

İç Ortam Hava Kalitesinin Tahmini için Makine Öğrenmesi Tabanlı Modellerin Karşılaştırılması

Comparative Analysis of Machine Learning-Based Models for Indoor Air Quality Prediction

Nurullah ÖZTÜRK^{1*}, Ramazan ALDEMİR²

Öz

Hava kirliliği, günümüzde çevre ve halk sağlığı açısından önemli bir tehdit oluşturmaktadır ve özellikle iç ortamlarda maruziyetin sürekli olması riskleri artırmaktadır. Bu nedenle, iç ortam hava kalitesinin güvenilir biçimde izlenmesi ve tahmin edilmesi büyük önem taşımaktadır. Bu çalışmada, iç ortam hava kirliliğine neden olan CO, NH₃, NO₂ ve PM_{2.5} kirleticilerinin tahmini amacıyla Karar Ağaçları, Rastgele Orman, Destek Vektör Makineleri, Doğrusal Regresyon, Yapay Sinir Ağları ve bu modellerin ağırlıklandırılmasıyla oluşturulan hibrit model karşılaştırmalı olarak değerlendirilmiştir. Çalışmada, Mendeley kütüphanesinde yer alan "Dataset of Indoor Air Pollutants using Low-Cost Sensor" veri seti kullanılmıştır. Veriler, kronolojik yapı korunarak %70 eğitim ve %30 test olacak şekilde ayrılmış; tüm modellerin hiperparametreleri yalnızca eğitim verisi üzerinde uygulanan 5-katlı çapraz doğrulama ve sistematik grid arama yöntemiyle optimize edilmiştir. Model performansları *RMSE*, *MSE*, *MAE* ve *R²* metrikleri ile değerlendirilmiş, yöntemler arasındaki farkların istatistiksel anlamlılığı Friedman ve Wilcoxon testleri ile analiz edilmiştir. Sonuçlar, CO değişkeni için hibrit modelin en dengeli ve başarılı yöntemi sunduğunu, NH₃, NO₂ ve özellikle PM_{2.5} değişkenlerinde ise tüm yöntemlerin test seti üzerinde sınırlı başarımlar sergilediğini göstermektedir. Bununla birlikte, hibrit modelin tüm kirleticiler için tekil modellere kıyasla daha tutarlı hata dağılımları ürettiği ve istatistiksel olarak anlamlı performans farkları sağladığı ortaya konmuştur.

Anahtar Kelimeler: İç ortam hava kalitesi, Makine öğrenmesi, Hibrit model, Düşük maliyetli sensörler.

Abstract

This Air pollution presents a considerable challenge to both environmental preservation and public well-being, especially within indoor settings where prolonged exposure amplifies the associated hazards. Consequently, reliable monitoring and forecasting of indoor air quality are of great importance. In this study, Decision Trees, Random Forest, Support Vector Machines, Linear Regression, Artificial Neural Networks, and a hybrid model developed through the weighted integration of these methods, to predict indoor air pollutants: CO, NH₃, NO₂, and PM_{2.5}. The experiments were performed utilising the "Dataset of Indoor Air Pollutants using Low-Cost Sensor," obtained from the Mendeley data repository. The dataset was chronologically divided into 70% to training and 30% to testing. Hyperparameter optimization of all models was conducted solely on the training data, employing 5-fold cross-validation and a systematic grid search. Model performance was assessed through *RMSE*, *MSE*, *MAE*, and *R²* metrics; furthermore, the statistical significance of performance disparities among the methodologies was examined via the Friedman and Wilcoxon tests. The findings suggest that the hybrid model demonstrates the most balanced and robust performance for the CO pollutant. Conversely, for NH₃, NO₂, and especially PM_{2.5}, all methods displayed restricted predictive capabilities on the test set. Nevertheless, the hybrid model conducts more stable error distributions compared to individual models, and achieves statistically significant performance distinctions across all pollutants.

Keywords: Indoor air quality, Machine learning, Hybrid model, Low-Cost sensors.

¹Kayseri Üniversitesi, Mühendislik Mimarlık ve Tasarım Fakültesi, Bilgisayar Mühendisliği Bölümü, Kayseri, Türkiye

²Kayseri Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Elektronik ve Otomasyon Bölümü, Kayseri, Türkiye

*Sorumlu Yazar/Corresponding Author: nurullahozturk@kayseri.edu.tr

1. Giriş

Hava kirliliği, endüstriyel faaliyetlerin artışı, fosil yakıt kullanımı ve kentleşme gibi etkenlerle birlikte günümüzde küresel ölçekte önemli bir çevresel ve halk sağlığı sorunu haline gelmiştir (Adekoya vd., 2021; Liu vd., 2020). Kirleticiler arasında karbon monoksit (CO), nitrojen oksitler (NO_x), metan (CH₄) ve partikül maddeler (PM), hem doğrudan hava kalitesini bozmakta hem de sera etkisiyle iklim değişikliğine katkı sağlamaktadır (IPCC 2014; Yaman, 2024). Kirleticilerin atmosferdeki konsantrasyonu; sıcaklık, basınç, nem, rüzgâr ve atmosferik karışım yüksekliği gibi meteorolojik değişkenlerle şekillenmekte, bu durum hava kalitesinin değerlendirilmesini çok boyutlu bir süreç haline getirmektedir (Srivastava vd., 2014; Lazaridis, 2011).

Genel hava kalitesi üzerine yapılan çalışmalar büyük ölçüde dış ortam kirliliğine odaklansa da, insanların günlük yaşamlarının yaklaşık %90'ını kapalı mekanlarda geçirmesi (Chatzidiakou vd., 2012), iç ortam hava kalitesini en az dış ortam kadar önemli kılmaktadır. Ev, okul, ofis gibi kapalı alanlardaki kötü hava kalitesi; baş ağrısı, yorgunluk, konsantrasyon kaybı gibi kısa vadeli etkilerin yanı sıra, uzun vadede kardiyovasküler hastalıklar, solunum yolu rahatsızlıkları ve kanser gibi ciddi sağlık problemleriyle ilişkilendirilmektedir (Bakó-Biró vd., 2012; WHO 2018; Rajagopalan vd., 2018).

Covid-19 pandemi döneminde, kapalı alanların hava kalitesinin enfeksiyon yayılımı üzerindeki etkisini açıkça ortaya koymuş; bu alandaki bilimsel araştırmaları hızlandırmıştır (El-Tanbouly vd., 2021; Megahed vd., 2021; Mata vd., 2022). Özellikle çocuklar, yaşlılar ve bağışıklık sistemi zayıf bireyler, iç ortam kirliliğine karşı daha duyarlıdır (Amoatey vd., 2018; Peters vd., 2019). Ülkemizde, öğrencilerin okul ortamlarında uzun süreler geçirdiği göz önünde bulundurulduğunda, sınıf havasının yetersiz kalması; öğrencilerin dikkat, öğrenme performansı ve sağlığı üzerinde olumsuz etkiler doğurabilmektedir (Amoatey vd., 2018). Ayrıca son dönemde yapılan çalışmalar, kirleticilere uzun süreli maruz kalmanın Alzheimer hastalığı ve diğer nörolojik hastalıkların riskini artırdığını göstermiştir (Hahad vd., 2020).

Bu bağlamda, iç ortam hava kalitesinin izlenmesi, kirletici kaynakların belirlenmesi ve ileri modelleme yöntemleri ile tahmin edilmesi hem bireysel hem toplumsal sağlık açısından kritik öneme sahiptir. İç ortam hava kalite sistemleri çoğunlukla verileri anlık olarak sunarken, son yıllarda makine öğrenmesi ve yapay sinir ağları gibi ileri veri işleme yöntemlerinin entegrasyonu ile daha doğru tahmin ve değerlendirme yapılabilen uygulamalar geliştirilmiştir. İç ortam hava kalitesinin güvenilir bir şekilde izlenmesi halk sağlığı açısından kritik öneme sahiptir. Ancak, mevcut sistemlerin yüksek maliyetli olması, karmaşık kurulum gereksinimleri ve genellikle sabit noktalarda sınırlı kapsama alanı sunması uygulamada çeşitli kısıtlamalara neden olmaktadır. Son yıllarda, düşük maliyetli sensörler, nesnelerin interneti (IoT) ve kablosuz sensör ağları (WSN) teknolojilerindeki gelişmeler sayesinde

daha yaygın, taşınabilir ve ölçeklenebilir çözümler sunulmaktadır. Bu sistemler; sıcaklık, nem ve CO₂ gibi temel parametrelerin yanı sıra, sistemin amacına bağlı olarak farklı kirletici türlerinin ölçümünü de yapmaktadır.

Literatürde, iç ortam hava kalitesini izlenmesi ve tahminine yönelik makine öğrenmesi, derin öğrenme ve nesnelerin interneti tabanlı çok sayıda çalışma vardır (Yadav vd., 2024; Chaturvedi, 2024; Aram vd., 2023). Pietraru ark. (2024) yaptıkları çalışmada, otomatik havalandırma sistemi bulunmayan iç ortam yaşam alanları için düşük maliyetli ve kullanıcı dostu bir iç hava kalitesi izleme sistemi geliştirmişlerdir. Çalışmada elde edilen verilerle gerçek zamanlı kirlilik tahmini ve görsel analizler sunmuşlardır. Othman ark. (2024), CO₂, sıcaklık, nem, VOC'ler, basınç ve ışık gibi parametreleri düşük maliyetli sensörler ile iç ortam hava kalite değerlerini anlık değerlendiren sistem tasarımı yapmışlardır. Sonuçları web tabanlı olarak kullanıcıya aktarmışlardır. Berkani ark. (2023), iç ortam hava kalitesini izlemek ve günlük yaşam aktivitesini tanımak amacıyla derin öğrenme tabanlı bir sistem önermişlerdir. Altı farklı sensör verisiyle yüksek doğruluk oranları elde etmişlerdir. Zhang ve ark. (2022), gerçek zamanlı PM_{2.5} tahmini için görüntü verilerine dayalı dikkat mekanizmalı LSTM ve regresyon modelinden oluşan yenilikçi bir hibrit model geliştirmişlerdir. Saravanan ve Kumar (2023), hava kirliliğinin tespiti ve izlenmesi amacıyla IoT tabanlı analiz ve sinir ağı adaptasyonunu kullanan düşük maliyetli bir sistem önermişlerdir. Sensör ağlarıyla geniş alanlarda izleme yapılabilen sistemde, çift yönlü geri beslemeli sinir ağı kullanılarak zamansal bağımlılıklarla birlikte hava kirliliği tahmini yapmışlardır.

Bu çalışmada, literatürdeki çalışmalardan farklı olarak, iç ortam hava kalitesi tahmininin de makine öğrenmesi yöntemlerinin yanı sıra eğitim verisi üzerindeki hata performansına duyarlı ağırlıklandırma stratejisine sahip bir hibrit model uygulanmaktadır. Ayrıca tüm modeller için yalnızca eğitim verisi kullanılarak gerçekleştirilen sistematik hiperparametre optimizasyonu uygulanmaktadır. Ek olarak elde edilen sonuçların istatistiksel olarak anlamlılığını değerlendiren Friedman ve Wilcoxon testleri bütünlük bir çerçevede sunmaktadır. Bu yönleriyle çalışma, iç ortam hava kalitesi tahmininde model seçiminin kirletici türüne bağlı değişimini ortaya koymaktadır.

