

Kentsel İçme Suyu Şebekelerinde Suyun Kalite Seviyesinin Yapay Zekâ Yaklaşımları Kullanılarak Belirlenmesi

Abdullah ŞENER^{1*}, Burhan ERGEN², Hamit MIZRAK³

¹ Bingöl Üniversitesi, Genç Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Bingöl, Türkiye

² Firat Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Elazığ, Türkiye

³ Malatya Turgut Özal Üniversitesi, Malatya, Türkiye

*¹ asener@bingol.edu.tr, ² bergem@firat.edu.tr, ³ hamitmizrak@gmail.com

(Geliş/Received: 02/01/2022;

Kabul/Accepted: 11/02/2022)

Öz: Su, bütün canlıların yaşam dengesini sağlayan ve faaliyetlerini ayakta tutan hayati bir öneme sahiptir. İçme suyunun kalitesi suyun kimyasal, biyolojik, fiziksel ve radyolojik ölçütleriyle bağlantılıdır. Su kalitesi insan sağlığını ve ekolojik sistemi doğrudan etkileyen bir faktördür. Kentsel şebekelerde içme suyunun kalitesini ölçebilmek için birçok teknik ve model geliştirilmiştir. Dünya nüfusunun hızlı artışı ve yaşamsal alanlardaki içme suyu şebekelerinin sayısını da artırmaktadır. Bu tür ihtiyaçlara anlık cevap verebilmek insanlar için zaman alıcı bir süreçtir. Son zamanlardaki teknolojik gelişmeler bu tür problemlerin çözümünde etkin rol üstlenmektedir. Bu çalışmada su bileşenlerini içeren açık veri kümesi kullanılarak yapay zekâ tabanlı analizler gerçekleştirilmiştir. Gerçekleştirilen analizlerde suyun kaliteli olup olmadığı tespit edilmiştir. Deneysel analizlerde makine öğrenme yöntemleri ve topluluk öğrenme algoritmaları kullanılmıştır. En iyi performans sonucu torbalama yöntemi ile elde edilmiştir. Bu yöntem ile %96,44 oranında genel doğruluk başarıları sağlanmıştır.

Anahtar kelimeler: Yapay Zekâ, Makine Öğrenme, Su Kalite Ölçümü, Sınıflandırma.

Determining Water Quality in Urban Drinking Water Networks Using Artificial Intelligence Approaches

Abstract: Water has vital importance as it provides balance to all living things and keeps their activities alive. The quality of drinking water depends on the chemical, biological, physical and radiological criteria of water. Water quality is a factor that directly affects human health and ecological system. Many techniques and models have been developed to measure the quality of drinking water in urban networks. With the rapid increase in the world population, the number of drinking water networks in vital areas is also increasing. It is a time-consuming process for people to be able to respond to such needs instantly. Recent technological developments are playing an active role in solving such problems. In this study, artificial intelligence-based analyzes were performed on an open data set with water components. The analyzes performed determined whether the water was of good quality. Machine learning methods and ensemble learning algorithms were used in the experimental analysis. The best performance result was obtained using the bagging method. With this method, an overall accuracy of 96.44% was achieved.

Key words: Artificial Intelligence, Machine Learning, Water Quality Measurement, Classification.

1. Giriş

Su, canlıların yaşam döngüsünü sağlayan, tatsız ve kokusuz bir bileşiktir [1]. Tatlı su, çevreyle ilgili sistemin gereksinimini karşılayan ve tarımsal kalkınmanın sürdürülebilirliğini sağlayan sınırlı bir kaynaktır. İnsanlar tarafından tüketilen suyun kalitesinin düşük olması birçok hastalığın oluşmasına neden olabilir [2]. Canlılar yaşam kalitelerini koruyabilmeleri ve sağlıklı bir şekilde sürdürebilmeleri için tükettikleri suyun kalitesi önemlidir. Denizler, göller, nehirler gibi su yataklarının kalite seviyeleri belirli işlem adımlarından geçirilerek ölçülebilmektedir. Bununla ilgili birçok teknolojik altyapı sistemi tasarlanmış ve kullanılmıştır. Suyun kalitesi, suyun biyolojik, fiziksel, kimyasal gibi etmenlerine bağlıdır [3]. Suyun kalite ölçümü son derece hayati önem taşımaktadır ve kalite standardına uymayan içme suları insan sağlığını doğrudan olumsuz etkileyebilmektedir [4]. Kalitesiz suyun sonuçları çevreyi, alt yapı sistemlerini veya insanları doğrudan etkileyebilir. Nitekim Birleşmiş Milletler 'in raporuna göre her yıl 1,5 milyon insan kirli ve kalitesi düşük olan içme sularını kullandığından dolayı çeşitli hastalıklara yakalanmakta ve bu hastalıklardan dolayı ölmektedir [5].

