



Askıda Katı Madde Konsantrasyonunun Tahmininde Yığma Modelleri ve Hiperparametre Optimizasyonunun Etkisi: Filyos Çayı Örneği

The Impact of Stacking Models and Hyperparameter Optimization on the Prediction of Suspended Solid Concentration: A Case Study of Filyos Stream

Rukiye Uzun Arslan¹ , Berna Aksoy^{2,*} , İrem Şenyer Yapıcı³

¹Zonguldak Bülent Ecevit Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Zonguldak, Türkiye

²Zonguldak Bülent Ecevit Üniversitesi, Mühendislik Fakültesi, İnşaat Mühendisliği Bölümü, Zonguldak, Türkiye

³Zonguldak Bülent Ecevit Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Zonguldak, Türkiye

Öz

Canlı yaşamının devamlılığı ve ekosistem dengesinin sürdürülebilirliği için hayati bir unsur olan su, çevresel ve insan kaynaklı faktörlerin etkisiyle fiziksel ve kimyasal özelliklerinde değişimlere uğrayabilmektedir. Su kalitesinin temel göstergelerinden biri olan askıda katı madde (AKM) konsantrasyonunun doğru bir şekilde tahmin edilmesi, sürdürülebilir su yönetimi açısından kritik öneme sahiptir. Bu çalışmada, Batı Karadeniz Havzası'ndaki Filyos Çayı Alt Havzası'na ait su kalitesi verileri kullanılarak, AKM konsantrasyonunun tahmini için yığma modeli temelli bir yaklaşım önerilmiştir. Modelde, temel model olarak rasgele orman (RO) ve gradyan artırma (GA) algoritmaları, meta-model olarak ise K-en yakın komşu (KNN), çok katmanlı algılayıcı (ÇKA), kategorik artırma (CatBoost), uyarlamalı artırma (AdaBoost) ve Torbalama (Bagging) algoritmaları kullanılmıştır. Model performansı, ortalama karesel hata (MSE), kök ortalama kare hatası (RMSE), ortalama mutlak hata (MAE), ortalama mutlak yüzde hata (MAPE) ve korelasyon katsayısı (R^2) gibi metriklerle değerlendirilmiştir. Sonuçlara göre, en yüksek R^2 değeri (%87) RO-GA+CatBoost modeli tarafından elde edilmiştir. Bu model, diğer kombinasyonlara kıyasla daha düşük hata değerleri (MSE: 143.85, RMSE: 11.99, MAE: 9.03, MAPE: 0.21) ile en iyi tahmin performansını sergilemiştir. Hiperparametre optimizasyonu sürecinde Izgara Arama (Grid Search) ve Rasgele Arama (Randomized Search) yöntemleri kullanılarak modellerin performansı daha da iyileştirilmiştir. Özellikle RO-GA+CatBoost modeli, optimizasyon sonrası R^2 değerini %89'a çıkararak en yüksek performansı göstermiştir. Bu çalışma, doğru algoritma seçimi ve hiperparametre optimizasyonunun model performansını artırmada kritik bir rol oynadığını ortaya koymuştur. CatBoost ve KNN algoritmaları, uygun hiperparametre optimizasyon teknikleri (örneğin Grid Search veya Randomized Search) ile entegre edildiğinde, modelin tahmin doğruluğu üzerinde anlamlı bir iyileşme sağlamaktadır. Gerçekleştirilen optimizasyon süreçleri, modelin karar sınırlarını veri setinin yapısına daha iyi uyarlamasına olanak tanıyarak genelleme performansını artırmakta ve aşırı öğrenme riskini azaltmaktadır. Elde edilen bulgular, benzer regresyon problemlerinde algoritma seçimi ve optimizasyon adımlarının titizlikle planlanması gerektiğini ortaya koymaktadır.

Anahtar Kelimeler: Askıda katı madde (AKM), hiperparametre optimizasyonu, makine öğrenmesi, regresyon analizi, su kalitesi.

Abstract

Water, a vital element for the survival of life and sustainability of ecosystem balance, can undergo changes in its physical and chemical properties under the influence of environmental and anthropogenic factors. Accurate estimation of suspended solids (SSM) concentration, one of the basic indicators of water quality, is of critical importance for sustainable water management. In this study, a stacking model-based approach is proposed for the estimation of SSM concentration using water quality data from the Filyos

*Sorumlu yazarın e-posta adresi: berna.oaksoy@beun.edu.tr

Rukiye Uzun Arslan orcid.org/0000-0002-2082-8695

Berna Aksoy orcid.org/0000-0001-6925-1594

İrem Şenyer Yapıcı orcid.org/0000-0003-0655-340X



Stream Sub-basin in the Western Black Sea Basin. In the model, random forest (RF) and gradient boosting (GB) algorithms are used as base models, and K-nearest neighbor (KNN), multilayer perceptron (MLP), categorical boosting (CatBoost), adaptive boosting (AdaBoost), and Bagging algorithms are used as meta-models. Model performance was evaluated with metrics such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and correlation coefficient (R^2). According to the results, the highest R^2 value (87%) was obtained by the RF-GB+CatBoost model. This model exhibited the best prediction performance with lower error values (MSE: 143.85, RMSE: 11.99, MAE: 9.03, MAPE: 0.21) compared to other combinations. The performance of the models was further improved by using Grid Search and Randomized Search methods in the hyperparameter optimization process. In particular, the RF-GB+CatBoost model showed the highest performance by increasing the R^2 value to 89% after optimization. This study revealed that the correct algorithm selection and hyperparameter optimization play a critical role in improving model performance. CatBoost and KNN algorithms, integrated with appropriate hyperparameter optimisation techniques (e.g. Grid Search or Randomised Search), provide a significant improvement on the prediction accuracy of the model. The optimisation processes performed allow the model to better adapt its decision boundaries to the structure of the dataset, improving generalisation performance and reducing the risk of overlearning. The findings suggest that algorithm selection and optimisation steps should be carefully planned for similar regression problems.

Keywords: Suspended solids (SS), hyperparameter optimization, machine learning, regression analysis, water quality.

1. Giriş

Nüfus artışı, sanayileşme ve kentleşme gibi insan kaynaklı faaliyetler, su kaynakları üzerinde giderek artan bir baskı oluşturmaktadır. Bu durum, su kalitesinin düşmesine ve ekosistem dengelerinin bozulmasına yol açmaktadır. Özellikle yeterli arıtma işlemlerinden geçirilmeden doğrudan alıcı ortamlara deşarj edilen atık sular, su kaynaklarının kirlenmesine ve biyolojik çeşitliliğin azalmasına neden olmaktadır. Su kalitesindeki bozulmaların çok boyutlu etkileri, yalnızca çevresel dengeleri tehdit etmekle kalmayıp aynı zamanda halk sağlığı riskleri oluşturmakta ve ekonomik istikrarı derinden etkileyebilmektedir. Bu nedenle, su kaynaklarının sürdürülebilir yönetimi ve su kalitesinin korunması günümüzde küresel ölçekte öncelikli bir araştırma konusu haline gelmiştir.