Bu çalışma kapsamında Mendeley kütüphanesinde yer alan "Dataset of Indoor Air Pollutants using Low-Cost Sensor" veri seti üzerinde farklı makine öğrenmesi modelleri uygulanmış ve iç ortam hava kalite analizi ile kirlilik seviyelerinin tahmini gerçekleştirilmiştir. Bu doğrultuda, iç ortam hava kalitesini güvenilir biçimde tahmin edebilen etkili ve uygulanabilir modellerin geliştirilmesi amaçlanmaktadır.

2. Materyal ve Metot

Düşük maliyetli sensörler kullanılarak iç ortam hava kirleticilerine ait verileri içeren Mendeley veri seti ve iç ortam hava kirliliği seviyelerinin tahmininde kullanılan yöntemler bu bölümde açıklanmaktadır. Çalışmada Karar Ağaçları (DT), Rastgele Orman (RF), Destek Vektör Makinesi (SVM), Lineer Regresyon (LR), Yapay Sinir Ağları (ANN) ve model performansına duyarlı ağırlıklandırma stratejisine dayalı hibrit yöntem uygulanmaktadır. Bu çalışmada, farklı makine öğrenimi modellerinin iç ortam hava kirliliği seviyelerini tahmin etmedeki başarımları karşılaştırmalı olarak değerlendirilmiştir.

2.1. Veri Seti

Mendeley veri seti, 2020 Kasım - 2022 Temmuz tarihleri arasında Hindistan'ın Pune şehrinde, tüm mevsimleri kapsayacak şekilde düşük maliyetli sensörler kullanılarak toplanmış 173.468 adet iç ortam hava kalitesi ölçüm verisi yer almaktadır. Veri setinde, solunabilir partikül madde (PM2.5) (ug/m³), azot dioksit (NO₂), amonyak (NH₃), karbon monoksit (CO, ppm), Ozon (O₃, ppb) ile sıcaklık (°C), nem (RH), basınç (hPa) gibi çeşitli kirleticilere ve çevresel faktörlere ait ölçümler yer almaktadır. Kaydedilen parametreler ve ilgili ölçüm birimleri Tablo 1'de sunulmuştur. Bu veriler GP2Y1010AU0F, MICS-6814, MQ131, BME280 sensörleri kullanılarak elde edilmiştir (Sonawani vd., 2022).

Tablo 1. Veri setindeki bulunan parametreler ve ölçü birimleri (Sonawani vd., 2022).

Ölçülen değerler	Ölçü Birimi
NH ₃	ppm
NO ₂	ppm
CO	ppm
PM2.5	ug/m ³
Temp	°C
Pressure	hPa
Humidity	RH
O ₃	ppb
Date	yyyy-MM-dd HH:mm:ss.SSS

2.2. Makine Öğrenmesi Yöntemleri

Bu bölümde hava kalitesi tahmini amacıyla kullanılan Rastgele Orman (RF), Karar Ağaçları (DT), Destek Vektör Makinesi (SVM), Lineer Regresyon (LR) ve Yapay Sinir Ağları (ANN) gibi

makine öğrenmesi yöntemleri ve bu modellerin performansına dayalı olarak uygulanan hibrit yöntem ayrıntılı şekilde açıklanmaktadır

2.2.1. Rastgele Orman Algoritması

Karar ağaçları temeline dayanan rastgele orman (Random Forest, RF) algoritması, makine öğrenmesi alanında sınıflandırma ve regresyon problemlerinde sıklıkla tercih edilen bir topluluk öğrenme yöntemidir. Bu algoritma, eğitim verisinden rastgele seçilen öznelik alt kümeleri kullanarak çok sayıda karar ağacı oluşturur. Her bir ağacının ürettiği tahminler doğrultusunda nihai karar, sınıflandırma problemlerinde çoğunluk oylaması, regresyon problemlerinde ise tahminlerin ortalama alınarak belirlenmektedir. Rastgele öznelik seçimi ve çoklu ağaç yapısı eğiliminden kaynaklanabilecek aşırı öğrenme (overfitting) riskini önemli ölçüde azaltarak genelleme başarımını artırmaktadır (Aram vd., 2024).

2.2.2. Karar Ağaçları

Karar ağacı (Decision Tree, DT), kurallara dayalı karar verme süreciyle çalışan denetimli bir makine öğrenmesi yöntemidir (Kamiński vd., 2018). Bu yöntemde, veri kümesindeki öznelikler kullanılarak ikili veya çoklu sorgular oluşturulur ve bu sorgular aracılığıyla veri kümesi ardışık biçimde homojen alt gruplara ayrılır. Bu yapı, ağaç şeklinde bir organizasyonu ifade eder. En üstte yer alan ve ilk bölünmeyi gerçekleştiren düğüm "kök düğüm" olarak adlandırılırken, sınıflandırma sonucunun elde edildiği uç noktalar "yaprak düğümler" olarak tanımlanır. Her bir sorgu sonucunda yeni bir dal oluşur ve bu dallar, öznelik uzayını ayrık bölgeler hâlinde parçalayarak sınıflandırma işlemini tamamlar (Nanfack vd., 2023).

2.2.3. Destek Vektör Makinesi

Destek vektör makineleri (Support Vector Machine, SVM), verileri sınıflandırmak veya regresyon problemlerinde tahmin yapmak amacıyla N boyutlu bir özellik uzayında ayırt edici bir hiper düzlem belirlemeye çalışan denetimli bir makine öğrenmesi algoritmasıdır. SVM'nin temel amacı, farklı sınıfa ait veri noktaları arasındaki mesafeyi en üst düzeye çıkararak en iyi ayırım sınırını elde etmektir. Bu algoritma hem doğrusal olarak ayrılabilir hem de doğrusal olarak ayrılamayan veri kümelerine uygulanabilmektedir. Doğrusal olmayan durumlarda, veri uzayı uygun bir çekirdek (kernel) fonksiyonu aracılığıyla daha yüksek boyutlu bir özellik uzayına dönüştürülmekte ve bu yeni uzayda optimal ayırım düzlemi oluşturmaktadır. SVM bağlamında yaygın olarak kullanılan çekirdek

fonksiyonları arasında doğrusal, polinom ve radyal tabanlı fonksiyonlar yer almaktadır (Dun vd., 2020).

2.2.4. Lineer Regresyon

Doğrusal regresyon (Linear Regression, LR), bağımlı ve bağımsız değişkenler arasındaki ilişkiyi kullanarak en uygun doğrusal denklemi belirlemeye çalışan denetimli bir makine öğrenmesi yöntemidir. Bu yaklaşımda, bağımlı değişkenin değeri bağımsız değişkenlerin doğrusal bir kombinasyonu olarak modellenir. Modelin temel amacı, tahmin edilen değerler ile gerçek gözlemler arasındaki farkı ifade eden artık (residual) değerlerin kareleri toplamını en aza indiren doğrusal denklemi belirlemektir. Bu optimizasyon süreci genellikle en küçük kareler (least squares) yöntemi kullanılarak gerçekleştirilir ve böylece veriye en iyi uyum sağlayan regresyon doğrusu elde edilir. Doğrusal regresyon modeli, basit yapısı ve yorumlanabilir olmasından dolayı yaygın olarak kullanılmaktadır. Ayrıca değişkenler arasındaki ilişkinin doğrusal olmadığı durumlarda sınırlı bir tahmin performansı sergileyebilmektedir (Chaturvedi, 2024).

2.2.5. Yapay Sinir Ağları

Yapay sinir ağları (Artificial Neural Networks, ANN), insan beynindeki biyolojik nöral yapıları taklit edecek şekilde tasarlanmıştır. Bu yapılar; girdi katmanı, gizli katmanlar ve çıktı katmanından oluşan birbirine bağlı çok katmanlı bir mimariyi içerir. Her katman, veri kümesinden anlamlı özellikler çıkarmayı amaçlayan ve nöron ya da düğüm olarak adlandırılan doğrusal olmayan işlem birimlerinden oluşur (Maleki vd., 2019). Girdi değişkenleri modele verildiğinde, her nöron kendisine bağlı ağırlıklar ve bias değerleri aracılığıyla bir çıktı üretir. Elde edilen bu çıktı ile hedef değer arasındaki hata, geri yayılım (backpropagation) algoritması kullanılarak hesaplanmakta ve bu hata doğrultusunda ağ ağırlıkları iteratif olarak güncellenmektedir. Bu öğrenme süreci, hata fonksiyonu belirli bir eşik değerinin altına düşene veya model yakınsama sağlayana kadar devam etmektedir (Dong vd., 2023).

2.2.6. Uygulanan Hibrit Yöntem

Uygulanan hibrit yöntem, ek bir meta-öğretici kullanmadan, uygulanan regresyon yöntemlerinin tahminlerini, her modelin eğitim verisindeki hata performansına bağlı olarak ağırlıklandırılarak birleştirmektedir. Bu sayede aşırı uyum riski azaltılırken, regresyon modellerinin

katkıları açık ve izlenebilir hale getirilmektedir. Hibrit yöntemde, her model için eğitim veri seti üzerinde kök ortalama karesel hata ($RMSE$) değerleri hesaplanmaktadır.

$$RMSE_m = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{i,m})^2} \quad (1)$$

Burada m , uygulanan regresyon model indeksini (DT, RF, LR, SVM, ANN) ifade etmektedir. Daha sonra her modelin uygulanan hibrit modeldeki ağırlığı, eğitim hatasının tersi ile oluşturulmaktadır.

$$w_m = \frac{\frac{1}{RMSE_m}}{\sum_{k=1}^M \frac{1}{RMSE_k}} \quad (2)$$

Yöntemde, düşük $RMSE$ değeri yüksek ağırlığa, yüksek $RMSE$ değeri düşük ağırlığa sahip olmaktadır. Bu sayede manuel ayarlama olmadan, öğrenme verisine bağlı olarak zayıf modeller baskılanarak güçlü modellerin öne çıkarılması sağlanmaktadır. Uygulanan her model test seti için tahmin değeri (\hat{y}_m^{test}) üretmektedir. Uygulanan hibrit yöntemin tahmin değeri aşağıda Denklem 3'te verilmektedir.

$$\hat{y}_{hibrit} = \sum_{m=1}^M w_m \cdot \hat{y}_m^{test} \quad (3)$$

Uygulanan hibrit yöntem, DT, RF, LR, SVM ve ANN modellerinden elde edilen tahminler doğrusal ağırlıklı toplam yaklaşımıyla birleştirilmiştir. Bu yapı, her bir modelin nihai tahmine katkısının açık biçimde izlenebilmesine olanak tanıyarak şeffaf, yorumlanabilir ve hesaplanabilir bir ensemble mimarisi sunmaktadır. Ağırlıklar, modellerin eğitim verisi üzerindeki hata performanslarına dayalı olarak otomatik biçimde belirlenmiş; böylece düşük hata üreten modellerin etkisi artırılırken, görece zayıf modellerin katkısı baskılanmıştır. Bu yaklaşım, model çeşitliliğinden yararlanarak bireysel yöntemlerin sınırlılıklarını dengelemeyi ve daha tutarlı bir genelleme başarımı elde etmeyi amaçlamaktadır.

