Research Journal of Biology Sciences
E-ISSN: 1308-0262 15(2): 39-49, 2022
Research Article

SU KALİTE İNDEKSİNİN MAKİNE ÖĞRENME YÖNTEMLERİ İLE TAHMİNİ

Arzu Altın Yavuz[1,a,*]

[1]*Eskişehir Osmangazi Üniversitesi, Fen Fakültesi,, İstatistik Bölümü, 26480, Eskisehir, Türkiye*

**Corresponding Author:*
E-mail: aaltin@ogu.edu.tr

(Received 17th July 2022; accepted 23th September 2022)

a: ORCID 0000-0002-3277-740X

ÖZET. Hayatın ve kalkınmanın en önemli parçası olan su, giderek artan kirlilik kaynakları nedeniyle kullanılabilir olma özelliğini yitirmektedir. Canlıların sağlıkla yaşamını sürdürmesi ve ekonomilerin büyümesi için temiz su kaynaklarının artması gerekir. Oysa, su kaynakları sınırlıdır ve temiz su kaynaklarının artması su kirliliğinin azaltılması ile mümkün olabilir. Günümüzde hemen her ülkede su kalitesi bozulmuş durumdadır. Bazı ülkelerde bu problem su kıtlığı olarak kendini göstermektedir. Kullanılabilir su miktarının arttırılabilmesi için düzenli ve sürekli olarak su kalitesinin takibi yapılmalıdır. Sürdürülebilir bir su yönetimi için sorunun kaynağında çözümü önem taşımaktadır. Nehirler başlıca tatlı su kaynaklarıdır. Buna rağmen doğal deşarj alanı olarak kullanıldıklarından kirliliğe çok fazla maruz kalmaktadırlar. Yüzey su kalitesinin takibinde tüm su kalite karakteristikleri yerine su kalite indekslerinin kullanılması pratiklik sağlamaktadır. Bu çalışmada yüzey su kalitesinin sensörlerle takibi için makine öğrenme yöntemlerinin performansı karşılaştırılmıştır. Su kalitesi karakteristiklerinin sensörleri pahalı olduğundan, minimum sayıda parametre kullanımı amaçlanmıştır. Su kalite indeksinde yer alan değişkenlerin indirgenerek, daha az sensör kullanımıyla yerinde su kalite takibini mümkün kılan bir yaklaşım sunulmaya çalışılmıştır. Porsuk nehri su kalite takibi için çözünmüş oksijen, fekal koliform, bulanıklık ve toplam katı madde değişkenleri kullanılarak su kalite indeksi tahmin edilmiştir. Bu değişkenlerin sensörleri ve XGBoost yöntemi kullanılarak, daha az maliyetle su kalite indeksinin etkin ve sürekli izlenebileceği belirlenmiştir.

Anahtar Kelimeler: *su kalite indeksi, makine öğrenme yöntemleri, destek vektör makineleri, XGBoost*

GİRİŞ

Doğal kaynaklar içerisinde insan yaşamının ve kalkınmanın temel kaynağı olan su, çevre ve sağlık üzerinde güçlü etkilere sahiptir. Giderek artan nüfus, hızlı kentleşme, sanayileşme ve vahşi tarımsal sulama kullanılabilir su kaynaklarının azalmasına ve su kalitesinin hızla düşmesine neden olmaktadır. Dünyanın yaklaşık %70'ini oluşturan su, ekosistemdeki tüm türlerin yaşam kaynağıdır. Buna karşın, içilebilir tatlı su kaynakları sınırlıdır. Başta içme ve sulama amaçlı olarak kullanılan yüzey suları birincil tatlı su kaynaklarıdır [1, 2, 3]. Nehir, dere ve çay gibi temel yüzey su kaynaklarının kalitesi, yağmur, erozyon gibi doğal olaylar, kentsel, tarımsal ve endüstriyel faaliyetler sonucu kirlilikle karşı karşıya kalmaktadır [4]. Birleşmiş Milletler Gelişme Programı'na göre dünya üzerindeki hastalıkların % 50'sinin kaynağı sudur [5]. Bu nedenle yüzey su kaynaklarının kalite değerlendirmesi, su kaynakları yönetim planlarını geliştirmek için en önemli görevlerden biridir. Su kalitesinin doğru tahmini, su yönetimini ve kirlilik kontrolünü iyileştirmenin anahtarıdır. Su kaynakları sonsuz olmadığından, su kirliliğinin azaltılması kullanılabilir su kaynaklarının artmasına neden olmaktadır.

Birçok ülkede olduğu gibi ülkemizde de yüzey su kalitesinin takibi için su kalitesi gözlem istasyonlarından yararlanılmaktadır. Bu istasyonlarda su kalitesini etkileyen

çeşitli faktörlerin ölçümü düzenli aralıklarla gerçekleştirilmektedir. Sonrasında elde edilen bulgular yardımıyla su kirliliği engellenmeye çalışılmaktadır. Ancak bu uygulama zaman alıcı ve pahalı laboratuvar analizleri gerektirmektedir. Ayrıca su kalitesindeki değişimlerin hızlı bir şekilde tespitine olanak sağlamamaktadır. Giderek artan su ihtiyacı nedeniyle su kaynaklarının kirlilik kontrolünün daha hızlı ve güvenilir bilimsel yöntemlerle yapılması büyük önem taşımaktadır. Bu amaçla su kaynaklarının sürekli takibi doğal su kaynaklarının korunması açısından zorunludur.