Özellikle akarsu ve göl gibi yüzeysel su kaynaklarında gözlemlenen kalite değişimleri, etkili yönetim stratejileri geliştirilmesini zorunlu kılmaktadır. Bu süreçlerin karmaşıklığı, geleneksel doğrusal modelleme yöntemlerinin yetersiz kalmasına neden olmaktadır. Örneğin, akarsulardaki katı madde taşınımının neden olduğu morfolojik değişimler, hem su kalitesinin bozulmasına hem de mühendislik yapılarında hasara yol açabilmektedir (Mete vd. 2022). Bu tür karmaşık sistemlerin modellenmesinde karşılaşılan temel zorluk, parametreler arasındaki doğrusal olmayan dinamik ilişkilerin geleneksel stokastik ve deterministik yöntemlerle tam olarak temsil edilememesidir. Söz konusu yöntemlerin doğrusal ilişki varsayımına dayanması, gerçek dünyada gözlemlenen karmaşık hidrolojik süreçlerin modellenmesini sınırlandırmaktadır (Gu vd. 2016, Li vd. 2018, Lu ve Ma 2020, Varol 2020). Bu sınırlamalar, son yıllarda makine öğrenmesi (MÖ) tabanlı yaklaşımların, su kalitesi modelleme ve öngörü süreçlerinde giderek daha fazla tercih edilmesine yol

açmıştır. Özellikle yapay sinir ağları (YSA), destek vektör makineleri (DVM) ve rastgele orman (RO) gibi algoritmalar; doğrusal olmayan etkileşimleri tanımlayabilme kapasiteleri, yüksek boyutlu veri kümelerinde başarılı tahmin performansı sunmaları ve ölçeklenebilirlik açısından sağladıkları avantajlar sayesinde, su kalitesi parametrelerinin tahminde geleneksel yöntemlere kıyasla daha yüksek doğruluk oranlarıyla öne çıkmaktadır (Yadav vd. 2018, Adiat vd. 2020, Tiyasha vd. 2020, Vijay ve Kamaraj 2021, Gaur vd. 2021, Ewusi vd. 2021, Rizal vd. 2022). Ayrıca, bu modellerin farklı çevresel senaryolara kolaylıkla uyarlanabilir olması ve gerçek zamanlı tahmin yetenekleri, su kaynaklarının sürdürülebilir yönetimine yönelik karar destek sistemlerinde önemli katkılar sağlamaktadır. Bu çerçevede, toplam çözünmüş madde, kimyasal oksijen ihtiyacı, toplam sertlik, askıda katı madde (AKM), pH seviyesi, elektriksel iletkenlik, tuzluluk ve toplam azot gibi çeşitli su kalite parametreleri, MÖ teknikleri aracılığıyla etkin biçimde tahmin edilebilmektedir. Özellikle AKM, suyun fiziksel özelliklerini doğrudan yansıtmaması ve çevresel faktörlerden yüksek oranda etkilenmesi nedeniyle, literatürde sıklıkla incelenen ve tahmin edilmesine yönelik çalışmaların yoğunlaştığı başlıca parametrelerden biri olarak öne çıkmaktadır. Literatürde gerçekleştirilen çok sayıda çalışmada, farklı MÖ algoritmalarının AKM parametresine ilişkin tahminlerde yüksek doğruluk düzeylerine ulaştığı çeşitli örneklerle ortaya konulmuştur (Bayram ve Kanal 2015, Kisi ve Zounemat-Kermani 2016, Aksoy vd. 2019, Baghapour vd. 2020, Samantaray ve Sahoo 2021).

Bununla birlikte, farklı coğrafi bölgelerde, değişken çevresel koşullar altında ve özgün veri kümeleri kullanılarak gerçekleştirilecek yeni uygulamalara duyulan ihtiyaç devam etmektedir. Bu çalışmada, söz konusu boşluğu doldurmak amacıyla, Batı Karadeniz Havzası'nda yer alan Filyos Nehri'ne ait

ken, meta-model ise bu temel modellerin çıktılarını birleştirerek nihai tahmini üretmektedir. Bu iki katmanlı yapı, temel modellerin bireysel hatalarını dengeleyerek, modelin genel performansını artırmayı ve aşırı öğrenme riskini azaltmayı amaçlamaktadır (Akkur ve Öztürk 2023). Gerçekleştirilen bu çalışmada, temel model olarak RO ve GA algoritmaları seçilmiştir. Bu algoritmalar, veri setindeki farklı özellikleri ve örüntüleri yakalama yetenekleri nedeniyle tercih edilmiştir. Temel modeller tarafından üretilen tahminler, meta-modele girdi olarak kullanılmadan önce, K-en yakın komşu (KNN), çok katmanlı algılayıcı (ÇKA), kategorik artırma (CatBoost), uyarlamalı artırma (AdaBoost) ve Torbalama (Bagging) algoritmalarıyla birleştirilmiştir. Bu sayede, modelin farklı MÖ algoritmalarıyla entegrasyonuna olanak tanıyan esnek ve uyarlanabilir bir yapı geliştirilmiştir.

2.3. Regresyon Algoritmaları

Bu bölümde, yığma modelinin oluşturulmasında kullanılan regresyon algoritmaları ve bunların temel özellikleri ayrıntılı olarak açıklanmaktadır.

2.3.1. Gradyan Arttırma

GA, zayıf tahmin modellerini birleştirerek güçlü bir regresyon modeli oluşturan bir MÖ yöntemidir. Bu yöntemde, her bir modelin eğitimi, bir önceki modelin oluşturduğu hata terimlerini minimize etmeye odaklanmakta, bu sayede öğrenme süreci aşamalı bir şekilde optimize edilmektedir. Ek bir hiperparametre optimizasyon sürecine gerek duymadan yüksek doğrulukta tahminler üretebilen GA, özellikle yüksek boyutlu ve karmaşık yapıya sahip veri setlerinde regresyon problemleri için etkili bir çözüm sunmaktadır (Al-Haddad vd. 2024).

2.3.3. Rasgele Orman

RO, karar ağaçlarını temel alan denetimli bir topluluk öğrenme algoritmasıdır. Bu algoritmada, birden fazla KA eğitilmekte ve her bir ağacın ürettiği tahminlerin ortalaması alınarak nihai çıktı belirlenmektedir. RO, modelin genelleme yeteneğini artırmak ve aşırı öğrenme riskini azaltmak için Torbalama (Bagging, bootstrapaggregating) ve rastgele alt uzay yöntemlerini kullanmaktadır. Bu yöntemler sayesinde, paralel olarak oluşturulan bağımsız KA hem modelin doğruluğunu artırmakta hem de hesaplama maliyetlerini optimize etmektedir (Shanmugasundar 2021, Arslan ve Yapıcı 2024).

2.3.4. K-En Yakın Komşu

KNN algoritması, regresyon analizinde sıklıkla tercih edilen, hesaplama açısından pratik ve etkili bir yöntemdir. Bu

algoritma, yeni bir veri noktasının en yakın K komşusunu belirleyerek, ilgili komşuların hedef değişken değerlerinin ortalamasını almak suretiyle tahmin yapmaktadır. Komşuların belirlenmesinde Öklid, Manhattan ve Minkowski gibi farklı mesafe ölçütleri kullanılmaktadır. KNN'nin parametrik olmayan ve kolay uygulanabilir bir yapıya sahip olması önemli bir avantaj sağlarken, büyük veri kümelerinde artan bellek ve hesaplama maliyeti dezavantaj oluşturmaktadır (Özcan 2020, Yapıcı ve Arslan 2024).

2.3.5. Çok Katmanlı Algılayıcı

ÇKA, bir giriş katmanı, bir çıkış katmanı ve en az bir gizli katmandan oluşan ileri beslemeli bir YSA'dır. En yaygın kullanılan ÇKA modeli, tek bir gizli katmana sahip üç katmanlı bir yapıdır. Bu ağlarda katmanlar tamamen bağlantılıdır; her bir nöron, önceki ve sonraki katmandaki tüm nöronlarla etkileşim halindedir. Bu yapıda bilgi akışı yalnızca ileri yönde gerçekleşmekte olup, geri besleme mekanizması bulunmamaktadır (Yapıcı vd. 2021, Zhang 2024).