Günümüzde suyun kalitesini ölçebilecek birçok model ve yaklaşım geliştirilmiştir. Son zamanlarda teknolojik tabanlı sistemler, yapay zekâ yaklaşımları ile birleştirilerek analizler yapılmaya başlanmıştır [3]. Bu durum karar verme süreçlerinin performansını olumlu etkilemiştir. Literatürde suyun kalite ölçümünü yapay zekâ yaklaşımları

* Sorumlu yazar: asener@bingol.edu.tr. Yazarların ORCID Numarası: ¹ 0000-0002-8927-5638, ² 0000-0003-3244-2615, ³ 0000-0002-4795-3007

ile gerçekleştiren birçok çalışma bulunmaktadır. Bu çalışmalardan bazıları incelenirse; Aldhyani ve ark. [3], su kalitesini sınıflandırma işleminde doğrusal olmayan otoregresif sinir ağı modeli ile uzun-kısa süreli bellek (LSTM) modelini kullanarak analizler gerçekleştirdi. Ayrıca makine öğrenme yöntemleri de deneysel analizlerde kullandılar. Aldhyani ve ark. [3] çalışmasında en iyi performansı doğrusal olmayan otomatik regresif sinir ağı modeli ile sağladılar ve bu modelden elde ettikleri genel doğruluk başarıları ise %96,17'dir. Khan ve ark. [6], su kalitesini sınıflandırma işleminde temel bileşenler analizi (PCA) yöntemi ile gradyan artırma yöntemini kullandılar. Khan ve ark. [6] çalışmalarında sırasıyla %95 ve %100 genel doğruluk başarıları elde ettiler. Radhakrishnan ve ark. [7] su kalitesini sınıflandırma sürecinde makine öğrenme yöntemlerini kullandılar. Radhakrishnan ve ark. [7] destek karar makineleri (DVM), karar ağaçları (KA), naif bayes (NB) yöntemlerini kullanarak analizlerini gerçekleştirdiler. En iyi sınıflandırma sonucunu KA yöntemi ile sağladılar ve bu yöntem ile %98,50 oranında genel doğruluk başarıları elde ettiler. Venkata ve ark. [8] su kalitesini sınıflandırma sürecinde sinir ağı modellerinden yararlandılar. Venkata ve ark. [8] tekrarlayan sinir ağı (RNN) ve LSTM modellerini deneysel analizlerde kullandılar. En iyi performans sonucunu LSTM yöntemi ile elde ettiler ve bu yöntem ile %94 oranında genel doğruluk başarıları sağladılar.

Bu çalışma şehir şebeke sularının kalite seviyesinin ölçülmesinde makine öğrenme yöntemleri ve topluluk öğrenme algoritmalarının etkinliğini ölçebilmek için gerçekleştirilmiştir. Çalışmanın diğer bölümleri hakkında özet bilgiler şu şekildedir; materyal, makine öğrenme yöntemleri ve topluluk öğrenme algoritmaları hakkında bilgiler ikinci bölümde yer almıştır. Deneysel analizler ve sonuçları hakkında açıklamalara üçüncü bölümde yer verilmiştir. Yazarların önerdikleri yaklaşım hakkındaki yorumlarına tartışma bölümünde yer verilmiştir. Analiz sonuçlarının değerlendirilmesi, gelecek çalışmalar hakkındaki bilgilere sonuç bölümünde yer verilmiştir.

2. Materyal ve Yapay Zekâ Yaklaşımları

Bu bölümde veri kümesi, makine öğrenme yöntemleri ve topluluk öğrenme algoritmaları hakkında genel bilgiler verilmiştir.

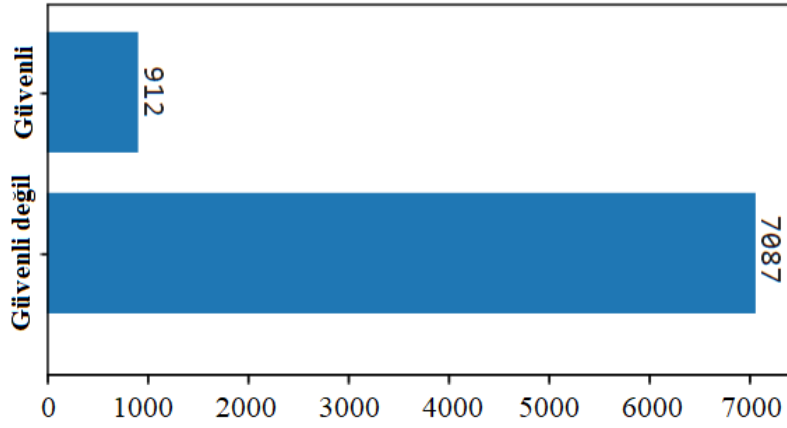
2.1. Veri kümesi

Veri kümesi açık erişimlidir ve 'csv' uzantılı dosyadan oluşmaktadır. Veri kümesinde toplam 21 öznitelik sütunu yer almaktadır ve kentsel bir yapıda su kalitesini ölçmeye ilişkin verilerden oluşmaktadır. Verilerin tümü sayısal değerlerden oluşmaktadır. 21. öznitelik "güvenli-1" ve "güvenli değil-0" sayısal etiketlerinden oluşmaktadır. Toplam 7999 örnek numune değerinden oluşmaktadır ve veri kümesindeki her bir öznitelik parametresindeki değerler bir litre içerisindeki su seviyesi dikkate alınarak oluşturulmuştur [9]. Diğer öznitelikler ve açıklamaları Tablo 1'de verilmiştir.

Tablo 1. Veri kümesinde yer alan öznitelikler ve açıklamaları.