2.3. İç Ortam Hava Kalitesinin Analizi

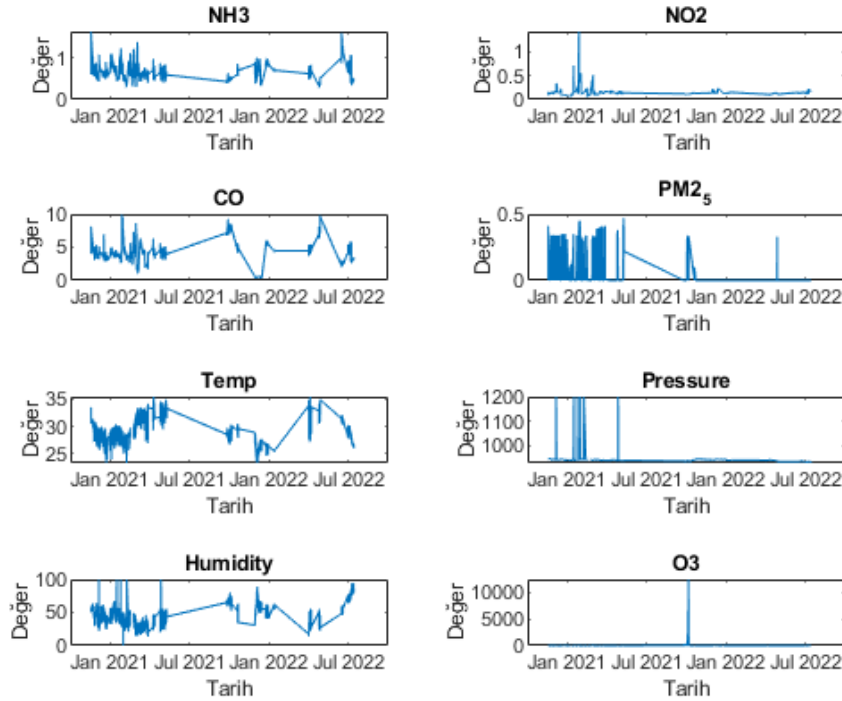
Mendeley veri setindeki değerler üzerinde öncelikle keşif veri analizi çalışması yapılmıştır. Veri seti üzerinde eksik veriler tespit analizi yapılmıştır. 173468 veri içerisinde NH₃, NO₂, CO, PM_{2.5} değerlerinden 3 adet ve O₃ değerinden 4 adet eksik veri tespit ve 14 adet hatalı sıcaklık verisi

edilmiştir. Eksik ve hatalı veriler, 5 komşu değer varsayımı ile tamamlanmıştır. Verilerin ortalama, medyan, min, max standart sapma değerleri hesaplanmış Tablo 2’de sonuçlar sunulmuştur.

Tablo 2. Verilerin ortalama, medyan, min, max standart sapma değerleri

Metrik	NH3	NO2	CO	PM2.5	Sıcaklık	Basınç	Nem	O3
Ortalama	0.67018	0.15491	4.3396	0.057456	29.426	941.26	46.559	15.862
Medyan	0.65	0.14	4.19	0	29.1	942.02	43.5	9.54
Minimum	0.29	0.06	0.29	0	23.28	929.75	0	1.34
Maksimum	1.63	1.42	9.92	0.47	35.48	1202.8	100	12413
Std	0.17797	0.076458	1.6156	0.1161	1.9806	4.4255	16.97	99.268

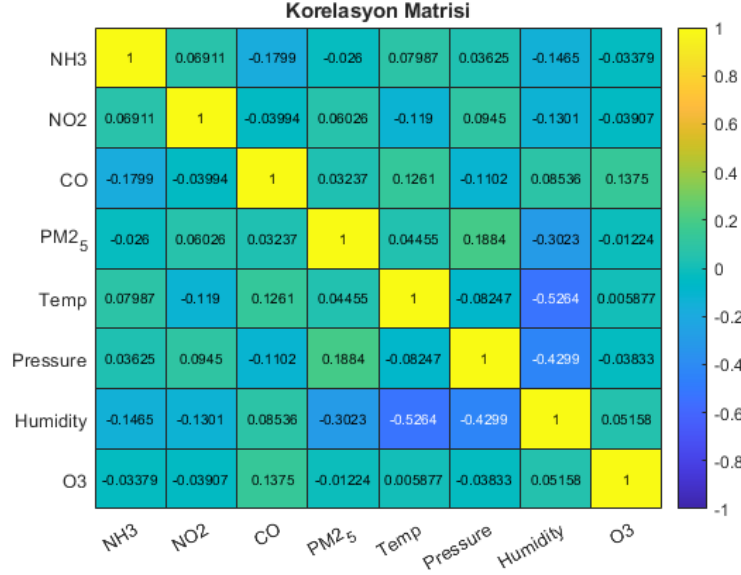
İç ortamdaki kirleticilerin zaman trendlerini daha iyi analiz etmek için zaman seri grafikleri çizdirilmiştir. Şekil 1’de zaman seri grafikleri gösterilmiştir.



Şekil 1. Kirleticilerin zaman seri grafikleri.

İç ortam kirleticilerinin arasındaki ilişkinin yönünü ve şiddetini belirlemek için korelasyon analizi yapılmıştır. Şekil 2’de korelasyon matrisinin ısı haritası verilmiştir. Korelasyon matrisi analiz edildiğinde sıcaklık, nem ve basınç değişkenleri arasında anlamlı bir ilişkiler bulunduğu ve bu değişkenlerin birbirlerini etkiledikleri görülmektedir. PM2.5, O3, CO, NO2 ve NH3 ise bağımsız değişkenler olarak kabul edilebilmektedir. Bu çalışmada ilgili hava kirleticilerinin tahmini amacıyla çok değişkenli regresyon tabanlı makine öğrenmesi yöntemleri uygulanmıştır. Veri seti, zaman bağımlılığını koruyacak şekilde kronolojik olarak ilk %70’lik bölümü eğitim, son %30’luk bölümü test verisi olarak bölünmüştür. Bu yaklaşım, gerçek hayattaki ileriye dönük tahmin senaryolarını

yansıtmak ve bilgi sızıntısını (data leakage) önlemek amacıyla tercih edilmiştir. Tüm ön işleme ve normalizasyon işlemleri yalnızca eğitim verisi üzerinde gerçekleştirilmiş, test verisine aynı parametreler uygulanmıştır.



Şekil 2. Kirlenmelerin korelasyon matrisi.

Çalışmada kullanılan regresyon modelleri arasında DT, RF, LR, SVM, ANN ve hibrit yöntem yer almaktadır. DT, RF ve SVM modellerine ait hiperparametreler, yalnızca eğitim verisi üzerinde uygulanan K-katlı çapraz doğrulama (K-fold cross-validation) ve sistematik grid arama (grid search) yöntemiyle optimize edilmiştir. RF modeli için ağaç sayısı (NumTrees) ve minimum yaprak boyutu (MinLeafSize), SVM modeli için ceza parametresi (BoxConstraint) ve çekirdek ölçeği (KernelScale), DT modeli için ise minimum yaprak boyutu hiperparametreleri optimize edilmiştir. LR modeli referans amaçlı olarak varsayılan ayarlarda kullanılmıştır. ANN modelinde tek gizli katmanlı yapı tercih edilmiş, gizli katman nöron sayısı 10 ve maksimum iterasyon sayısı 100 olarak belirlenmiştir. Performansa duyarlı ağırlıklandırma stratejisine dayalı uygulanan hibrit yöntemde her bir temel modelin eğitim verisi üzerindeki *RMSE* hesaplanmış ve bu hata değerlerinin tersine orantılı olacak şekilde model ağırlıkları belirlenmiştir. Elde edilen ağırlıklar kullanılarak, test seti üzerindeki model tahminleri doğrusal ağırlıklı ortalama yöntemiyle birleştirilmiştir.

Model performansları *RMSE*, *MAE*, *MSE* ve R^2 metrikleri ile değerlendirilmiş; modeller arasındaki performans farklarının istatistiksel anlamlılığı Friedman testi ve Wilcoxon işaretli sıralar testi kullanılarak analiz edilmiştir.

3. Bulgular ve Tartışma

Bu bölümde DT, RF, LR, SVM, ANN ve hibrit yöntem modelleri karşılaştırmalı olarak analiz edilmiştir. Uygulanan modellerin başarımları Kök Ortalama Kare Hatası (*RMSE*), Ortalama Karesel Hata (*MSE*), Ortalama Mutlak Hata (*MAE*) ve Determinasyon Katsayısı (R^2) metrikleri ile değerlendirilmiştir. *RMSE*, *MSE*, *MAE* metrik değerlerinin küçük olması ilgili yöntemin başarımının daha yüksek olduğunu; R^2 değerinin ise 1'e yakın olması, yöntemin veriyi iyi açıkladığını göstermektedir. *RMSE* metriği aşağıdaki gibi hesaplanmaktadır (Cabuk vd., 2024; Hodson, 2022).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

Burada i değişkeni, n veri nokta sayısını, y_i , gerçek gözlem zaman serisini ve \hat{y}_i ise tahmini zaman serisini ifade etmektedir. *MSE* metriği aşağıdaki gibi hesaplanmaktadır (Cabuk vd., 2024, Hodson, 2022):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

MAE metriği aşağıdaki gibi hesaplanmaktadır (Cabuk vd., 2024):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

R^2 metriği aşağıdaki gibi hesaplanmaktadır (Chicco vd., 2021):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Burada \bar{y} , gerçek verilerin ortalamasını ifade etmektedir. \bar{y} parametresi aşağıdaki gibi hesaplanmaktadır:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (8)$$

İç ortam hava kalitesinin değerlendirilmesinde CO önemli bir parametredir. Renksiz, kokusuz ve tatsız bir gaz olması nedeniyle bireyler bulunduğu ortamda CO yoğunluğunu fark edememektedir. Ancak CO'nun hemoglobine bağlanarak oksijen taşınımını engellemesi, uzun süreli ortamda

kalınması durumunda bağı ağrısı yorgunluk baş dönmesi ve bulantı gibi semptomlara yol açmakta; yüksek konsantrasyonlarda ise ciddi sağlık sorunlarına ve ölümlere neden olabilmektedir. Bu nedenle CO konsantrasyonunun güvenilir şekilde tahmin edilmesi büyük önem taşımaktadır. CO değişkeni için kullanılan regresyon modellerinin hiperparametreleri, yalnızca eğitim verisi üzerinde uygulanan 5-katlı çapraz doğrulama ve sistematik grid arama yöntemi ile optimize edilmiştir. Elde edilen en iyi hiperparametreler ve karşılık gelen çapraz doğrulama *RMSE* değerleri Tablo 3'te sunulmuştur. RF modeli için ağaç sayısı 100 ve minimum yaprak boyutu 5 olarak belirlenirken, SVM modelinde BoxConstraint değeri 10 ve KernelScale değeri 1 olarak seçilmiştir. DT modeli için minimum yaprak boyutu 10 olarak belirlenmiştir. ANN modelinde ise hesaplama maliyetini sınırlamak amacıyla gizli katman nöron sayısı 10 ve maksimum iterasyon sayısı 100 olarak sabitlenmiştir. ANN modeli için hesaplama maliyetini sınırlamak amacıyla gizli katman nöron sayısı ve iterasyon değeri sabit tutulmuş, bu nedenle çapraz doğrulama temelli bir *CV – RMSE* değeri raporlanmamıştır.