2.3.6. CatBoost

CatBoost, kategorik ve sayısal verilerle çalışan denetimli bir MÖ algoritmasıdır. Özellikle kategorik değişkenlerin işlenmesini iyileştirmek ve modele entegrasyonunu optimize etmek üzere geliştirilmiştir. Bu doğrultuda algoritma, verinin iki değerli forma dönüştürülmesi (binarize edilmesi), kategorik değişkenlerin sayısal forma dönüştürülmesi ve karar ağaçlarıyla özellik birleştirme işlemlerinden oluşan üç aşamalı bir süreç izlemektedir. Bu süreç, modelin verimliliğini artırırken aşırı öğrenme riskini azaltarak genelleme performansını iyileştirmektedir (Singh vd. 2024).

2.3.7. AdaBoost

AdaBoost, zayıf öğrencilerin ardışık olarak eğitilmesi yoluyla hata oranını minimize etmeyi amaçlayan bir topluluk öğrenme yöntemidir. Modelleme sürecinde, önceki aşamalarda yüksek hata oranına sahip örnekler belirlenerek bu örnekler daha yüksek ağırlık atanmaktadır. Bu yaklaşım, sonraki öğrencinin hata payı yüksek örnekler üzerinde daha yüksek doğrulukla genelleme yapmasını sağlamaktadır. Modelin performansı, tahmin edilen ve gerçek değerler arasındaki hata göz önünde bulundurularak iteratif bir şekilde iyileştirilmekte ve süreç, belirlenen hata eşiğine ulaşılan kadar devam etmektedir. İteratif bir yapıya sahip olan AdaBoost, hata oranını kademeli olarak azaltarak modelin genelleme kapasitesini artırmakta ve tahmin doğruluğunu optimize etmektedir (Shanmugasundar 2021).

2.3.8. Torbalama

Torbalama algoritması, birden fazla zayıf öğrenicinin farklı veri alt kümeleri üzerinde paralel olarak eğitilmesi ve tahminlerin birleştirilmesi prensibine dayanan bir topluluk öğrenme yöntemidir. Bu yaklaşım, bootstrap örnekleme ve değişken alt kümelerinin kullanımı yoluyla model çeşitliliğini artırmakta ve aşırı öğrenme riskini azaltmaktadır. Torbalama, öngörülen değerlerin ortalamasını alarak tahmin doğruluğunu artırmakta ve bu sayede modelin genelleme performansını iyileştirmektedir (Moeini 2024).

2.4. Performans Kriterleri

Bu çalışmada, model performansını değerlendirmek için çeşitli hata ve doğruluk metriklerinden yararlanılmıştır. Değerlendirme sürecinde, temel performans ölçütleri olarak ortalama karesel hata (Mean Square Error, MSE), kök ortalama kare hatası (Root Mean Square Error, RMSE), ortalama mutlak hata (Mean Absolute Error, MAE) ve ortalama mutlak yüzde hata (Mean Absolute Percentage Error, MAPE) tercih edilmiştir. Ayrıca, modelin açıklayıcılık gücünü belirlemek amacıyla korelasyon katsayısı (R^2) da önemli bir karar ölçütü olarak ele alınmıştır. İlgili metriklere ait matematiksel eşitsizlikler aşağıda verilmiştir:

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

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

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

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

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

Burada y_i i. gözleme ait gerçek (gözlemlenen) değerini, $\hat{y}_i$ i. gözleme ait model tarafından tahmin edilen değerini, $\bar{y}$ tüm gözlemlerinin aritmetik ortalamasını ve n ise toplam gözlem sayısını temsil etmektedir.

Model performansının değerlendirilmesinde, R^2 değerinin yüksek olması modelin açıklayıcılık gücünün yüksek olduğunu ve hedef değişken üzerindeki varyansı iyi açıkladığını göstermektedir. Buna karşılık, MSE, RMSE, MAE ve MAPE gibi hata ölçütlerinin düşük değerler alması, modelin tahmin hatalarının minimal düzeyde olduğunu ve dolayısıyla genel performansının daha yüksek olduğunu ortaya koymaktadır (Gültepe 2019).

3. Bulgular ve Tartışma

Bu çalışmada, Batı Karadeniz Havzası'nın en büyük alt havzası olan Filyos Çayı Alt Havzası'nda yer alan Filyos Çayı'na ait su kalitesi verileri kullanılarak, sudaki AKM konsantrasyonunun tahmini için MÖ tabanlı bir yığma modeli önerilmiştir. Önerilen modelde, iki farklı algoritma temel model, bir algoritma ise yığma meta-model olarak kullanılmıştır. Temel model olarak RO ve GA tercih edilirken, meta-model olarak KNN, ÇKA, CatBoost, AdaBoost ve Torbalama algoritmalarından yararlanılmıştır. Model değerlendirme sürecinde RO-GA+KNN, RO-GA+ÇKA, RO-GA+CatBoost, RO-GA+AdaBoost ve RO-GA+Torbalama şeklindeki algoritma kombinasyonları incelenmiştir.

Modelleme sürecinde kullanılan veri seti, Filyos nehri üzerinde belirlenen 5 farklı istasyondan alınan su örneklerine ait fiziksel ve kimyasal parametreleri içermekte olup, iletkenlik, sıcaklık ($^{\circ}\text{C}$), pH, çözülmüş oksijen (ÇO), kimyasal oksijen ihtiyacı (KOİ), bulanıklık, toplam organik karbon (TOC), amonyum (NH_4^+), kalsiyum (Ca), magnezyum (Mn), sertlik, fosfat (PO_4^{3-}), nitrit, nitrat, alüminyum (Al), mangan (Mn), demir (Fe), krom (Cr), kurşun (Pb) ve çinko (Zn) değişkenleri giriş (bağımsız değişken) olarak kullanılmış; AKM değeri ise hedef (bağımlı değişken) değişken olarak modellenmiştir. Çizelge 2'de, F1 istasyonundan alınan veri setinin bir kesit verilmiştir. Ayrıca F1 istasyonuna ait verilerin mevsimsel olarak minimum, maksimum ve ortalama değerleri Çizelge 3'te sunulmuştur.

Veri seti, modelin daha sağlıklı bir öğrenme süreci geçirmesi için *Standard Scaler* yöntemi ile ölçeklendirilmiş ve %80'i eğitim, %20'si test olmak üzere iki alt veri kümesine ayrılmıştır. Modelin performansı, MSE, RMSE, MAE, MAPE ve R^2 gibi performans ölçütleri ile değerlendirilmiştir. Şekil 2'de, incelenen algoritma kombinasyonlarına ait R^2 değerleri sunulmuştur.