No	Öznitelik	Tehlike eşik değeri	No	Öznitelik	Tehlike eşik değeri
1	alüminyum	2,8	11	virüsler	0
2	amonyak	32,5	12	kurşun	0,015
3	arsenik	0,01	13	nitratlar	10
4	baryum	2	14	nitritler	1
5	kadmiyum	0,005	15	cıva	0,002
6	kloramin	4	16	perklorat	56
7	krom	0,1	17	radyum	5
8	bakır	1,3	18	selenyum	0,5
9	florür	1,5	19	gümüş	0,1
10	bakteri	0	20	uranyum	0,3

Tablo 1'de özniteliklerin açıklaması dikkate alındığında, alüminyum değeri 1 litre suyun içerisinde 2,8 ve üzeri bir değere sahipse tehlikeli bir seviyededir. Bu durumda su kalitesi güvenli değildir. Diğer özniteliklerin tehlike seviyelerini belirten eşik değerleri Tablo 1'de verilmiştir. Güvenli örneklerin sayısı 717'dir ve güvenli olmayan örneklerin sayısı 7087'dir. Sınıf türünü ve sayısını gösteren grafik Şekil 1'de gösterilmiştir.



Şekil 1. Su kalitesi veri kümesinin sınıf türü ve sınıf türüne ait örnek sayısı.

2.2. Makine öğrenme yöntemleri

2.2.1. Lojistik regresyon

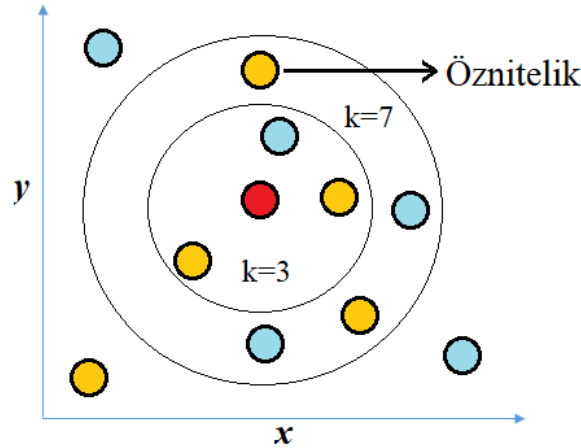
Lojistik regresyon (LR), birden fazla bağımsız değişken parametresini, bağımlı değişkenler üzerinden ilişkisini kurarak tahmin yürütmeyi sağlayan bir makine öğrenme yöntemidir. Tahmin yürütme işleminde ikili bağımlı değişken kullanılır. Yani, iki farklı kategorik değer alabilen bir sonuç üretilir. Örneğin; “doğru-yanlış”, “iyi-kötü”, “1 veya 0” gibi iki değer tahminini yürütür. Burada çıkışın sıfır veya bir olduğu varsayılırsa, bağımsız değişkenin değerinin bir olması bağımlı değişkenin de değerinin bir olma olasılığını artırmaktadır. LR yöntemi doğrusal regresyonun bir uzantısıdır. LR yönteminde çıkış değeri ikiden fazla ise her bir bağımsız değer sırasıyla sabit tutularak ikili sınıflandırma gerçekleştirilir ve çoklu sınıflandırma sürecinde ise tüm bağımsız değer sonuçları karşılaştırılarak girdi verisinin sınıflandırma işlemi gerçekleştirilir [10,11].

Olasılıksal bir değer üretme işleminde bağımsız değişken ile bağımlı bir değişken arasında sigmoidal bir ilişki vardır. Bu sebeple girdi verileri için üretilen tahmin değeri 0 ile 1 arasındadır. Bu durumun matematiksel işlemi Denklem 1'e göre gerçekleştirilir. Denklem 1 incelendiğinde; olasılık (p), regresyon katsayıları (b_0 , b_1), bağımsız değişken (X) ile temsil edilmektedir [10].

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 \times X \quad (1)$$

2.2.2. K-En yakın komşu

En yakın komşu (kNN) yöntemi, girdi verisini kendisinden önceki girdi verilerinin yakınlık derecesine bağlı olarak sınıflandırma ve regresyon işlemini gerçekleştirebilen bir makine öğrenme yöntemidir. kNN yöntemi mevcut girdi verilerinin durumunu saklar, yeni veri durumlarını da mevcut veri durumlarını dikkate alarak benzerlik ölçütlerini elde eder. Benzerlik olasılığı hesaplanırken özniteliklerin bulunduğu uzayda iki öznitelik arasındaki mesafeler hesaplanır. Ardından komşu öznitelikler ile mesafeler hesaplanır ve özniteliklerin sınıflandırma sürecinde komşu öznitelikler arasında mesafe uzunluğu en yakın olan özniteliklerin sınıf etiketine ataması gerçekleştirilir. İki öznitelik arasındaki mesafe uzunluğu hesaplanırken genellikle Öklid yöntemi tercih edilir. kNN yönteminde k değeri $\{2,3,5,\dots\}$ gibi değerlerden biri seçilerek sınıflandırma işlemi gerçekleştirilir. Örneğin; $k = 3$ olarak tercih edildiğinde uzay kümesinde mevcut özniteliklerin etrafındaki en yakın üç özniteliklere göre uzaklık mesafeleri ölçülür ve mevcut öznitelik kendisine en yakın özniteliklerin sınıf türüyle etiketlenir [12,13]. k değerinin ikili sınıflandırma işleminde seçimi Şekil 2'de gösterilmiştir.



Şekil 2. kNN yönteminde k değerine göre özniteliklerin seçimi.