Tablo 3. CO değişkeni için 5-katlı çapraz doğrulama ile belirlenen hiperparametreleri

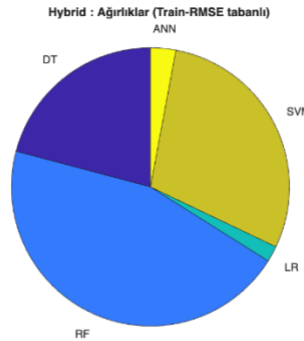
Model	HiperParametreler	<i>CV – RMSE</i>
RF	NumTree=100, MinLeafSize=5	0.0562
SVM	BoxConstraint = 10, KernelScale = 1	0.0719
DT	MinLeafSize = 10	0.1105
ANN	Hidden = 10, İterasyon s = 100	-

CO değişkeninde uygulanan yöntemlerin test seti üzerindeki performans sonuçları Tablo 4'te sunulmuştur. Sonuçlar incelendiğinde, uygulanan hibrit yöntemin, diğer regresyon modellerine kıyasla en düşük *RMSE* (2.417) ve *MSE* (5.843) değerlerini ve pozitif R^2 (0.004) değerini elde ederek genel olarak en başarılı yöntem olduğu görülmektedir. SVM ve RF modelleri, benzer hata değerleri ($RMSE \approx 2.43$) sergileyerek daha başarılı sonuçlar üretmiştir. Bununla birlikte, her iki model için de R^2 değerlerinin sıfıra yakın ve negatif olması, CO değişkeninin karmaşık yapısının tekil modeller tarafından yeterince temsil edilemediğini göstermektedir. DT modeli, RF ve SVM'ye kıyasla daha yüksek hata değerleri üretmiş, ancak aşırı başarısız bir davranış sergilememiştir. ANN modeli, hata metrikleri açısından orta düzeyde bir performans göstermiştir. LR modeli, en yüksek hata değerlerine ($RMSE = 4.690$) ve en düşük R^2 değerine (-2.749) sahip olup, CO değişkeninin doğrusal varsayımlar altında modellenmesinin uygun olmadığını açıkça ortaya koymaktadır. Bu sonuçlar, CO konsantrasyonunun çevresel ve kimyasal değişkenlerle olan ilişkisinin belirgin biçimde doğrusal olmayan bir yapı sergilediğini göstermektedir. Genel olarak elde edilen bulgular, DT, RF, LR, SVM, ANN regresyon modellerinin CO değişkenini sınırlı başarıyla tahmin edebildiğini, buna karşın eğitim performansına dayalı ağırlıklandırma stratejisi kullanan hibrit yöntemin, model hatalarını dengeleyerek daha tutarlı ve genellenebilir bir tahmin sunduğunu ortaya koymaktadır.

Tablo 4. CO değişkeni için modellerin performans değerlendirmesi

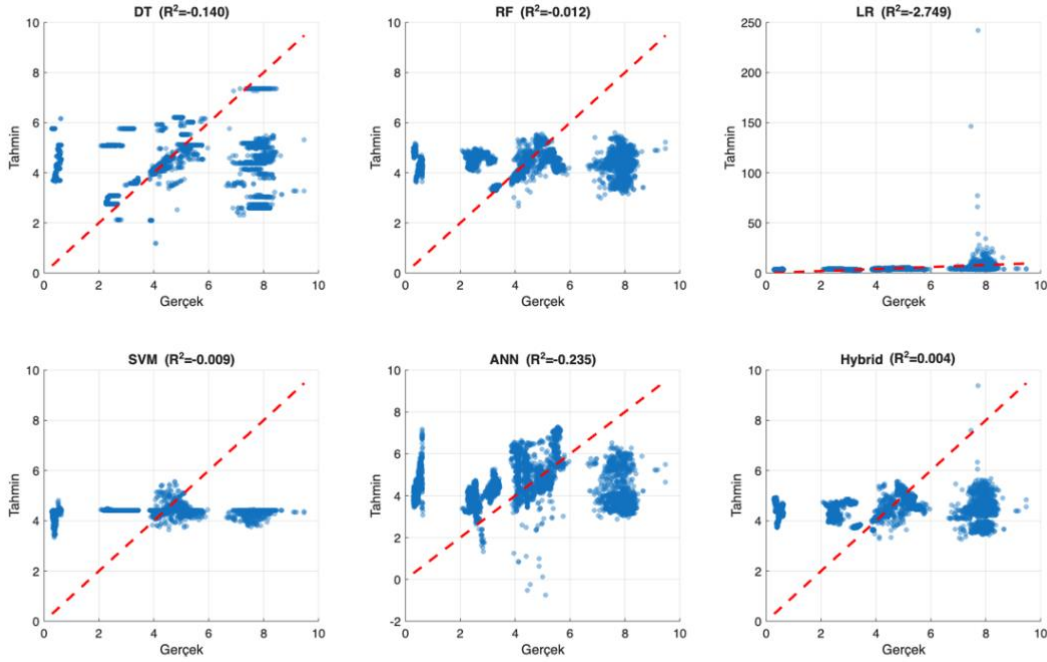
	<i>RMSE</i>	<i>MSE</i>	<i>MAE</i>	R^2
DT	2.586	6.690	1.940	-0.140
RF	2.437	5.940	1.901	-0.012
LR	4.690	22.001	1.748	-2.749
SVM	2.433	5.920	2.000	-0.009
ANN	2.693	7.250	2.189	-0.235
Hibrit Yöntem	2.417	5.843	1.897	0.004

CO için uygulanan hibrit yöntemde kullanılan ağırlıkların dağılımı Şekil 3'te sunulmuştur. Ağırlıklar incelendiğinde, RF ve SVM modellerinin hibrit yapı içerisinde en yüksek katkıyı sağladığı görülmektedir. Bu durum, söz konusu modellerin eğitim verisi üzerinde daha düşük hata değerleri üretmesiyle doğrudan ilişkilidir. Buna karşılık, LR ve ANN modellerinin görece yüksek eğitim hataları nedeniyle hibrit yapı içerisindeki katkıları oldukça düşük kalmıştır. DT modeli ise orta düzeyde bir katkı sağlamıştır. Bu dağılım, önerilen hibrit yaklaşımın model katkılarını veri temelli ve otomatik bir biçimde belirlediğini ve zayıf modellerin etkisini baskıladığını göstermektedir.

**Şekil 3.** CO değişkeni için uygulanan hibrit modelde *RMSE* tabanlı ağırlıkların dağılımı

CO değişkeni için test seti üzerinde elde edilen gerçek ve tahmin edilen değerlerin dağılımı Şekil 4'te sunulmuştur. DT modeli, Noktalar parçalı yatay şeritler hâlinde yaprak bazlı sabit tahminler nedeniyle tahminleri belirli bantlarda yoğunlaştırmış ve varyans açıklama gücü sınırlı kalmıştır. RF modeli, DT'ye kıyasla daha dengeli tahminler üretmiş; ancak tahminler ideal doğrultudan belirgin sapmalar göstermiştir. LR modeli ürettiği aşırı tahminler, CO değişkeninin doğrusal varsayımlar altında modellenemeyeceğini açık biçimde göstermektedir. SVM modeli hata metrikleri açısından rekabetçidir. Ancak tahminler ideal doğrultu etrafında yeterince dağılmadığından R^2 değeri sınırlı kalmıştır. ANN modeli doğrusal olmayan ilişkileri kısmen yakalayabilmektedir. Fakat tahminlerdeki

yüksek saçılma genelleme başarımını sınırlamıştır. Hibrit model, bireysel modellerin uç tahminlerini dengeleyerek daha tutarlı bir dağılım üretmiş ve pozitif R^2 değerine ulaşan tek yöntem olmuştur.



Şekil 4. CO değişkeni test seti üzerinde uygulanan yöntemler için gerçek tahmin değerlerinin dağılımı

NH₃, keskin ve tahriş edici kokusu olan renksiz bir gazdır. İç ortamlarda özellikle temizlik ürünlerinden kaynaklanabilen önemli bir kirletici parametredir. NH₃'ün düşük konsantrasyonlarda dahi solunum yollarını tahriş edici etkiler gösterebilmekte, gözlerde burunda ve boğazda yanma oluşturabilmektedir. Astım ve bronşit gibi solunum yolu hastalıkları olan kişiler için risklidir. NH₃ değişkeni için kullanılan regresyon modellerinin hiperparametreleri, yalnızca eğitim verisi üzerinde uygulanan 5-katlı çapraz doğrulama ve sistematik grid arama yöntemi ile optimize edilmiştir. Elde edilen en iyi hiperparametreler ve karşılık gelen çapraz doğrulama *RMSE* değerleri Tablo 5'te sunulmuştur. CO değişkenindeki hiperparametre değerleri ile sabit tutulmuştur.

Tablo 5. NH₃ değişkeni için 5-katlı çapraz doğrulama ile belirlenen en iyi hiperparametreleri

Model	HiperParametreler	<i>CV – RMSE</i>
RF	NumTree=100, MinLeafSize=5	0.00664
SVM	BoxConstraint = 10, KernelScale = 1	0.05585
DT	MinLeafSize = 10	0.01015
ANN	Hidden = 10, İterasyon s = 100	-

NH₃ değişkeni için elde edilen performans sonuçları Tablo 6’te sunulmuştur. Sonuçlar incelendiğinde, tüm regresyon modellerinin test seti üzerinde negatif R^2 değerleri ürettiği ve hata oranlarının önceki CO değişkenine kıyasla görece yüksek olduğu görülmektedir. Bu durum, NH₃ konsantrasyonunun kullanılan çevresel ve kimyasal değişkenler ile olan ilişkisinin, mevcut özellik kümesi altında yeterince temsil edilemediğini göstermektedir. DT ve RF modelleri, diğer yöntemlere kıyasla daha düşük hata değerleri üretmiş olmakla birlikte ($RMSE \approx 0.24-0.26$), her iki model için de R^2 değerlerinin sıfıra yakın ve negatif olması, açıklanan varyansın sınırlı kaldığını ortaya koymaktadır. ANN modeli, hata metrikleri açısından orta düzey bir performans sergilemiş; ancak genelleme başarımı sınırlı kalmıştır. SVM ve LR modelleri ise yüksek hata değerleri ve belirgin biçimde negatif R^2 değerleri ile NH₃ değişkeninin bu modeller altında etkili biçimde tahmin edilemediğini göstermiştir. Özellikle LR modelinin çok düşük R^2 değeri (-9.211), NH₃ değişkeninin doğrusal varsayımlar altında modellenmesinin uygun olmadığını açıkça ortaya koymaktadır. Eğitim performansına dayalı ağırlıklandırma stratejisi kullanılarak oluşturulan hibrit yöntem, bireysel modellere kıyasla daha dengeli hata değerleri üretmiş olsa da ($RMSE = 0.261$), negatif R^2 değeri nedeniyle NH₃ değişkeni için güçlü bir genelleme başarımı sunamamıştır. Bu durum, hibrit yaklaşımın model hatalarını kısmen dengelemesine rağmen, mevcut veri yapısı ve özellik kümesi altında NH₃ değişkeninin güvenilir biçimde tahmin edilmesinin zor olduğunu göstermektedir.

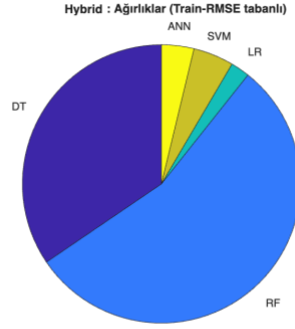
Yapılan çalışma sonucu, NH₃ konsantrasyonunun CO değişkenine kıyasla daha düzensiz ve gürültülü bir davranış sergilediğini ve yalnızca çevresel parametreler kullanılarak modellenmesinin sınırlı kaldığını ortaya koymaktadır.

