önerilen yaklaşımın PM_{2.5} değerini etkili bir şekilde tahmin edebildiğini göstermektedir [33].

Bu çalışmada ise, Çevre, Şehircilik ve İklim Değişikliği Bakanlığı Ulusal Hava Kalitesi İzleme Ağından elde edilen, Sakarya ili Hava Kalitesi Merkez İstasyonu verilerini kullanarak derin öğrenme yöntemleri ile ileriye dönük, azot dioksit (NO₂), azotoksitler (NO_x), partikül maddeler (PM₁₀) ve karbon monoksit (CO) konsantrasyonlarının tahmin edilmesi amaçlanmıştır.

3. MATERYAL VE YÖNTEM

Çalışma kapsamında kullanılacak hava kalitesi verileri, Çevre ve Şehircilik Bakanlığı'nın hava kalitesi izleme sisteminden temin edilmiştir [34]. Çalışma için Sakarya Merkez, ilçesinde yer alan hava kalitesi izleme istasyonundan alınan veriler kullanılmış olup derin öğrenme algoritmalarıyla tahmin performansı yüksek bir model oluşturulması planlanmıştır. Bu çalışmada hava kirliliği parametrelerinin tahmininin öneminin artması nedeniyle sinir ağı kullanılarak doğru ve etkili bir tahmin modeli geliştirmek amaçlanmıştır. Derin öğrenme yaklaşımına dayalı önerilen modelin, sığ makine öğrenmesi modellerine ve istatistiksel modellere göre daha iyi sonuç ve performans ortaya çıkardığını göstermek amaçlamaktadır. Son olarak, derin öğrenme tekniklerini kullanarak hava kalitesi tahminlerinin iyileştirilmesine yönelik tedbirler üzerinde durulması planlanmıştır.

Çalışmada, Çevre ve Şehircilik Bakanlığının, Sakarya'da bulunan istasyonlarından biri olan, Sakarya-Merkez istasyonundan ölçülüp www.havaizleme.gov.tr adresine aktarılan hava kalitesi verileri kullanılmıştır. Bu istasyon Sakarya ili merkezinde yer alacak bir konumdadır ve trafikten kaynaklı kirliliği ölçmektedir.

Bu çalışma bilgisayar üzerinden belirlenen bir takım yazılım araçları kullanılarak uygulanmıştır. Yazılım araçları, hava kalitesi veri setleri ve kullanılan yöntemler hakkında detaylı olarak bu bölümde başlıklar halinde açıklanacaktır.

3.1. Yazılım Araçları

Çalışmada kullanılan uygulamalar, yazılım kütüphaneleri LSTM modeli için gerekli materyaller bu kısımda detaylandırılmıştır. Bu materyaller, kullanım kolaylığı ve çalışma için uygun olduğundan dolayı tercih edilmiştir.

3.1.1. Python

Python; 1991 yılında Guido van Rossum tarafından geliştirilmiştir. Python son yıllarda çalışmalarda oldukça aktif olarak kullanılan bir yazılımdır. Python, uzun yıllardır kullanılan Java ve C'yi geride bırakarak, 2021 yılından itibaren bu yazılımlardan daha

popüler hale gelmiştir. Sahip olduğu geniş kütüphanesi onun tercih edilmesinin önemli bir nedenidir. 380.000’de fazla kütüphaneye sahiptir [35].

Python yukarıda bahsi geçen özelliklerinden dolayı birçok alanda kullanılmaktadır. 2000’li yıllardan bu yana, mühendislik, finans, bilimsel araştırmalar, veri analizi ile ilgili çalışmalar vb. gibi alanlarda kullanılmaya başlanmıştır [36].

Bu çalışmada hava kalitesi verileri üzerinde değerlendirmeler yapılabilmesi için kullanılmıştır. Sakarya Merkez İstasyonundan alınan verilerin işlenmesi ve çalıştırması amacıyla bu programlama diline yer verilmiştir

3.1.2. Kullanılan kütüphaneler

3.1.2.1. Matplotlib

Dağınık olan verileri düzenleyerek kolay anlaşılabilir ve yorumlanabilir hale getirmek amacıyla çalışmada Matplotlib kütüphanesinden yararlanılmıştır. Sonuçları en iyi şekilde analiz edebilmek amacıyla esnek bir raporlama imkânı sunan bu kütüphane ile çalışmada yer alan veriler çeşitli grafikler halinde görselleştirilerek sunulmuştur.

3.1.2.2. Keras ve tensorflow

Tez kapsamında gerçekleştirilen derin öğrenme çalışmalarında ise Keras ve Tensorflow kütüphanelerinden yararlanılmıştır. Günümüzde veri bilimi çalışmalarından oldukça yaygın bir kullanıma sahip bu kütüphaneler ile derin öğrenme metotları ile çalışma daha anlaşılır ve kolay hale gelmektedir. Açık kaynaklı birer kütüphane olan Keras ve Tensorflow, tez çalışmasında verilerin eğitilmesi için kurulan derin sinir ağlarında kullanılmıştır.

3.1.2.3. Pandas

Bilindiği üzere makine öğrenmesi veya derin öğrenme yöntemlerini kullanırken analizlere başlamadan önce veriyi analize uygun hale getirmek oldukça önemlidir. Bu amaçla kullanılan açık kaynaklı kütüphanelerden en yaygın kullanıma sahip olan hiç şüphesiz Pandas kütüphanesidir. Derin öğrenme algoritmaları ile ağı eğitime başlamadan önce gerekli olan veri ön işleme ve veri temizleme işlemleri oldukça fazla zaman almaktadır. Pandas kütüphanesi ile veriler organize edilerek analizlerin daha kolay ve hızlı yapılabilmesi mümkün olmaktadır. Bu özelliği ile Pandas, Python’ın en önemli kütüphanelerinden biridir.

3.1.2.4. Numpy

Bilimsel hesaplamalar yaparken kullanılan bir matematik kütüphanesidir ve temelini numpy dizleri oluşturur. Numpy dizeleri Python listeleri gibidir ancak daha kullanışlıdır. Numpy dizileri homojen yapıda ve aynı türde veri tiplerinden oluşmaktadır.

3.1.2.5. Keras-tuner

Hiperparametre aramanın sorunlu bölümlerini çözeabilen, kolay kullanımlı, ölçeklenebilir bir hiperparametre optimizasyon çerçevesidir. Arama alanını çalışmaya göre tanımlı bir sözdizimi ile yapılandırır, ardından modele en uygun olan hiperparametre değerlerini bulmak için mevcut arama algoritmalarından birini kullanır. KerasTuner, Bayes Optimizasyonu, Hiperbant ve Rastgele Arama algoritmalarıyla birlikte kullanılır.

3.2. Veri Setleri

Son yıllarda meydana gelen, teknolojik ve bilgi sistemleri gelişimlerin etkisiyle, kamu kurum, kuruluşlarının ve çeşitli işletmelerin veri tabanında çok sayıda veri toplanmaktadır. Bu veriler işlenmediği sürece, anlamsız veri deposu yığınının ibarettir [37].

Yazılım alanındaki gelişimler toplanan bu verilerin gereksiz olarak kalmasının önüne geçmiş, verilerin analizine ve değerlendirilmesine olanak sağlamıştır. Günümüzde birçok kurum ve kuruluş topladıkları verileri yararlı olacak şekilde işleyip değerlendirmektedir. Veri toplamının ve sadece geçmişe yönelik bilgi toplamının sağlamış olduğu yararlar dikkat çekmekte ve veri madenciliğinin önemini arttırmaktadır [38].

Çalışmada kullanılan veriler hava kalitesi ölçümü için kullanılan, Sakarya-Merkez İstasyonundan temin edilmiştir. 2020 Ocak ve 2022 Eylül arasındaki saatlik hava kalitesi verilerini içermektedir. Veri kütüphanesi olarak kullanılan istasyon trafikten kaynaklanan kirletici miktarının ölçümünde görevli olduğu için burada dört adet kirletici parametre ölçümü yapılmaktadır. Çevre ve Şehircilik Bakanlığının resmi internet sayfasından saatlik ya da günlük olarak çekilebilmektedir. Araştırmacıların kullanımına açık olarak sunulmakta <https://sim.csb.gov.tr/> uzantısından temin edilebilmektedir [34].

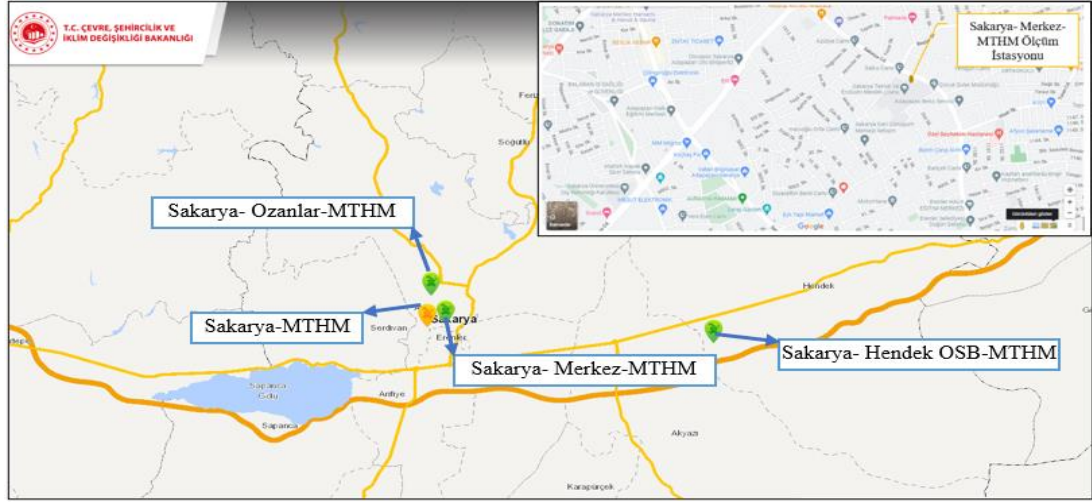
Hava kalitesi ölçüm istasyonları dört kategoriye ayrılmakta ve kentsel, trafik, endüstri ve kırsal olarak sınıflandırılmaktadır. Her biri farklı parametreler üzerinde çalışmaktadır. İstasyonlarda, kükürt dioksit, partikül madde (PM₁₀ ve PM_{2.5}), azot oksitler, ozon, karbon monoksit ve meteorolojik parametrelerin ölçümü yapılmaktadır. İlimizde hava kalitesi ölçümlerinin yapılabilmesi için, Ulusal Hava Kalitesi İzleme Ağı (UHKİ) kapsamında kurulan ölçüm istasyonları farklı yıllarda faaliyete geçmiştir. Sakarya istasyonu 2007 yılında, Sakarya Ozanlar istasyonu ve Sakarya Merkez istasyonu 2013 yıllarında, Sakarya Hendek istasyonu ise 2017 yılında ölçüm yapmaya başlamıştır. Her bir istasyon ve ölçüm yaptığı parametreler ve istasyon koordinatları Tablo 3.1.'de gösterilmiştir.

Tablo 3.1. Sakarya Hava Kalitesi İstasyon konumları ve ölçülen parametreler

İstasyon İsimleri	Koordinatlar (Enlem, Boylam)	Hava Kirleticileri					
		PM ₁₀ µg /m ³	PM _{2.5} µg /m ³	SO ₂ µg /m ³	NO _x µg /m ³	CO µg /m ³	O ₃ µg /m ³
Meteoroloji Md.	40.766976, 30.393851	✓	✓	✓	✓	✗	✓
Ozanlar Mah.	40.790711, 30.396915	✓	✓	✓	✓	✗	✓
Sakarya Caddesi	40.769188, 30.410228	✓	✗	✗	✓	✓	✗
Hendek OSB	40.754949, 30.650214	✗	✓	✓	✓	✗	✓

Çalışmamızda kullandığımız istasyon Sakarya Merkez istasyonudur. Tablo 3.1.'de Sakarya Caddesi olarak isimlendirilmiş, konumu Şekil 3.1.'de gösterilmiştir. Sakarya Caddesi Endüstri Meslek Lisesinin yanında, en yakın meskene uzaklığı 3.5m, yola 0.5m uzaklıkta konumlanmaktadır.

Bu istasyon, trafikten kaynaklanan hava kirliliğini ölçmek amaçlı kurulmuştur. Mart 2013 tarihi itibarıyla sürekli ölçüm yapmaya başlamıştır ve ölçüm sonuçları anlık olarak www.havaizleme.gov.tr internet sitesinden paylaşılmaktadır. Sakarya'da hava kalitesi ölçümü için kullanılan diğer istasyonların konumları Şekil 3.1.'de gösterilmiştir.

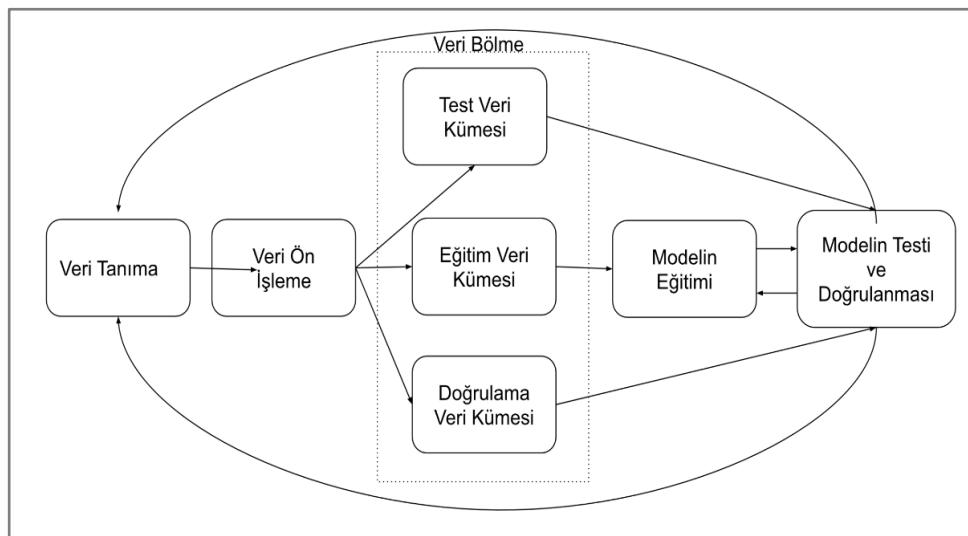


Şekil 3.1. Sakarya merkez ve diğer istasyonların harita üzerinde konumu

3.3. Veri Ön İşleme

Veri işleme, veriler elde edildikten sonra çalışma için uygun bir formata dönüştürme işlemlerini kapsamaktadır. Veri ön işleme, makine öğrenimi algoritmasının genelleme yapabilmesi amacıyla gerekli olan işlemleri içeren bir adımdır [39]. Veri ön işleme, genellikle eksik veri atamasını, olağandışı gözlemleri kaldırmayı veya değiştirmeyi, veri dönüştürme ve öznitelik çıkarma işlemlerini kapsamaktadır [40].

Araştırmalar gösteriyor ki, istatistiki çalışmalarda zamanın büyük bir kısmını “veriyi tanıma ve ön işleme” kısmına harcanmaktadır. Şekil 3.2.’de veri hazırlama ve tahmin için yapılan adımlar gösterilmektedir.