2.2.3. Karar ağacı

Karar ağacı (KA) yöntemi, karar ve yaprak düğümlerinden oluşan sınıflandırma ve regresyon işlemi için entropi değerlerini hesaplayan grafiksel tabanlı şematik bir makine öğrenme yöntemidir. Bu yöntemde sınıflandırma ve regresyon işlemleri için (ID3, C4.5, CART, vb.) algoritmalar kullanılır. Algoritma şemasında sırasıyla kök, düğüm ve yapraklar yer alır. Verilerin belirsizlikleri hesaplarken entropi (E) değerleri dikkate alınır. Entropi (E) değeri olasılık değerlerinin hesaplanmasında kullanılan önemli bir parametredir. Entropi parametresinin hesaplanmasında Denklem 2’de belirtilen formül kullanılır. Denklem 2 incelendiğinde, N değişkeni toplam öznitelik sayısını temsil eder ve P değişkeni ise i . özniteliğin olasılık değerini temsil eder [14,15]. Sınıflandırma sürecinde entropi değeri minimum olan yaprak tercih edilir ve etiketlenir.

$$E = - \sum_{i=1}^N P_i \log_2 P_i \quad (2)$$

2.2.4. Rastgele Orman

Rastgele orman (RO) yöntemi, birden fazla karar ağacının bir araya gelmesi ile oluşturulmuş, sınıflandırma ve regresyon işlemlerinde kullanılan bir makine öğrenme yöntemidir. RO yöntemi, KA yöntemine göre algoritmasında birden fazla karar ağacı algoritması içerdiği için hata oranını daha doğru tahmin eder. Bir araya getirilen karar ağaçları “orman” topluluğunu oluşturur ve ağaçlar arasında korelasyon daha düşük seviyededir. Orman topluluğu içerisinde rastgele düğüm seçimi gerçekleştirilir ve ardından en iyi düğümü belirlemek için *Gini* fonksiyonu kullanılır. *Gini* fonksiyonu düğümlerin homojenlik değerini ölçer. Ağaç yapısındaki bir dalın alt düğümünün *Gini* değeri üst düğümünün *Gini* değerinden daha düşükse, düğümlerin bulunduğu dal başarılı sayılır. *Gini* fonksiyonun hesaplanmasında Denklem 3 kullanılır. Bu denklemde, toplam öznitelik sayısı N değişkeni ile temsil edilir ve seçilmiş öznitelik değeri n değişkeni gösterilir. Seçilmiş özniteliğin kendisinden küçük öznitelik sayısının kendisinden büyük öznitelik sayısına bölümünün karesi ise p_i değişkeni ile temsil edilir [16,17].

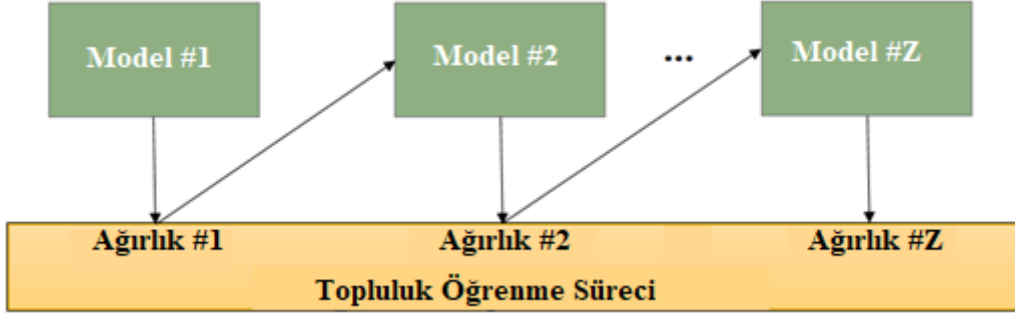
$$Gini(N) = 1 - \sum_{i=1}^n p_i^2 \quad (3)$$

2.3. Topluluk öğrenme algoritmaları

2.3.1. Uyarlanabilir artırma

Uyarlanabilir artırma (AdaBoost) algoritmasında ağırlık değerleri her bir girdi verisi için yeniden oluşturulmaktadır. Yanlış sınıflandırılan girdi verilerine ise daha yüksek değerlere sahip ağırlıklar atanmaktadır.

Bu işlem döngüsel olarak sonlanana kadar devam etmektedir ve ağırlık değerlerinin artırılması gerçekleştirildiği için bu algoritmaya uyarlanabilir artırma denilmektedir. AdaBoost algoritmasında "Z" sayıda karar ağacı olduğunu varsayarsak, Z sayıda model ve Z sayıda ağırlıklar oluşturulur. Birinci modelde yanlış sınıflandırılan girdi verileri tespit edilir ve ikinci modele kayıtları aktarılır. Bu işlem 3., 4., ..., Z. modele kadar devam eder ve her model arası ağırlık aktarımları da artırılarak gerçekleştirilir. Z sayıda modelin her biri eğitim sürecini güçlendirmek ve performansı artırabilmek için tasarlanmıştır [18]. Bu süreci gösteren tasarım Şekil 3'te gösterilmiştir.



Şekil 3. AdaBoost algoritmasının modeller arasında ağırlık aktarımı.