Tablo 6. NH₃ değişkeni için modellerin performans değerlendirilmesi

	<i>RMSE</i>	<i>MSE</i>	<i>MAE</i>	R^2
DT	0.244	0.059	0.178	-0.007
RF	0.264	0.070	0.206	-0.183
LR	0.776	0.603	0.257	-9.211
SVM	0.658	0.433	0.567	-6.349
ANN	0.360	0.129	0.240	-1.196
Hibrit Yöntem	0.261	0.068	0.201	-0.158

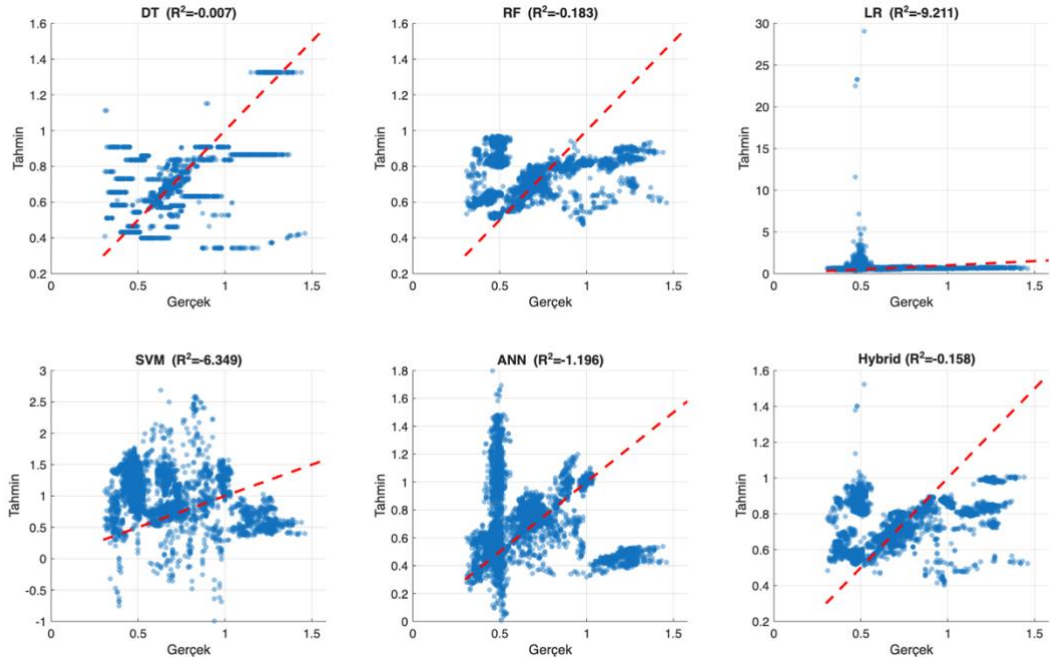
Şekil 5’te sunulan hibrit model ağırlıkları incelendiğinde, RF ve DT modellerinin hibrit yapı içerisinde baskın katkıyı sağladığı görülmektedir. RF modeli yaklaşık %55’lik ağırlık ile en yüksek katkıyı sağlarken, DT modeli yaklaşık %35 oranında hibrit yapıya katkıda bulunmuştur. Buna karşılık, LR, SVM ve ANN modellerinin eğitim verisi üzerindeki görece yüksek hata değerleri nedeniyle hibrit yapı içerisindeki katkıları sınırlı kalmıştır. Bu dağılım, NH₃ değişkeninin karar ağacı

tabanlı yöntemler tarafından daha etkin biçimde temsil edilebildiğini ve önerilen hibrit yaklaşımın model katkılarını veri temelli ve otomatik olarak belirlediğini göstermektedir.



Şekil 5. NH₃ değişkeni için uygulanan hibrit modelde *RMSE* tabanlı ağırlıkların dağılımı

NH₃ değişkeni için test seti üzerinde elde edilen gerçek ve tahmin edilen değerlerin dağılımları Şekil 6'da verilmiştir. Grafikler incelendiğinde, tüm modellerde tahminlerin ideal doğrultu etrafında yeterli yoğunlaşma göstermediği ve noktaların belirgin bir saçılma sergilediği görülmektedir. Bu durum, NH₃ değişkeninin mevcut özellik kümesi altında güvenilir biçimde modellenmesinin zor olduğunu göstermektedir.



Şekil 6. NH₃ değişkeni test seti üzerinde uygulanan yöntemler için gerçek tahmin değerlerinin dağılımı

DT modeli, yaprak bazlı karar yapısı nedeniyle tahminleri belirli seviyelerde yoğunlaştırmış ve noktaların yatay bantlar hâlinde kümelenmesine neden olmuştur. Bu yapı, düşük hata değerlerine rağmen modelin varyansı yeterince açıklayamadığını ve genelleme başarımının sınırlı kaldığını göstermektedir ($R^2 \approx -0.007$). RF modeli, DT'ye kıyasla daha yumuşak ve sürekli tahminler üretmiş; ancak tahminlerin ideal doğrultudan sistematik biçimde sapması nedeniyle açıklanan varyans düşük kalmıştır ($R^2 \approx -0.183$). Bu durum, RF'nin NH3 için yerel örüntüleri kısmen yakalayabildiğini ancak genel ilişkiyi yeterince temsil edemediğini göstermektedir. LR modeli, ürettiği aşırı ve dengesiz tahminler nedeniyle gerçek değerlerle anlamlı bir uyum sergileyememiştir. Özellikle düşük gerçek değerler için yüksek tahminler üretmesi, doğrusal varsayımların NH3 değişkeni için geçerli olmadığını açık biçimde ortaya koymuştur. Bu durum, oldukça düşük R^2 değeri (-9.211) ile desteklenmektedir. SVM modeli, doğrusal olmayan ilişkileri modelleme potansiyeline sahip olmasına rağmen, tahminlerin geniş bir aralıkta dağılması ve ideal doğrultu etrafında yoğunlaşmaması nedeniyle düşük bir genelleme başarımı göstermiştir ($R^2 \approx -6.349$). Bu durum, NH3 değişkenindeki gürültü ve düzensizliğin SVM tarafından yeterince bastırılmadığını düşündürmektedir. ANN modeli, doğrusal olmayan ilişkileri kısmen yakalayabilmiş olsa da, tahminlerde gözlenen yüksek saçılma ve bazı bölgelerde dikey kümelenmeler, modelin kararlı bir genelleme sağlayamadığını göstermektedir ($R^2 \approx -1.196$). Hibrit model, bireysel modellerin uç ve dengesiz tahminlerini belirli ölçüde dengelemiş ve daha kompakt bir dağılım üretmiştir. Ancak tahminlerin ideal doğrultu etrafında yeterince yoğunlaşmaması nedeniyle hibrit model de NH3 değişkeni için pozitif bir R^2 değerine ulaşamamıştır ($R^2 \approx -0.158$). Bu durum, hibrit yaklaşımın model hatalarını kısmen azaltmasına rağmen, NH3 değişkeninin mevcut veri ve özellik yapısı altında güvenilir biçimde tahmin edilmesinin zor olduğunu göstermektedir.

NO2, özellikle fosil yakıtların yanması sonucu oluşan bir gazdır. Düşük yoğunlukta bile insan sağlığını olumsuz etkileyebilmektedir. Solunum yolu enfeksiyonlara sebep olabilmektedir. İç ortamda ozon ve PM2.5 gibi zararlı kirleticilerin oluşumuna katkı sağlayabilmektedir. NO2 değişkeni için kullanılan regresyon modellerinin hiperparametreleri, yalnızca eğitim verisi üzerinde uygulanan 5-katlı çapraz doğrulama ve sistematik grid arama yöntemi ile optimize edilmiştir. Elde edilen en iyi hiperparametreleri ve karşılık gelen çapraz doğrulama *RMSE* değerleri Tablo 7'de sunulmuştur. CO değişkenindeki hiperparametre değerleri ile sabit tutulmuştur.

Tablo 7. NO2 değişkeni için 5-katlı çapraz doğrulama ile belirlenen en iyi hiperparametreleri

Model	HiperParametreler	<i>CV - RMSE</i>
RF	NumTree=100, MinLeafSize=5	0.00370
SVM	BoxConstraint = 10, KernelScale = 1	0.06656
DT	MinLeafSize = 10	0.00336
ANN	Hidden = 10, İterasyon s = 100	-

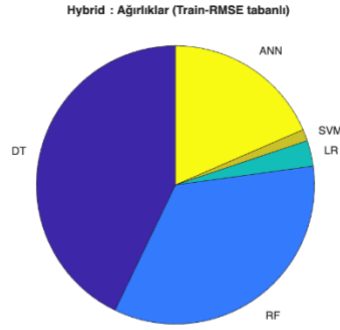
NO₂ değişkeninde uygulanan yöntemlerin performans sonuçları Tablo 8’de verilmiştir. Sonuçlar incelendiğinde, ANN modelinin diğer yöntemlere kıyasla belirgin biçimde daha başarılı bir performans sergilediği görülmektedir. ANN modeli, en düşük *RMSE* (0.0212), *MSE* (0.00045) ve *MAE* (0.0113) değerlerini elde etmiş ve pozitif R^2 (0.518) değerine ulaşarak, test verisi üzerindeki varyansın anlamlı bir bölümünü açıklayabilen tek yöntem olmuştur. Bu durum, hedef değişken ile kullanılan özellikler arasındaki ilişkinin doğrusal olmayan ve karmaşık bir yapı sergilediğini ve bu yapının ANN tarafından daha etkin biçimde öğrenilebildiğini göstermektedir. DT modeli, hata metrikleri açısından görece düşük değerlere sahip olmasına rağmen (*RMSE* = 0.0423), negatif R^2 değeri (-0.914) nedeniyle açıklanan varyans bakımından sınırlı bir genelleme başarımı sunmuştur. RF modeli ise DT’ye kıyasla daha yüksek hata değerleri üretmiş ve oldukça düşük bir R^2 değeri (-4.565) sergileyerek bu hedef değişken için uygun bir modelleme yapısı sunamamıştır. LR modeli, en yüksek hata değerlerini (*RMSE* = 0.2438) ve aşırı derecede negatif R^2 değerini (-62.706) üretmiş olup, hedef değişkenin doğrusal varsayımlar altında modellenmesinin uygun olmadığını açık biçimde ortaya koymuştur. Benzer şekilde, SVM modeli de yüksek hata değerleri ve oldukça düşük R^2 değeri (-21.191) ile hedef değişkenin bu yöntem altında yeterince temsil edilemediğini göstermiştir. Hibrit yöntem, bireysel modellerin bazı uç tahminlerini dengeleyerek DT ve RF modellerine kıyasla daha düşük hata değerleri üretmiş olsa da (*RMSE* = 0.0363), negatif R^2 değeri (-0.410) nedeniyle ANN modelinin genelleme başarımına ulaşamamıştır. Bu durum, hibrit yaklaşımın ANN modeline göre mutlak performansı artırmada sınırlı kalabildiğini göstermektedir.