Şekil 3.2. Veri sirkülasyonu [41]

Bu çalışmada Python programlama dili kullanılarak veri temizleme çalışması yapılmıştır. Veri setinde boş değer olup olmadığı kontrol edildikten sonra eksik değerler ortalama ve en yakın komşu (K-Nearest Neighbours- k_{NN}) yöntemi ile tamamlanmıştır. Hava kalitesi parametrelerinden PM_{10} , CO, NO_x , NO_2 değerlendirmeye alınmıştır. Sakarya ili için çalışma kapsamında kullanılacak veriler, Çevre ve Şehircilik Bakanlığı'nın hava kalitesi izleme istasyonlarından temin edilmiştir. Sakarya'da 4 adet hava kalitesi ölçüm istasyonu bulunmakta ve çalışmada bu istasyonlardan veri detayı açısından en zengin istasyon olan Sakarya Merkez hava kalitesi izleme istasyonu seçilmiştir. Merkez istasyona ait her 4 parametre için saatlik olarak Ocak 2020-Eylül 2022 tarihleri arasındaki ölçüm sonuçları kullanılmıştır. Belirlenen tarih aralığında her bir parametre için 24.282 olmak üzere 97.128 veri olması gerekmektedir. Fakat sensör hataları ve benzeri durumlara bağlı olarak kısmi veri eksiklikleri gözlemlenmiştir. Tablo 3.2.'de parametre bazlı gözlemlenen hataların giderilmesi için veri ön işleme ile doldurulan eksik veri adetleri sunulmuştur.

Tablo 3.2. Sakarya Merkez İstasyon PM_{10} , CO, NO_x , NO_2 ölçüm değerleri

İstasyon	Parametre	Birim	Ortalama Değer	Eksik Veri Sayısı	Eksik Veri Yüzdesi
Sakarya – Merkez – MTHM	PM_{10}	$\mu g/m^3$	35.8	1402	%5.77
Sakarya – Merkez – MTHM	CO	$\mu g/m^3$	1177.78	1615	%6.65
Sakarya – Merkez – MTHM	NO_2	$\mu g/m^3$	31.76	1647	%6.78
Sakarya – Merkez – MTHM	NO_x	$\mu g/m^3$	65.80	1888	%7.77

Ayrıca kullanılan parametrelerin her biri için Ocak 2020-Eylül2022 tarihleri arasında en yüksek ve en düşük olduğu noktalar değerlendirilmiştir. Buna istinaden oluşturulan Tablo 3.3.'de özetlenmiştir.

Tablo 3.3. Sakarya Merkez İstasyon PM_{10} , CO, NO_x , NO_2 için minimum ve maximum değerleri

İstasyon	Parametre	Birim	Min. Veri	Min. Veri Tarihi	Maksi. Veri	Maksi. Veri Tarihi
Sakarya – Merkez – MTHM	PM_{10}	$\mu g/m^3$	0.05	19.07.2020	385.59	09.12.2020
Sakarya – Merkez – MTHM	CO	$\mu g/m^3$	-20.15	06.01.2022	7058.43	07.02.2021
Sakarya – Merkez – MTHM	NO_2	$\mu g/m^3$	0.08	07.01.2022	275.42	21.09.2022
Sakarya – Merkez – MTHM	NO_x	$\mu g/m^3$	0.95	13.01.2022	842.74	09.12.2021

3.4. Veri Tahmini Yöntemleri, Performans Ölçümleri ve Hiperparametreler

Çalışmada hava kalitesi veri setlerinin eğitimi ve tahminini için kullanılan yöntemler, performans ölçümleri ve hiperparametreler hakkında detaylı açıklamalara yer verilmiştir. Veri Tahmini için tekrarlayan sinir ağlarına dayalı bir tahmin yöntemi olan LSTM (Long-Short Term Memory) kullanılmıştır. Veri ön işleme sonrası elde edilen veriler LSTM Algoritması ile eğitilmiştir.

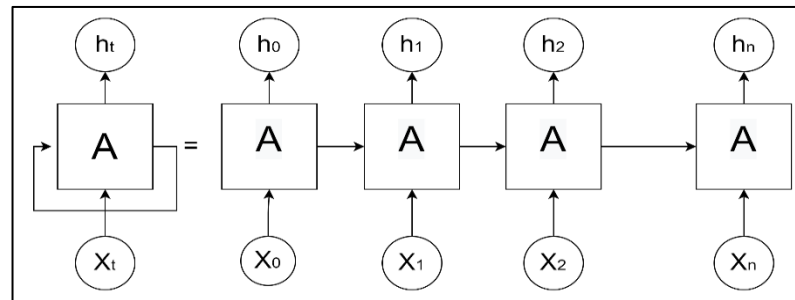
3.4.1. LSTM (long short term memory) modeli

Sakarya'da hava kirliliği tahmini yapmak için uzun-kısa süreli bellek (LSTM) algoritmaları ile başarılı sonuçlar elde edileceği düşünülmüştür. LSTM algoritması, özellikle zaman serileri verileri gibi zamana bağlı olan verilerin işlenmesinde oldukça etkilidir.

LSTM, tahmin için kullanılan tekrarlayan sinir ağlarına (RNN) dayalı bir öğrenme türüdür. RNN'ler döngülerden oluşmakta ve yapısında giriş, unutma ve çıkış kapıları bulunmaktadır. Giriş kapısında veri kümesinden veri okuma işlemi yapılmakta, unutma kapısında ise veriler depolanmaktadır. Son olarak çıkış kapısı ise yazma amaçlı kullanılmaktadır.

Bu kapılar, nöronal bilgi iletim sisteminden esinlenerek ortaya çıkarılmıştır. Mevcut nöron verilerin tahmininde kullanılmakta ve nöron verilerin sayısı bir sonrakine iletilmektedir [42].

RNN'ler aynı ağın birden fazla kopyası olarak düşünülebilir. Şekil 3.3.'te açılmış bir RNN Modeli gösterilmiştir.



Şekil 3.3. Açılmış bir RNN modeli

Şekil 3.3.'te yer alan diyagramda “A”, giriş X, Çıkışı h ifade etmektedir. RNN'ler önceki verileri kullanarak, güncel görev ile ilişkilendirmekte oldukça iyidirler. Aynı tip benzer döngülerden oluşmaktadır. RNN'lerin içinde var olan bu döngüsel formları

ile oldukça kullanışlı bir model olarak karşımıza çıkmakta ve konuşma tanıma, dil modellemesi, çeviri ve resim çözümlenmesi gibi alanlarda yaygın kullanılmaktadır. Öğrenme ve uygulama arası uzun bir süreyi kapsamıyorsa geçmiş bilgiyi kullanarak tahmin konusunda oldukça kabiliyetlidirler. Ancak veriler arasında boşluklar var ve uzun süreli öğrenme gerekiyorsa RNN'lerin öğrenmesi zorlaşmaktadır. Bu noktada, bu sorunun çözümü için RNN tabanlı bir tür olan LSTM modeli türetilmiştir [43].

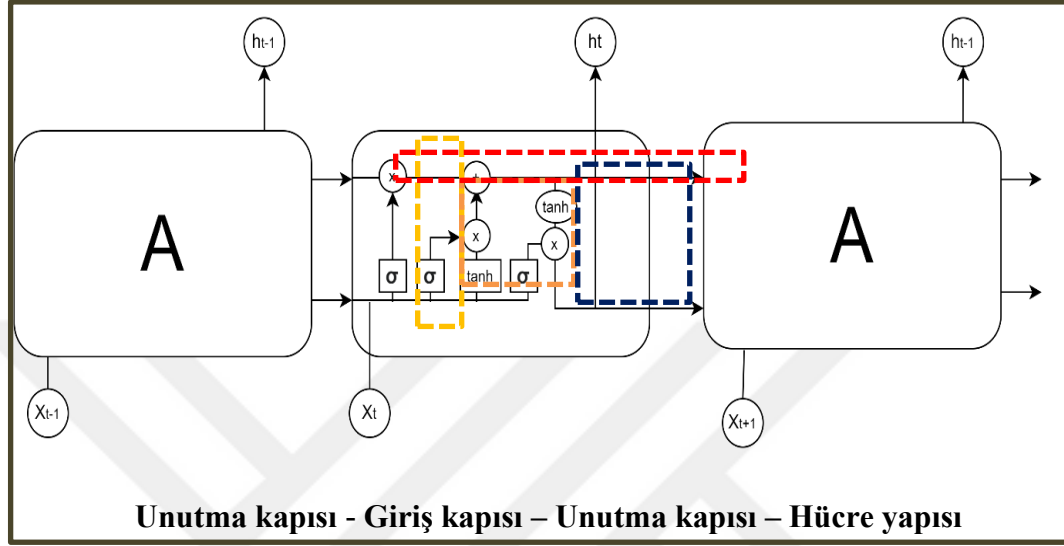
RNN'ler yinelenen ağ modeli ile doğrusal olmayan yalnızca bir işlev kullanmakta, LSTM'ler ise bilgileri uzun süre hatırlayabilmek için birden fazla doğrusal işlev kullanmaktadır. LSTM modeli bu haliyle RNN'ye göre daha karmaşık bir belleğe sahiptir. LSTM, bilgiyi depolar ve tahmin sırasında hangi bilgilerin saklanıp hangi hücrelerin kullanılacağına ve hangilerinin unutulacağına karar verir. Giriş ve çıkış kapılarında, bilgiler işlevine göre geçmekte ya da engellenmektedir. RNN'ler ile tahminde, veriler arasında kısa vadeli boşluklar var ise iyi bir başarı elde edilebilecekken, uzun vadeli öğrenme konusunda kabiliyetli olmama durumuna uzun vadeli bağımlılık problemi olarak isimlendirilmektedir. LSTM'ler ise temelde bu problemin çözümü için tasarlanmıştır. LSTM, akımın girişini, önceki durumunu ve hücrenin hafızasını birleştirerek uzun vadeli bağımlılıkları belirleyebilir [44].

RNN modelinde uzun vadeli bağımlılık problemi dışında büyük verilerin yığılınarak artmasından kaynaklanan artan gradyan patlaması problemi oluşmaktadır. LSTM tabanlı modeller, RNN modelindeki gradyan patlaması ve gradyan kaybolması problemlerini önemli ölçüde çözer. LSTM yöntemi tahmin için avantajlı ve yüksek doğruluk bir model olarak bu özellikleriyle öne çıkar. Hava kalitesi tahmini içinde, bu özellikleriyle önemli ölçüde tercih edilebilir olmaktadır [45].

LSTM ağları uzun kısa süreli bellek olarak dilimize çevrilmektedir. LSTM modeli, kapı olarak da adlandırılan dört adet katman içerir. Bilgiler bu kapılar vasıtasıyla depolanarak, hafıza hücresine yazılır ve okunur. Kapılar bilginin hücre tarafından ne zaman depolanacağı, ne zaman silinip ve ne zaman yazılacağı konusunda karar mekanizmasıdır. Giriş kapısı ve çıkış kapısı bilgilerin eğitilerek geçmesine ya da engellenmesine karar verir.

LSTM'lerin de art arda birbiri takip eden yapıları vardır. Ancak bir sonraki parça farklı bir yapıya sahiptir. Tek bir sinir ağı katmanı içermez. LSTM çalışma prensibi hafıza hücreleri ve kullandığı çeşitli kapılardır. Hücre yapısı tahmin için gerekli olan anlamlı

bilgileri taşıyan iletim hattıdır. Sistemin hafızasıdır. Bu yapısı sayesinde, problemler çözülerek eski veriler ağ zinciri boyunca taşınabilmektedir. Hücre yapısının bu yolculuğu süresince taşınması gerekmekte olan bilgiler kapılar tarafından belirlenir. Birbiriyle etkileşimde olan bu kapılar ve hücre yapısı dört parçaya sahiptir. Şekil 3.4.'te LSTM'nin bilgi akış diyagramı sunulmuştur [46].



Şekil 3.4. LSTM bilgi akışı diyagramı [46]

Unutma kapısı olan bir LSTM hücresinin ileri geçişi için kullanılan denklem yapısı;

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad (3.1)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad (3.2)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad (3.3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (3.4)$$

$$h_t = o_t \odot \sigma_h(c_t) \quad (3.5)$$

Burada $c_0 = 0$ ve $h_0 = 0$ başlangıç değerlerini belirlerken t ise zaman adımlarını indekslemede kullanılmaktadır.

Temel LSTM mimarisinde giriş katmanı (Denklem 3.2), çıkış (Denklem 3.3), unut kapısı (Denklem 3.1) ve hafıza hücrelerinden (Denklem 3.4) oluşmaktadır. Denklemdeki i , o , f , c , W terimleri giriş, çıkış, unut kapısını ve hücre aktivasyon vektörünü, ağırlık matrisini ifade etmektedir. Denklemde kullanılan $\odot$, İki vektörün skaler çarpımını göstermektedir. σ_x standart bir sigmoid fonksiyonunun tanımlanmasını gösterir.

LSTM'nin ilk adımı hücreden hangi bilginin unutulacağına karar vermektir. Bu kararı “unutma kapısı” olarak adlandırılan bir sigmoid tabakası verir. Hücre durumunda her sayı için 0 ve 1 arasında bir çıkış değeri oluşur. Çıkış değeri 1 ise bu bilginin tutulacağını, çıkış değeri 0 ise bu mevcut bilginin tamamen unutulacağını göstermektedir.

Çalışmada, LSTM algoritması kullanılarak Sakarya'da hava kirliliği tahmini yapılmış, PM₁₀, CO, NO_x, NO₂ gibi hava kirliliği parametreleri kullanılmıştır. Tahminler, gerçek veriler ile kıyaslanmıştır. Daha sonra LSTM modelinin performansını ölçmek için bazı metrikler kullanılmıştır. Bu metrikler bir sonraki başlıkta açıklanmıştır.

3.4.2. Performans ölçümleri

Tahmin yöntemlerinin performansı, birtakım metrikler ile ölçülür. Modelde, yüksek hata değeri var ise, kullanılan modelin tahmin doğruluğunun düşük olduğu anlamına gelir.

Çalışma kapsamında elde edilen tahmin sonuçlarının değerlendirilmesi için kullanılan belirlilik katsayısı R², hata kareler ortalamasının karekökü (Root Mean Squared Error (RMSE)), hata kareler ortalaması (Mean Squared Error (MSE)), hatanın mutlak ortalaması (Mean Absolute Error (MAE)) metriklerine ilişkin formüller aşağıda listelenmiştir.

MSE, hata kareler (Mean square error) ortalamasıdır. Bir regresyon eğrisinde nokta ile dizi mesafe arasındaki mesafeyi açıklar. MSE tahmin mekanizmasının performansını ölçmede kullanılır. Sonucu 0'a yakın olması tercih edilir, sıfıra ne kadar yakınsa performansı o kadar iyidir. Denklem 3.6'da metriğe ait denkleme yer verilmiştir.

$$MSE = \frac{\sum_{t=1}^n (u)^2}{n} \quad (3.6)$$

RMSE, hata kareler ortalamasının karekökü (Root Mean Squared Error): MSE'nin karekökü alınarak elde edilir (Denklem 3.7). Bir makine öğrenmesi modelinde tahmin edilen değer ile gerçek değer arasındaki uzaklığın bulunmasını sağlayan bir ölçümdür.

Tahmin hatalarının standart sapmasını açıklamak için kullanılır. RMSE 0 değeri alamaz ve genel anlamı ile değeri düşük olması tercih edilir. Model karşılaştırmalarında tercih edilir.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (u)^2}{n}} \quad (3.7)$$

MAE, hatanın mutlak ortalaması (Mean Absolute Error) olarak isimlendirilir. Sürekli değişkenler arasındaki farkı ölçer. MAE sonucu kolay yorumlanabilir olduğu için tercih edilir. Denklem 3.8.'de MAE hesaplaması gösterilmiştir.

$$MAE = \frac{\sum_{t=1}^n |u_t|}{n} \quad (3.8)$$

Bu metriklere göre, hangi arama algoritmasının kullanılmasının daha uygun olacağı ile ilgili yorumlama yapabilmek mümkün olmaktadır.