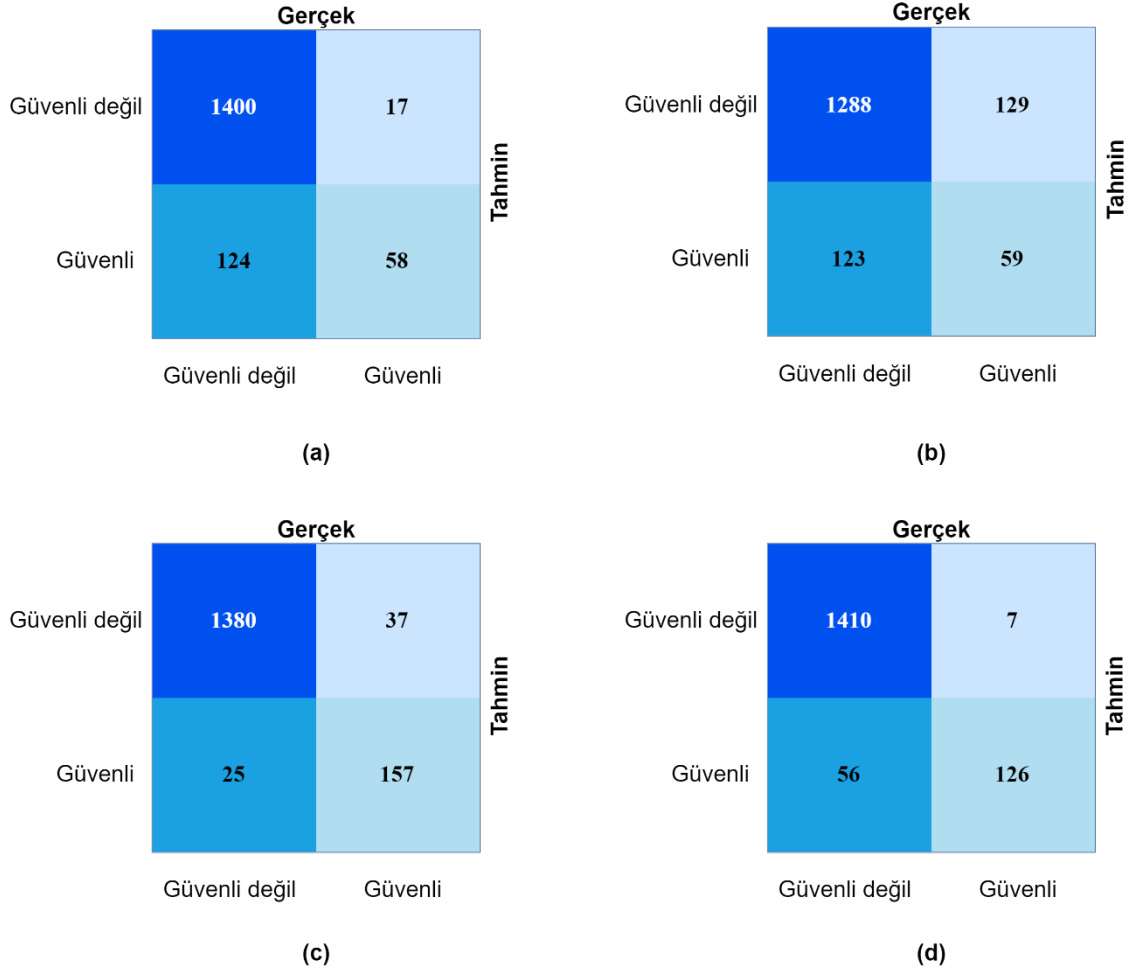
2.3.2. Aşırı gradyan artırma

Aşırı gradyan artırma (XGBoost), sınıflandırma ve regresyon tahminlerini gerçekleştiren modelleri içerisinde barındıran topluluk algoritmasıdır. Bu algoritmada AdaBoost algoritması gibi karar ağaçlarını kullanarak her model adımında yanlış girdi verilerini tespit ederek bir sonraki modele aktarılmasını sağlar. Bu işlem süresinde modellerin ağırlıkları gradyan tabanlı artırma ile sürekli güncellenir. XGBoost algoritması, AdaBoost algoritmasının zorlandığı ve hatalı verileri tespit etmede yaşanan güçlükleri tedarik edebilmek için tasarlanmıştır. XGBoost algoritmasının model mimarisi farklıdır ve modellerden çıkabilecek hata kaybını en aza indirebilmek için gradyan iniş algoritmasını kullanmıştır [19].