Tablo 8. NO₂ değişkeni için modellerin performans değerlendirmesi

	<i>RMSE</i>	<i>MSE</i>	<i>MAE</i>	R^2
DT	0.0423	0.0017	0.0279	-0.914
RF	0.0721	0.0051	0.0468	-4.565
LR	0.2438	0.0594	0.0454	-62.706
SVM	0.1439	0.0207	0.1356	-21.191
ANN	0.0212	0.0004	0.0113	0.518
Hibrit	0.0363	0.0013	0.0246	-0.410
Yöntem				

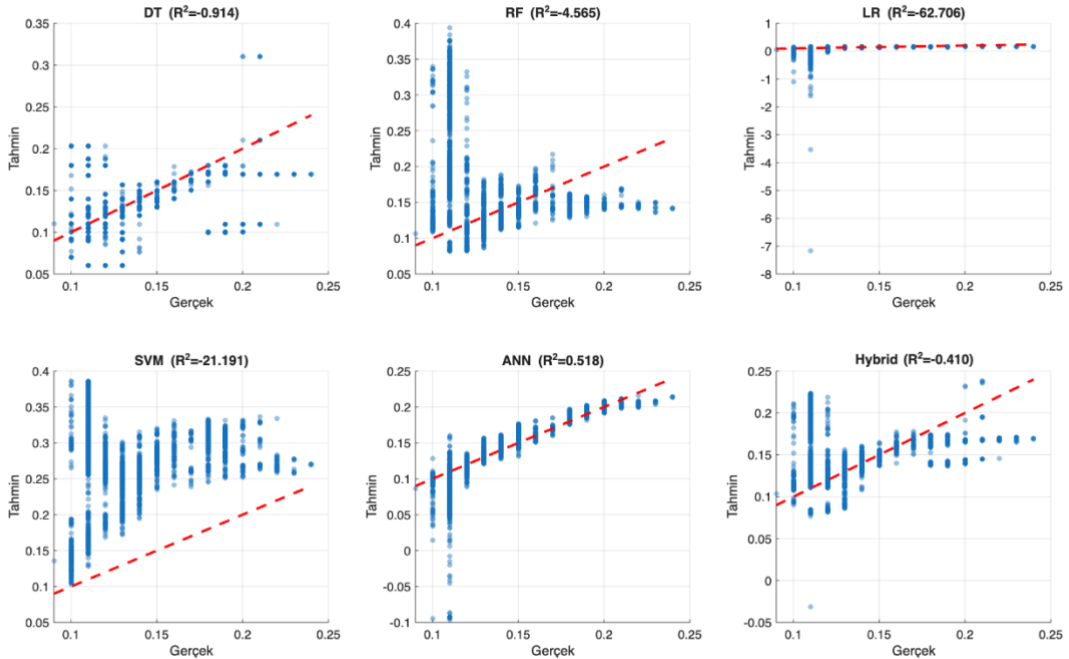
Şekil 7’de verilen hibrit model ağırlık dağılımı incelendiğinde, Karar DT ve RF modellerinin hibrit yapı içerisinde en yüksek katkıyı sağladığı görülmektedir. Bu durum, NO₂ değişkeninin eğitim verisi üzerinde ağaç tabanlı modeller tarafından daha düşük hata değerleri ile temsil edilebilmesiyle doğrudan ilişkilidir. ANN modeli, DT ve RF’ye kıyasla daha sınırlı ancak anlamlı bir ağırlık almıştır. Bu durum, ANN’nin doğrusal olmayan ilişkileri kısmen yakalayabildiğini ancak NO₂ değişkeninin yapısının derin öğrenme tabanlı temsillere kıyasla ağaç tabanlı modellere daha uygun olduğunu göstermektedir. Buna karşılık, SVM ve LR modellerinin hibrit yapı içerisindeki ağırlıkları oldukça

düşük seviyede kalmıştır. Elde edilen ağırlık dağılımı, önerilen hibrit yaklaşımın model katkılarını tamamen veri temelli, performansa dayalı ve otomatik bir biçimde belirlediğini ortaya koymaktadır. Ayrıca bu sonuçlar, NO₂ değişkeninin tahmininde ağaç tabanlı modellerin daha uygun olduğunu ve zayıf performans sergileyen modellerin hibrit yapı içerisinde etkin biçimde baskılandığını göstermektedir.



Şekil 7. NO₂ değişkeni için uygulanan hibrit modelde *RMSE* tabanlı ağırlıkların dağılımı

Şekil 8'te NO₂ değişkeni için test seti üzerinde uygulanan regresyon modellerinin gerçek ve tahmin edilen değerleri arasındaki ilişki sunulmaktadır. Grafikler incelendiğinde, modeller arasında belirgin performans farklılıkları olduğu açıkça görülmektedir.



Şekil 8. NO₂ değişkeni test seti üzerinde uygulanan yöntemler için gerçek tahmin değerlerinin dağılımı

DT modeli, tahminlerin belirli bantlarda yoğunlaştığı parçalı bir yapı sergilemekte olup, yaprak tabanlı sabit tahminler nedeniyle gerçek değerlerle olan uyum sınırlı kalmıştır. Bu durum, DT modelinin hata metriklerinin görece düşük olmasına rağmen negatif R^2 değeri (-0.914) üretmesiyle tutarlıdır ve genelleme başarımının yetersiz olduğunu göstermektedir. RF modeli, DT'ye kıyasla daha dengeli tahminler üretmiştir. Ancak tahminlerin ideal doğrultu etrafında toplanmadığı ve gerçek değerlerle olan ilişkinin zayıf kaldığı görülmektedir. RF modelinin oldukça düşük R^2 değeri (-4.565), NO₂ değişkeninin bu model tarafından yeterince temsil edilemediğini ortaya koymaktadır. LR modeli, grafik üzerinde belirgin biçimde gerçek değerlerden kopuk ve aşırı sapmalı tahminler üretmiştir. Tahminlerin büyük ölçüde yatay bir bant üzerinde toplanması ve bazı uç değerlerin gözlenmesi, LR modelinin doğrusal varsayımlar altında NO₂ değişkenini modelleyemediğini açıkça göstermektedir. Bu durum, son derece düşük R^2 değeri (-62.706) ile de desteklenmektedir. SVM modeli, tahminlerde yüksek saçılma sergilemiş ve gerçek değerlerle olan doğrusal ilişkiyi yakalayamamıştır. Tahminlerin ideal doğrultudan sistematik biçimde uzaklaşması, modelin genelleme başarımını ciddi şekilde sınırlamış ve bu durum negatif R^2 değeri (-21.191) ile nicel olarak da ortaya konmuştur. Ancak ANN modeli, diğer tüm yöntemlere kıyasla belirgin biçimde daha iyi bir performans sergilemiştir. Tahminlerin ideal doğrultu etrafında yoğunlaştığı ve gerçek değerlerle güçlü bir doğrusal ilişki gösterdiği görülmektedir. ANN modelinin pozitif R^2 değeri (0.518), test verisi üzerindeki varyansın anlamlı bir bölümünü açıklayabildiğini ve NO₂ değişkeninin doğrusal olmayan yapısının bu model tarafından daha etkin biçimde öğrenildiğini göstermektedir. Hibrit model, bireysel modellerin bazı uç tahminlerini dengeleyerek DT, RF ve SVM'ye göre daha tutarlı bir dağılım üretmiştir. Ancak hibrit modelin R^2 değerinin negatif kalması (-0.410), bu hedef değişken için baskın ve güçlü bir temel model olan ANN'nin performansını aşmakta sınırlı kaldığını göstermektedir. Bu durum, hibrit yaklaşımların her problem için tek başına en iyi çözüm olmayabileceğini, bazı değişkenler için tekil modellerin daha yüksek genelleme başarımı sunabildiğini ortaya koymaktadır.

PM_{2.5}, çapı 2.5 mikrometreden küçük olan ince partikül maddeleri ifade etmektedir. Sigara dumanı, sobalar ve pişirme süreçleri sonucu meydana çıkmaktadır. Yeterli bir havalandırma bulunmadığında solunum sisteminde ve kardiyovasküler etkilere sebep olabilmektedir. Dünya sağlık örgütü PM_{2.5}'i insan sağlığı açısından kanserojen olarak sınıflandırmaktadır. PM_{2.5} değişkeni için kullanılan regresyon modellerinin hiperparametreleri, yalnızca eğitim verisi üzerinde uygulanan 5-katlı çapraz doğrulama ve sistematik grid arama yaklaşımı ile optimize edilmiştir. Elde edilen en iyi hiperparametreleri ve karşılık gelen çapraz doğrulama *RMSE* değerleri Tablo 9'da verilmiştir. CO değişkenindeki hiperparametre değerleri ile sabit tutulmuştur.

Tablo 9. PM2.5 değişkeni için 5-katlı çapraz doğrulama ile belirlenen en iyi hiperparametreler

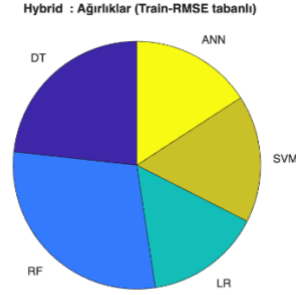
Model	HiperParametreler	$CV - RMSE$
RF	NumTree=100, MinLeafSize=5	0.0775
SVM	BoxConstraint = 10, KernelScale = 1	0.0995
DT	MinLeafSize = 10	0.0900
ANN	Hidden = 10, İterasyon s = 100	-

PM2.5 değişkeninde uygulanan yöntemlerin performans sonuçları Tablo 10'da sunulmuştur. Sonuçlar incelendiğinde, genel olarak tüm modellerin negatif R^2 değerleri üretmesi, PM2.5 konsantrasyonunun mevcut öznelikler kullanılarak tahmin edilmesinin oldukça zor olduğunu ve değişkenin yüksek düzeyde karmaşık bir yapıya sahip olduğunu göstermektedir. DT ve RF modelleri, diğer yöntemlere kıyasla daha düşük hata değerleri üretmiştir. RF modeli, $RMSE = 0.0498$ ve $MSE = 0.00248$ değerleri ile DT modeline göre hafif bir üstünlük sergilemiştir. Bununla birlikte, her iki model için de R^2 değerlerinin oldukça düşük ve negatif olması (DT: -4.959, RF: -4.529), açıklanan varyansın sınırlı kaldığını ve bu modellerin PM2.5 değişkeninin test verisi üzerindeki dağılımını yeterli düzeyde temsil edemediğini ortaya koymaktadır. LR modeli, açık ara en düşük performansı sergileyen yöntem olmuştur. Çok yüksek hata değerleri ($RMSE = 0.1719$) ve aşırı derecede negatif R^2 değeri (-64.971), PM2.5 değişkeninin doğrusal varsayımlar altında modellenmesinin uygun olmadığını açık biçimde göstermektedir. Bu durum, PM2.5 konsantrasyonunun çevresel faktörlerle olan ilişkisinin belirgin biçimde doğrusal olmayan bir yapıya sahip olduğunu desteklemektedir. SVM ve ANN modelleri, DT ve RF'ye kıyasla daha yüksek hata değerleri üretmiş ve negatif R^2 değerleri sergilemiştir. Bu sonuçlar, söz konusu modellerin PM2.5 değişkeninin gürültülü ve değişken yapısını genelleme açısından yeterince yakalayamadığını göstermektedir. Hibrit yöntem, $RMSE = 0.0496$ ve $MSE = 0.00246$ değerleri ile tüm yöntemler arasında en düşük hata değerlerine sahip model olmuştur. Ancak hibrit yöntemin R^2 değerinin de negatif kalması (-4.499), her ne kadar hata metrikleri açısından iyileşme sağlansa da PM2.5 değişkeni için test seti üzerindeki belirleyici gücün sınırlı olduğunu göstermektedir. Bu durum, hibrit yaklaşımın bireysel modellerin hatalarını dengeleyerek daha istikrarlı tahminler üretebildiğini, ancak veri setindeki özneliklerin PM2.5 değişkenini açıklamada yetersiz kaldığını ortaya koymaktadır.