3.4.3. Hiperparametreler

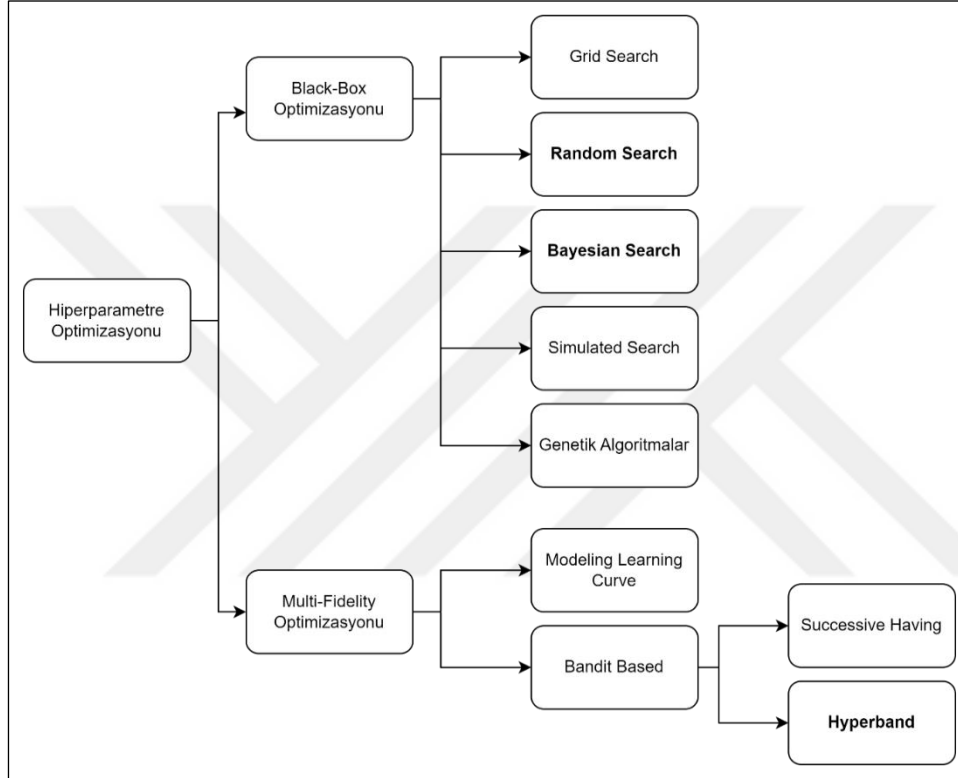
Çalışmada kullanılan, sinir ağı mimarisinin özel bir türü olan Long Short Term Memory'nin (LSTM) düzenlenmesi, optimizasyonu ve performansının artırılabilmesi için hiperparametre optimizasyonu yapılmıştır.

Bir sinir ağı modelinde değerleri, geçmiş verilerden tahmin edilen değişken türüne parametre, model parametrelerini kontrol edebilen model dışı parametrelere ise, hiperparametre denir. Hiperparametreler kontrol parametreleridir. Modelin eğitilmesi için doğru hiperparametreyi seçmek modelin performansını doğrudan etkileyecektir. Hiperparametre seçimleri manuel olarak kontrol edilebilmektedir. Ancak bu durum fazlaca zaman harcanmasına sebep olmaktadır. Bu durumun sonucu olarak hiperparametre optimizasyonu seçenek olarak sunulmaktadır.

Hiperparametre optimizasyonu (HPO), kullanılan modelin eğitim sürecinde uygun bir hiperparametre seçimini temel almaktadır. Bir sinir ağı mimarisinin eğitimi ve gelişimi için önemli bir bileşen olmakla birlikte mukayeseyi de kolaylaştırmaktadır [47].

HPO 1990'lı yıllardan günümüze kadar üzerinde çalışılan bir konudur. HPO işlemi, araştırmacıları, makine öğrenimi döngüsünden çıkarmak için makine öğrenim modelinin hiperparametrelerini otomatik olarak optimize etmektedir. HPO, son zamanlarda derin öğrenme modellerinin geliştirilmesinde en iyi sonucu bulabilmek için tercih edilmekte ve gereklilik arz etmektedir. Özellikle büyük ölçekli modellerde, hiperparametrelerini manuel olarak ayarlanabilmesi için çok fazla zamana ve hesaplama kaynağına ihtiyaç duyulmaktadır. Gerekli hesaplama kaynakları

incelendiğinde önemli bir yere sahip olan hiperparametreler, HPO aşamasında öncelikli olarak tercih edilir. Eğitim sırasında ağırlıklar üzerinde daha çok etkili olan hiperparametreler sinir ağı eğitimde önemlidir. Hiperparametreler model eğitimi ve model tasarımı olmak üzere iki farklı kullanım alanına ayrılmakta, model eğitimi ile ilgili olan hiperparametreler sinir ağının daha hızlı öğrenmesini ve gelişmiş bir performans elde etmesini sağlamaktadır [48]. Şekil 3.5.'te hiperparametre optimizasyon türleri verilmiştir [49].



Şekil 3.5. Hiperparametre optimizasyon türleri [49]

Bu çalışmada hiperparametre arama algoritmalarından Random Search, Bayesian Search ve Hyperband Search kullanılmıştır.

3.4.3.1. Random search

Geliştirilen model için en etkili ve doğru hiperparametre değerini rastgele kombinasyonları ile arar. Rastgele arama bazen arama alanındaki önemli noktaları gözden kaçırabilmektedir. Ancak zaman yönetimi açısından idealdir [50].

3.4.3.2. Bayesian search

Bir amaç fonksiyonunu minimize veya maksimize etmek, farklı hiperparametrelerin kombinasyonunu bulmak için Bayes teoremini kullanmaktadır. Karmaşık, gürültülü

ve pahalı objektif işlevler için uyarlanmıştır. Amaç fonksiyonunun çalışması gereken sayısını azaltarak çalışır. Bayes optimizasyonu için iki ana bileşen mevcuttur.

Amaç Fonksiyonu, vekil modelin tahmin için kullandığı, tahminin gerçekçiliği için verimli ve güvenilir olması gereken fonksiyondur. Amaç fonksiyonu sonuçları gelecekteki örneklemeyi yönlendirmek için kullanılabilir

Alım fonksiyonu, amaç fonksiyonunun değerlendirilmesi ile üretilen yeni değeri ölçer. Yalnızca en iyi sonuç veren hiperparametre setini seçerek arama alanındaki örneklem noktalarını önerir.

Bayes yöntemi, daha sonra, amaç işlevini değerlendirmek için kullanılan vekil işlev üzerinde en iyi performansı gösteren hiperparametrelerin kombinasyonunu bularak çalışmasını sürdürür [51].

3.4.3.3. Hyperband search

Ardışık yarıya indirme tekniğini genişleterek arama alanının optimize eden bir tekniktir. Manuel hiperparametre ayarlaması yapılırken, bir hiperparametre yapılandırması kötü sonuç verdiğinde eğitimi durdurup farklı bir yapılandırma denemesi yapılmaktadır. Hyperband search optimizasyonu da bu işlemi yapmaktadır.

Hyperband optimizasyonu; performans açısından, en iyi genelleme elde etmek için birkaç hiperparametre denemesiyle yineleme yapar. Erken durdurmayı kullanarak, kötü bir performans gösteren konfigürasyonları eğitmeyi durdurur, ancak daha iyi performansa sahip gelecek vaat eden konfigürasyonlara daha fazla kaynak tahsis eder [49].



4. DENEYSEL ÇALIŞMA VE SONUÇLARI

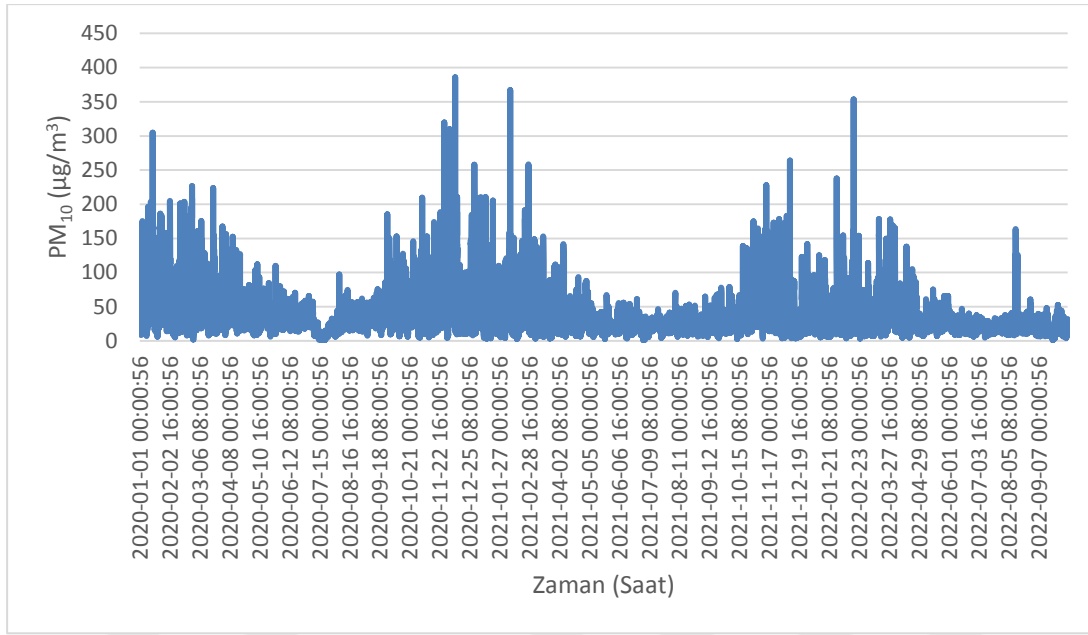
Bu çalışmada öncelikle, sensör hataları ya da işletmeden kaynaklı hatalar sebebiyle eksik olan verilerin kNN ve ortalama yöntemleriyle doldurulması işleminden sonra, hangi doldurma yönteminin daha iyi sonuç verdiği araştırılmıştır. Ortalama ve kNN ile doldurulan hava kirleticilerine ait veri setleri LSTM algoritması ile eğitilmiş ve MSE sonuçları kaydedilmiştir. Geliştirilen modeller rastgele üretilen sonuçlar içermemesi adına farklı rassal sayı üreteçleri ile 10 kez çalıştırılmıştır.

Elde edilen sonuçlara göre komşu arama algoritmasının dağılımı ve ortalaması NO_x dışında daha iyi olduğu gözlemlenmiştir. NO_x parametresinin ise ortalama ile doldurulduğunda daha iyi sonuç vermektedir. Böylece ikinci aşamada kullanılacak olan veri setleri NO_x için ortalama ile doldurulan, NO₂, PM₁₀ ve CO için ise kNN algoritması ile doldurulan yöntem olarak belirlenmiştir.

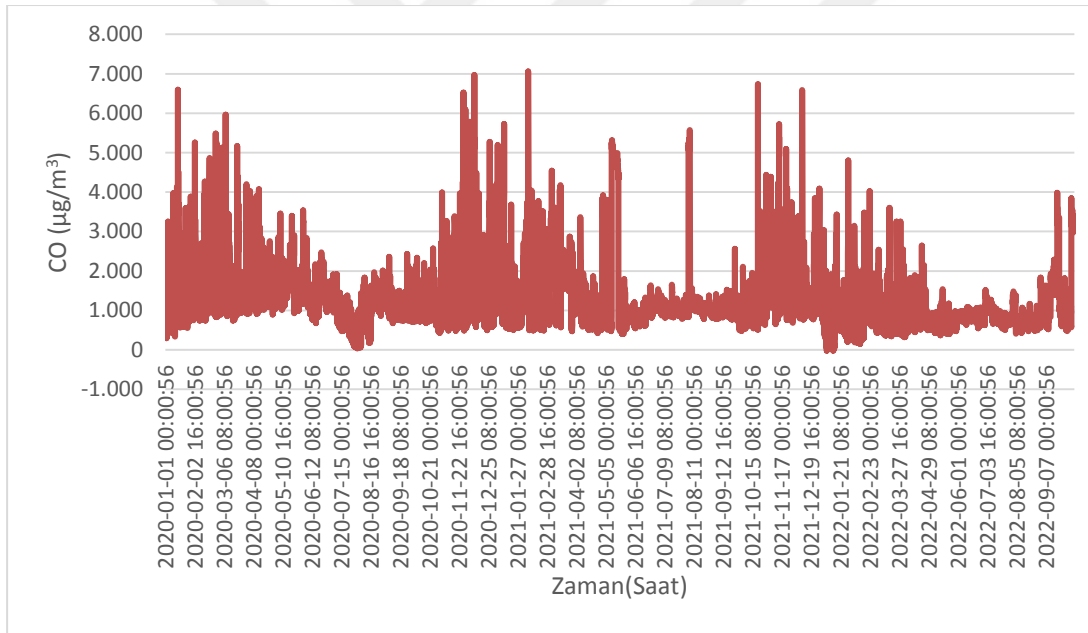
4.1. Veri Ön İşleme Sonrasında Elde Edilen Sonuçlar

Sakarya Merkez istasyonu, trafikten kaynaklanan hava kirliliği ölçümünün yapılabilmesi amacıyla kurulmuştur. Bu istasyonda PM₁₀, Toplam NO_x ve CO, azotdioksit (NO₂) ölçümü yapan cihazlar bulunmaktadır. Bu sebeple bu çalışmada, söz konusu dört adet kirletici üzerinden değerlendirmeler yapılmıştır.

Sakarya merkez istasyonunda, PM₁₀, NO₂, NO_x, CO veri değerleri saatlik olarak ele alınmıştır. Bu veriler ön işleme işleminden geçirildikten sonra tamamlanarak 97.128 adet veri elde edilmiştir. 2020-2022 yılları arasında her bir parametre (CO, PM₁₀, NO₂, NO_x) için elde edilen 24.282 adet veri üzerinden hava kalitesinin zamana bağlı değişimini gösteren grafikler Şekil 4.1., Şekil 4.2., Şekil 4.3. ve Şekil 4.4.'te verilmiştir.