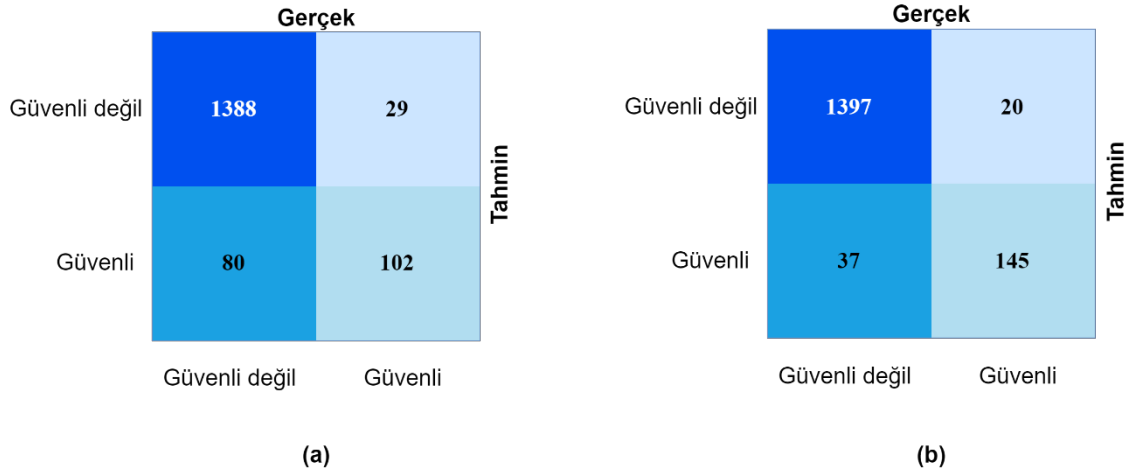
3. Deneysel Analizi ve Sonuçları

Bu çalışmanın deneysel analizleri Python yazılım dili ve Jupyter Notebook arayüzü kullanılarak gerçekleştirilmiştir. Yazılım kodlarının derlenmesinde donanım desteği olarak Google Colaboratory sunucusu kullanılmıştır. Google Colaboratory sunucusunun sunmuş olduğu donanım özellikleri; işlemci türü Intel® Xeon® CPU @2.20 GHz, ekran kartı Tesla K80, hafıza kartı 8 GB. Ekran kartının GPU destekleme özelliğinin olması, çalışmaların deneysel analizlerini daha hızlı gerçekleşmesini sağlamıştır [20,21]. Analiz sonuçlarının ölçülmesinde karmaşıklık matrisi kullanılmıştır. Karmaşıklık matrisi hesaplanırken; kesinlik (precision), geri çağırma (recall), F1-skor ve doğruluk (accuracy) metrikleri kullanılmıştır [22–24]. Makine öğrenme yöntemleri ve topluluk öğrenme algoritmaları Sklearn Kütüphanesi tarafından desteklendi ve bu yöntemler için varsayılan değerler kabul edilerek kullanılmıştır [25,26].

Deneysel analizlerin tümünde veri kümesinin %20'si test verisi olarak kullanılmıştır. Veri kümesinin %80'i eğitim verisi olarak kullanılmıştır. Veri kümesi sırasıyla LR, kNN, KA, RO makine öğrenme yöntemleri ile analiz edilmiştir. Ardından AdaBoost, XGBoost topluluk öğrenme algoritmaları kullanılarak analizler gerçekleştirilmiştir. Makine öğrenme yöntemlerindeki genel doğruluk başarıları sırasıyla %91.18, %84.24, %96.12, %96.06 elde edilmiştir. Topluluk öğrenme algoritmalarından elde edilen genel doğruluk başarıları sırasıyla %93.18, %96.44'tür. Makine öğrenme yöntemleri ve topluluk öğrenme algoritmalarından elde edilmiş karmaşıklık matris grafikleri Şekil 4 ve Şekil 5'te sırasıyla gösterilmiştir.



Şekil 4. Makine öğrenme yöntemlerinden elde edilmiş karmaşıklık matrisleri; a) LR, b) kNN, c) KA, d) RO.



Şekil 5. Topluluk öğrenme algoritmalarından elde edilmiş karmaşıklık matrisleri; a) AdaBoost, b) XGBoost.

Karmaşıklık matrislerin metrik sonuçları Tablo 2’de verilmiştir. Öğrenme yöntemleri arasında en iyi performansı topluluk öğrenme algoritmalarında yer alan XGBoost vermiştir. XGBoost algoritması ile %96,44 oranında genel doğruluk başarısı elde edilmiştir. Ayrıca veri kümesinin sınıfları arasında dengesiz dağılımdan

dolayı yöntemlerin F1 skorları da karşılaştırılmıştır. F1-skorları arasında en iyi performansı XGBoost algoritması ile edilmiştir ve elde edilmiş F1-skor oranı %98'dir.

Tablo 2. Deneysel analizlerden elde edilmiş karmaşıklık matrislerin metrik sonuçları.

Yöntem	Sınıf	Kesinlik(%)	Geri çağırma(%)	F1-skor(%)	Doğruluk (%)
LR	Güvenli değil	92	99	95	91,18
	Güvenli	77	32	45	
kNN	Güvenli değil	91	91	91	84,24
	Güvenli	31	32	32	
KA	Güvenli değil	98	97	98	96,12
	Güvenli	81	86	84	
RO	Güvenli değil	96	100	98	96,06
	Güvenli	95	69	80	
AdaBoost	Güvenli değil	95	98	96	93,18
	Güvenli	78	56	65	
XGBoost	Güvenli değil	97	99	98	96,44
	Güvenli	88	80	84	

4. Tartışma

Yapılan çalışmada suyun kalite değerini sınıflandırabilen yapay zekâ tabanlı yöntemler kullanılmıştır. Çalışmada veri kümesindeki veri sınıflarının dengesiz dağılımı F1-skorların önemini ön plana çıkarmıştır. Nitekim F1-skorları arasında da XGBoost algoritması genel doğruluk başarısında olduğu gibi daha iyi sonuç vermiştir. Ayrıca veri kümesindeki özniteliklerin sayısının 20 olması ön işlem adımlarının (öznitelik seçme, özellik çıkarma, vb.) uygulanması açısından sağlıklı olmayacağı düşünülmüştür. Sınıflandırma performansına katkı sağlayabilmek için belki evrimsel bir sinir ağı tasarlanabilirdi ve geliştirilmiş derin öğrenme modeline bu çalışmada kullanılmış makine öğrenme yöntemleri sınıflandırma aşamasında kullanılabilirdi. Bu çalışmaya benzer analizler ve sonuçları Tablo 3'te verilmiştir. Bu analizler incelenirse;

Tablo 3. Literatürde yer alan benzer çalışmalar ile bu çalışmanın karşılaştırılması.