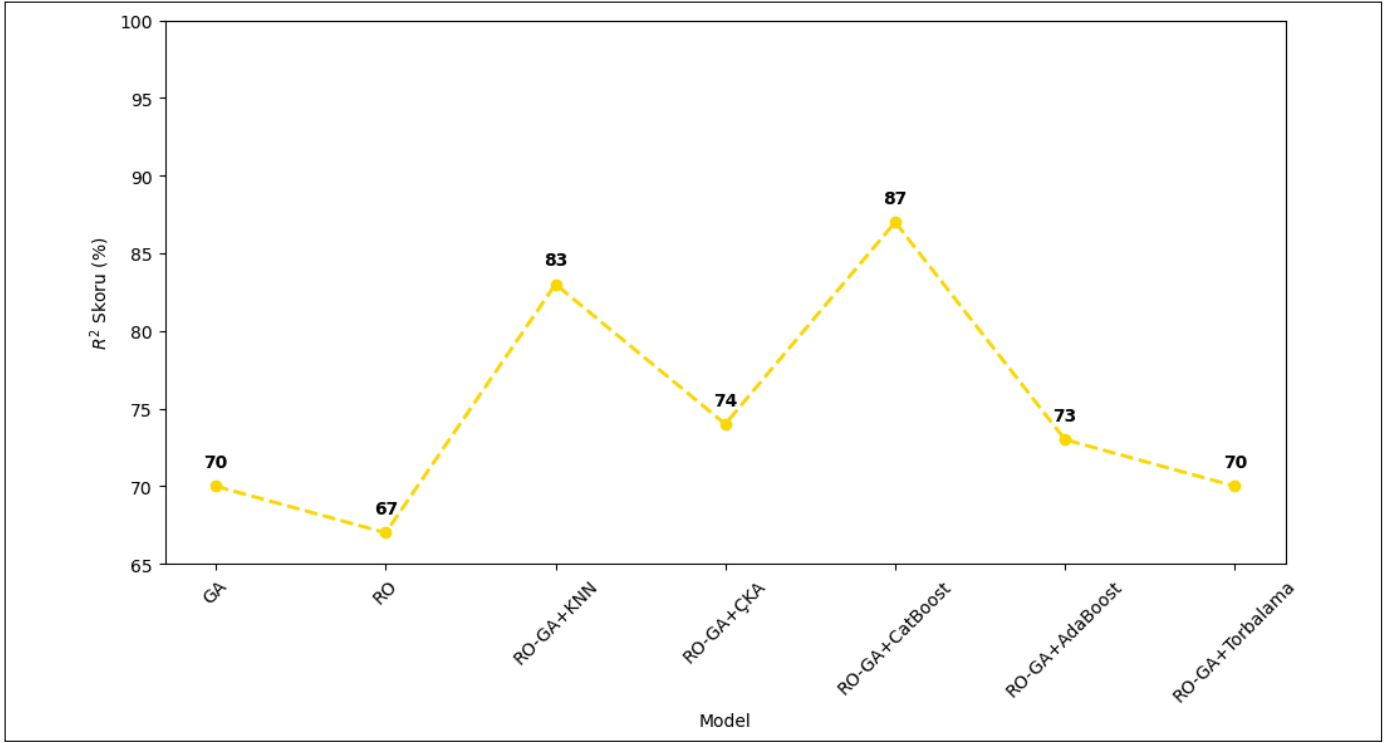
Şekil 2'de, çeşitli algoritma kombinasyonlarına ilişkin R^2 değerleri karşılaştırıldığında, modeller arasında istatistiksel olarak anlamlı performans farkları olduğu dikkati çekmektedir. Özellikle, RO-GA+CatBoost modeli %87'lik R^2 değeri ile en yüksek doğrusal uyum başarımını sergilemiştir. Bu sonuç, RO-GA+CatBoost modelinin tahmin performansı açısından diğer kombinasyonlara göre daha üstün olduğunu göstermektedir. RO-GA+KNN modeli %83'lük R^2 değeri ile yüksek bir tahmin performansı ortaya koymuştur. RO-GA+ÇKA ve RO-GA+AdaBoost modelleri ise sırasıyla %74 ve %73'lük R^2 değerlerine ulaşmış ve diğer modellere kıyasla daha düşük bir performans sergilemiştir.

Çizelge 2. Filyos nehri üzerindeki F1 istasyondan alınan su örneklerine ait fiziksel ve kimyasal parametreler.

Parametre												
İletkenlik	0.52	0.57	0.52	0.52	0.52	0.53	0.53	0.41	0.48	0.48	0.55	0.55
Sıcaklık	23.40	21.00	19.40	15.00	10.10	9.80	14.00	16.40	25.70	26.80	23.40	23.30
pH	8.03	8.13	8.26	8.21	8.06	8.35	8.25	8.41	8.35	8.15	8.42	8.40
ÇO	5.28	3.11	4.14	3.42	3.94	3.98	5.35	9.77	7.94	7.81	9.19	9.24
AKM	62.00	116.00	30.00	52.00	61.00	49.00	48.00	75.00	120.00	162.00	110.00	109.00
KOİ	126.00	92.00	97.00	94.00	96.00	91.00	93.00	102.86	73.47	39.84	88.00	91.00
Bulanıklık	98.60	86.40	95.60	78.40	68.30	69.10	46.70	28.12	63.25	35.72	17.09	64.12
TOC	2.27	2.31	2.87	2.98	3.46	3.41	2.86	2.95	5.03	3.86	3.49	5.26
Amonyum	1.09	1.07	1.28	1.13	0.46	0.28	0.19	0.17	0.96	1.25	0.99	1.06
Kalsiyum	52.31	50.68	51.64	52.31	54.86	55.16	55.13	55.23	55.25	55.31	53.48	51.68
Magnezyum	21.30	23.41	25.14	22.57	21.63	20.45	22.46	20.36	26.42	26.41	20.49	21.46
Sertlik	347.89	352.46	321.48	364.21	355.41	352.16	351.36	359.69	384.69	384.93	354.26	345.35
Fosfat	0.53	0.32	0.45	0.28	0.43	0.51	0.49	0.41	0.78	0.11	0.42	0.32
Nitrit	1.28	1.34	1.38	1.43	1.56	1.62	1.68	1.79	1.73	1.54	1.38	1.24
Nitrat	6.91	7.06	6.81	6.53	5.87	5.64	5.32	5.28	6.73	6.16	6.54	6.82
Al	7.65	10.96	12.80	17.50	25.40	15.24	4.33	4.64	2.14	5.53	4.43	8.61
Mn	0.22	0.15	0.35	0.44	0.37	0.37	0.33	0.11	0.64	0.02	0.09	0.15
Fe	22.45	20.61	18.46	13.52	19.45	21.05	66.85	74.08	41.02	10.96	12.05	11.68
Cr	0.09	0.14	0.43	0.26	0.32	0.36	0.13	0.03	0.28	0.11	0.05	0.07
Pb	0.02	0.01	0.02	0.02	0.01	0.01	0.02	0.00	0.05	0.01	0.08	0.02
Zn	0.25	0.55	0.37	0.25	0.21	0.13	0.59	0.66	0.39	0.59	0.29	0.10

Çizelge 3. F1 istasyonuna ait verilerin mevsimsel olarak istatistiksel analizi.