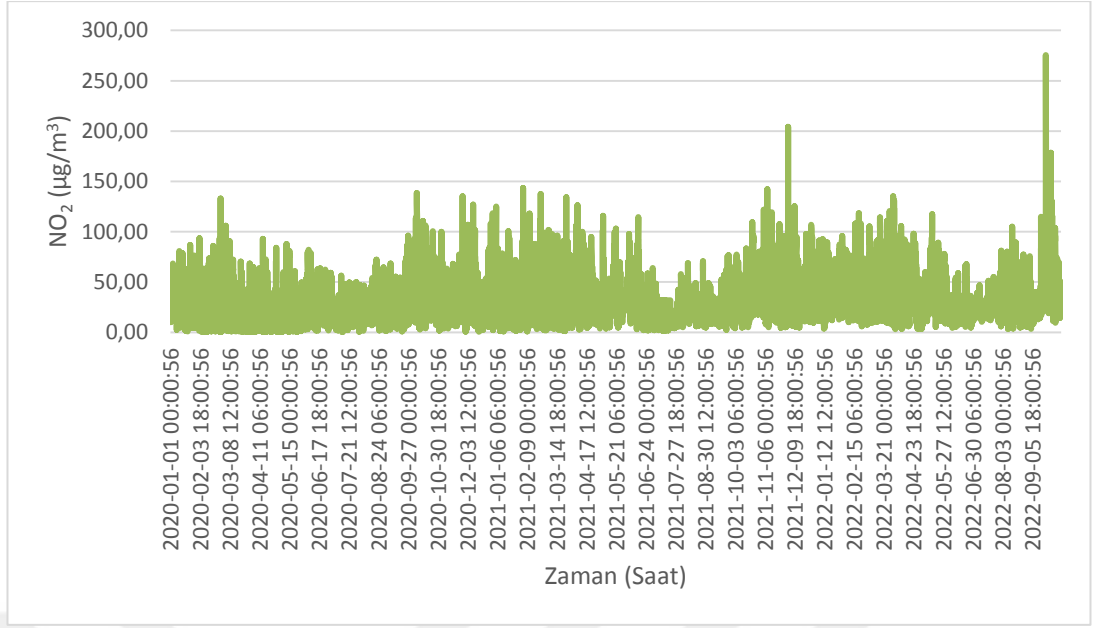


Şekil 4.1. PM₁₀'nun zamana bağlı değişimi

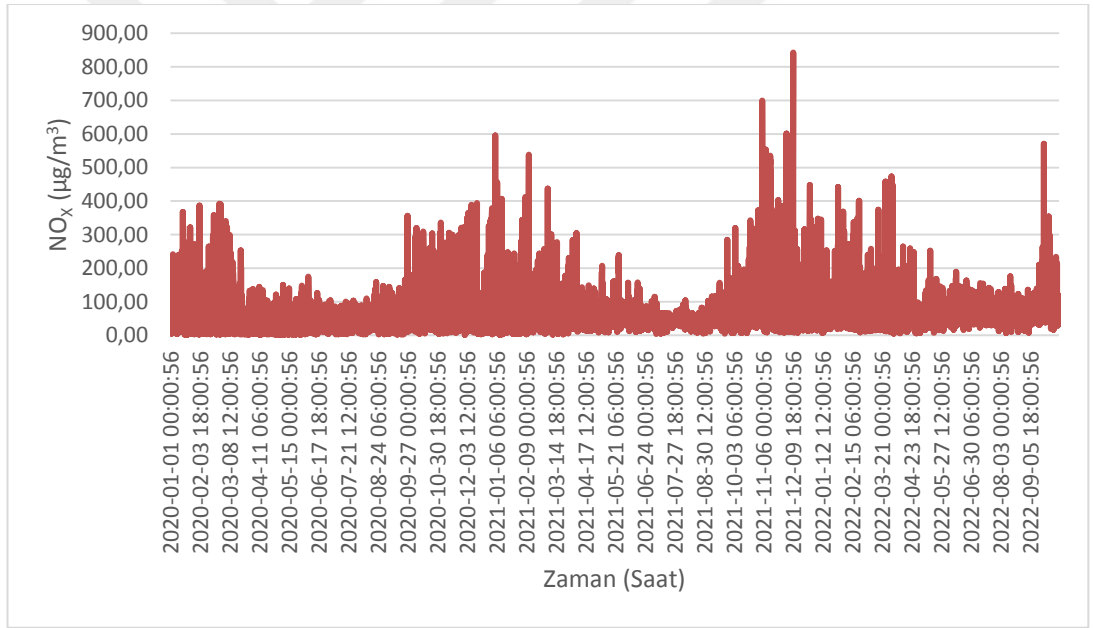


Şekil 4.2. CO'nun zamana bağlı değişimi

Burada değerlendirilen kirleticiler yalnızca tek bir istasyon verilerini içermekte fakat il genelini de kapsamaktadır. Hava kirleticilerinin noktasal olmadığı, küresel önem taşıdığı ilk bölümlerde detaylandırılmıştır. Grafikler, hava kirletici parametrelerin saatlik ölçümünden elde edilen verilerin, kNN ve ortalama yöntemleri ile eksik verilerin tamamlanması sonrası elde edilmiştir.



Şekil 4.3. NO₂'nin zamana bağlı değişimi



Şekil 4.4. NO_x'in zamana bağlı değişimi

Grafiklerdeki artış ve azalış noktaları izlendiğinde birtakım yorumlamalar yapmak mümkündür. Örneğin, COVID-19 pandemisi sırasında, sosyal hayatta kısıtlamaya gidilen tarih aralıklarında hava kalitesi parametrelerine ait verilerde azalış gözlenmektedir. 2020 Nisan ayından itibaren ülke genelinde uygulanmaya başlayan tedbirler, sokağa çıkma kısıtlamaları, özel ve devlete bağlı işletmelerin evden çalışmaya geçmesi, birçok eğlence mekânının kapanması vb. gibi durumlarında etkisi

ile kirletici emisyonlara sebep olan yaşamsal işlemlerin yavaşlamasına sebep olmuştur.

Ayrıca çalışmada kullanılan 2020 Ocak-2021 Aralık ayı için elde edilen saatlik verilerin ortalaması alınmış ve Tablo 4.1.'de sunulmuştur. Tablo 4.1. incelendiğinde günün bazı saatlerinde her bir parametre için artış gözlemlenmektedir. Örneğin 16.00'da değerlerler genelde düşüş göstermektedir. 00.00'da ise verilerde yükseliş gözlemlenmektedir. Bu sonuçlar göz önüne alındığında bu saatlerde trafik yoğunluğu ile ilgili yorumlama yapmak mümkün olmaktadır.

Tablo 4.1. 2020-2021 yılları arasında hava kirleticilerinin saatlik değişimlerin ortalama değerleri

Zaman	Kirletici Parametreler			
	PM ₁₀ (µg/m ³)	CO (µg/m ³)	NO ₂ (µg/m ³)	NO _x (µg/m ³)
00:00	54,7	1543,4	34,5	74,1
01:00	52,9	1534,7	33,1	69,6
03:00	48,3	1388,6	24,3	49,1
05:00	38,9	1241,8	18,8	35,6
08:00	34,6	1248,3	24,4	61,4
10:00	46,3	1335,6	31,8	85,1
12:00	45,1	1132,3	26,6	56,7
14:00	38,2	1026,6	22,4	42,1
16:00	35,6	1074,5	25,5	46,5
18:00	38,1	1123,5	31,0	58,4
20:00	46,0	1376,7	39,4	78,6
22:00	53,7	1521,1	37,7	77,7

Genel olarak 2020, 2021 ve 2022 yılları için ayrı ayrı ortalama aldığımızda Tablo 4.2.'de görülen sonuçlar ortaya çıkmaktadır.

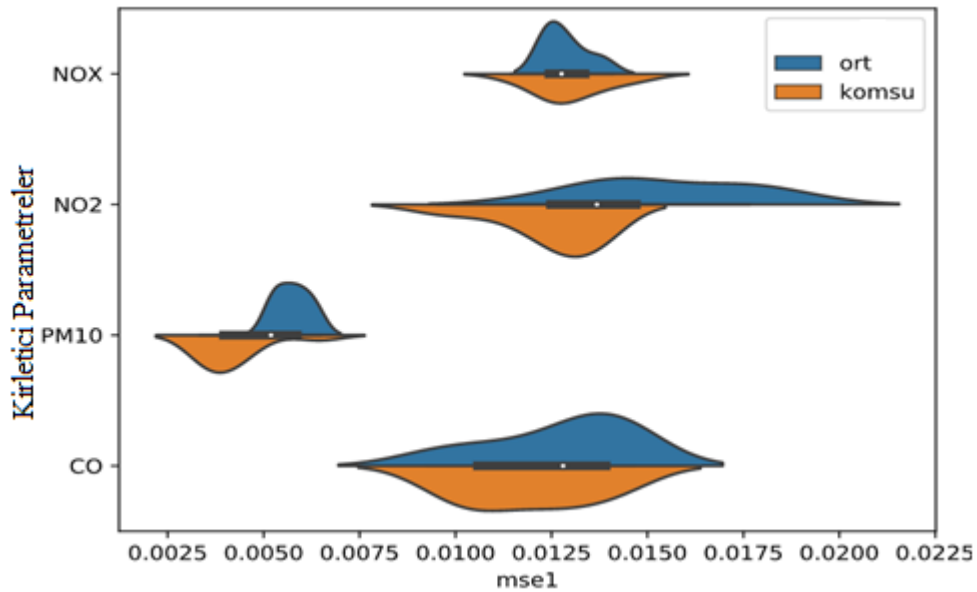
Tablo 4.2. 2020-2022 yılları kirletici konsantrasyonu ortalamaları

Yıl	PM ₁₀ (µg/m ³)	CO (µg/m ³)	NO ₂ (µg/m ³)	NO _x (µg/m ³)
2020	44,06	1278,99	29,67	62,93
2021	33,50	1308,59	30,89	64,03
2022	28,04	875,66	35,63	71,83

2022 yılında NO_x ve NO₂'de artış gözlemlenmekte fakat CO miktarında ciddi seviyede azalma görülmektedir. Buna bağlı olarak NO_x ve NO₂ salınımını indirmek için çalışmalar yapılmasının gerekliliği konusunda öneride bulunmak mümkündür.

4.2. Hiperparametre Optimizasyonu

2020 Ocak-2022 Eylül arasındaki ölçüm değerlerinin yer aldığı veri setindeki her bir parametreden 24.282, olmak üzere toplamda 97.128 örnek elde edildiği çalışmamızda, ortalama ve kNN yöntemleriyle doldurulan veri setleri için, MSE sonuçları göz önüne alınarak oluşturulan violin grafiği oluşturulmuştur (Şekil 4.5.).



Şekil 4.5. Kirleticilerin MSE analizi

Şekil 4.5.'deki grafikte, MSE sonuçları için; elde edilen sonuçlara göre komşu arama algoritmasının dağılımı ve ortalaması, NO_x dışında daha iyi olduğu gözlemlenmiştir. NO_x parametresinin ise ortalama ile doldurulduğunda daha iyi sonuç vermektedir. Böylece ikinci aşamada kullanılacak olan veri setleri NO_x için ortalama yöntemi ile doldurulan, NO₂, PM₁₀ ve CO için ise kNN algoritması ile doldurulan yöntem olarak belirlenmiştir.

Veri ön işleme sonrası her kirletic parametreye ait 24.282 adet veri seti elde edilmiş ve train-split veri setinin %80'i (19.423 örnek) eğitim, %20'si (4.859 örnek) test için ayrılmıştır. Veri setindeki korelasyonlar incelendiğinde, PM₁₀ ile NO₂ ve NO_x, CO kirleticileri arasında güçlü bir ilişki olduğu söylenebilmektedir.

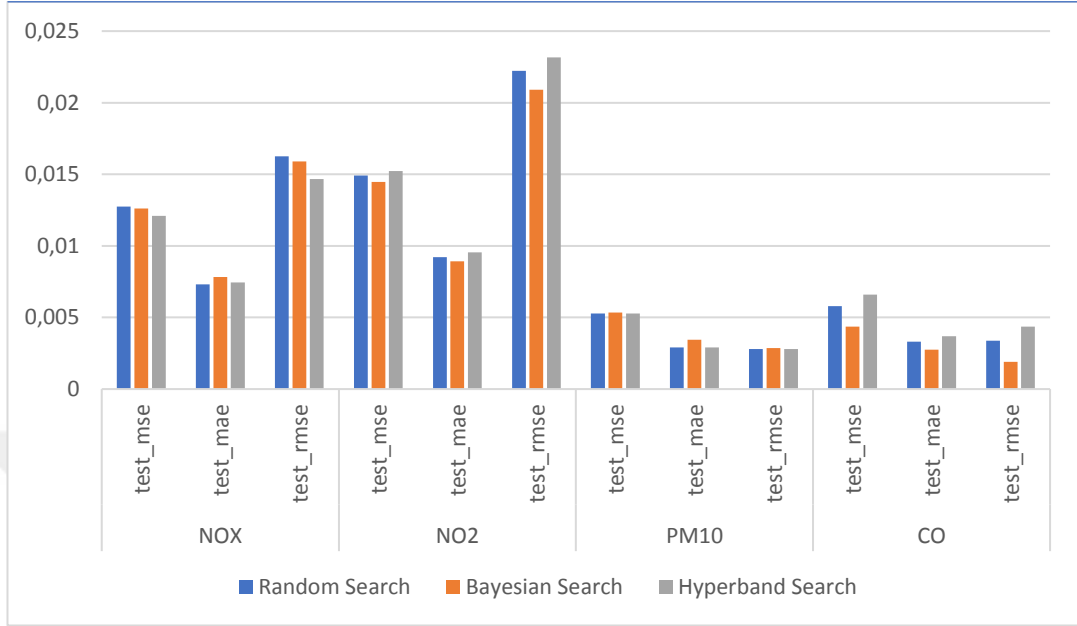
Çalışmada tahmine dayalı analiz için, derin öğrenme yöntemlerinden LSTM (Long Short Term Memory) tercih edilmiştir. Çalışmada kullanılan LSTM'nin tahmin performansını artırtmak için hiperparametre optimizasyonu uygulanmıştır. Hiperparametre arama algoritmalarından, random search, bayesian search, hyperband search algoritmaları uyarlanmıştır. Hava kalitesine ait her parametre için (NO₂, NO_x, PM₁₀, CO) doğru hiperparametre seçimi hiperparametre optimizasyonu ile sağlanmıştır. Hiperparametre optimizasyonuna ait değerler elde edilen sonuçlar Tablo 4.3.'de gösterilmiştir.

Tablo 4.3. Hiperparametre kıyaslanması

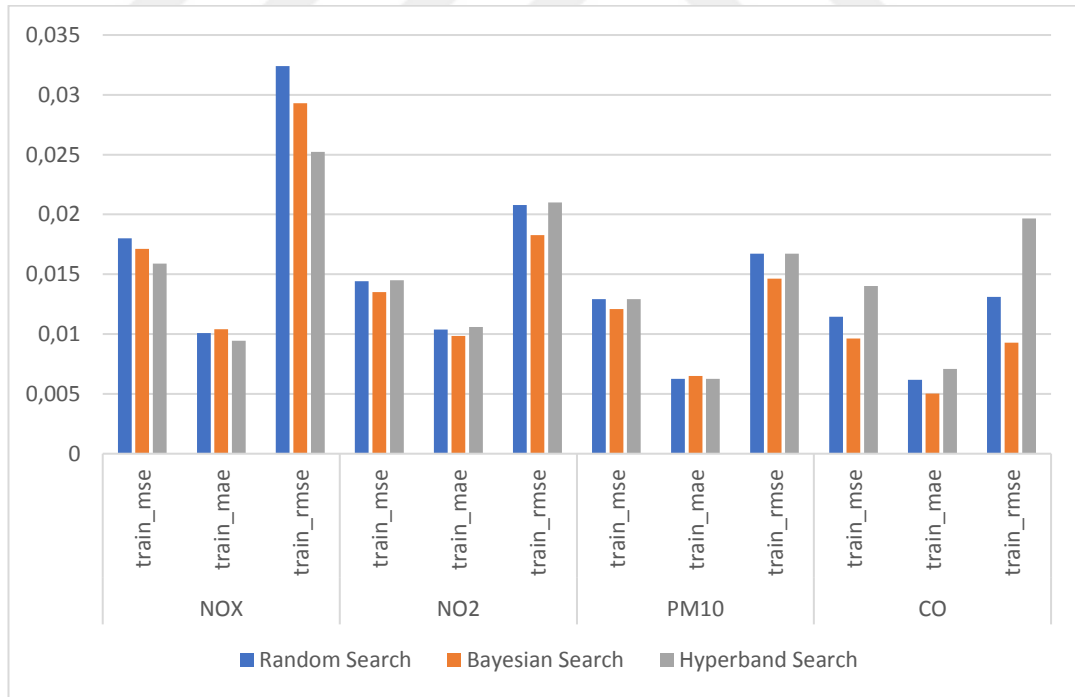
Hava Kirletic Parametre	Perfor mans Parametreleri	Eğitim			Test		
		Random Search	Bayesian Search	Hyperband Search	Random Search	Bayesian Search	Hyperband Search
NO _x	RMSE	0.000324	0.000293	0.000252	0.000163	0.000159	0.000147
	MSE	0.018004	0.017119	0.015885	0.012748	0.012611	0.012108
	MAE	0.010095	0.010395	0.009443	0.007309	0.007818	0.007456
	R2	0.923399	0.930745	0.940371	0.937072	0.93841	0.943226
NO ₂	RMSE	0.000208	0.000183	0.00021	0.000222	0.000209	0.000232
	MSE	0.014421	0.013516	0.014488	0.01491	0.014458	0.01522
	MAE	0.010366	0.009837	0.010597	0.009216	0.008919	0.009548
	R2	0.960247	0.965077	0.959877	0.962825	0.965043	0.961261
PM ₁₀	RMSE	0.000167	0.000146	0.000167	2.8E-05	2.86E-05	2.8E-05
	MSE	0.012933	0.012096	0.012933	0.005289	0.005352	0.005289
	MAE	0.006248	0.006488	0.006248	0.002908	0.003453	0.002908
	R2	0.972448	0.975901	0.972448	0.983574	0.983181	0.983574
CO	RMSE	0.000131	9.29E-05	0.000197	3.37E-05	1.9E-05	4.36E-05
	MSE	0.011453	0.009636	0.014022	0.005803	0.004354	0.006606
	MAE	0.006176	0.005034	0.007078	0.003309	0.002742	0.003686
	R²	0.988553	0.991897	0.982842	0.993442	0.996308	0.991502

Tablodaki veriler bir çubuk grafiğinde Şekil 4.6.'da görselleştirilmiş, sonuçların daha iyi analiz edilebilmesi sağlanmıştır. Çubuk grafiğinde yer verilebilmesi için RMSE değeri 100 ile çarpılarak grafiğe dâhil edilmiştir.