Makale	Yıl	Yöntem	Doğruluk (%)
Sillberg ve ark. [27]	2021	DVM	95
Shahra ve ark. [28]	2021	Yapay Sinir Ağı (YSA)	94
Al-Adhaileh ve ark. [29]	2021	Nöro-bulanık çıkarım sistemi, İleri beslemeli sinir ağı	100
Bu çalışma	2022	XGBoost	96,44

Sillberg ve ark. [27] su kalitesi indeksi verilerini kullanarak nehir suyunun kalite düzeyini sınıflandırmışlardır. Sillberg ve ark. [27] önerdikleri yaklaşımda öznitelik gerçekleştirme ve DVM yöntemini hibrit bir model tasarlayarak kullanmışlardır. Sınıflandırma sonucunda %95 oranında genel doğruluk başarısı elde etmişlerdir. Sillberg ve ark. [27] DVM dışında farklı sınıflandırma yöntemlerini kullanabilselerdi belki de performans sonucunu artırabilirdi. Shahra ve ark. [28] içme suyu verilerini kullanarak suyun kaliteli olup olmadığını tespit edebilmek için yapay zekâ yaklaşımlarını kullanmışlardır. Shahra ve ark. [28] önerdikleri yaklaşımda YSA modelini ve DVM yöntemini kullanmışlardır. Sınıflandırma sonucunda YSA model ile en iyi performansı elde etmişlerdir ve %94 oranında genel doğruluk başarısı elde etmişlerdir. Shahra ve ark. [28] derin

ağlar ile bir YSA modeli tasarımları performans sonucunu olumlu etkilemiştir. Al-Adhaileh ve ark. [29] içme ve atık su arıtma verilerini kullanarak suyun kaliteli olup olmadığını tespit eden analizler gerçekleştirmişlerdir. Al-Adhaileh ve ark. [29] nöro-bulanık çıkarım sistemi algoritmasını kullandılar ve sınıflandırma sürecinde ileri beslemeli sinir ağı ile kNN yöntemini kullandılar. En iyi performansı nöro-bulanık çıkarım sistemi ile birlikte kullandıkları ileri beslemeli sinir ağından elde etmişlerdir. Al-Adhaileh ve ark. [29] çalışmalarında %100 genel doğruluk başarısı elde etmişlerdir. Onların çalışması bizim çalışmamıza kıyasla daha özgün yaklaşımlar içermiştir ve bunun karşılığında performansı olumlu katkı sağlamıştır.

5. Sonuç

Su sağlığımız ve yaşam kalitemiz için önemlidir. Tüketilen suyun kalitesinin düşük ve kirli olması insanların sağlığını tehdit etmektedir. Suyun kalitesinin ölçümü çeşitli teknoloji tabanlı modeller kullanılarak ölçülebilmektedir. Bu çalışmada suyun kalite değerini sınıflandırabilen yapay zekâ tabanlı yöntemler kullanılmıştır. Çalışmanın analizlerinde elde edilen performans sonuçları umut vericidir. Yapılan çalışma sayesinde kentsel içme suyu şebekelerinde suyun kalite seviyesinin belirlenmesinde doğruluk oranı yüksek sonuçlar elde edilmektedir.

Gelecek çalışmada, çeşitli veri kümeleri kullanılarak yeni bir derin öğrenme modeli geliştirilecektir. Derin öğrenme modelinde ön işlem adımları ve son işlem adımlarına yer verilerek performans artışı sağlanacaktır.