Tablo 10. PM2.5 değişkeni için modellerin performans değerlendirilmesi

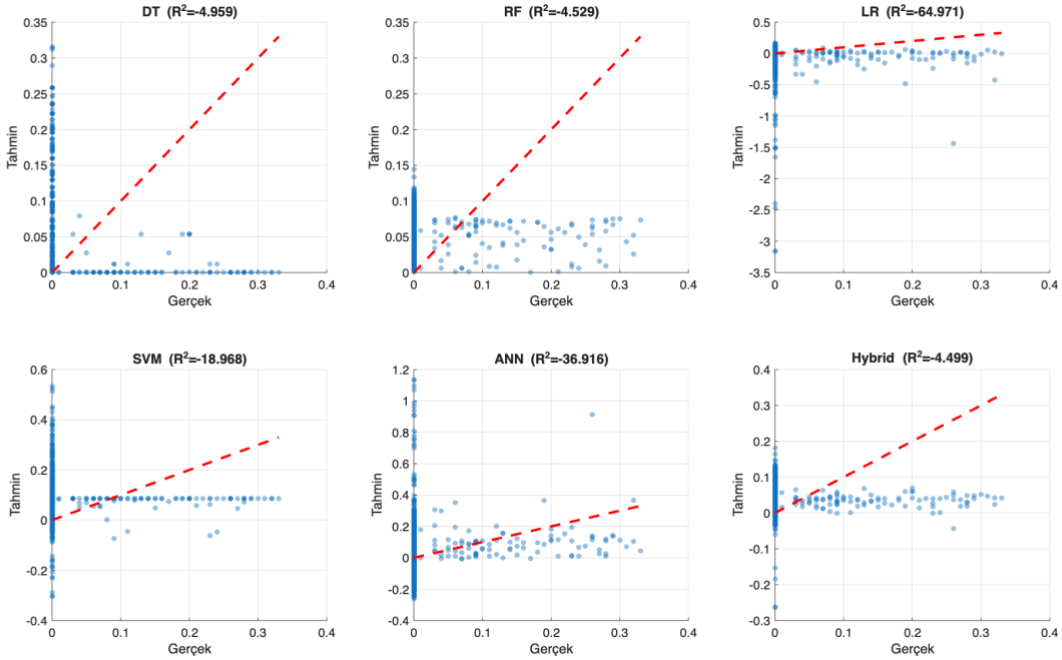
	$RMSE$	MSE	MAE	R^2
DT	0.0517	0.0026	0.0194	-4.959
RF	0.0498	0.0025	0.0431	-4.529
LR	0.1718	0.0295	0.0733	-64.971
SVM	0.0946	0.0089	0.0881	-18.968
ANN	0.1303	0.0169	0.1018	-36.916
Hibrit	0.0496	0.0024	0.0390	-4.499
Yöntem				

Şekil 9’da verilen hibrit model ağırlık dağılımı incelendiğinde, RF ve DT modellerinin en yüksek katkıyı sağladığı, bunu SVM ve ANN modellerinin izlediği, LR modelinin ise görece daha sınırlı bir etkiye sahip olduğu görülmektedir.



Şekil 9. PM2.5 değişkeni için uygulanan hibrit modelde *RMSE* tabanlı ağırlıkların dağılımı

Şekil 10’da PM2.5 değişkeni için test seti üzerinde uygulanan regresyon ve hibrit modellerin gerçek ve tahmin edilen değerleri arasındaki ilişki sunulmaktadır. Grafikler genel olarak incelendiğinde, PM2.5 değişkeninin tüm modeller için tahmin edilmesi zor bir yapı sergilediği ve tahminlerin ideal doğrultu etrafında yeterli düzeyde toplanmadığı görülmektedir.



Şekil 10. PM2.5 parametresinde uygulanan yöntemlerin sonuçları

DT modeli, tahminlerin büyük ölçüde belirli sabit bantlar üzerinde yoğunlaştığı ve düşük PM2.5 değerleri için aşırı yoğunlaşma gösterdiği parçalı bir yapı sergilemektedir. Bu durum, yaprak tabanlı sabit tahminlerin genelleme kabiliyetini sınırladığını ve DT modelinin negatif R^2 değeri (-4.959) üretmesine neden olduğunu göstermektedir. RF modeli, DT'ye kıyasla daha dengeli bir dağılım sunmasına rağmen, tahminlerin gerçek değerlerle olan doğrusal ilişkisi zayıf kalmıştır. Tahminlerin ideal doğrultudan sistematik biçimde uzaklaşması, RF modelinin de negatif R^2 değeri (-4.529) üretmesine yol açmıştır. Bu sonuç, PM2.5 değişkeninin ağaç tabanlı modellerle kısmen temsil edilebildiğini ancak test verisi üzerinde yeterli genelleme başarımına ulaşamadığını göstermektedir. LR modeli, tahminlerin gerçek değerlerden belirgin biçimde koptuğu ve bazı uç değerlerin gözlemlendiği en başarısız yöntem olmuştur. Grafik üzerinde tahminlerin yatay bir bant hâlinde toplanması, modelin PM2.5 değişkeninin doğrusal olmayan yapısını yakalayamadığını açıkça ortaya koymaktadır. Bu durum, son derece düşük R^2 değeri (-64.971) ile de nicel olarak desteklenmektedir. SVM modeli, tahminlerde yüksek saçılma ve sistematik sapmalar sergilemiş, gerçek değerlerle olan ilişkiyi yakalayamamıştır. Tahminlerin ideal doğrultudan uzaklaşması ve negatif R^2 değeri (-18.968), SVM modelinin PM2.5 değişkeni için genelleme başarımının sınırlı kaldığını göstermektedir. ANN modeli, doğrusal olmayan ilişkileri öğrenme potansiyeline rağmen, tahminlerde yüksek saçılma ve kararsız bir yapı sergilemiştir. Özellikle düşük PM2.5 değerlerinde gözlenen aşırı saçılma, modelin test verisi üzerindeki genelleme yeteneğini sınırlamış ve bu durum negatif R^2 değeri (-36.916) ile sonuçlanmıştır. Hibrit model, bireysel modellerin bazı uç tahminlerini kısmen dengelemiş ve tahminlerin ideal doğrultu etrafında yeterince toplanamaması nedeniyle negatif R^2 değeri (-4.499) üretmiştir. Bu sonuç, hibrit yaklaşımın hata metriklerini düşürmede etkili olabildiğini ancak PM2.5 değişkeni için belirleyici gücün sınırlı kaldığını göstermektedir.

Modeller arasındaki performans farklarının istatistiksel olarak anlamlı olup olmadığını değerlendirmek amacıyla, test seti mutlak hata değerleri kullanılarak Friedman testi ve ikili karşılaştırmalar için Wilcoxon işaretli sıralar testi uygulanmıştır. Elde edilen sonuçlar, CO, NH₃, NO₂ ve PM2.5 değişkenlerinin tamamı için Friedman testine ait p-değerlerinin 0.05 anlamlılık düzeyinin oldukça altında olduğunu göstermiştir ($p < 0.001$). Bu bulgu, karşılaştırılan regresyon modellerinin performansları arasında istatistiksel olarak anlamlı farklar bulunduğunu ortaya koymaktadır. Friedman testini takiben gerçekleştirilen Wilcoxon ikili karşılaştırmaları, hibrit model ile tüm tekil modeller (DT, RF, LR, SVM ve ANN) arasındaki farkların tamamının istatistiksel olarak anlamlı olduğunu göstermiştir. Bu sonuçlar, hibrit yaklaşımın yalnızca hata metrikleri açısından daha dengeli sonuçlar üretmekle kalmadığını, aynı zamanda istatistiksel olarak da tekil modellere kıyasla anlamlı bir performans farklılığı sunduğunu doğrulamaktadır. Özellikle anlaşılabilirliğin sınırlı kaldığı kirleticilerde (NH₃ ve PM2.5), hibrit yapının hata dağılımlarını dengeleme yeteneği

istatistiksel testlerle desteklenmiş; CO ve NO₂ gibi değişkenlerde ise hibrit yaklaşımın sağladığı performans kazanımlarının rastgele olmadığını ortaya konmuştur. Bu bağlamda, önerilen hibrit yöntemin güvenilirliği ve genellenebilirliği, yalnızca deterministik metriklerle değil, istatistiksel anlamlılık analizleriyle de desteklenmiştir.

4. Sonuçlar ve Öneriler

Bu çalışmada, iç ortam hava kalitesi kapsamında CO, NH₃, NO₂, PM_{2.5} kirleticilerinin makine öğrenmesi modellerinden DT, RF, LR, ANN, SVM ve hibrit model karşılaştırmalı olarak değerlendirilmiştir. Tüm modeller için hiperparametreler yalnızca eğitim verisi üzerinde uygulanan 5-katlı çapraz doğrulama ve sistematik grid arama yöntemiyle belirlenmiş, test verisi değerlendirme sürecine dâhil edilmemiştir. Elde edilen sonuçlar, CO değişkeni için tekil regresyon modellerinin sınırlı belirginliğe sahip olduğunu, buna karşın hibrit yaklaşımın en düşük hata değerlerine (*RMSE* ve *MSE*) ve pozitif R^2 değerine ulaşarak genel olarak en dengeli performansı sunduğunu göstermektedir. RF ve SVM modelleri hata metrikleri açısından rekabetçi sonuçlar üretmiştir, negatif veya sıfıra yakın R^2 değerleri bu modellerin CO değişkenindeki karmaşık ve doğrusal olmayan yapıyı tek başına yeterince temsil edemediğini ortaya koymuştur. LR modeli ise yüksek hata değerleri ve oldukça düşük R^2 sonuçlarıyla, doğrusal varsayımların CO tahmini için uygun olmadığını açık biçimde göstermiştir. NH₃ değişkeni için elde edilen sonuçlar, tüm modellerin önceki çalışmalara kıyasla daha düşük belirginliğe sahip olduğunu ve test setinde R^2 değerlerinin genel olarak negatif kaldığını göstermektedir. Bu durum, NH₃ konsantrasyonunun mevcut öznitelik setiyle yeterince temsil edilemediğine işaret etmektedir. Hibrit yaklaşım, tekil modellere kıyasla hata dengeleme açısından avantaj sağlamış; ancak anlaşılabilirlik açısından belirgin bir iyileşme sunamamıştır. Buna rağmen, hibrit yapı içerisinde RF ve DT modellerinin daha yüksek ağırlıklar alması, bu yöntemlerin görece olarak daha güvenilir tahminler ürettiğini göstermektedir. NO₂ değişkeni için yapılan değerlendirmelerde, ANN modelinin pozitif R^2 değeri elde eden tek model olduğu, hibrit yaklaşımın ise hata değerlerini düşürmesine rağmen anlaşılabilirlik açısından sınırlı kaldığı görülmüştür. Hibrit ağırlık dağılımında DT, RF ve ANN modellerinin baskın rol üstlenmesi, NO₂ değişkeninde farklı model tiplerinin tamamlayıcı özellikler sunduğunu göstermektedir. Bu sonuçlar, NO₂ tahmininde doğrusal olmayan ilişkilerin varlığını ve derin öğrenme tabanlı yapıların potansiyel avantajını ortaya koymaktadır. PM_{2.5} değişkeni için ise tüm modellerin tahmin performansı belirgin biçimde düşük kalmış, test setinde elde edilen R^2 değerleri negatif olmuştur. PM_{2.5} konsantrasyonunun yüksek frekanslı dalgalanmalar ve ani değişimler içeren yapısı, kullanılan meteorolojik ve gaz tabanlı özniteliklerle yeterince açıklanamamış; bu durum hem tekil modellerin hem de hibrit yaklaşımın genelleme başarımını sınırlamıştır. Hibrit model, hata metrikleri açısından RF ve DT modellerine

yakın sonuçlar üretmiştir. Ancak anlaşılabilirlik bakımından belirgin bir üstünlük sağlayamamıştır. Genel olarak elde edilen bulgular, önerilen eğitim-performansı-tabanlı hibrit yaklaşımın, tekil modellerin zayıf yönlerini dengeleyerek daha tutarlı ve istikrarlı hata performansı sunduğunu; ancak kirlenici türüne bağlı olduğunu göstermektedir. Özellikle CO değişkeni için hibrit yaklaşım en başarılı yöntem olarak öne çıkarken, NH₃, NO₂ ve PM_{2.5} değişkenlerinde veri yapısı ve öznitelik sınırlamaları tahmin başarısını kısıtlamıştır.