Parametre	Sonbahar			Kış			İlkbahar			Yaz		
	Min.	Ort.	Mak.	Min.	Ort.	Mak.	Min.	Ort.	Mak.	Min.	Ort.	Mak.
İletkenlik	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Sıcaklık	19.4	21.4	23.4	9.8	12.4	15.0	14.0	19.9	25.7	23.3	25.1	26.8
pH	8.0	8.1	8.3	8.2	8.3	8.4	8.3	8.3	8.4	8.2	8.3	8.4
ÇO	4.1	4.7	5.3	3.4	3.7	4.0	5.4	6.6	7.9	7.8	8.5	9.2
AKM	30.0	46.0	62.0	49.0	50.5	52.0	48.0	84.0	120.0	109.0	135.5	162.0
KOİ	97.0	111.5	126.0	91.0	92.5	94.0	73.5	83.2	93.0	39.8	65.4	91.0
Bulanıklık	95.6	97.1	98.6	69.1	73.8	78.4	46.7	55.0	63.3	35.7	49.9	64.1
TOC	2.3	2.6	2.9	3.0	3.2	3.4	2.9	3.9	5.0	3.9	4.6	5.3
Amonyum	1.1	1.2	1.3	0.3	0.7	1.1	0.2	0.6	1.0	1.1	1.2	1.2
Kalsiyum	51.6	52.0	52.3	52.3	53.7	55.2	55.1	55.2	55.2	51.7	53.5	55.3
Magnezyum	21.3	23.2	25.1	20.5	21.5	22.6	22.5	24.4	26.4	21.5	23.9	26.4
Sertlik	321.5	334.7	347.9	352.2	358.2	364.2	351.4	368.0	384.7	345.4	365.1	384.9
Fosfat	0.5	0.5	0.5	0.3	0.4	0.5	0.5	0.6	0.8	0.1	0.2	0.3
Nitrit	1.3	1.3	1.4	1.4	1.5	1.6	1.7	1.7	1.7	1.2	1.4	1.5
Nitrat	6.8	6.9	6.9	5.6	6.1	6.5	5.3	6.0	6.7	6.2	6.5	6.8
Al	7.7	10.2	12.8	15.2	16.4	17.5	2.1	3.2	4.3	5.5	7.1	8.6
Mn	0.2	0.3	0.4	0.4	0.4	0.4	0.3	0.5	0.6	0.0	0.1	0.2
Fe	18.5	20.5	22.5	13.5	17.3	21.1	41.0	53.9	66.9	11.0	11.3	11.7
Cr	0.1	0.3	0.4	0.3	0.3	0.4	0.1	0.2	0.3	0.1	0.1	0.1
Pb	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
Zn	0.2	0.3	0.4	0.1	0.2	0.3	0.4	0.5	0.6	0.1	0.3	0.6



Şekil 2. Modellerin R² skorlarının karşılaştırılması.

Çizelge 4. Modellere ait regresyon performans metrikleri.

MODELLER	MSE	RMSE	MAE	MAPE
RO	372.91	19.31	16.34	32.78
GA	343.96	18.54	15.38	32.66
RO-GA+KNN	189.99	13.78	11.13	0.23
RO-GA+ÇKA	294.10	17.14	13.56	0.30
RO-GA+CatBoost	143.85	11.99	9.03	0.21
RO-GA+AdaBoost	295.83	17.19	12.53	0.20
RO-GA+Torbalama	336.69	18.34	15.74	0.26

GA ve RO modelleri ise %70 ve %67'lik R² değerleri ile temel yöntemler arasında yer almış ve daha sınırlı bir uyum başarımı sunmuştur. Öte yandan, yalnızca kombinasyon yöntemleri dikkate alındığında, en düşük R² değeri %70 ile RO-GA+Torbalama modelinde elde edilmiştir. Elde edilen bulgular, yığma modelinde tercih edilen algoritma kombinasyonlarının önerilen modelin performansı üzerinde belirleyici bir etkiye sahip olduğunu ortaya koymaktadır. Bununla birlikte, modellerin tahmin performanslarını daha ayrıntılı bir şekilde incelemek amacıyla farklı algoritma kombinasyonları için çeşitli regresyon hata metrikleri hesaplanmış ve sonuçlar Çizelge 4'te özetlenmiştir.

Çizelge 4'te sunulan regresyon performans metrikleri incelendiğinde, temel yöntemler olan RO ve GA modellerinin hata değerlerinin, kombinasyon yöntemlerine kıyasla daha yüksek olduğu görülmüştür. Özellikle, RO modeli 372.91 MSE, 19.31 RMSE, 16.34 MAE ve 32.78 MAPE değerleri ile en yüksek hata oranlarına sahip olurken, GA modeli 343.96 MSE, 18.54 RMSE, 15.38 MAE ve 32.66 MAPE değerleri ile RO'ya göre bir miktar daha iyi bir performans sergilemiştir. Buna karşılık, RO ve GA yöntemlerinin farklı algoritmalarla kombinasyonu sonucunda tahmin performansında belirgin bir artış gözlemlenmiştir. Analiz sonuçlarına göre, en düşük hata değerlerine sahip olan model, 143.85 MSE, 11.99 RMSE, 9.03 MAE ve 0.21 MAPE de-

Çizelge 5. Grid Search ve Randomized Search yöntemleri ile RO ve GA modelleri için hiperparametre değerleri.

	RO	GA
Grid Search	n_estimators: [300, 400, 500] max_depth: [12, 15, 20] min_samples_split: [2, 5] min_samples_leaf: [1, 2]	n_estimators: [300, 400, 500] learning_rate: [0.01, 0.05, 0.1] max_depth: [3, 5, 7]
Randomized Search	n_estimators: [300, 400, 500, 600] max_depth: [12, 15, 20, 25] min_samples_split: [2, 4, 6] min_samples_leaf: [1, 2, 3]	n_estimators: [300, 400, 500, 600] learning_rate: [0.01, 0.05, 0.1] max_depth: [3, 5, 7, 9]

ğerleri ile RO-GA+CatBoost modeli olarak belirlenmiştir. Bu bulgular, RO-GA+CatBoost modelinin, diğer modellere kıyasla daha yüksek bir tahmin performansı sergilediğini ve veri setindeki karmaşık ilişkileri daha etkili bir şekilde modelleyebildiğini ortaya koymaktadır. RO-GA+KNN modeli ise 189.99 MSE, 13.78 RMSE, 11.13 MAE ve 0.23 MAPE değerleri ile nispeten iyi bir performans göstermiştir. Ancak, bu modelin hata metrikleri RO-GA+CatBoost modeline kıyasla daha yüksek olduğundan, tahmin doğruluğu açısından ikinci sırada yer aldığı görülmektedir. RO-GA+KNN modelinin performansı, özellikle büyük ölçekli veri setlerinde hesaplama karmaşıklığı ve model esnekliği açısından sınırlı kalabileceğini düşündürmektedir. Diğer taraftan, RO-GA+ÇKA ve RO-GA+AdaBoost modelleri sırasıyla 294.10 ve 295.83 MSE değerleri ile daha yüksek hata oranlarına sahip olmuştur. Bu modellerin MAE ve MAPE değerleri de (sırasıyla 13.45 ve 0.28 için RO-GA+ÇKA; 13.52 ve 0.29 için RO-GA+AdaBoost) RO-GA+CatBoost ve RO-GA+KNN modellerine kıyasla daha düşük performans sergilediğini göstermektedir. Bu durum, söz konusu modellerin veri setindeki karmaşık ilişkileri modellemede yetersiz kaldığını veya hiperparametre optimizasyonunun yeterince etkin yapılamadığını düşündürmektedir. En yüksek hata değerleri ise 336.69 MSE, 18.34 RMSE, 15.74 MAE ve 0.26 MAPE ile RO-GA+Torbalama modelinde gözlemlenmiştir. Bu modelin performansının diğer modellere kıyasla önemli ölçüde düşük olması, tahmin doğruluğu açısından sınırlı bir kapasiteye sahip olduğunu göstermektedir.

Model performansını daha da iyileştirmek ve genelleme yeteneğini artırmak amacıyla, hiperparametre optimizasyonu süreci gerçekleştirilmiştir. Bu süreç, modellerin eğitim aşamasında kullanılan parametrelerin sistematik olarak ayarlanmasını ve en uygun kombinasyonların belirlenmesini içermektedir. Bu amaçla, Grid Search ve Randomized Search gibi yaygın olarak kullanılan optimizasyon yöntemleri tercih edilmiştir. Grid Search yöntemi, belirli bir hiperpara-

metre uzayını kapsamlı bir şekilde tarayarak en iyi kombinasyonu bulmayı hedeflerken, Randomized Search yöntemi ise rastgele örnekleme yoluyla daha geniş bir hiperparametre uzayını daha verimli bir şekilde keşfetmeyi amaçlamaktadır.