Kaynaklar

- [1] A. Boretti, L. Rosa, Reassessing the projections of the World Water Development Report, *Npj Clean Water*. 2 (2019) 15. doi:10.1038/s41545-019-0039-9.
- [2] R.J. Hogeboom, The Water Footprint Concept and Water's Grand Environmental Challenges, *One Earth*. 2 (2020) 218–222. doi:https://doi.org/10.1016/j.oneear.2020.02.010.
- [3] T.H.H. Aldhyani, M. Al-Yaari, H. Alkahtani, M. Maashi, Water Quality Prediction Using Artificial Intelligence Algorithms, *Appl. Bionics Biomech.* 2020 (2020) 6659314. doi:10.1155/2020/6659314.
- [4] M. Allaire, H. Wu, U. Lall, National trends in drinking water quality violations, *Proc. Natl. Acad. Sci.* 115 (2018) 2078–2083. doi:10.1073/pnas.1719805115.
- [5] Water | United Nations, UN Water. (2021). <https://www.un.org/en/global-issues/water> (accessed December 29, 2021).
- [6] M.S. Islam Khan, N. Islam, J. Uddin, S. Islam, M.K. Nasir, Water quality prediction and classification based on principal component regression and gradient boosting classifier approach, *J. King Saud Univ. - Comput. Inf. Sci.* (2021). doi:https://doi.org/10.1016/j.jksuci.2021.06.003.
- [7] N. Radhakrishnan, A.S. Pillai, Comparison of Water Quality Classification Models using Machine Learning, in: 2020 5th Int. Conf. Commun. Electron. Syst., 2020: pp. 1183–1188. doi:10.1109/ICCES48766.2020.9137903.
- [8] D. Venkata Vara Prasad, L. Y Venkataramana, P.S. Kumar, G. Prasannamedha, K. Soumya, P. A.J., Water quality analysis in a lake using deep learning methodology: prediction and validation, *Int. J. Environ. Anal. Chem.* (2020) 1–16. doi:10.1080/03067319.2020.1801665.
- [9] P. Smarty, Water Quality Dataset, Kaggle. (2021). <https://www.kaggle.com/mssmartypants/water-quality> (accessed December 31, 2021).
- [10] P. Schober, T.R. Vetter, Logistic Regression in Medical Research, *Anesth. Analg.* 132 (2021) 365–366. doi:10.1213/ANE.0000000000005247.
- [11] M. Toğaçar, B. Ergen, M.E. Sertkaya, Zatürre Hastalığının Derin Öğrenme Modeli ile Tespiti, *Fırat Üniversitesi Mühendislik Bilim. Derg.* 31 (2019) 223–230.
- [12] M. Jawthari, V. Stoffová, Predicting students' academic performance using a modified kNN algorithm, *Pollack Period.* 16 (2021) 20–26. doi:10.1556/606.2021.00374.
- [13] M.E. Sertkaya, B. Ergen, M. Togacar, Diagnosis of Eye Retinal Diseases Based on Convolutional Neural Networks Using Optical Coherence Images, in: 2019 23rd Int. Conf. Electron., 2019: pp. 1–5. doi:10.1109/electronics.2019.8765579.
- [14] A. Topîrceanu, G. Grossecck, Decision tree learning used for the classification of student archetypes in online courses, *Procedia Comput. Sci.* 112 (2017) 51–60. doi:https://doi.org/10.1016/j.procs.2017.08.021.
- [15] H. Polat, M. Turkoglu, O. Polat, Deep network approach with stacked sparse autoencoders in detection of DDoS attacks on SDN-based VANET, *IET Commun.* 14 (2020) 4089–4100. doi:10.1049/iet-com.2020.0477.
- [16] V. Tümen, B. Ergen, Intersections and crosswalk detection using deep learning and image processing techniques, *Phys. A Stat. Mech. Its Appl.* 543 (2020) 123510. doi:10.1016/j.physa.2019.123510.
- [17] M. Schonlau, R.Y. Zou, The random forest algorithm for statistical learning, *Stata J. Promot. Commun. Stat. Stata.* 20 (2020) 3–29. doi:10.1177/1536867X20909688.
- [18] V. Chang, T. Li, Z. Zeng, Towards an improved Adaboost algorithmic method for computational financial analysis, *J. Parallel Distrib. Comput.* 134 (2019) 219–232. doi:10.1016/j.jpdc.2019.07.014.
- [19] A. Ibrahim Ahmed Osman, A. Najah Ahmed, M.F. Chow, Y. Feng Huang, A. El-Shafie, Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia, *Ain Shams Eng. J.* (2021).

- doi:<https://doi.org/10.1016/j.asej.2020.11.011>.
- [20] T. Carneiro, R.V.M. Da Nóbrega, T. Nepomuceno, G. Bian, V.H.C. De Albuquerque, P.P.R. Filho, Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications, *IEEE Access*. 6 (2018) 61677–61685. doi:10.1109/access.2018.2874767.
- [21] E. Başaran, Z. Cömert, A. Şengür, Ü. Budak, Y. Çelik, M. Toğaçar, Chronic Tympanic Membrane Diagnosis based on Deep Convolutional Neural Network, in: 2019 4th Int. Conf. Comput. Sci. Eng., 2019: pp. 1–4. doi:10.1109/ubmk.2019.8907070.
- [22] M. Toğaçar, B. Ergen, Deep Learning Approach for Classification of Breast Cancer, in: 2018 Int. Conf. Artif. Intell. Data Process., 2018: pp. 1–5. doi:10.1109/idap.2018.8620802.
- [23] M. Toğaçar, B. Ergen, Biyomedikal Görüntülerde Derin Öğrenme ile Mevcut Yöntemlerin Kıyaslanması, *Fırat Üniversitesi Mühendislik Bilim. Derg.* 31 (2019) 109–121.
- [24] K.-S. Cheng, J.-Y. Ling, T.-W. Lin, Y.-T. Liu, Y.-C. Shen, Y. Kono, Quantifying Uncertainty in Land-Use/Land-Cover Classification Accuracy: A Stochastic Simulation Approach, *Front. Environ. Sci.* 9 (2021) 46. doi:10.3389/fenvs.2021.628214.
- [25] Scikit-learn developers, Machine learning, Scikit-Learn. (2019). https://scikit-learn.org/stable/supervised_learning.html#supervised-learning (accessed January 2, 2022).
- [26] Scikit-learn developers, Ensemble methods, Scikit-Learn. (2019). <https://scikit-learn.org/stable/modules/ensemble.html> (accessed January 2, 2022).
- [27] C. Sillberg, P. Kullavanijaya, O. Chavalparit, Water Quality Classification by Integration of Attribute-Realization and Support Vector Machine for the Chao Phraya River, *J. Ecol. Eng.* 22 (2021) 70–86. doi:10.12911/22998993/141364.
- [28] E.Q. Shahra, W. Wu, S. Basurra, S. Rizou, Deep Learning for Water Quality Classification in Water Distribution Networks, in: 2021: pp. 153–164. doi:10.1007/978-3-030-80568-5_13.
- [29] M. Hmoud Al-Adhaileh, F. Waselallah Alsaade, Modelling and Prediction of Water Quality by Using Artificial Intelligence, *Sustainability*. 13 (2021) 4259. doi:10.3390/su13084259.