RMSE değeri için düşük olan hiperparametrenin uygun olmasını bekleriz. MSE değerinin ise 0'a yakın olmasını beklenen ve tercih edilen bir durumdur. Bu bilgiler ışında, görseldeki grafiği yorumlayabilmek mümkün kılınmıştır.



Şekil 4.6. Hiperparametre algoritmalarının kıyaslanması (Test seti)



Şekil 4.7. Hiperparametre algortimalarının kıyaslanması (Eğitim seti)

Elde edilen grafikte düşük olmasını istediğimiz hiperparametre değerleri incelendiğinde, NO_x için Hyberband, NO₂, PM₁₀, CO için Bayesian, hiperparametrelerinin uygun olduğu görülmektedir.

Uyarlamalar sonrası, elde edilen sonuçlar, aşağıda tablolar halinde gösterilmiştir. Tablo 4.4.'te ortak parametreler listelenmiş, Tablo 4.5.'te parametre arama uzayı, Tablo 4.6.'da arama algoritmalarının elde ettiği en iyi parametreler verilmiştir.

Tablo 4.4. Ortak parametreler

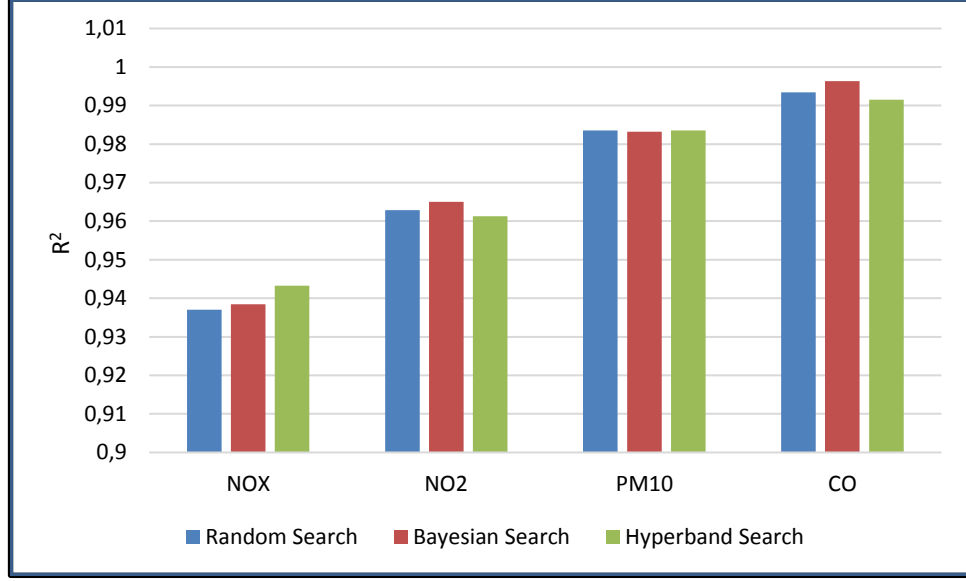
BATCH_SIZE	128
BUFFER_SIZE	150
EPOCHS	20
EVALUATION_INTERVAL	100
MAX_TRIAL	30
SEED	2022
TRAIN_SPLIT	19423
Number of hidden layer	2
Optimizer	Adam
Loss	MSE (Mean Squared Error)

Tablo 4.5. Parametre arama uzayı

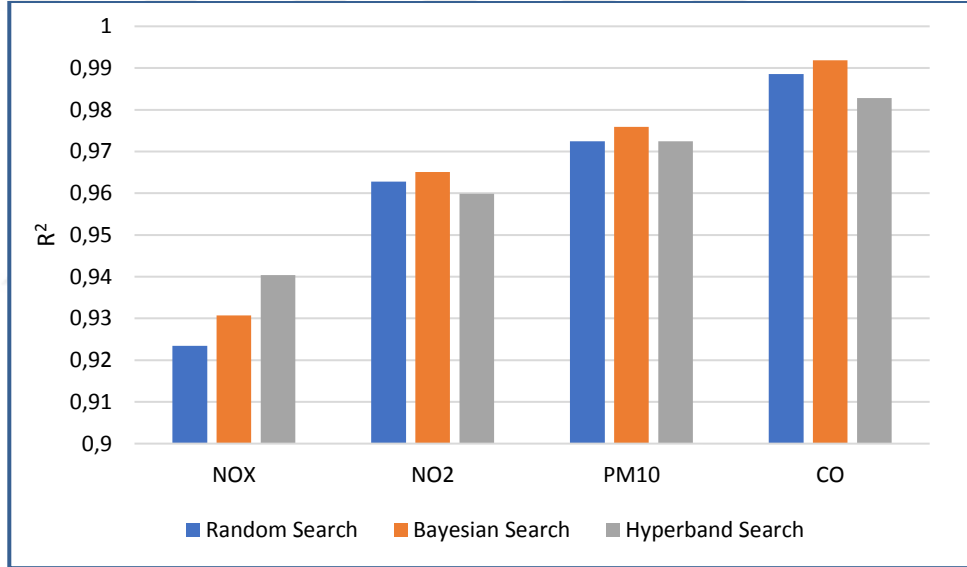
Birinci katmandaki nöron sayısı	'units1', min_value=50, max_value=200, step=50
İkinci katmandaki nöron sayısı	'units2', min_value=50, max_value=200, step=50
Dropout oranı	min_value=0.0, max_value=0.5, step=0.1
Learning Rate	[1e-1,1e-2, 1e-3, 1e-4,1e-5, 1e-6]

Tablo 4.6. Arama algoritmalarının elde ettiği en iyi parametreler

best_hyperparameters_hb_nox (Hyperband ile optimize edilmiş NO _x kirleticisi için en iyi parametreler)	{'units1': 150, 'units2': 200, 'rate': 0.1, 'learning_rate': 0.001, 'tuner/epochs': 20, 'tuner/initial_epoch': 7, 'tuner/bracket': 1, 'tuner/round': 1, 'tuner/trial_id': '0019'}
best_hyperparameters_bayes_no2 (Bayesian arama ile optimize edilmiş NO ₂ kirleticisi için en iyi parametreler)	{'units1': 150, 'units2': 50, 'rate': 0.4, 'learning_rate': 0.001}
best_hyperparameters_bayes_pm10	{'units1': 100, 'units2': 200, 'rate': 0.1, 'learning_rate': 0.001}
best_hyperparameters_bayes_co	{'units1': 200, 'units2': 100, 'rate': 0.0,'learning_rate': 0.001}



Şekil 4.8. Hiperparametre algoritmaları için R^2 kıyaslaması (Test seti)



Şekil 4.9. Hiperparametre algoritmaları için R^2 kıyaslaması (Eğitim seti)

Şekil 4.8. ve Şekil 4.9.'da hiperparametreler için R^2 kıyaslamasına ait grafiklerde, NO_x için en uygun hiperparametrenin diğer metriklerde olduğu gibi hyperband, diğer parametreler için ise bayesian hiperparametresi olduğu görülmüştür. NO_x için test tahmin oranı (R^2 0.941), NO₂ için (R^2 0.967) PM₁₀ için (R^2 0.973) ve CO için (R^2 0.995) olarak yüksek performansla tahmin edilmiştir.

Sonuç olarak modelde R değerlerinin 1'e yakın olması regresyon modelinin iyi olduğunu göstermektedir. Bununla birlikte, aralarında farklar olmasına rağmen

kullanılan modeldeki tüm hiperparametreler için R^2 istenen değer aralığında yüksek tahmin yeteneğine sahip olarak değerlendirilebilir.

Ayrıca çalışma kapsamında değerlendirilen tüm hiperparametreler için çalışma süreleri elde edilmiş ve Tablo 4.7.'de gösterilmiştir.

Tablo 4.7. Arama algoritması çalışma süreleri

Hava Kirleticisi	Algoritma	CPU(s)
CO	Bayesian Search	11253.75
CO	Random Search	6131572
CO	Hyperband Search	4007367
NO ₂	Bayesian Search	7957383
NO ₂	Random Search	5156578
NO ₂	Hyperband Search	3751802
NO _x	Bayesian Search	7306751
NO _x	Random Search	5109111
NO _x	Hyperband Search	3499959
PM ₁₀	Bayesian Search	10592.38
PM ₁₀	Random Search	5759.5
PM ₁₀	Hyperband Search	4242168

4.3. Veri Setine Ait Sonuçlar

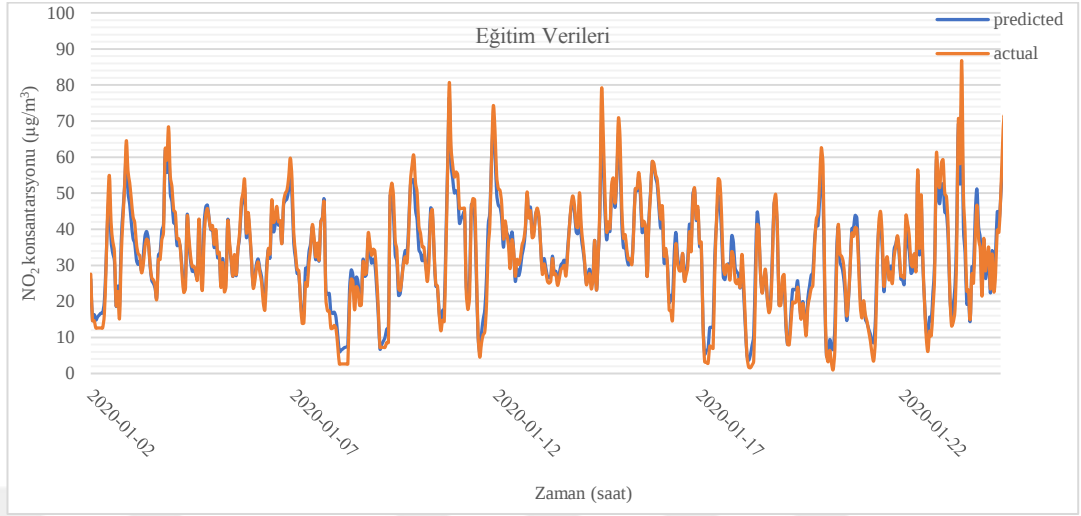
Bu kısımda LSTM modelinin komşu ve ortalama algoritmalarıyla doldurulup oluşturulan veri seti üzerindeki sonuçlarına yer verilmiştir. Veri setinin %80'i eğitim %20'si test olarak ayrılmıştır. En iyi arama algoritması ile tespit edilen hiperparametre değerleri yeniden eğitim için kullanılmıştır. Bu aşamadan sonra oluşturulan tahminler ayrı ayrı başlıklar altında değerlendirilmiştir.

4.3.1. NO₂ parametresi için model sonuçlarının değerlendirmesi

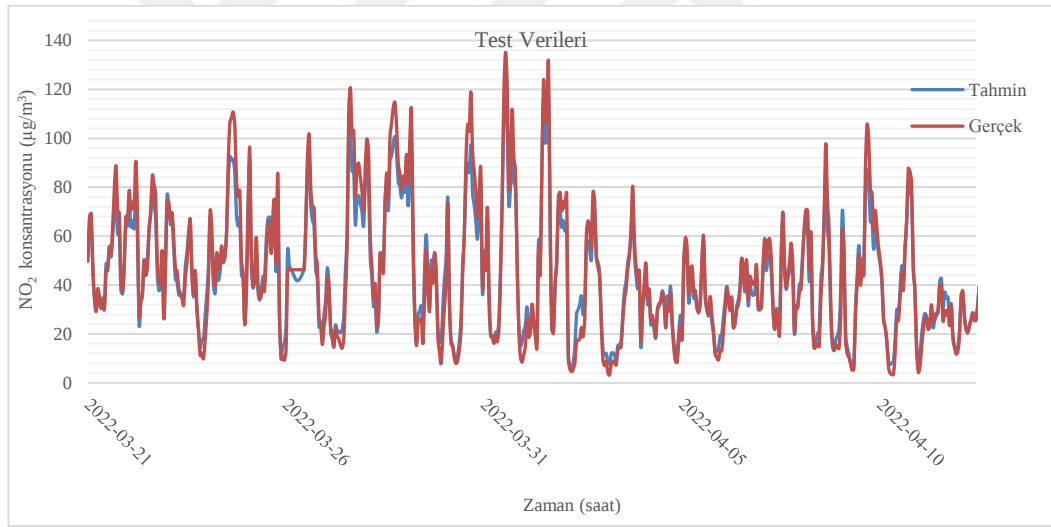
LSTM algoritması; NO₂'ye ait, 21.03.2022-8.10.2022 tarihleri arasında, toplam 200 günlük saatlik veri seti kullanılarak test edilmiştir. Test sonucunda algoritmanın güvenilirliği mukayese edilmiş olup Şekil 4.10., Şekil 4.11.'de yer alan grafikte gösterilmiştir, grafikler incelendiğinde eğitim ve test verilerinin güvenilirlik düzeyi yüksek olarak tespit edilmiştir.

Sonuç olarak NO₂'nin tahmini için kullanılan LSTM algoritmasının ve hiperparametre optimizasyonunun yüksek tahmin oranına imkân sağladığı görülmüştür. Buna bağlı olarak NO₂ parametresi için algoritma 10 kez çalıştırılmış ve hata metrikleri ortaya çıkarılmış, deney sırasına göre Tablo 4.8.'de gösterilmiştir. Ortalama ve komşu

yöntemi için ayrı ayrı değerlendirilerek eğitim ve test verileri karşılaştırmalı olarak Tablo 4.8’de gösterilmiştir. Predicted tahmin verilerini, actual gerçek verileri gösterir.