RO ve GA için belirlenen hiperparametre aralıkları ve değerleri, Çizelge 5'te ayrıntılı olarak sunulmuştur. Bu parametrelerden *n_estimators*, modelde oluşturulacak toplam ağaç sayısını belirlerken; *max_depth*, her bir ağacın alabileceği maksimum derinliği sınırlayarak modelin karmaşıklığını kontrol etmektedir. *min_samples_split* ve *min_samples_leaf* parametreleri ise, her bir bölünme ve yaprak düğümünde bulunması gereken minimum örnek sayısını düzenleyerek aşırı uyum riskini azaltmayı amaçlamaktadır. GA modellerinde *learning_rate* parametresi, her bir zayıf öğrencinin model üzerindeki katkı oranını ayarlayarak genel öğrenme sürecinin hızını ve performansını optimize etmektedir. Hiperparametrelerin belirlenmesinde kullanılan Randomized Search yöntemi ise, klasik Grid Search yöntemine kıyasla daha geniş bir parametre uzayını rastgele örnekleyerek daha hızlı ve verimli bir arama süreci sağlamaktadır. Belirlenen bu hiperparametreler, modellerin karmaşıklığı, öğrenme dinamikleri ve genelleme yeteneği üzerinde doğrudan etkili olup model performansını belirleyen temel unsurlar arasında yer almaktadır.

Hiperparametre optimizasyonu sürecinin ardından, modellerin hata metriklerinde önemli iyileşmeler kaydedilmiş ve bu sonuçlar Çizelge 6'da verilmiştir. Özellikle RO-GA+CatBoost ve RO-GA+KNN modellerinde, optimizasyon sonrası MSE, RMSE, MAE ve MAPE değerlerinde belirgin düşüşler gözlemlenmiştir. Bu bulgular, hiperparametre optimizasyonunun model performansını artırmada kritik bir rol oynadığını ve hata oranlarını minimize etmede etkili bir yöntem olduğunu göstermektedir.

Çizelge 6'da sunulan sonuçlar incelendiğinde, Grid Search yöntemi ile optimize edilen modeller arasında en iyi per-

Çizelge 6. Grid Search ve Randomized Search yöntemlerinin model performansları üzerindeki etkisi.

	MODEL	MSE	RMSE	MAE	MAPE	R ²
Grid Search	RO-GA+KNN	185.15	13.60	11.10	0.20	0.83
	RO-GA+ÇKA	321.72	17.93	14.21	0.32	0.71
	RO-GA+CatBoost	126.14	11.23	9.71	0.21	0.89
	RO-GA+AdaBoost	228.93	15.13	12.78	0.22	0.79
	RO-GA+Torbalama	271.09	16.46	13.12	0.24	0.75
Randomized Search	RO-GA+KNN	185.15	13.60	11.10	0.20	0.83
	RO-GA+ÇKA	312.80	17.68	13.91	0.31	0.72
	RO-GA+CatBoost	188.66	13.73	12.98	0.23	0.83
	RO-GA+AdaBoost	295.17	17.18	13.66	0.22	0.73
	RO-GA+Torbalama	299.24	17.29	13.51	0.24	0.73

formansın CatBoost algoritması ile elde edildiği görülmektedir. Bu model, R² değeri olarak 0.89 ile diğer modellere kıyasla belirgin bir üstünlük sergilemiştir. Ayrıca, CatBoost modelinin hata metrikleri de oldukça düşük değerler göstermiştir: MSE (126.14), RMSE (11.23) ve MAE (9.71). Ancak Randomized Search uygulandığında, CatBoost'un performansında dikkat çekici bir düşüş gözlemlenmiştir (MSE: 188.66, R²: 0.83). Bu sonuçlar, CatBoost'un hem doğrusal hem de doğrusal olmayan ilişkileri etkili bir şekilde modelleyebildiğini ve Grid Search ile hiperparametre optimizasyonu sonrasında tahmin performansını önemli ölçüde artırabildiğini ortaya koymaktadır.

Öte yandan, Grid Search ile optimize edilen modeller arasında en düşük performans, ÇKA modelinde gözlemlenmiştir. Bu model, R² değeri olarak 0.71, MSE değeri olarak 321.72 ve RMSE değeri olarak 17.93 ile diğer modellere kıyasla daha yüksek hata oranlarına sahip olmuştur. Randomized Search'te ise ÇKA modelinin performansında hafif bir iyileşme görülmüştür (MSE: 312.80, R²: 0.72), bu da bazı modellerde rastgele aramanın daha iyi sonuçlar verebileceğini göstermektedir. Bu durum, ÇKA modelinin veri setindeki karmaşık ilişkileri yakalamada yetersiz kaldığını veya hiperparametre optimizasyonunun model performansı üzerindeki etkisinin algoritmaya bağımlı olduğunu göstermektedir.

Randomized Search yöntemi ile optimize edilen modellerin performansları incelendiğinde ise en yüksek doğruluk oranlarına KNN (R² = 0.83, MSE = 185.15, RMSE = 13.60) ve CatBoost (R² = 0.83, MSE = 188.66, RMSE = 13.73) algoritmalarının sahip olduğu görülmektedir. Her iki model de benzer R² değerleri ve hata metrikleri ile nispeten yüksek bir tahmin performansı sergilemiştir. Bununla birlikte, Rando-

mized Search ile optimize edilen modeller arasında en düşük performans yine ÇKA modelinde gözlemlenmiştir. Bu model, R² değeri olarak 0.72, MSE değeri olarak 312.80 ve RMSE değeri olarak 17.68 ile diğer modellere kıyasla daha zayıf sonuçlar vermiştir.

Grid Search ve Randomized Search optimizasyon yöntemlerinin uygulanması, genel olarak tüm modellerde tahmin doğruluğunu artırmış ve R² değerlerinde anlamlı iyileşmelere katkı sağlamıştır. Özellikle Grid Search yöntemi, hata metriklerini azaltarak modellerin doğrusal uyum performansında belirgin bir artış sağlamıştır. Bu kapsamda, RO-GA+CatBoost modeli, Grid Search sonrası elde ettiği 0.89 R² skoru ile en yüksek başarıyı göstermiştir. Bununla birlikte, Randomized Search yöntemi uygulandığında, RO-GA+CatBoost modelinde hata metriklerinde kısmi bir artış ve R² değerinde yaklaşık %4 oranında bir azalma gözlemlenmiştir. Ancak bu durum, çalışmanın genel bulgularını değiştirmemekte; hiperparametre optimizasyon yönteminin seçiminde, model ve veri seti özelliklerine uygun stratejilerin belirlenmesinin performans üzerinde doğrudan etkili olduğunu göstermektedir. Elde edilen sonuçlar, önerilen yöntemlerin uygun hiperparametre optimizasyonları ile desteklendiğinde, model başarımının istikrarlı bir şekilde artırılacağına ortaya koymaktadır.