Elde edilen sonuçlar, önerilen hibrit yaklaşımın özellikle CO gibi doğrusal olmayan ve sağlık riski yüksek kirlenicilerin tahmininde, tekil modellere kıyasla daha güvenilir ve dengeli sonuçlar sunduğunu göstermektedir. Bu durum, okul, ofis, hastane ve konut gibi kapalı alanlarda sürekli izleme gerektiren ortamlarda, erken uyarı ve karar destek sistemleri için yöntemin pratik olarak uygulanabilir olduğunu ortaya koymaktadır. Düşük maliyetli sensörler ve IoT tabanlı izleme altyapıları ile entegre edildiğinde, hibrit modelin gerçek zamanlı hava kalitesi tahmini ve otomatik havalandırma kontrolü gibi uygulamalarda etkin biçimde kullanılabilmesi değerlendirilmektedir.

Gelecek çalışmalarda, daha zengin öznitelik setleri, zamansal bağımlılıkları açık biçimde modelleyen derin öğrenme mimarileri (ör. LSTM, Temporal CNN) ve otomatik hiperparametre optimizasyonu yaklaşımlarının kullanılması, özellikle PM_{2.5} ve NH₃ gibi değişkenler için tahmin performansını artırma potansiyeli taşımaktadır. Ayrıca, farklı ağırlıklandırma stratejilerine sahip adaptif hibrit yapılar ve çoklu veri kaynaklarının entegrasyonu, iç ortam hava kalitesinin daha güvenilir biçimde modellenmesine katkı sağlayabilir.

Yazarların Katkısı

Tüm yazarlar çalışmaya eşit katkıda bulunmuştur.

Çıkar Çatışması Beyanı

Yazarlar arasında herhangi bir çıkar çatışması bulunmamaktadır.

Araştırma ve Yayın Etiği Beyanı

Yapılan çalışmada araştırma ve yayın etiğine uyulmuştur.

Kaynaklar

- Adekoya, O. B., Olabode, J. K., Rafi, S. K. (2021). Renewable energy consumption, carbon emissions and human development: Empirical comparison of the trajectories of world regions. *Renewable Energy*, 179, 1836–1848. <https://doi.org/10.1016/j.renene.2021.08.019>
- Amoatey, P., Omidvarborna, H., Baawain, M. S., Al-Mamun, A. (2018). Indoor air pollution and exposure assessment of the gulf cooperation council countries: A critical review. *Environment international*, 121, 491–506. <https://doi.org/10.1016/j.envint.2018.09.043>
- Aram, S. A., Nketiah, E. A., Saalidong, B. M., Wang, H., Afitiri, A. R., Akoto, A. B., Lartey, P. O. (2024). Machine learning-based prediction of air quality index and air quality grade: A comparative analysis. *International Journal of Environmental Science and Technology*, 21(2), 1345–1360. <https://doi.org/10.1007/s13762-023-05016-2>
- Bakó-Biró, Z., Clements-Croome, D. J., Kochhar, N., Awbi, H. B., Williams, M. J. (2012). Ventilation rates in schools and pupils' performance. *Building and Environment*, 48, 215–223. <https://doi.org/10.1016/j.buildenv.2011.08.018>
- Berkani, M. R. A., Chouchane, A., Himeur, Y., Ouamane, A., Amira, A. (2023). An intelligent edge-deployable indoor air quality monitoring and activity recognition approach. *arXiv preprint arXiv:2311.03920*. <https://doi.org/10.48550/arXiv.2311.03920>
- Chaturvedi, P. (2024). Air Quality Prediction System Using Machine Learning Models. *Water, Air, & Soil Pollution*, 235(9), 578. <https://doi.org/10.1007/s11270-024-07390-0>
- Chatzidiakou, L., Mumovic, D., Summerfield, A. J. (2012). What do we know about indoor air quality in school classrooms? A critical review of the literature. *Intelligent Buildings International*, 4(4), 228–259. <https://doi.org/10.1080/17508975.2012.725530>
- Chicco, D., Warrens, M. J., Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj computer science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Dong, J., Goodman, N., Rajagopalan, P. (2023). A review of artificial neural network models applied to predict indoor air quality in schools. *International Journal of Environmental Research and Public Health*, 20(15), 6441. <https://doi.org/10.3390/ijerph20156441>
- Dun, M., Xu, Z., Chen, Y., Wu, L. (2020). Short-term air quality prediction based on fractional grey linear regression and support vector machine. *Mathematical Problems in Engineering*, 2020, Article ID 8914501. <https://doi.org/10.1155/2020/8914501>
- El-Tanbouly, R., Hassan, Z., El-Messeiry, S. (2021). The role of indoor plants in air purification and human health in the context of COVID-19 pandemic: a proposal for a novel line of inquiry. *Frontiers in Molecular Biosciences*, 8, 709395. <https://doi.org/10.3389/fmolb.2021.709395>
- Hahad, O., Lelieveld, J., Birklein, F., Lieb, K., Daiber, A., Münzel, T. (2020). Ambient air pollution increases the risk of cerebrovascular and neuropsychiatric disorders through induction of inflammation and oxidative stress. *International journal of molecular sciences*, 21(12), 4306. <https://doi.org/10.3390/ijms21124306>
- Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 2022, 1–10
- Intergovernmental Panel on Climate Change (IPCC). (2014). *Climate change 2014: Synthesis report*. In Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (p. 151). IPCC. <https://www.ipcc.ch/report/ar5/syr/>
- Kamiński, B., Jakubczyk, M., Szufel, P. (2018). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*, 26(1), 135–159. <https://doi.org/10.1007/s10100-017-0479-6>
- Lazaridis, M. (2011). *First principles of meteorology and air pollution*. Springer. https://doi.org/10.1007/978-94-007-0162-5_2
- Liu, H., Li, J., Sun, Y., Wang, Y., Zhao, H. (2020). Estimation method of carbon emissions in the embodied phase of low carbon building. *Advances in Civil Engineering*, 2020, Article ID 8846917. <https://doi.org/10.1155/2020/8853536>
- Maleki, H., Sorooshian, A., Goudarzi, G., Baboli, Z., Tahmasebi Birgani, Y., Rahmati, M. (2019). Air pollution prediction by using an artificial neural network model. *Clean Technologies and Environmental Policy*, 21, 1341–1352. <https://doi.org/10.1007/s10098-019-01709-w>

- Mata, T. M., Martins, A. A., Calheiros, C. S., Villanueva, F., Alonso-Cuevilla, N. P., Gabriel, M. F., Silva, G. V. (2022). Indoor air quality: a review of cleaning technologies. *Environments*, 9(9), 118. <https://doi.org/10.3390/environments9090118>
- Megahed, N. A., Ghoneim, E. M. (2021). Indoor Air Quality: Rethinking rules of building design strategies in post-pandemic architecture. *Environmental research*, 193, 110471. <https://doi.org/10.1016/j.envres.2020.110471>
- Nanfack, G., Temple, P., Frénay, B. (2023). Learning customised decision trees for domain-knowledge constraints. *Pattern Recognition*, 142, 109610. <https://doi.org/10.1016/j.patcog.2023.109610>
- Othman, H., Azari, R., Guimarães, T. (2024). Low-cost IoT-based indoor air quality monitoring. *Technology | Architecture + Design*, 8(2), 250–270. <https://doi.org/10.1080/24751448.2024.2405403>
- Peters, A., Dockery, D. W., Muller, J. E. (2019). Air pollution and cardiovascular disease: A scientific statement from the American Heart Association. *Circulation*, 116(1), 123–133. <https://doi.org/10.1161/CIRCULATIONAHA.116.024620>
- Pietraru, R. N., Olteanu, A., Adochiei, I. R., Adochiei, F. C. (2024). Reengineering indoor air quality monitoring systems to improve end-user experience. *Sensors*, 24(8), 2659. <https://doi.org/10.3390/s24082659>
- Rajagopalan, S., Al-Kindi, S. G., Brook, R. D. (2018). Air pollution and cardiovascular disease: JACC state-of-the-art review. *Journal of the American College of Cardiology*, 72(17), 2054–2070. <https://doi.org/10.1016/j.jacc.2018.07.099>
- Saravanan, D., Kumar, K. S. (2023). Improving air pollution detection accuracy and quality monitoring based on bidirectional RNN and the Internet of Things. *Materials Today: Proceedings*, 81, 791–796. <https://doi.org/10.1016/j.matpr.2021.04.239>
- Serefoglu Cabuk, K., Cengiz, S. K., Guler, M. G., Topcu, H., Cetin Efe, A., Ulas, M. G., Poslu Karademir, F. (2024). Chasing the objective upper eyelid symmetry formula; R2, RMSE, POC, MAE, and MSE. *International Ophthalmology*, 44(1), 303. <https://doi.org/10.1007/s10792-024-03157-y>
- Sonawani, S., Patil, K. (2022). Dataset of indoor air pollutants using low-cost sensors. *Mendeley Data*, V1. <https://doi.org/10.17632/2r232jpfb2.1>
- Srivastava, R. K., Sarkar, S., Beig, G. (2014). Correlation of various gaseous pollutants with meteorological parameter (temperature, relative humidity and rainfall). *Global Journal of Science Frontier Research*, 14(6), 56–65. <https://journalofscience.org/index.php/GJSFR/article/view/1452>
- World Health Organization. (2018). *Air quality and health*. <https://www.who.int/publications/i/item/9789241565585> (Erişim tarihi: 10 Şubat 2025)
- Yadav, V., Yadav, A. K., Singh, V., Singh, T. (2024). Artificial neural network an innovative approach in air pollutant prediction for environmental applications: A review. *Results in Engineering*, 102305 <https://doi.org/10.1016/j.rineng.2024.102305>
- Yaman, C. (2024). A review on the process of greenhouse gas inventory preparation and proposed mitigation measures for reducing carbon footprint. *Gases*, 4(1), 18–40. <https://doi.org/10.3390/gases4010002>
- Zhang, B., Geng, Z., Zhang, H., Pan, J. (2022). Densely connected convolutional networks with attention long short-term memory for estimating PM2.5 values from images. *Journal of Cleaner Production*, 333, 130101. <https://doi.org/10.1016/j.jclepro.2021.130101>