Şekil 4.10. NO₂ konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (eğitim seti)



Şekil 4.11. NO₂ konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (test seti)

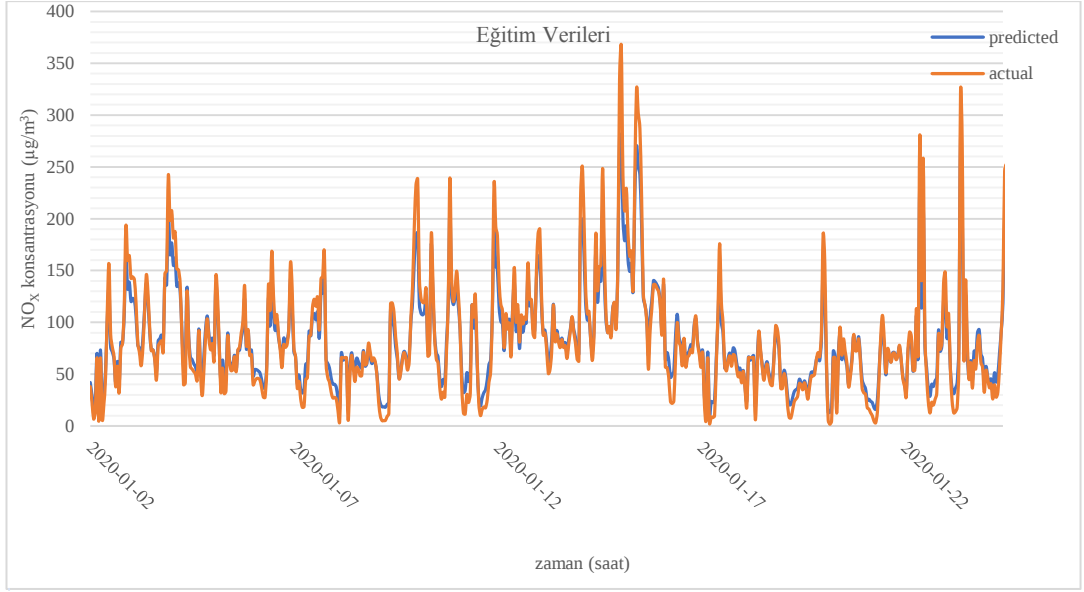
Tablo 4.8. NO₂ parametresi için en iyi arama algoritması performans sonuçları

NO ₂		Test Seti					Eğitim Seti				
Deney No	Eksik Veri Doldurma Yöntemi	rmse ₁	mse ₁	mae ₁	mape ₁	r ² ₁	rmse ₂	mse ₂	mae ₂	mape ₂	r ² ₂
1	Ortalama Yöntemi	0,0001	0,012	0,009	0,104	0,972	0,0001	0,012	0,009	0,104	0,972
2		0,0003	0,018	0,011	0,114	0,941	0,0003	0,018	0,011	0,114	0,941
3		0,0004	0,019	0,013	0,133	0,932	0,0004	0,019	0,013	0,133	0,932
4		0,0002	0,015	0,009	0,093	0,96	0,0002	0,015	0,009	0,093	0,96
5		0,0002	0,015	0,01	0,108	0,956	0,0002	0,015	0,01	0,108	0,956
6		0,0003	0,018	0,012	0,136	0,941	0,0003	0,018	0,012	0,136	0,941
7		0,0003	0,017	0,011	0,105	0,945	0,0003	0,017	0,011	0,105	0,945
8		0,0002	0,014	0,009	0,09	0,963	0,0002	0,014	0,009	0,09	0,963
9		0,0002	0,014	0,009	0,08	0,963	0,0002	0,014	0,009	0,08	0,963
10		0,0002	0,014	0,01	0,099	0,96	0,0002	0,014	0,01	0,099	0,96
1	Komşu Yöntemi	0,0001	0,011	0,007	0,059	0,98	0,0001	0,011	0,007	0,059	0,98
2		0,0002	0,013	0,009	0,101	0,97	0,0002	0,013	0,009	0,101	0,97
3		0,0002	0,013	0,008	0,08	0,974	0,0002	0,013	0,008	0,08	0,974
4		0,0002	0,013	0,008	0,088	0,971	0,0002	0,013	0,008	0,088	0,971
5		0,0002	0,014	0,009	0,085	0,968	0,0002	0,014	0,009	0,085	0,968
6		0,0001	0,012	0,008	0,084	0,975	0,0001	0,012	0,008	0,084	0,975
7		0,0002	0,014	0,008	0,079	0,969	0,0002	0,014	0,008	0,079	0,969
8		9,00E-05	0,01	0,007	0,086	0,985	9,00E-05	0,01	0,007	0,086	0,985
9		0,0002	0,014	0,009	0,094	0,969	0,0002	0,014	0,009	0,094	0,969
10		0,0002	0,012	0,008	0,082	0,974	0,0002	0,012	0,008	0,082	0,974

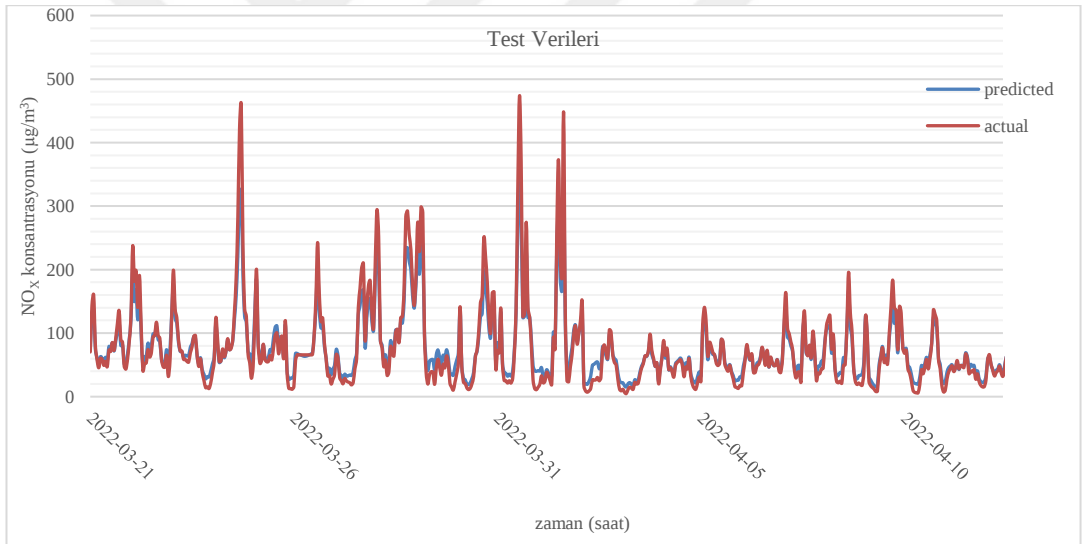
4.3.2. NO_x parametresi için model sonuçlarının değerlendirilmesi

LSTM algoritması; NO_x'ye ait, 21.03.2022-8.10.2022 tarihleri arasında, toplam 200 günlük saatlik veri seti kullanılarak test edilmiştir. Test sonucunda algoritmanın güvenilirliği mukayese edilmiş olup Şekil 4.12. ve Şekil 4.13.'te yer alan grafikte gösterilmiştir, grafikler incelendiğinde eğitim ve test verilerinin güvenilirlik düzeyi yüksek olarak tespit edilmiştir.

Sonuç olarak NO_x'nin ileriye dönük tahmini için kullanılan LSTM algoritması ve hiperparametre optimizasyonunun yüksek tahmin oranına imkân sağladığı görülmüştür. Buna bağlı olarak NO_x parametresi için 10 kez çalıştırılmıştır ve hata metrikleri ortaya çıkarılmış, deney sırasına göre Tablo 4.9.'de gösterilmiştir. Ortalama ve komşu yöntemi için ayrı ayrı değerlendirilerek tahmin işlemi sonuçlandırılmıştır. Predicted tahmin verilerini, actual gerçek verileri gösterir.



Şekil 4.12. NO_x konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (eğitim seti)



Şekil 4.13. NO_x konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (test seti)

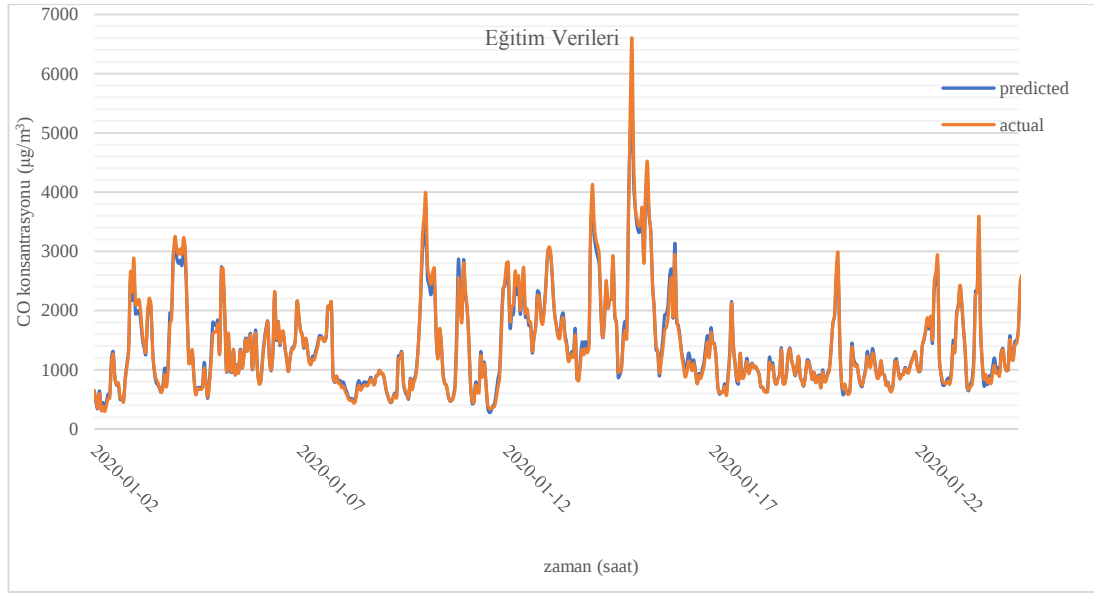
Tablo 4.9. NO_x parametresi için en iyi arama algoritması performans sonuçları

NO _x		Test Seti					Eğitim Seti				
Deney No	Eksik Veri Doldurma Yöntemi	rmse ₁	mse ₁	mae ₁	mape ₁	r ² ₁	rmse ₂	mse ₂	mae ₂	mape ₂	r ² ₂
1	Ortalama Yöntemi	0,0002	0,012	0,008	0,152	0,94	0,0002	0,012	0,008	0,152	0,94
2		0,0002	0,014	0,009	0,144	0,925	0,0002	0,014	0,009	0,144	0,925
3		0,0002	0,012	0,008	0,159	0,94	0,0002	0,012	0,008	0,159	0,94
4		0,0002	0,014	0,008	0,135	0,928	0,0002	0,014	0,008	0,135	0,928
5		0,0002	0,013	0,009	0,154	0,934	0,0002	0,013	0,009	0,154	0,934
6		0,0002	0,012	0,008	0,133	0,942	0,0002	0,012	0,008	0,133	0,942
7		0,0002	0,012	0,008	0,134	0,941	0,0002	0,012	0,008	0,134	0,941
8		0,0002	0,013	0,009	0,168	0,939	0,0002	0,013	0,009	0,168	0,939
9		0,0002	0,013	0,008	0,145	0,937	0,0002	0,013	0,008	0,145	0,937
10		0,0002	0,013	0,008	0,135	0,934	0,0002	0,013	0,008	0,135	0,934
1	Komşu Yöntemi	0,0002	0,013	0,009	0,166	0,942	0,0002	0,013	0,009	0,166	0,942
2		0,0002	0,014	0,008	0,11	0,93	0,0002	0,014	0,008	0,11	0,93
3		0,0002	0,012	0,008	0,154	0,945	0,0002	0,012	0,008	0,154	0,945
4		0,0002	0,013	0,008	0,129	0,943	0,0002	0,013	0,008	0,129	0,943
5		0,0002	0,015	0,01	0,157	0,923	0,0002	0,015	0,01	0,157	0,923
6		0,0002	0,013	0,009	0,164	0,943	0,0002	0,013	0,009	0,164	0,943
7		0,0001	0,011	0,008	0,137	0,954	0,0001	0,011	0,008	0,137	0,954
8		2,00E-04	0,014	0,009	0,155	0,934	2,00E-04	0,014	0,009	0,155	0,934
9		0,0002	0,013	0,009	0,154	0,938	0,0002	0,013	0,009	0,154	0,938
10		0,0002	0,012	0,009	0,172	0,947	0,0002	0,012	0,009	0,172	0,947

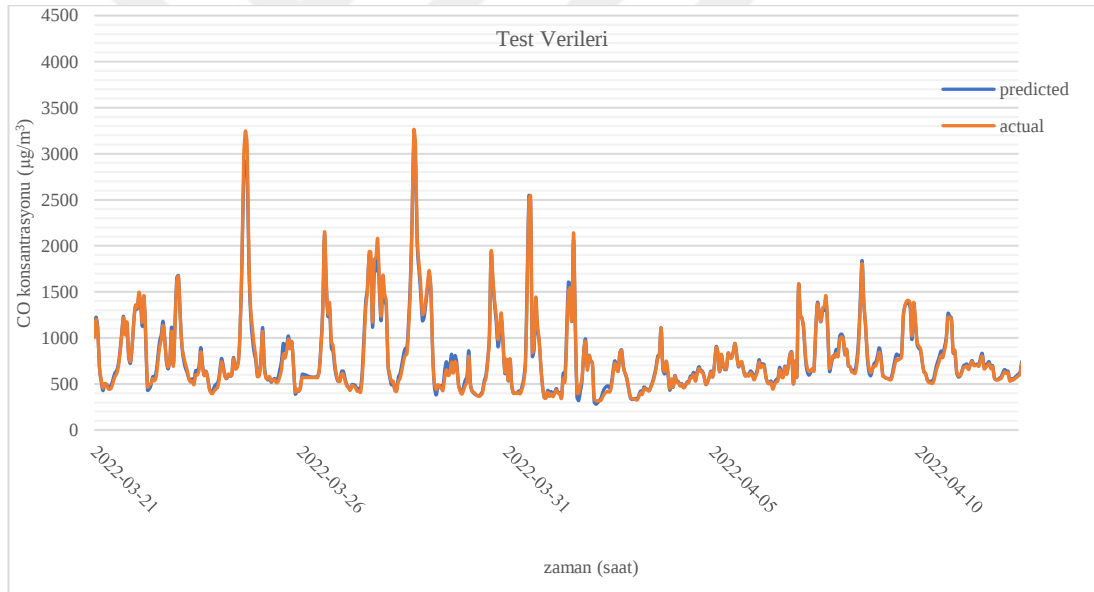
4.3.3. CO parametresi için model sonuçlarının değerlendirilmesi

Çalışmada kullandığımız LSTM algoritması; CO'ya ait, 21.03.2022-8.10.2022 tarihleri arasında, toplam 200 günlük saatlik veri seti kullanılarak test edilmiştir. Test sonucunda algoritmanın güvenilirliği mukayese edilmiş olup Şekil 4.14. , Şekil 4.15.'te yer alan grafikte gösterilmiştir, grafikler incelendiğinde eğitim ve test verilerinin güvenilirlik düzeyi yüksek olarak tespit edilmiştir.