Bu bulgular, yığma modelinde kullanılan algoritma kombinasyonlarının ve uygulanan optimizasyon süreçlerinin, tahmin performansı üzerinde doğrudan ve önemli bir etkiye sahip olduğunu göstermektedir. Özellikle CatBoost ve KNN gibi algoritmalar, uygun hiperparametre optimizasyon yöntemi ile birlikte kullanıldığında, tahmin doğruluğunu önemli ölçüde artırabilmektedir. Diğer taraftan, ÇKA gibi modellerin performansının görece düşük kalması, bu

tür modellerin hiperparametre optimizasyonu süreçlerinde daha detaylı bir ayarlama gerektirdiğini veya alternatif modelleme yaklaşımlarının değerlendirilmesi gerektiğini belirtmektedir.

Sonuç olarak, bu çalışma, doğru algoritma seçimleri ve uygun hiperparametre optimizasyonu stratejilerinin, model performansını artırmada kritik bir rol oynadığını ortaya koymaktadır. Özellikle Grid Search ve Randomized Search gibi sistematik optimizasyon yöntemlerinin kullanılması, modellerin genelleme yeteneklerini geliştirerek hata oranlarını minimize etmekte etkili olmuştur. Bu bulgular, benzer regresyon problemlerinde algoritma seçimi ve optimizasyon süreçlerinin titizlikle planlanması gerektiğini vurgulamaktadır.

4. Sonuç ve Öneriler

Bu çalışmada, Filyos Çayı Alt Havzası'ndaki su kalitesi verileri kullanılarak, AKM konsantrasyonunun tahmini için yığma modeli temelli bir yaklaşım önerilmiştir. Elde edilen sonuçlar, farklı algoritma kombinasyonlarının ve hiperparametre optimizasyonu süreçlerinin model performansı üzerinde önemli etkileri olduğunu ortaya koymuştur. Çalışmada, temel öğrenici olarak RO ve GA algoritmaları, meta öğrenici olarak ise KNN, ÇKA, CatBoost, AdaBoost ve Bagging algoritmaları kullanılmıştır. En yüksek performans, RO-GA+CatBoost kombinasyonu tarafından elde edilmiştir. Bu model, %87'lik R^2 değeri ve düşük hata metrikleri (MSE: 143.85, RMSE: 11.99, MAE: 9.03, MAPE: 0.21) ile diğer modellere kıyasla belirgin bir üstünlük sergilemiştir. CatBoost'un bu başarısı, hem doğrusal hem de doğrusal olmayan ilişkileri etkili bir şekilde modelleyebilme yeteneğinden kaynaklanmaktadır. RO-GA+KNN modeli de %83'lük R^2 değeri ile iyi bir performans göstermiştir. Ancak, KNN algoritması, büyük ölçekli veri setlerinde hesaplama maliyeti açısından yüksek karmaşıklığa sahip olup, özellikle yüksek boyutlu verilerde performans açısından sınırlamalar göstermektedir.

Diğer modeller (RO-GA+ÇKA, RO-GA+AdaBoost ve RO-GA+Torbalama) ise daha düşük performans sergilemiştir. Özellikle ÇKA modelinin düşük performansı, bu modelin veri setindeki karmaşık ilişkileri yakalamada yetersiz kaldığını veya hiperparametre optimizasyonunun yetersiz etkili olmadığını düşündürmektedir. Bu durum, ÇKA gibi modellerin daha karmaşık veri setlerinde etkili bir şekilde kullanılabilmesi için daha detaylı hiperparametre optimizasyonu süreçlerine ihtiyaç duyulduğunu göstermektedir.

Öte yandan hiperparametre optimizasyonu çalışmaları, model performansını artırmada yöntem seçiminin önemini ortaya koymuştur. Özellikle, RO-GA+CatBoost modelinde Grid Search yönteminin $R^2=0.89$ ve MSE=126.14 gibi üstün performans metrikleri elde ederken, aynı modelde Randomized Search yönteminin $R^2=0.83$ ve MSE=188.66 değerleriyle nispeten daha düşük performans sergilemesi, bu durumu açıkça göstermektedir. Ayrıca, farklı modeller üzerinde gerçekleştirilen optimizasyon sonuçları da bu bulguları desteklemektedir. Randomized Search yöntemi ile RO modeli için en iyi hiperparametreler {'n_estimators': 400, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_depth': 20} olarak, GB modeli için ise {'n_estimators': 600, 'max_depth': 3, 'learning_rate': 0.01} olarak belirlenmiştir. Öte yandan, Grid Search yöntemi kullanıldığında RO modelinin en iyi hiperparametreleri {'max_depth': 12, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 300} şeklinde; GB modelinin ise {'learning_rate': 0.01, 'max_depth': 3, 'n_estimators': 300} şeklinde olduğu tespit edilmiştir. Bu sonuçlar, hiperparametre optimizasyonunun model başarımını artırmada ve hata oranlarını azaltmada etkili bir yaklaşım olduğunu bir kez daha ortaya koymaktadır. Bu bulgu, hiperparametre optimizasyonunun model başarımını artırmada ve hata oranlarını azaltmada etkili bir yaklaşım olduğunu göstermektedir.

Bu çalışma, doğru algoritma seçimi ve uygun hiperparametre optimizasyonu stratejilerinin, model performansını artırmada kritik bir rol oynadığını göstermektedir. Özellikle CatBoost ve KNN gibi algoritmalar, Grid Search veya Randomized Search gibi sistematik hiperparametre optimizasyon teknikleriyle birlikte kullanıldığında, tahmin doğruluğunda kayda değer artışlar sağlayabilmektedir. Bu artış, modelin parametre uzayında en uygun yapılandırmalara ulaşması sayesinde daha iyi genelleme yeteneği kazanmasıyla açıklanabilmektedir. Diğer taraftan, ÇKA gibi modellerin performansının görece düşük kalması, bu tür modellerin hiperparametre optimizasyonu süreçlerinde daha detaylı bir ayarlama gerektirdiğini veya alternatif modelleme yaklaşımlarının değerlendirilmesi gerektiğini belirtmektedir.

Sonuç olarak, bu çalışma, su kalitesi tahmini gibi karmaşık regresyon problemlerinde yığma modelleri ve hiperparametre optimizasyonu süreçlerinin etkinliğini ortaya koymuştur. Gelecekteki araştırmalar, farklı algoritma kombinasyonlarının ve optimizasyon yöntemlerinin daha geniş veri setleri üzerinde incelenmesini gerektirmektedir. Ayrıca, özellikle ÇKA gibi modellerin performansını iyileştirmek amacıyla daha kapsamlı hiperparametre optimizasyonu çalışmalarının yapılması önerilmektedir. Bu tür araştırmalar, model

performansını artırarak, su kalitesi yönetimi ve diğer çevresel uygulamalarda daha güvenilir tahminler yapılmasına olanak tanıyacaktı.

Yazar katkısı: Bu çalışmanın tüm aşamalarına (tasarım, veri analizi, yazım ve revizyon dahil) tüm yazarlar eşit düzeyde katkı sağlamıştır.

Etik Beyan: Bu çalışma kapsamında etik kurul onayı gerektiren herhangi bir insan veya hayvan deneyi yürütülmemiştir. Bu nedenle, çalışmada etik beyan yer almamaktadır.

5. Kaynaklar

- Adiat, KAN., Ajayi, OF., Akinlalu, AA., Tijani, IB. 2020. Prediction of groundwater level in basement complex terrain using artificial neural network: a case of Ijebu-Jesa, southwestern Nigeria. *Applied Water Science*, 10:1-14. DOI: 10.1007/s13201-019-1094-6
- Akkur, E., Öztürk, AC. 2023. Chronic kidney disease prediction with stacked ensemble-based model. *Alfa Mühendislik ve Uygulamalı Bilimler Dergisi*, 1(1):50-61.
- Aksoy, B., Özölçer, İH., Dündar, O., İlhan, S. 2019. Filyos Nehri'ndeki askıda katı madde miktarının yapay sinir ağları ile belirlenmesi. *SETSCI-Conference Proceedings*, 9:404-407. DOI: 10.36287/setsci.4.6.103
- Al-Haddad, LA., Jaber, AA., Hamzah, MN., Fayad, MA. 2024. Vibration-current data fusion and gradient boosting classifier for enhanced stat or fault diagnosis in three-phase permanent magnet synchronous motors. *Electrical Engineering*, 106(3):3253-3268. DOI: 10.1007/s00202-023-02148-z
- Arslan, RU., Yapıcı, İŞ. 2024. Farklı örnekleme tekniklerine ve farklı sınıflandırıcılara dayanarak kalp yetmezliği tanılı hastaların sağlık durumlarının incelenmesi. *EMO Bilimsel Dergi*, 14(2):35-47.
- Baghapour, MA., Shooshtarian, MR., Zarghami, M. 2020. Process mining approach of a new water quality index for long-term assessment under uncertainty using consensus-based fuzzy decision support system. *Water Resources Management*, 34:1155-1172. DOI: 10.1007/s11269-020-02489-5
- Bayram, A., Kankal, M. 2015. Artificial neural network modeling of dissolved oxygen concentration in a Turkish Watershed. *Polish Journal of Environmental Studies*, 24(4).
- Ewusi, A., Ahenkorah, I., Aikins, D. 2021. Modelling of total dissolved solids in water supply systems using regression and supervised machine learning approaches. *Applied Water Science*, 11(2):1-16. DOI: 10.1007/s13201-020-01352-7
- Gaur, S., Mishra, A., Gupta, A., Jain, A., Dave, A., Eslamian, S., ..., Graillot, D. 2021. Application of artificial neural network model for the prediction of suspended sediment load in the Large River. *Water Resources*, 48:565-575. DOI: 10.1134/S0097807821040163
- Gu, Q., Zhang, Y., Ma, L., Li, J., Wang, K., Zheng, K., ..., Sheng, L. 2016. Assessment of reservoir water quality using multivariate statistical techniques: a case study of Qiandao Lake, China. *Sustainability*, 8(3):243. DOI: 10.3390/su8030243
- Gültepe, Y. 2019. Makine öğrenmesi algoritmaları ile hava kirliliği tahmini üzerine karşılaştırmalı bir değerlendirme. *Avrupa Bilim ve Teknoloji Dergisi*, 16:8-15. DOI: 10.31590/ejosat.530347
- Kisi, O., Zounemat-Kermani, M. 2016. Suspended sediment modeling using neuro-fuzzy embedded fuzzy c-means clustering technique. *Water resources management*, 30:3979-3994. DOI: 10.1007/s11269-016-1405-8
- Li, Z., Wang, G., Wang, X., Wan, L., Shi, Z., Wanke, H., ..., Uahengo, CI. 2018. Groundwater quality and associated hydrogeochemical processes in Northwest Namibia. *Journal of Geochemical Exploration*, 186:202-214. DOI: 10.1016/j.gexplo.2017.12.015
- Lu, H., Ma, X. 2020. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, 249:126169. DOI: 10.1016/j.chemosphere.2020.126169
- Mete, B., Baki, OT., Bayram, A. 2022. Sera Deresi Havzasında (Trabzon) Askıda Katı Madde Hareketinin İncelenmesi ve Değerlendirilmesi. *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 27(1): 419-436. DOI: 10.17482/uumfd.970338
- Moeini, M. 2024. Hyperparametertuning of supervised bagging ensemble machine learning model using Bayesian optimization forestimating storm water quality. *Sustainable Water Resources Management*, 10(2):83. DOI: 10.1007/s40899-024-01064-9
- Onan, A. 2018. Parçacık sürüşü eniyilemesine dayalı yığılmuş genelleme yöntemi ve metin sınıflandırma üzerinde uygulanması. *Academic Platform-Journal of Engineering and Science*, 6(2):134-141.
- Özcan, T. 2020. Yığıllanmış özdevinimli kodlayıcılar ile göğüs kanserinin sınıflandırılması ve klasik makine öğrenme metotları ile performans karşılaştırması. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*, 36(2):151-160.
- Rizal, NNM., Hayder, G., Yusof, KA. 2022. Water quality predictive analytics using an artificial neural network with a graphical user interface. *Water*, 14(8):1221. DOI: 10.3390/w14081221
- Samantaray S. Sahoo A. 2021. Prediction of suspended sediment concentration using hybrid SVM-WOA approaches. *Geocarto International*, 37(19):5609-5635. DOI: 10.1080/10106049.2021.1920638
- Shanmugasundar, G., Vanitha, M., Çep, R., Kumar, V., Kalita, K., Ramachandran, M. 2021. A comparative study of linear, random forest and adaboost regressions for modeling non-traditional machining. *Processes*, 9(11):2015. DOI: 10.3390/pr9112015

- Singh, P., Hasija, T., Ramkumar, KR. 2024.** Leveraging ML with XGBoost, CatBoost and LGBost classifiers to optimize water quality assessment and prediction. 2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), pp. 1-6, Karnataka, India. DOI: 10.1109/ICITEICS61368.2024.10625322
- Tiyasha, T., Tran, TM, Yaseen, ZM. 2020.** A survey on river water quality modelling using artificial intelligence models: 2000-2020. *Journal of Hydrology*, 585(3731):124670. DOI: 10.1016/j.jhydrol.2020.124670
- Varol, M. 2020.** Use of water quality index and multivariate statistical methods for the evaluation of water quality of a stream affected by multiple stressors: A case study. *Environmental Pollution*, 266:115417. DOI: 10.1016/j.envpol.2020.115417
- Vijay, S., Kamaraj, K. 2021.** Prediction of water quality index in drinking water distribution system using activation functions based Ann. *Water Resources Management*, 35(2):535-553. DOI: 10.1007/s11269-020-02729-8
- Yadav, A., Chatterjee, S., Equeenuddin, SM. 2018.** Prediction of suspended sediment yield by artificial neural network and traditional mathematical model in Mahanadi river basin, India. *Sustainable Water Resources Management*, 4:745-759. DOI: 10.1007/s40899-017-0160-1
- Yapıcı, İŞ., Arslan, RU. 2024.** Gebelikte anne sağlığı risk gruplarının tahminine yönelik makine öğrenmesi tabanlı bir karar destek sistem tasarımı. *Black Sea Journal of Engineering and Science*, 7(3):509-520. DOI: 10.34248/bsengineering.1455473
- Yapıcı, İS., Erkeymaz, O., Arslan, RU. 2021.** A hybrid intelligent classifier to estimate obesity levels based on ERG signals. *Physics Letters A*, 399:127281. DOI: 10.1016/j.physleta.2021.127281
- Yildirim, P., Birant, KU., Radevski, V., Kut, A., Birant, D. 2018.** Comparative analysis of ensemble learning methods for signal classification. 26th Signal Processing and Communications Applications Conference (SIU), pp. 1- 4, Turkey. DOI: 10.1109/SIU.2018.8404601
- Zhang, W., Shen, X., Zhang, H., Yin, Z., Sun, J., Zhang, X., Zou, L. 2024.** Feature importance measure of a multilayer perceptron based on the presingle-connection layer. *Knowledge and Information Systems*, 66(1):511-533. DOI: 10.1007/s10115-023-01959-7