Nehirler atık su deşarj yeri olarak kullanıldığından, en kolay kirletilebilen doğal su kaynaklarıdır. Sürdürülebilir su kalitesi için son yıllarda birçok gelişmiş ülke nehirler üzerine sensörlere dayalı gözlem istasyonları kurmaktadır [6, 7]. Bu durum özellikle ABD ve Kanada'da gibi gelişmiş ülkeleri 2030 yılına kadar yapay zekâ ile sürdürülebilir su yönetimine 6.3 milyar dolarlık yatırım yapmaya teşvik etmiştir [8]. Makine öğrenme yöntemleri, sensörlerden gelen bilgileri hızlı ve güvenilir bir şekilde analiz ederek, su kirliliğinin hızlı bir şekilde önlenmesine yardımcı olmaktadır [9]. Bilimsel araştırmaların sonucunda makine öğrenme yöntemlerinin su kalitesinin değerlendirilmesinde oldukça yüksek güvenilirlikte sonuçlar verdiği belirlenmiştir [10, 11].

Su kirliliği kontrolü için temel bir gösterge olarak su kalite indekslerinden (Water Quality Index-WQI) yararlanılmaktadır. Su kaynağının değerlendirilmesi için seçilen çeşitli parametrelere göre farklı su kalite indeksleri mevcuttur. WQI çok sayıda faktörün ağırlıklandırılmasıyla elde edilen tek bir ölçüdür. İndeks değeri sonucuna göre suyun kalitesi farklı sınıflara atanmaktadır. Su kaynaklarındaki kalite değişiminin tespiti için gözlemlerin sürekli yapılabilmesi gerekmektedir. WQI değeri havzadaki su kalitesine ilişkin sürdürülebilir bir ölçüdür ve diğer su kaynakları ile kalite karşılaştırılmasına olanak sağlar.

Bu çalışmada su kalitesinin gerçek zamanlı olarak etkin bir şekilde tahmin edilebilmesi için makine öğrenmesine dayalı bir yöntem önerilmesi amaçlanmıştır. Porsuk nehri için su kalite indeksinin hesaplanmasında Biyolojik Oksijen İhtiyacı (BOI-mg/L), çözünmüş oksijen (mgO2 /L), pH (Alkalinite), Sıcaklık (T-°C), Toplam Çözünmüş Katılar (TDS - mg/L), Fekal Koliform, Toplam Fosfor (mg/L), Nitrat (mg/L) ve bulanıklık değişkenleri ele alınmıştır. WQI değerinin hesaplanmasında kullanılan bu değişkenlerin tamamının sensörler yardımıyla tespiti yüksek maliyete neden olmaktadır. Bu nedenle WQI değerini etkileyen daha az sayıda bir değişken grubu belirlenerek, çeşitli makine öğrenme yöntemleri ile WQI değerleri tahmin edilmiştir. İndeks hesabında kullanılan bu değişkenlerin tamamının sensörlerle takibi yerine, indirgenmiş değişkenlere dayalı izleme prosedürü daha düşük maliyetle, daha hızlı ve güvenilir sonuç elde etmeye neden olacaktır. Böylelikle pahalı ve zahmetli laboratuvar analizlerine gerek kalmadan sürdürülebilir bir su yönetimi sağlanabilecektir.

MATERYAL VE METOD

Çalışma Alanı ve Veri Toplama

Porsuk Çayı yaklaşık 460 km uzunluğu ile Sakarya Nehri'nin önemli bir koludur. 11.188 km2'lik bir havza alanına sahiptir. Kütahya il sınırı içerisinde yer alan Murat dağından doğarak, Kocaçay, Kokarçay, Avşar ve Kureyşler derelerinin birleşmesiyle oluşmaktadır [12]. Porsuk çayı Eskişehir ilinin içme suyunu karşılamaktadır. Ayrıca geçtiği alanlarda tarımsal sulama amaçlı kullanılmaktadır. Kütahya azotlu gübre

fabrikası, şeker fabrikası, manyezit fabrikası, Seyitömer termik santrali gibi fabrikalar, evsel ve endüstriyel faaliyetler nedeniyle birçok kirleticiye maruz kalmaktadır.

Bu çalışmada Porsuk çayının su kalite indeksinin makine öğrenme yöntemleri ile tahmini konusu ele alınmıştır. 2007-2017 yılları arasında Devlet Su İşleri tarafından yapılan kalite çalışmaları sonucunda elde edilen verilerden yararlanılmıştır. Bu yıllar arasında aylık olarak düzenli ölçümü yapılan Biyolojik Oksijen İhtiyacı (BOI-mg/L), çözünmüş oksijen (mgO2 /L), pH (Alkalinite), Sıcaklık (T-°C), Toplam Çözünmüş Katılar (TDS -mg/L), Fekal Koliform, Toplam Fosfor (mg/L), Nitrat (mg/L) ve bulanıklık değerleri yardımıyla su kalite indeksi hesaplanmıştır.

Su kalite indeksi için çok sayıda indeks hesabı bulunmaktadır. Bu çalışmada ABD Ulusal Sanitasyon Vakfı Su Kalite İndeks (NFS-WQI) hesabı ele alınmıştır. İndeks hesabında kullanılan su kalite karakteristikleri ve etki faktörleri Tablo 1'de verilmiştir.

Tablo 1. *NFS-WQI indeksinin hesaplanılmasında kullanılan parametreler ve etki değerleri*

Parametreler	**Etki Faktörü**
Çözünmüş Oksijen	0.17
Fekal Koliform	0.16
pH	0.11
BOİ	0.11
Nitrat	0.10
Sıcaklık	0.10
Fosfat	0.10
Bulanıklık	0.08
Toplam Katı Madde	0.07
TOPLAM	1.00

Tablo 1'de verilen etki faktörü değerleri ve her bir su kalite karakteristiği için Q değerleri kullanılarak Ulusal Sanitasyon Vakfı Su Kalite İndeksi Eşitlik (1)'deki gibi hesaplanılmaktadır [13]. Su kalite indeksi hesaplanırken her bir parametre için kullanılacak Q değerleri farklıdır. Bazı faktörlere ilişkin Q değerleri için grafikler Figür 1' de verilmiştir.

$$\mathrm{WQI} = \sum_{i=1}^{n} Q_i W_i \qquad \textbf{Eqn. 1}$$

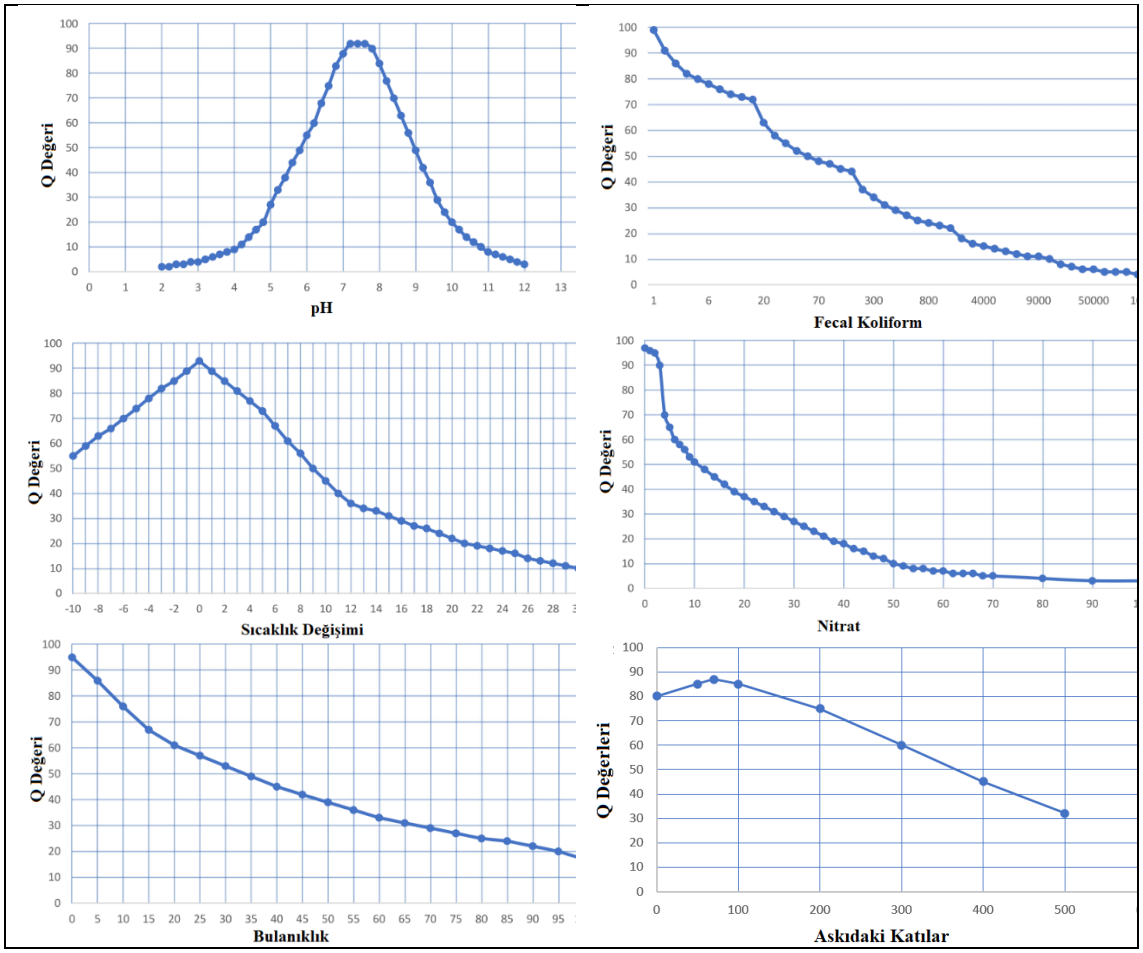


Fig. 1. *NFS-WQI hesabında kullanılan su kalite parametrelerinin Q değerleri*

NSF-WQI indeksi sonucunda elde değerlere göre su kalite sınıfları oluşturulmaktadır. NSF-WQI değerlerine göre su kalite sınıfları Tablo'2 de verilmiştir.

Tablo 2. *NFS-WQI indeksi sonucuna göre su kalite sınıfları*

NSF-WQI Değerleri	Kalite Sınıfı
90.1-100	Mükemmel
70.1-90	İyi
50.1-70	Orta
26-50	Kirli
0-25	Çok Kirli

Metod

Bu çalışmada Porsuk nehri su kalitesi indeksinin makine öğrenme yöntemleri ile gerçek zamanlı tahmini ele alınmıştır. Daha az parametre ile güvenilir sonuçlar verecek bir sistemin kurulabilmesi için su kalite indeksinin etkin ve yüksek doğruluklu tahminlerinin elde edilmesi gerekir. Bu nedenle çalışmada çoklu doğrusal regresyon, yapay sinir ağları, destek vektör makineleri, gradyan arttırma algoritması ve extreme

gradyan arttırma algoritması kullanılmıştır. Analizler R programlama dili ve MATLAB programı kullanılarak gerçekleştirilmiştir.

Çoklu Doğrusal Regresyon

Uygulama çalışmalarında en yaygın kullanıma sahip olan regresyon analizinde, bir bağımlı değişkenin birden fazla bağımsız değişken tarafından açıklanması çoklu doğrusal regresyon modeli olarak adlandırılmaktadır. Neden-sonuç ilişkilerini açıklamak amacıyla kullanılan en eski yöntemdir. Y bağımlı değişkeni ve $X_1, X_2, \dots, X_p$ ise bağımsız değişkenleri göstermek üzere çoklu doğrusal regresyon modeli,

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + e_i \qquad \textbf{Eqn. 2}$$

şeklindedir. Burada β_j'ler regresyon katsayılarını, e_i ise hata terimini ifade etmektedir [14].

Yapay Sinir Ağları

İnsan beynindeki sinir hücrelerinin yapısını temel alan bir işleyişe sahip olan yapay sinir ağları (YSA) güçlü bir makine öğrenme yöntemidir. Sinir hücrelerinin oluşturduğu katmanlara benzeyen bir algoritma ile neden-sonuç ilişkisini açıklamayı amaçlar. Figür 2'de de belirtildiği üzere YSA, genel olarak 5 temel unsurdan oluşmaktadır. Bunlar girdiler, ağırlık değerleri, toplam ve aktivasyon fonksiyonu ve çıktılardır.

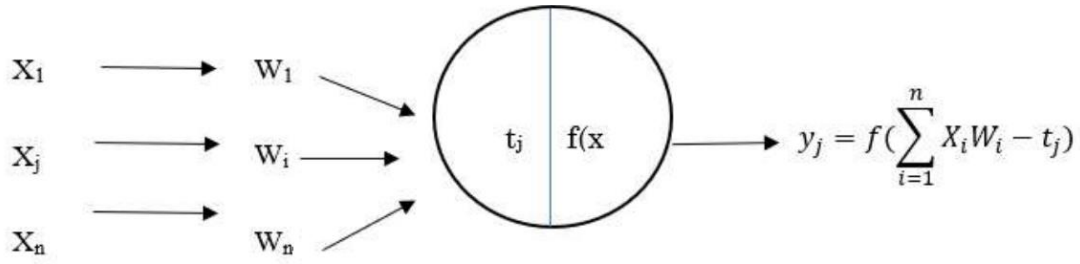


Fig. 2. *Yapay Sinir Ağlarının Bileşenleri*

Burada, W_i ağırlıklar, X_i'ler girdiler, t_j'ler sınır değerleri, $f(x)$ aktivasyon fonksiyonunu göstermektedir [15]. Genel olarak bir YSA mimarisi Figür 3'te verilmiştir.

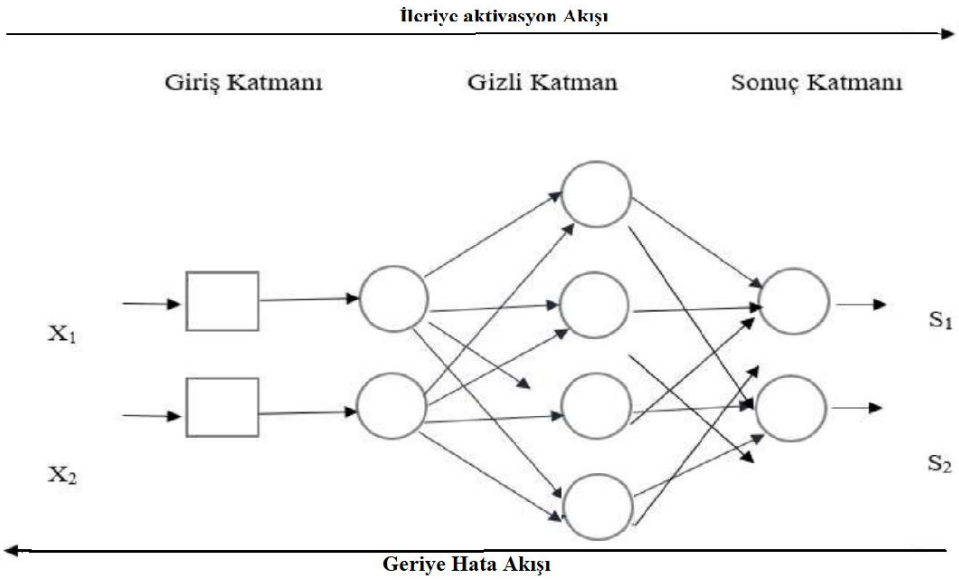


Fig. 3. *YSA Mimarisi*

YSA'nın çok katmanlı olması ve bu katmanlar arasında bağlantılar bulunması, ileri ve geri beslemeli olarak öğrenebilen bir sistem olması en önemli özelliğidir. Öğrenme işlevinde tahmin hatasını minimize etmek amacıyla ağırlık değerlerinden yararlanır [16].

Destek vektör makineleri

Sınıflandırma ve regresyon amaçlı olarak kullanılabilen Destek vektör makineleri (DVM) makine öğrenme yöntemleri içerisinde oldukça popülerdir. DVM sonucu elde edilen modelin genelleme yeteneği ve tahmin doğruluğu oldukça yüksektir. Ayrıca eğitim hızı yüksek olması nedeniyle özellikle görüntü tanıma çalışmalarında oldukça popülerdir. Örneklem hacminin sınırlı olduğu küçük çaplı veri setlerinde dahi elde edilen sonuçlar oldukça iyidir [17]. Ayrıca, doğrusal ve doğrusal olmayan veri yapıları için modelleme başarısı oldukça yüksektir. Bağımlı değişkenin sürekli olması durumunda regresyon amaçlı kullanılan DVM, destek vektör regresyonu (DVR) olarak adlandırılmaktadır [18].

Gradyan Artırma Algoritması (Gradient Boosting)

Friedman tarafından sunulan gradyan artırma (gradient boosting) algoritması, topluluk öğrenme algoritmalarından biridir [19]. GA algoritmasında, iteratif olarak tahmin edilen hatanın iyileştirilmesi amaçlanmaktadır. Bu amaçla dereceli azaltma sonuçları fonksiyona ilave edilir. Bir başka ifade ile GA algoritmasında zayıf olan öğreniciler bir araya getirilerek güçlü bir öğrenici elde edilmeye çalışılır. GA regresyon ve sınıflandırma amaçlı kullanılan bir makine öğrenme yöntemidir [20].

Extreme Gradient Boosting (XGBoost)

GA algoritmasının bazı zayıflıklarının giderilmesi sonucu eXtreme Gradient Boosting (XGBoost) algortiması geliştirilmiştir. XGBoost algoritması eksik veri ve aşırı öğrenme problemlerine çözüm oluşturması, hızlı ve tahmin gücünün yüksek olması nedeniyle son yıllarda makine öğrenme yöntemleri arasında oldukça dikkat çekmektedir. GA algoritmasında olduğu gibi hem regresyon hem de sınıflandırma problemlerinde kullanılabilmektedir [21,22].

Performans Ölçütü

İstatistiksel modelleme çalışmalarında yapılan tahminlerin etkinliğini değerlendirebilmek amacıyla yaygın olarak Hata Kareler Ortalaması (Mean Squared Error-MSE) ve Ortalama Mutlak Hata (Mean Absolute Error-MAE) ölçütleri kullanılmaktadır.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad \textbf{Eqn. 3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \qquad \textbf{Eqn. 4}$$

Burada $\hat{y}_i$, y_i'nin tahmin değerini göstermektedir.

BULGULAR VE TARTIŞMA

Çalışmanın bu bölümünde Porsuk nehrine ait su kalite verileri kullanılarak NFS-WQI indeks değerinin modellenmesi üzerinde durulmuştur. Ele alınan yöntemler yardımıyla NFS-WQI indeks değerini en etkin şekilde tahmin eden modelin belirlenmesi için veri seti iki gruba ayrılmıştır. Verilerin %75'inin eğitim verisi, %25'inin test verisi olarak ayrılması genel kabul görmüş bir uygulamadır. Bu nedenle çalışmada eğitim ve test verileri bu oranlara göre belirlenmiştir. Eğitim veri setindeki değerler yardımıyla ele alınan yöntemlere ilişkin modeller belirlendikten sonra, bu modellerin etkinlikleri test veri setindeki değerler yardımıyla belirlenmektedir. Yukarıda ele alınan modellerin tahmin başarısının tespiti için kayan pencere yöntemi kullanılmıştır. Kayan pencere yönteminde, *t* zamanındaki gözlem değeri, *t-1, t-2, t-3, …, t-k* zamanındaki gözlem verileri kullanılarak tahmin edilmektedir. Bu uygulama *k* kayan pencere yöntemi olarak adlandırılmaktadır. Kayan pencere yönteminde, her bir tahmin değerinden sonra gözlemler birer birim kaydırılarak yeni tahmin değerleri elde edilir. Böylelikle *k* gözlem kullanılarak her seferinde bir tahmin değeri elde edilir.

Modelleme aşamasına geçmeden önce verilerin ölçü birimleri birbirinden farklı olduğundan, öncelikle her bir değişken için normalizasyon yapılmıştır. Normalizasyon yapılması oluşturulan modelin performansını ve eğitim sonuçlarını etkilemektedir. Daha sonra aykırı değer tespiti için kutu-bıyık grafikleri çizilmiştir. Kutu-bıyık grafiği aykırı değer tespiti için görsel bir sonuç vermektedir. Ele alınan 9 su kalite değişkenine ait kutu-bıyık grafikleri Figür 4'de verilmiştir. Figür 4 incelendiğinde Çözünmüş Oksijen, Fekal Koliform, BOI ve fosfat değişkenlerinde aykırı değer bulunmadığı, kalan değişkenlerde ise çok sayıda aykırı gözlemlerin olduğu tespit edilmiştir.

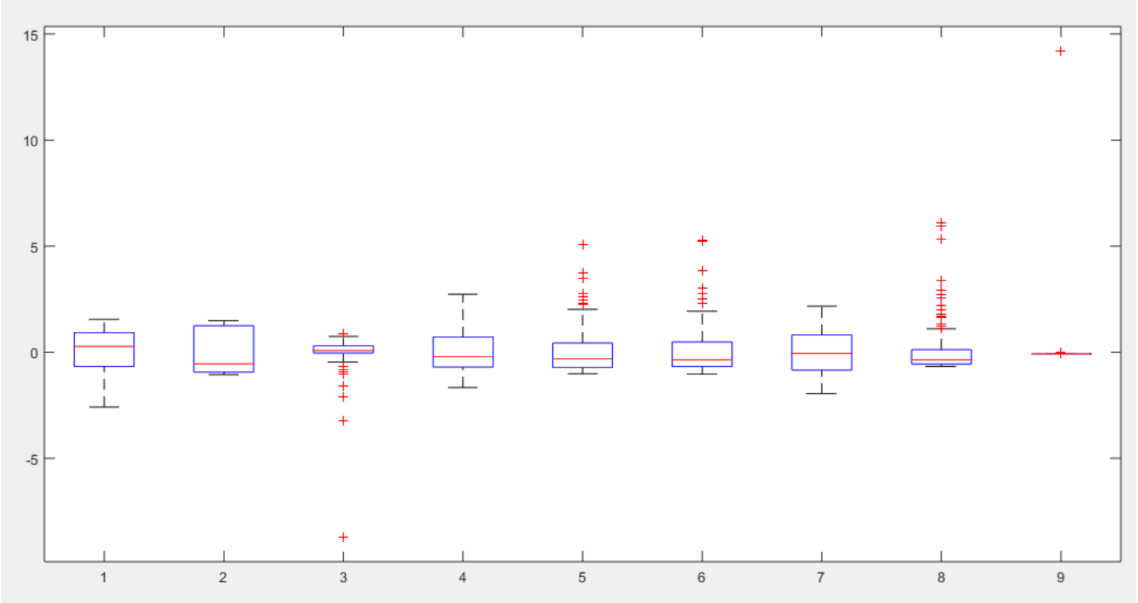


Fig. 4. *NFS-WQI indeks değeri için kullanılan su kalite karakteristiklerine ait kutu-bıyık grafikleri*

Çalışmanın devamında NFS-WQI indeks değerinin daha az parametre ile tahmini için değişkenler arası korelasyon değerleri hesaplanmıştır. Elde edilen korelasyon matrisi Figür 5'te verilmiştir. Son satırda yer alan NFS-WQI indeks değeri ile yüksek korelasyonlu olan Çözünmüş Oksijen, Fekal Koliform, bulanıklık ve toplam katı madde değişkenleri temel değişkenler olarak seçilmiştir.

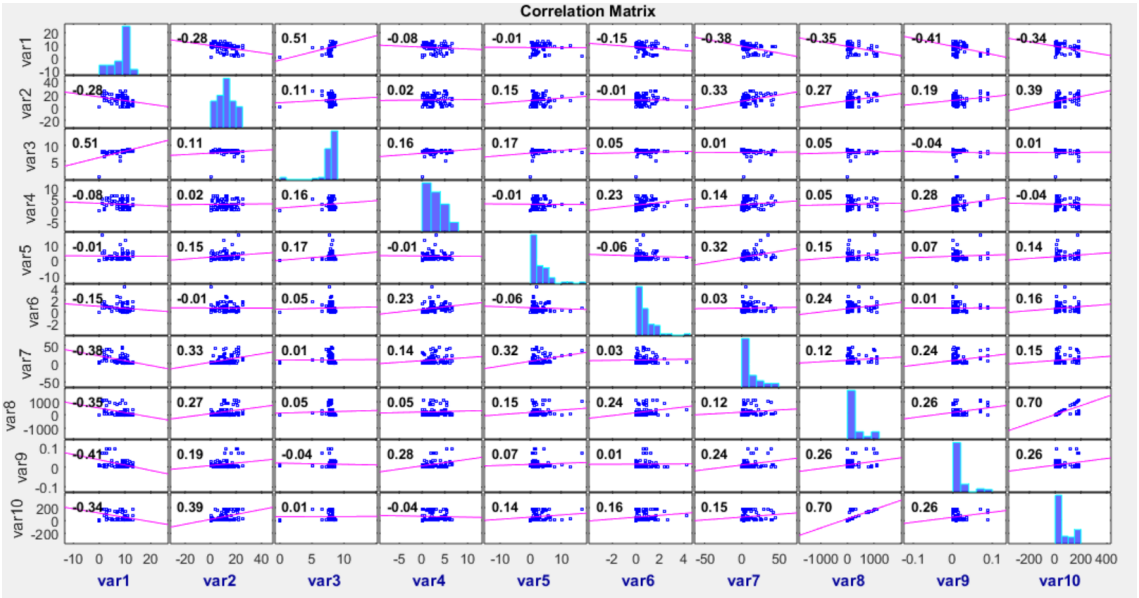


Fig. 5. *NFS-WQI indeks değeri ile açıklayıcı değişkenler arasındaki korelasyon matrisi*

NFS-WQI indeks değeri hesaplanırken kullanılan 9 kalite karakteristiğinin etki faktörleri farklıdır. Ancak suyun alındığı lokasyondaki kalite değişimini etkileyen unsurlar korelasyon analizi yardımı ile tespit edilmiştir. Makine öğrenmesine dayalı

olarak sensörler yardımıyla su kalitesinin tespiti üzerine yapılan çalışmalarda bulanıklık, sıcaklık ve pH gibi az sayıda değişkenle yapılmış çalışmalar mevcuttur [23]. Ancak sadece 3 parametre ile su kalitesinin tespiti oldukça sınırlıdır. Öte yandan su kalitesi için 25 farklı parametre kullanan çalışmalarda bulunmaktadır. İleriye ya da geriye doğru değişken seçimi yöntemleri ile değişken sayısı azaltılsa bile, fazla sayıda değişken olması sensör maliyetini arttıran bir faktördür. Bu durum ucuz ve gerçek zamanlı bir su kalite takibini olumsuz etkilemektedir. Sistem maliyetini azaltmak amacıyla, bu çalışmada su kalitesi indeksini tahmin etmek için çözünmüş oksijen, Fekal Koliform, bulanıklık ve toplam katı madde değişkenleri ele alınmıştır. Tüm değişkenler kullanılarak hesaplanan NFS-WQI indeks değerleri, sadece bu değişkenler yardımıyla tahmin edilmeye çalışılmıştır.

NFS-WQI indeks değeri tahmini için çoklu doğrusal regresyon, yapay sinir ağları, destek vektör makineleri, gradient boosting ve XGBoost yöntemleri kullanılarak modeller oluşturulmuş, test verisine ilişkin performans ölçütü sonuçları Tablo 3' te verilmiştir.

Tablo 3. *NFS-WQI tahmini için model performansları*

Yöntem	MAE	MSE
Çoklu Doğrusal Regresyon	4,1137	20,75809
Yapay Sinir Ağları	3,9218	19,42115
DVM	3,7866	18,56924
Gradient Boosting	3,7211	17,59897
XGBoost	3,6648	17,10683

Tablo 3 incelendiğinde hem MAE hem de MSE açısından en küçük değere sahip olan yöntem XGBoost yöntemidir. Geleneksel olarak kullanılan çoklu doğrusal regresyon analizi sonucunda elde edilen MAE ve MSE değerleri ise tüm yöntemlere göre daha yüksektir.

SONUÇ

Dünyanın temiz ve kullanılabilir su kaynaklarındaki azalma geleceğin en önemli problemlerinden bir tanesidir. Çevre, sağlık ve gelişmişlik üzerindeki etkileri düşünüldüğünde enerji probleminden çok daha ön plana çıkmaktadır. Enerji kaynaklarının alternatifi varken temiz su kaynaklarının alternatifi bulunmamaktadır. Temiz ve kullanılabilir su kaynaklarının arttırılması ancak bu kaynakların etkin kullanımı ve kirleticilere anında müdahale edilmesi ile mümkündür. Bu durum birçok gelişmiş ülkenin gelecek planlarında yer almaktadır.

Ülkemizde su kalitesi tespiti geleneksel olarak numune alımı ve laboratuvar analizleriyle gerçekleştirilmektedir. Bu durum birçok kirlilik kaynağının geç tespitine ve giderek su fakiri bir ülke olmamıza neden olmaktadır. Su kaynakları üzerine yerleştirilecek sensörler yardımıyla su kalitesinin takibi kirlilik kontrolü için zaman kazandıracaktır. Bununla birlikte tüm parametrelerin sensörlerle takibi hem pahalı hem de güç bir uygulamadır. Su kalitesi indeksleri su kalitesinin genel durumunu gösteren tek bir değerdir. Bu nedenle bu çalışmada su kalitesi indeksini belirlemek için az sayıda parametreye dayalı sürdürülebilir bir yöntem önerilmiştir.

Porsuk çayı su kalitesi tespiti için NFS-WQI indeks değerinin hesaplanmasında kullanılan 9 parametre yerine, indeks değeri ile yüksek korelasyonlu çözünmüş oksijen, Fekal Koliform, bulanıklık ve toplam katı madde değişkenleri kullanılarak NFS-WQI değeri tahmin edilmeye çalışılmıştır. Az sayıda değişkenle daha etkin sonuçlar elde edilebilmesi amacıyla XGBoost yönteminin kullanılmasının yararlı olacağı tespit edilmiştir.

Bu değişkenlerin belirlenmesi sonucunda az sayıda sensör yardımıyla sürdürülebilir bir su yönetimi gerçekleştirilebilir. Bu durum su kalitesinin sürdürülebilir olmasına, analizlerinin sıklıkla yapılmasına, yerinde ve hızlı bir çevre kirliliği kontrolünün yapılmasına olanak sağlamaktadır.

KAYNAKLAR

[1] Yanık, T., Aras, N. M. & Çiltaş, A. K. (2001): Su Kalitesi Ders Kitabı. Atatürk Üniversitesi Ziraat Fakültesi Ders Yayınları No:225, Atatürk Üniversitesi Ziraat Fakültesi Ofset Tesisi, Erzurum, 142 s.

[2] Akın, M. & Akın, G. (2007): Suyun Önemi, Türkiye'de Su Potansiyeli, Su Havzaları ve Su Kirliliği. Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi, 47(2): 105-118.

[3] Koranga, M., Pant, P., Pant, D., Bhatt, A. K., Pant, R. P., Ram, M. & Kumar, T. (2021): DVM Model to Predict the Water Quality Based on Physicochemical Parameters. International Journal of Mathematical, Engineering and Management Sciences, 6(2): 645-659.

[4] Asadollah, S.B.H.S., Sharafati, A., Motta, D. & Yaseen, Z.M. (2021): River water quality index prediction and uncertainty analysis: A comparative study of machine learning models, Journal of Environmental Chemical Engineering, 9: 1-14.

[5] Eren, B. & Çelebi, A. (2018): Sakarya Nehri Su Kalitesinin İstatistiksel Veri Değerleme Yöntemleri Kullanılarak Değerlendirilmesi, n 6th International Symposium on Innovative Technologies in Engineering and Science 09-11 November 2018 (ISITES2018 Alanya - Antalya - Turkey)

[6] Lu, H. & Ma, X. (2020): Hybrid decision tree-based machine learning models for short-term water quality prediction, Chemosphere, 249: 1-12.

[7] Tousi, E. G., Duan, J., Gundyb, P. M., Bright, K. & Gerba, C. P. (2021): Evaluation of E. coli in sediment for assessing irrigation water quality using machine learning. Science of the Total Environment, 799: 1-13.

[8] Yıldız, D. & Özgüler, H. (2020): Yapay Zeka ve Su Yönetimi. Su Politikaları Derneği, Rapor No: 30. Ankara.

[9] Bilali, A. & Taleb, A. (2020): Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. Journal of the Saudi Society of Agricultural Sciences, 19: 439-451.

[10] Castrillo, M. & García, A. L. (2020): Estimation of high frequency nutrient concentrations from water quality surrogates using machine learning methods. Water Research, 172: 1-9.

[11] Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H. & Kazakis, N. (2020): Improving prediction of water quality indices using novel hybrid machine-learning algorithms. Science of the Total Environment, 721: 1-16.

[12] Bakış, R., Koyuncu, H., Özkan, A., Banar, M., Yılmaz, G. & Yörükoğulları, E. (2011): Porsuk Havzası Yüzeysel ve Yer altı Suyu Kirlilik Düzeyinin Araştırılması. Anadolu Üniversitesi Bilim ve Teknoloji Dergisi-A. Uygulamalalı Bilimler ve Mühendislik, 12(2): 75-89.

[13] Tyagi, S., Sharma, B., Singh, P. & Dobhal, R. (2013): Water Quality Assessment in Terms of Water Quality Index. American Journal of Water Resources, 1(3): 34-38.

[14] Johnson,R. & Wichern, D. (2002): Applied Multivariate Statistical Analysis, Prentice Hall, Upper Saddle River, NJ, 5. ed edition.
[15] Haykin, S. (1999): Neural Networks: A Comprehensive Foundation. Prentice Hall, USA. Erişim adresi: https://dl.acm.org/doi/book/10.5555/521706
[16] Alizamir, M., Kisi, O., Ahmed, A. N., Mert, C., Fai, C. M., Kim, S., Kim, N. W. & El-Shafie, A. (2020): Advanced machine learning model for better prediction accuracy of soil temperature at different depths. PLoS ONE, 15(4): 1-25. https://doi.org/10.1371/journal.pone.0231055.
[17] Vapnik, V. (1998): Statistical Learning Theory. Wiley-Interscience, New York.
[18] Zhang, F. & O'Donnell, L. J. (2020): Support vector regression. In Machine Learning, Academic Press.
[19] Friedman, J. (2001): Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29(5): 1189-1232.
[20] Alpaydın, E. (2009): Introduction to Machine Learning (2nd ed). Cambridge: The MIT Press.
[21] Chen, T. & Guestrin, C. (2016): XGBoost: A Scalable Tree Boosting System. *Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939785
[22] Yeşilyurt, S. & Dalkılıç, H. (2021). Xgboost ve Gradient Boost Machine ile Günlük Nehir Akımı Tahmini. *Paper presented at the 3rd International Symposium of III Engineering Applications on Civil Engineering and Earth Sciences.*
[23] Shafi, U., Mumtaz, R., Anwar, H., Qamar, A. M. & Khurshid, H.: Surface Water Pollution Detection using Internet of Things. *In Proceedings of the 2018 15th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT)*, Islamabad, Pakistan, 8–10 October 2018; pp. 92–96.