Sonuç olarak CO'nun tahmini için kullanılan LSTM algoritması ve hiperparametre optimizasyonu yüksek tahmin oranına imkân sağladığı görülmüştür. Buna bağlı olarak NO_x parametresi için 10 kez çalıştırılmıştır ve hata metrikleri ortaya çıkarılmış, deney sırasına göre Tablo 4.10.'de gösterilmiştir. Ortalama ve komşu yöntemi için ayrı ayrı değerlendirilerek tahmin işlemi sonuçlandırılmıştır. Predicted tahmin verilerini, actual gerçek verileri gösterir.



Şekil 4.14. CO konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (eğitim seti)



Şekil 4.15. CO konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (test seti)

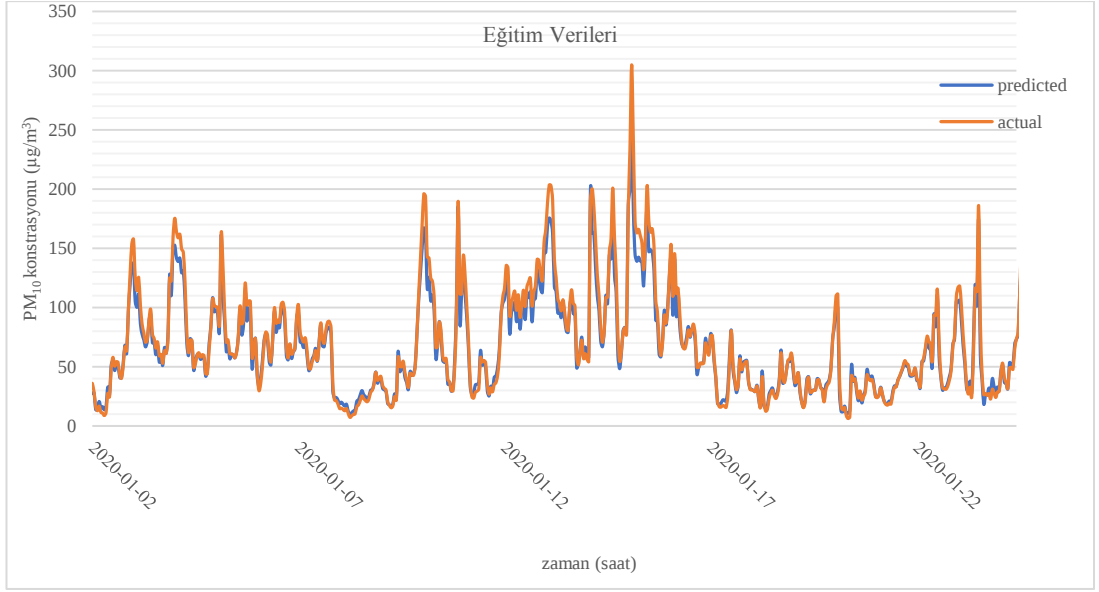
Tablo 4.10. CO parametresi için en iyi arama algoritması performans sonuçları

CO		Test Seti					Eğitim Seti				
Deney No	Eksik Veri Doldurma Yöntemi	rmse ₁	mse ₁	mae ₁	mape ₁	r ² ₁	rmse ₂	mse ₂	mae ₂	mape ₂	r ² ₂
1	Ortalama Yöntemi	0,0001	0,011	0,006	0,034	0,972	0,0001	0,011	0,006	0,034	0,972
2		0,0002	0,013	0,006	0,038	0,965	0,0002	0,013	0,006	0,038	0,965
3		0,0002	0,014	0,007	0,047	0,958	0,0002	0,014	0,007	0,047	0,958
4		0,0001	0,01	0,006	0,044	0,977	0,0001	0,01	0,006	0,044	0,977
5		9,00E-05	0,009	0,005	0,029	0,981	9,00E-05	0,009	0,005	0,029	0,981
6		0,0002	0,015	0,008	0,053	0,953	0,0002	0,015	0,008	0,053	0,953
7		0,0002	0,015	0,009	0,055	0,953	0,0002	0,015	0,009	0,055	0,953
8		0,0002	0,013	0,008	0,054	0,963	0,0002	0,013	0,008	0,054	0,963
9		0,0002	0,014	0,008	0,055	0,959	0,0002	0,014	0,008	0,055	0,959
10		0,0002	0,014	0,008	0,053	0,955	0,0002	0,014	0,008	0,053	0,955
1	Komşu Yöntemi	0,0001	0,012	0,006	0,033	0,973	0,0001	0,012	0,006	0,033	0,973
2		0,0002	0,013	0,006	0,035	0,967	0,0002	0,013	0,006	0,035	0,967
3		0,0002	0,014	0,007	0,043	0,961	0,0002	0,014	0,007	0,043	0,961
4		0,0001	0,01	0,006	0,04	0,979	0,0001	0,01	0,006	0,04	0,979
5		9,00E-05	0,01	0,005	0,027	0,982	9,00E-05	0,01	0,005	0,027	0,982
6		0,0002	0,013	0,007	0,047	0,967	0,0002	0,013	0,007	0,047	0,967
7		0,0001	0,011	0,005	0,034	0,978	0,0001	0,011	0,005	0,034	0,978
8		0,0002	0,014	0,008	0,051	0,96	0,0002	0,014	0,008	0,051	0,96
9		0,0001	0,012	0,007	0,052	0,971	0,0001	0,012	0,007	0,052	0,971
10		0,0001	0,01	0,005	0,029	0,979	0,0001	0,01	0,005	0,029	0,979

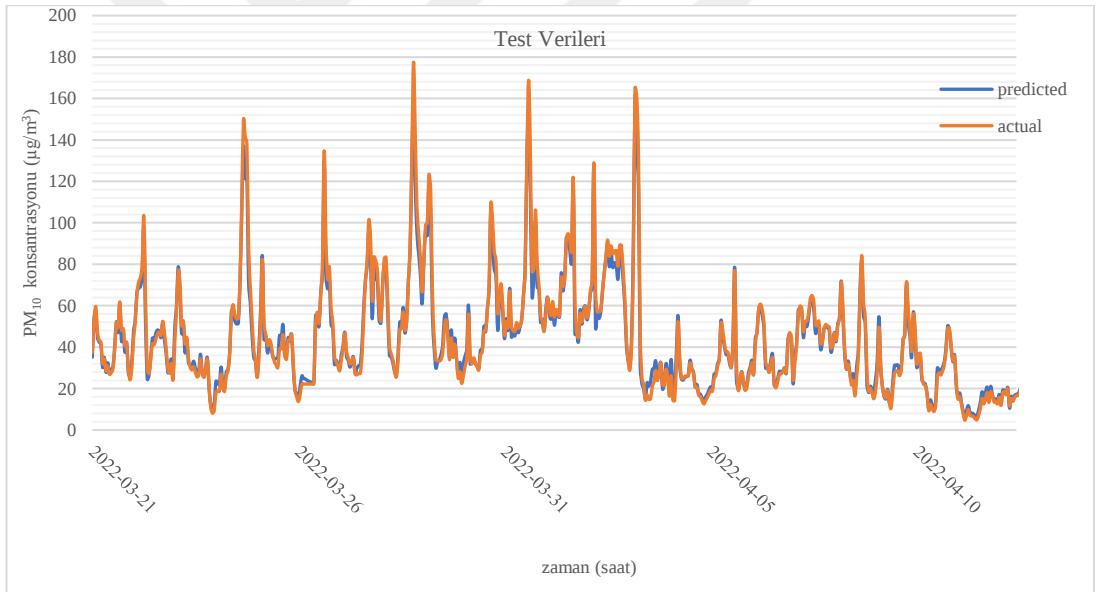
4.3.4. PM₁₀ parametresi için model sonuçlarının değerlendirilmesi

Çalışmada kullandığımız LSTM algoritması ile PM₁₀'a ait, 21.03.2022-8.10.2022 tarihleri arasında, 200 günlük, saatlik veri seti kullanılarak test edilmiş olup test sonucunda algoritmanın güvenilirliği mukayese edilmiş olup Şekil 4.16. , Şekil 4.17.'de yer alan grafikte gösterilmiştir. Grafikler incelendiğinde eğitim ve test verilerinin güvenilirlik düzeyi yüksek olarak tespit edilmiştir.

Sonuç olarak diğer kirlenici parametrelerdeki gibi PM₁₀'nun tahmini için de, kullanılan LSTM algoritması ve hiperparametre optimizasyonunun yüksek tahmin oranına imkân sağladığı görülmüştür. Buna bağlı olarak PM₁₀ parametresi için 10 kez çalıştırılmıştır ve hata metrikleri ortaya çıkarılmış, deney sırasına göre Tablo 4.11.'de gösterilmiştir. Predicted tahmin verilerini, actual gerçek verileri gösterir.



Şekil 4.16. PM₁₀ konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (eğitim seti)



Şekil 4.17. PM₁₀ konsantrasyonları için tahmin ve gerçek değerlerin karşılaştırılması (test seti)

Tablo 4.11. PM₁₀ parametresi için en iyi arama algoritması performans sonuçları

PM ₁₀		Test Seti					Eğitim Seti				
Deney No	Eksik Veri Doldurma Yöntemi	rmse ₁	mse ₁	mae ₁	mape ₁	r ² ₁	rmse ₂	mse ₂	mae ₂	mape ₂	r ² ₂
1	Ortalama Yöntemi	0,00003	0,006	0,0037	0,074	0,98	0,000033	0,0057	0,00369	0,074	0,981
2		0,00004	0,006	0,0039	0,072	0,98	0,000041	0,0064	0,00392	0,072	0,976
3		0,00003	0,005	0,0031	0,057	0,98	0,000027	0,0052	0,00305	0,057	0,984
4		0,00003	0,005	0,0036	0,072	0,98	0,000029	0,0054	0,00355	0,072	0,983
5		4,00E-05	0,006	0,0043	0,087	0,98	3,80E-05	0,0062	0,00427	0,087	0,978
6		0,00003	0,005	0,0031	0,053	0,98	0,00003	0,0055	0,00307	0,053	0,982
7		0,00003	0,006	0,004	0,083	0,98	0,000034	0,0059	0,004	0,083	0,98
8		0,00003	0,006	0,0041	0,086	0,98	0,000034	0,0059	0,00409	0,086	0,98
9		0,00003	0,005	0,0033	0,067	0,98	0,000027	0,0052	0,00334	0,067	0,984
10		0,00004	0,006	0,0044	0,092	0,98	0,000039	0,0063	0,00441	0,092	0,977
1	Komşu Yöntemi	0,00001	0,003	0,0017	0,026	0,99	0,000011	0,0034	0,00171	0,026	0,993
2		0,00001	0,004	0,0024	0,046	0,99	0,000013	0,0037	0,00236	0,046	0,992
3		0,00002	0,004	0,0019	0,028	0,99	0,000018	0,0043	0,00191	0,028	0,989
4		0,00002	0,004	0,0025	0,047	0,99	0,000019	0,0044	0,00249	0,047	0,989
5		4,00E-05	0,006	0,0045	0,095	0,98	4,20E-05	0,0065	0,00453	0,095	0,975
6		0,00001	0,004	0,0019	0,026	0,99	0,000015	0,0039	0,00189	0,026	0,991
7		0,00002	0,005	0,0035	0,071	0,99	0,000022	0,0047	0,00347	0,071	0,987
8		0,00001	0,004	0,0021	0,039	0,99	0,000013	0,0036	0,00208	0,039	0,992
9		0,00002	0,004	0,0024	0,046	0,99	0,000016	0,004	0,00245	0,046	0,991
10		0,00001	0,003	0,0014	0,021	0,99	0,000011	0,0034	0,00144	0,021	0,993

5. SONUÇ VE ÖNERİLER

Son yıllarda artan hava kirliliğinin kontrol altına alınması, insan ve çevre sağlığı açısından oldukça önemlidir. Hava kirliliği astım, bronşit, üst solunum yolu enfeksiyonları gibi birçok hastalığa ve bazı durumlarda yaşamı tehdit eden etkilere neden olmaktadır. Ayrıca hava kirliliğinin, doğal denge üzerinde derin ve rahatsız edici etkileri vardır.

Bu çalışma, hava kirliliğinin olumsuz sonuçlarına dikkat çekmek için, ayrıca havadaki bileşenlerin korunmasını sağlamak ve buna dair alınacak önlemlere yol gösterebilmek üzere yapılmıştır. Çevre bilimleri ve bilgisayar tabanlı bilim dallarını bir arada kullanarak, detaylı disiplinler arası bir çalışma tasarlanmıştır.

Sakarya ili için hava kalitesi tahmini yapılmıştır. Hava kalitesi verileri Çevre ve Şehircilik Bakanlığı'nın resmi internet sitesi www.havaizleme.gov.tr'den alınmıştır. Çalışma, Ocak 2020'den Eylül 2022'ye kadar olan dönemi kapsamaktadır. Hava kalitesi verileri, Ulusal Hava Kalitesi İzleme Ağı kapsamında kurulan ölçüm istasyonları tarafından toplanmıştır. Çalışmada Sakarya ilinde bulunan dört adet hava kalitesi ölçüm istasyonundan biri kullanılmıştır. Çalışmada kullanılan istasyon Sakarya Caddesi üzerinde Meslek Yüksekokulu yanında bulunan Sakarya Merkez istasyonudur. En yakın konuta 3,5 metre, yola 0,5 metre mesafededir. Bu istasyon, trafik kaynaklı hava kirliliğini ölçmek için kurulmuş olup, Mart 2013'ten beri sürekli olarak ölçüm yapmaktadır. Çalışmada hava kalitesinin ileriye dönük tahmini için PM_{10} , NO_2 , NO_x ve CO olmak üzere dört parametre kullanılmıştır. Merkez istasyonun dört parametresinin her biri için Ocak 2020 ve Eylül 2022 kullanılmıştır. Belirtilen tarih aralığında her parametre için 24.282 olmak üzere toplamda 97.128 veri olmalıdır. Ancak sensör hataları ve benzeri durumlardan dolayı kısmi veri eksiklikleri gözlemlenmiştir.

Veri setinde Python kullanılarak veri temizliği yapılmıştır. Veri setindeki eksik değerler ortalama ve en yakın komşu (kNN) yöntemleri ile tamamlanarak iki ayrı veri seti oluşturulmuştur. Verilerin ön işlenmesinden sonra veri setindeki eksik veriler tespit edilmiştir. Daha sonra veri setindeki değerler gözlemlenmiş ve kirlileti

parametrelerin deęerlerinin en yoęun olduęu dnemler ile en dřük olduęu dnemler ayrı ayrı deęerlendirilmiřtir. MSE sonuları dikkate alınarak ortalama ve kNN yntemleri ile doldurulan veri kmeleri iin bir keman grafięi oluřturulmuřtur. Bu grafięe dayanarak, ortalama ve kNN yntemleri arasında bir karřılařtırma yapılmıřtır. Elde edilen MSE sonularına gre komřu arama algoritmasının NO_x dıřında en iyi olduęu ortaya ıkmıřtır. NO_x parametresi ise ortalama yntemi ile doldurulduęunda daha iyi sonular vermiřtir. Bu nedenle ikinci ařamada NO_x iin ortalama ynteminin ve NO₂, PM₁₀ ve CO veri setlerindeki eksik deęerlerin doldurulması iin kNN algoritmasının kullanılmasına karar verilmiřtir. Veri n iřlemesinin ardından, her bir kirletici parametresi iin toplam 24.282 veri seti elde edilmiř ve veri setinin %80'i (19.423 rnek) eęitim, kalan %20'si (4.859 rnek) test iin ayrılmıřtır. Veri seti incelendięinde PM₁₀ ile NO₂, NO_x ve CO kirleticileri arasında gl bir iliřki gzlenmiřtir. alıřmada tahmin analizi iin derin ęrenme yntemlerinden biri olan LSTM (Long Short Term Memory) tercih edilmiřtir. alıřmada kullanılan LSTM'nin tahmin performansını artırmak iin hiperparametre optimizasyonu uygulanmıřtır. Hiperparametre optimizasyonu ile hava kalitesinin her bir parametresi iin doęru hiperparametre seimi saęlanmaktadır. Hiperparametre arama algoritmalarından, rastgele arama, bayesian arama ve hyberband arama algoritmaları kullanıřmıřtır. Her kirletici iin hiperparametre optimizasyonu iin RMSE ve MSE deęeri hesaplanmıř ve elde edilen sonular incelenerek hiperparametre seimi yapılmıřtır. NO_x iin Hyberband, NO₂ iin Bayesian, PM₁₀, CO iin hiperparametre uygun bulunmuřtur. En iyi arama algoritması tarafından belirlenen hiperparametre deęerleri yeniden eęitim iin kullanılmıřtır. Bu ařamadan sonra oluřturulan modelin doęruluęu, tahmin ve test veri setleri kullanılarak deęerlendirilmiřtir. Tahmin ve test iin oluřturulan metrikler karřılařtırıldıęında modelin yksek bir performans sergiledięi grlmřtir. Tahmin iin kullanılan LSTM yntemi, hava kalitesinin ileriye dnk tahmininde tercih edilen bir yntemdir.

Sonu olarak kullanılan modelin tahmin doęruluęu yksek bir yntem olduęu grlmřtir. Bu baęlamda, LSTM Algoritmasının hava kalitesi verilerini tahmin etmek iin yaygın olarak kullanılabileceęini gstermiřtir. Ayrıca hava kirlilięine dikkat ekmek iin alıřmanın doęru bir yol olduęu dřnlmřtir. Tahmin adımları geniřletilip geliřtirilirse daha verimli sonular retilbileceęi dřnlmektedir.

KAYNAKLAR

- [1] “Hava Kalitesi İndeksi – Çevre Portal”. <https://www.cevreportal.com/hava-kalitesi-indeksi/> (erişim 14 Aralık 2022).
- [2] “AQI Basics | AirNow.gov”. <https://www.airnow.gov/aqi/aqi-basics> (erişim 14 Aralık 2022).
- [3] “Mevzuat Türkiye Standartları | Hava Kalitesi İzleme Projesi - İstanbul Büyükşehir Belediyesi”. <https://havakalitesi.ibb.gov.tr/Icerik/mevzuat> (erişim 14 Aralık 2022).
- [4] “SİM (Sürekli İzleme Merkezi) | T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı”. <https://sim.csb.gov.tr/> (erişim 14 Aralık 2022).
- [5] H. G. Reichle Jr vd., “Middle and upper tropospheric carbon monoxide mixing ratios as measured by a satellite-borne remote sensor during November 1981”, J. Geophys. Res. Atmospheres, c. 91, sy D10, ss. 10865-10887, 1986.
- [6] A. B. Ishak, M. B. Daoud, ve A. Trabelsi, “Ozone concentration forecasting using statistical learning approaches”, J Mater Env. Sci, c. 8, sy 12, ss. 4532-4543, 2017.
- [7] S. Yun ve C. Yoo, “The effects of spring and winter blocking on PM10 concentration in Korea”, Atmosphere, c. 10, sy 7, s. 410, 2019.
- [8] K. Ramgolam vd., “Size-partitioning of an urban aerosol to identify particle determinants involved in the proinflammatory response induced in airway epithelial cells”, Part. Fibre Toxicol., c. 6, sy 1, ss. 1-12, 2009.
- [9] K. Maduna ve V. Tomašić, “Air pollution engineering”, Phys. Sci. Rev., c. 2, sy 12, 2017.
- [10] A. J. Ghio, M. S. Carraway, ve M. C. Madden, “Composition of air pollution particles and oxidative stress in cells, tissues, and living systems”, J. Toxicol. Environ. Health Part B, c. 15, sy 1, ss. 1-21, 2012.
- [11] D. Seng, Q. Zhang, X. Zhang, G. Chen, ve X. Chen, “Spatiotemporal prediction of air quality based on LSTM neural network”, Alex. Eng. J., c. 60, sy 2, 2021.
- [12] X. Li vd., “Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation”, Environ. Pollut., c. 231, ss. 997-1004, 2017.
- [13] C. Janiesch, P. Zschech, ve K. Heinrich, “Machine learning and deep learning”, Electron. Mark., c. 31, sy 3, ss. 685-695, 2021.

- [14] A. Aykut, "Deep Learning Nedir? • Turhost Blog", Turhost Blog, 26 Ekim 2020. <https://blog.turhost.com/deep-learning-nedir/> (erişim 15 Aralık 2022).
- [15] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, ve F. E. Alsaadi, "A survey of deep neural network architectures and their applications", *Neurocomputing*, c. 234, ss. 11-26, 2017.
- [16] Y. Bengio, "Learning Deep Architectures for AI", *Found. Trends® Mach. Learn.*, c. 2, sy 1, ss. 1-127, Kas. 2009, doi: 10.1561/22000000006.
- [17] P. Ongsulee, "Artificial intelligence, machine learning and deep learning", içinde 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE), Bangkok: IEEE, Kas. 2017, ss. 1-6. doi: 10.1109/ICTKE.2017.8259629.
- [18] "Makine öğrenimi Derin öğrenme Yapay zeka Algoritması, Makine Öğrenmesi, metin, logo, İş png | PNGWing". <https://www.pngwing.com/tr/free-png-tgvpc> (erişim 18 Aralık 2022).
- [19] M. T. Lei, J. Monjardino, L. Mendes, D. Gonçalves, ve F. Ferreira, "Macao air quality forecast using statistical methods", *Air Qual. Atmosphere Health*, c. 12, sy 9, ss. 1049-1057, Eyl. 2019, doi: 10.1007/s11869-019-00721-9.
- [20] J.-H. Jeong vd., "A novel statistical-dynamical method for a seasonal forecast of particular matter in South Korea", *Sci. Total Environ.*, c. 848, s. 157699, 2022.
- [21] E. Marinov, D. Petrova-Antonova, ve S. Malinov, "Time Series Forecasting of Air Quality: A Case Study of Sofia City", *Atmosphere*, c. 13, sy 5, s. 788, 2022.
- [22] Z. Zhao, J. Wu, F. Cai, S. Zhang, ve Y.-G. Wang, "A statistical learning framework for spatial-temporal feature selection and application to air quality index forecasting", *Ecol. Indic.*, c. 144, s. 109416, 2022.
- [23] M. Oprea, S. F. Mihalache, ve M. Popescu, "Applying Artificial Neural Networks to Short-Term PM2.5 Forecasting Modeling", içinde *Artificial Intelligence Applications and Innovations*, L. Iliadis ve I. Maglogiannis, Ed., içinde *IFIP Advances in Information and Communication Technology*, vol. 475. Cham: Springer International Publishing, 2016, ss. 204-211. doi: 10.1007/978-3-319-44944-9_18.
- [24] L. Montalvo, D. Fosca, D. Paredes, M. Abarca, C. Saito, ve E. Villanueva, "An Air Quality Monitoring and Forecasting System for Lima City With Low-Cost Sensors and Artificial Intelligence Models", *Front. Sustain. Cities*, c. 4, 2022, Erişim: 26 Mart 2023. [Çevrimiçi]. Erişim adresi: <https://www.frontiersin.org/articles/10.3389/frsc.2022.849762>

- [25] K. D. Karatzas, G. Papadourakis, ve I. Kyriakidis, "Understanding and Forecasting Air Pollution with the Aid of Artificial Intelligence Methods in Athens, Greece", içinde Tools and Applications with Artificial Intelligence, C. Koutsojannis ve S. Sirmakessis, Ed., içinde Studies in Computational Intelligence, vol. 166. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, ss. 37-50. doi: 10.1007/978-3-540-88069-1_4.
- [26] D. Mishra ve P. Goyal, "Development of artificial intelligence based NO₂ forecasting models at Taj Mahal, Agra", Atmospheric Pollut. Res., c. 6, sy 1, ss. 99-106, Oca. 2015, doi: 10.5094/APR.2015.012.
- [27] A. Murad, F. A. Kraemer, K. Bach, ve G. Taylor, "Probabilistic Deep Learning to Quantify Uncertainty in Air Quality Forecasting", Sensors, c. 21, sy 23, Art. sy 23, Oca. 2021, doi: 10.3390/s21238009.
- [28] C.-Y. Lin, Y.-S. Chang, ve S. Abimannan, "Ensemble multifeatured deep learning models for air quality forecasting", Atmospheric Pollut. Res., c. 12, sy 5, s. 101045, May. 2021, doi: 10.1016/j.apr.2021.03.008.
- [29] M. Alghieth, R. Alawaji, S. H. Saleh, ve S. Alharbi, "Air Pollution Forecasting Using Deep Learning", Int. J. Online Biomed. Eng. IJOE, c. 17, sy 14, Art. sy 14, Ara. 2021, doi: 10.3991/ijoe.v17i14.27369.
- [30] S. Du, T. Li, Y. Yang, ve S.-J. Horng, "Deep Air Quality Forecasting Using Hybrid Deep Learning Framework", IEEE Trans. Knowl. Data Eng., c. 33, sy 6, ss. 2412-2424, Haz. 2021, doi: 10.1109/TKDE.2019.2954510.
- [31] F. M. Awan, R. Minerva, ve N. Crespi, "Improving Road Traffic Forecasting Using Air Pollution and Atmospheric Data: Experiments Based on LSTM Recurrent Neural Networks", Sensors, c. 20, sy 13, Art. sy 13, Oca. 2020, doi: 10.3390/s20133749.
- [32] G. Mani ve R. Volety, "A comparative analysis of LSTM and ARIMA for enhanced real-time air pollutant levels forecasting using sensor fusion with ground station data", Cogent Eng., Tem. 2021, Erişim: 05 Nisan 2023. [Çevrimiçi]. Erişim adresi: <https://www.tandfonline.com/doi/abs/10.1080/23311916.2021.1936886>
- [33] Y.-T. Tsai, Y.-R. Zeng, ve Y.-S. Chang, "Air pollution forecasting using RNN with LSTM", içinde 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), IEEE, 2018, ss. 1074-1079.
- [34] "SİM (Sürekli İzleme Merkezi) | T.C ÇEVRE, ŞEHİRCİLİK VE İKLİM DEĞİŞİKLİĞİ BAKANLIĞI". <https://www.havaizleme.gov.tr/> (erişim 14 Aralık 2022).

- [35] H. Quan, J. Wang, B. Li, X. Du, K. Liu, ve L. Li, “Characterizing Python Method Evolution with PyMevol: An Essential Step Towards Enabling Reliable Software Systems”, içinde 2022 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), IEEE, 2022, ss. 81-86.
- [36] “Python Programlama Nedir? Ne İşe Yarar? Nerelerde Kullanılır? | Teknoloji Projeleri”. <https://teknolojiprojeleri.com/programlar/python-nedir-ne-ise-yarar-nerelerde-kullanilir> (erişim 18 Aralık 2022).
- [37] J. Han, M. Kamber, ve J. Pei, “4 - Data Warehousing and Online Analytical Processing”, içinde Data Mining (Third Edition), J. Han, M. Kamber, ve J. Pei, Ed., içinde The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann, 2012, ss. 125-185. doi: 10.1016/B978-0-12-381479-1.00004-6.
- [38] A. S. Bozkır, E. Sezer, ve G. Bilge, “Öğrenci Seçme Sınavında (Öss) Öğrenci Başarımını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti Determination Of The Factors Influencing Student’s Success In Student Selection Examination (Oss) Via Data Mining Techniques”.
- [39] S. B. Kotsiantis, D. Kanellopoulos, ve P. E. Pintelas, “Data Preprocessing for Supervised Learning”, c. 1, sy 1, 2006.
- [40] M. Castelli, F. M. Clemente, A. Popovič, S. Silva, ve L. Vanneschi, “A Machine Learning Approach to Predict Air Quality in California”, Complexity, c. 2020, ss. 1-23, Ağu. 2020, doi: 10.1155/2020/8049504.
- [41] D. KILINÇ, “Python ile Veri Ön İşlemeye Dalış”, Medium, 05 Kasım 2019. <https://medium.com/@denizkilinc/python-ile-veri> (erişim 18 Aralık 2022).
- [42] S. Gunasekar, G. J. R. Kumar, ve G. P. Agbulu, “Air Quality Predictions in Urban Areas Using Hybrid ARIMA and Metaheuristic LSTM.”, Comput Syst Sci Eng, c. 43, sy 3, ss. 1271-1284, 2022.
- [43] “Understanding LSTM Networks -- colah’s blog”. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> (erişim 18 Aralık 2022).
- [44] S. Hochreiter ve J. Schmidhuber, “Long short-term memory”, Neural Comput., c. 9, sy 8, ss. 1735-1780, 1997.
- [45] H. Chen, M. Guan, ve H. Li, “Air Quality Prediction Based on Integrated Dual LSTM Model”, IEEE Access, c. 9, ss. 93285-93297, 2021.
- [46] İ. DÖLEK, “LSTM”, Medium, 10 Ocak 2021. <https://ishakdolek.medium.com/lstm-d2c281b92aac> (erişim 18 Aralık 2022).
- [47] D. Sculley, J. Snoek, A. Rahimi, ve A. Wiltschko, “On Pace, Progress, And Empirical Rigor”, 2018.

- [48] S. Altun ve M. F. Talu, “Derin Sinir Ağları için Hiperparametre Metodlarının ve Kitlerinin İncelenmesi”, DÜMF Mühendis. Derg., Ara. 2020, doi: 10.24012/dumf.767700.
- [49] R. Khandelwal, “Hyperparameter Optimization with KerasTuner”, Geek Culture, 26 Temmuz 2021. <https://medium.com/geekculture/hyperparameter-optimization-with-kerastuner-cba8af446add> (erişim 18 Aralık 2022).
- [50] T. Ye ve S. Kalyanaraman, “A recursive random search algorithm for large-scale network parameter configuration”, içinde Proceedings of the 2003 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, içinde SIGMETRICS '03. New York, NY, USA: Association for Computing Machinery, Haz. 2003, ss. 196-205. doi: 10.1145/781027.781052.
- [51] Z. Guo, S. Wu, M. Ohno, ve R. Yoshida, “Bayesian Algorithm for Retrosynthesis”, J. Chem. Inf. Model., c. 60, sy 10, ss. 4474-4486, Eki. 2020, doi: 10.1021/acs.jcim.0c00320.



ÖZGEÇMİŞ

Ad-Soyad : Ayşegül ATALI

ÖĞRENİM DURUMU:

- Lisans** : 2011, Sakarya Üniversitesi, Mühendislik Fakültesi, Çevre Mühendisliği
- Yükseklisans** : 2023, Sakarya Üniversitesi, Çevre Mühendisliği

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2011-2012 yılları arasında KARÇED firmasında mühendis olarak çalıştı.
- 2013-Halen İŞKUR'da İş ve Meslek Danışmanı olarak çalışmakta.

TEZDEN TÜRETİLEN ESERLER:

- Atalı A., Eren B., Erden C. ve Atalı G. (2022, 21-23, Ekim). LSTM Derin Öğrenme Yaklaşımı ile Hava Kalitesi Verilerinin Tahmini: Sakarya Örneği. *International Symposium on Innovative Technologies in Engineering and Science*, Bursa, Turkey.

DİĞER ESERLER: