

ARITMA ÇAMURLARINDA POLİSİKLIK AROMATİK HİDROKARBONLARIN (PAH'LARIN) GİDERİMLERİNİN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE TAHMİNİ

Burcu ÇAĞLAR GENÇOSMAN*^{ID}
Gizem EKER ŞANLI**^{ID}

Alınma: 21.10.2020; düzeltme: 07.12.2020; kabul: 10.01.2021

Öz: Çevreye ve insan sağlığına olumsuz etkileri olan polisiklik aromatik hidrokarbonların (PAH'ların) atıksu arıtma çamurlarından gideriminde kullanılan yöntemlerden biri UV-C (ultraviyole-C) ışığı ve fotokatalizörler varlığında gerçekleştirilen fotoparçalanma uygulamalarıdır. PAH gideriminin sağlanıp sağlanmadığı, gerçekleştirilen deneylerden sonra ortaya çıkar ve bu durum zaman ve maliyeti artırır. Alternatif olarak veri madenciliği sınıflandırma yöntemleri ile deney girdi koşullarına göre PAH'ların giderimi tahmin edilebilir, böylece zaman ve maliyet tasarrufu sağlanabilir. Bu sayede, arıtma çamurlarındaki başlangıç PAH konsantrasyonları esas alınarak UV teknolojilerinin kullanımı kararı daha az maliyet ve çabayla verilebilir. Çalışmanın ilk aşamasında 12 PAH türünü içeren 4 farklı özellikteki arıtma çamurunda UV uygulamaları gerçekleştirilerek PAH giderimleri belirlenmiş, sonrasında ilk aşamadaki sonuçlar veri kümelerinde kullanılarak başlangıç PAH seviyelerine göre PAH'ların giderimleri tahmin edilmiştir. Çok katmanlı algılayıcı (ÇKA) ağı, k-en yakın komşu (k-NN), C4.5 karar ağacı (C4.5), rastgele orman (RO) ve torbalama yöntemleri gibi çeşitli sınıflandırma yöntemleri giderim tahmini için kullanılmıştır. Performans karşılaştırmaları için kesinlik⁺, duyarlılık, belirleyicilik, %doğruluk, AUC (Alıcı işlem karakteristikleri eğrisi) ve F-ölçütü esas alınmıştır. Ortalama doğruluk parametresine göre en başarılı üç yöntem sırasıyla RO (%95,730), k-NN (%95,588) ve ÇKA (%91,275) yöntemleridir. Azınlık sınıfı tahmininde ise ortalama AUC göz önüne alındığında RO (0,974), k-NN (0,944) ve Torbalama (0,939) yöntemleri diğer yöntemlerden daha iyi performans göstermiştir.

Anahtar Kelimeler: Kentsel/endüstriyel atıksu arıtma çamuru, UV-C ışığı, veri madenciliği sınıflandırma yöntemleri, aşırı-örnekleme yöntemleri, PAH giderim tahmini.

Prediction of Polycyclic Aromatic Hydrocarbons (PAHs) Removal in Wastewater Treatment Sludge with Data Mining Methods

Abstract: One of the methods used in the removal of polycyclic aromatic hydrocarbons (PAHs), which are known to have negative effects on the environment and human health, from wastewater treatment sludge, is photodegradation applications performed with UV-C (ultraviolet-C) light and photocatalysts. However, the PAH removal is revealed after the experiments performed and this increases the time and cost. Alternatively, with the data mining classification methods, the removal of PAHs can be estimated before the experiments are carried out; hence, the application of UV technologies is decided with less cost and effort. In this study, UV applications were performed on 4 types of treatment sludge containing 12 PAH types, and PAH removals were determined. Then the removal of PAHs was estimated regarding the initial PAH levels. Multi-layer perceptron (MLP) network, k-nearest neighbor (k-NN), C4.5 decision tree (C4.5), random forest (RF), and bagging were performed for the removal prediction. Precision⁺, recall, specificity, accuracy%, AUC (Area Under the ROC Curve), and F-measure were used for performance comparisons. Regarding the average accuracy, the three most successful methods are RO (95.730%), k-NN (95.588%)

* Bursa Uludağ Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 16059, Nilüfer, Bursa

** Bursa Uludağ Üniversitesi, Mühendislik Fakültesi, Çevre Mühendisliği Bölümü, 16059, Nilüfer, Bursa

İletişim Yazarı: Burcu Çağlar Gençosman (burcucaglar@uludag.edu.tr)

and MCA (91.275%), respectively. Considering the average AUC, RO (0.974), k-NN (0.944) and Bagging (0.939) methods performed better than other methods.

Keywords: Urban/industrial wastewater treatment sludge, UV-C light, data mining classification methods, over-sampling methods, PAH removal prediction.

1. GİRİŞ

Polisiklik aromatik hidrokarbonlar (PAH'lar), çoğunlukla fosil yakıtların tam yanmaması sonucu çevreye salınan yarı uçucu organik bileşiklerdir (YUOB'lerdir). PAH'ların çoğu çevrede uzun süre kalmaları ve birikimleri sonucu kirliliğe sebep olurlar ve biyolojik dengeyi bozarlar (Sprovieri ve diğ., 2007). Bu nedenle toprak, hava ve su gibi tüm çevresel ortamlardan giderilmesi veya kontrol altına alınması gerekir. İnsan faaliyetlerine bağlı olarak çok sayıda organik kirlilik atıksu yapısına geçmektedir. Yapılan araştırmalar, atıksularda 6000 organik bileşiğin bulunabileceğini göstermiştir (European Communities, 2001). PAH ve diğer YUOB'ler atıksu arıtma tesislerinde görülen kirleticilerdir. PAH'lar kanalizasyon sistemiyle taşınarak veya atmosferden çökerek atıksu arıtma tesisine gelmekte ve çamurun yapısına geçmektedir (Blanchard ve diğ., 2004; Stevens ve diğ., 2003). Arıtma çamurlarının bertarafında kullanılacak yöntemin seçiminde çamur karakterizasyonunun göz önünde bulundurulması gerekir (Filibeli, 1996). PAH gibi YUOB'leri içeren çamurların bertarafında dikkatli olunması ve gerekli ön arıtmaların yapılması uygun olacaktır. Ülkemizde arıtma çamurlarının çok miktarda oluştuğu ve bertaraf sıkıntısının yaşandığı göz önünde bulundurulduğunda, çamurlardaki PAH kirliliğinin tespiti ve giderim yöntemlerine yönelik araştırmaların artırılmasının önemi daha iyi anlaşılmaktadır.

PAH'ların atıksudan ve arıtma çamurlarından uzaklaştırılmasına yönelik literatürde çok sayıda araştırma çalışması (Ballesteros-Gómez ve diğ., 2019; Da Rocha ve diğ., 2013; Eker ve Hatipoglu, 2019; Eker ve diğ., 2019; Ghasemi ve diğ., 2019; Jing ve diğ., 2014; Karaca ve Tasdemir, 2013; Kong ve diğ., 2018; Manoli ve Samara, 2008; Oh ve diğ., 2016; Ren ve diğ., 2020; Saber ve diğ., 2020; Wang ve diğ., 2018; Ren ve diğ., 2020) ve derleme yayınları yer almaktadır (Gaur, ve diğ., 2018; Mojiri ve diğ., 2019; Syafiuddin ve Boopathy, 2020; Zhang ve diğ., 2019). PAH'ların gideriminde kullanılacak yöntemlerden bir tanesi fotoparçalanma mekanizmasının etkin olduğu UV uygulamalarıdır (Salihoglu ve diğ., 2012). Çeşitli bilim insanları, UV ışığının toprak/arıtma çamuru gibi katı matrikslerdeki PAH'ların gideriminde kullanılabileceğini (Dong ve diğ., 2010; Karaca ve Tasdemir, 2013; Salihoglu ve diğ., 2012; Zhang ve diğ., 2008) ve PAH'ların ışık etkisiyle parçalanma etkinliklerini artırmak için fotokatalizör/fotosensitizör olarak Titanyum dioksit (TiO₂) ve Dietilamin (DEA) gibi kimyasallardan faydalanılabileceğini ortaya koymuştur (Lin ve diğ., 2004; Rababah ve Matsuzawa, 2002).

Atıksu arıtma çamurundan fotoparçalanma yoluyla PAH'ların giderilmesi oldukça karmaşık ve maliyetli bir süreç olup çok sayıda deney yapmayı gerektirir. Bunun sebebi girdi laboratuvar koşullarının (sıcaklık, katkı maddesi türü ve miktarı vb.) çeşitlendirilmesi gerekliliğidir. Diğer taraftan, veri madenciliği sınıflandırma yöntemleri ile girdi deney koşulları göz önüne alınarak PAH gideriminin olup olmayacağı tahmin edebilir ve UV teknolojisinin etkin bir şekilde uygulanabilirliği konusunda bir öngörüle/tahminde bulunarak gerçekleştirilecek deney sayısı, iş yükü ve maliyet azaltılabilir. Bu yöntemler, bir girdi veri kümesinden sınıflandırma modelleri oluşturur ve bu modeller aracılığıyla giderim işlemini tahmin eder (Fan ve diğ., 2018). Sınıflandırma modelleri, öğrenme algoritmaları kullanılarak makine öğrenmesi yöntemleri içinde oluşturulabilir. Makine öğrenmesi yöntemleri kapsamındaki sınıflandırma yöntemleri, hem tarihsel verilere uyan hem de yeni (karşılaşılmamış) verilerin sınıf etiketlerini doğru bir şekilde tahmin eden başarılı genelleme kapasitesine sahip sınıflandırma modelleri oluşturmayı amaçlamaktadır. Büyük veri kümelerindeki yararlı bilgiler keşfedilebilir, yeni ve bilinmeyen

modeller sınıflandırma yöntemleriyle bulunabilir. Çeşitli araştırmacılar tarafından atıksulardaki kirletici miktarlarının/giderimlerinin tahmini için sınıflandırma yöntemleri kullanılmıştır (Farzin ve diğ., 2020; Salari ve diğ., 2005; Talebkeikhah ve diğ., 2020; Yaqub ve diğ., 2020; Yaqub ve Lee, 2020). Konu ile ilgili bilimsel yayınlardaki artış, veri madenciliği ve makine öğrenmesi araçlarının çevre kirliliği kontrolü için önemini göstermektedir (Bhagat ve diğ., 2020; Corominas ve diğ., 2018; Fan ve diğ., 2018; Hadjimichael ve diğ., 2016; Newhart ve diğ., 2019; Ye ve diğ., 2020; Zhao ve diğ., 2020).

Çeşitli araştırmacılar, arıtma çamurları için farklı amaçlarla yapay zekâ tabanlı modelleri ve yapay sinir ağları (YSA) yaklaşımlarını kullanmıştır. Ofman ve Struk-Sokołowska (2019), kanalizasyon çamurundaki azotlu bileşiklerin giderilmesinin tahmini için bir YSA yaklaşımı geliştirmiştir. Boztoprak ve diğ. (2016), aktif çamur tesisinde YSA yaklaşımı ile çamur hacim endeksini tahmin etmiştir. Picos-Benítez ve diğ. (2017), anaerobik çamur reaktörünün KOİ (kimyasal oksijen ihtiyacı) giderim verimliliğini optimize etmek için yapay zekâ tabanlı bir model geliştirmiştir. Literatürde atıksu arıtma çamurlarındaki PAH'ların giderim tahmininde sınıflandırma yöntemlerinin kullanımına yönelik herhangi bir çalışmaya rastlanmamıştır. Sunulan çalışma bu alanda yapılan ilk araştırmalardan biridir.

Bu çalışmanın amacı, kentsel ve endüstriyel arıtma çamurlarının giriş PAH konsantrasyonlarına göre UV uygulamalarıyla PAH gideriminin sağlanıp sağlanamayacağını tahmin edilmesinde başarıyla kullanılacak sınıflandırma yöntemlerini belirlemektir. Bu kapsamda, yapay sinir ağları (çok katmanlı algılayıcı-ÇKA) (McCulloch ve Pitts, 1943; Rosenblatt, 1958), k-en yakın komşu (k-NN) (Cover ve Hart, 1967), C4.5 karar ağaçları (C4.5) (Quinlan, 1993), rastgele orman (RO) (Breiman, 2001) ve torbalama (Breiman, 1996) yöntemleri önerilmiştir. Tahmin sürecini iyileştirmek için ana veri kümesi PAH bileşikleri ve PAH halkalarına (altı bileşik ve üç PAH halka grubu) göre ayrılmıştır. Altı bileşik Phe, Ant, Fl, Pyr, BaA ve Chr ve halka grupları 3 halkalı (Phe ve Ant), 4 halkalı (Fl, Pyr, BaA ve Chr) ve 5+6 halkalı (BbF, BkF, BaP, InP, DahA ve BghiP) şeklindedir. Veri kümeleri, rastgele aşırı örnekleme ve SMOTE ile genişletilmiş ve 30 farklı veri kümesi elde edilmiştir. 2019-2020 yıllarında gerçekleştirilen çalışmada farklı veri kümelerindeki PAH bileşiklerinin giderim tahmininde başarılı olan yöntemler belirlenmiş ve aşırı örnekleme yöntemlerinin tahmin kalitesi üzerindeki olumlu etkisi kanıtlanmıştır.

2. MATERYAL ve YÖNTEM

Bu bölümde çamur örnekleme, PAH giderim uygulamaları, PAH'ların analizi, veri madenciliği yaklaşımı aşamaları, veri hazırlama ve ön işleme aşamaları, aşırı-örnekleme yöntemleri ve veri kümelerinin özellikleri, PAH'ların gideriminin tahmini için önerilen sınıflandırma yöntemleri ve performans karşılaştırmaları için kullanılan değerlendirme parametreleri detaylı biçimde verilmiştir.

2.1. Çamur örnekleme ve PAH giderim uygulamaları

2.1.1. Çamur örnekleme

Dört farklı işletmeden arıtma çamuru örnekleri alınmış olup detaylar aşağıdaki gibidir. Çamur örnekleme çalışmalarında temsili örnek alınması önemli bir husustur. Sunulan çalışmada, bunu sağlamak için ilgili ünite çıkışından cam kaplarla yaklaşık 5'er kg'lık örnekler alınmıştır. Çamur örnekleri, alüminyum kapaklı cam kavanozlarda laboratuvara getirilerek yaklaşık 10 dakika cam çubuklarla karıştırılarak homojenize edilmiştir. Katı madde analizi ve ekstraksiyon dâhil olmak üzere tüm ön işlemler tamamlanmıştır.

Kentsel çamur: Bursa'da bulunan ve nüfusu 585.000'e eşdeğer bir kentsel atıksu arıtma tesisinin bant filtre pres çıkışından örnek alınmıştır. Tesiste atıksuyun %35'i endüstriyel, kalan %65'i evsel kaynaklıdır.

Gıda sanayi çamuru: Yıllık 20.000 ton konserve meyve ve sebze üretim kapasitesine sahip bir gıda fabrikasından örnek alınmıştır. Bu fabrika yılda 50 ton kuru çamur üretmektedir. Arıtma tesisinde sırasıyla dengeleme, nötralizasyon, havalandırma ve çöktürme üniteleri yer almaktadır. Çamur örnekleri belt filtre pres çıkışından alınmıştır.

Otomotiv sanayi çamuru: Yıllık 360.000 otomobil ve 450.000 motor üretim kapasitesine sahip bir otomotiv fabrikasının atıksu arıtma tesisinden örnek alınmıştır. Tesiste, dengeleme, fizikokimyasal arıtma, biyolojik arıtma ve son çöktürme havuzları yer almaktadır. Son çökeltimde oluşan çamur stabilizasyon işlemine tabi tutulup filtre preste suyu alınmaktadır. Bant filtre pres çıkışından çamur örneği alınmıştır.

Organize sanayi bölgesi (OSB) çamuru: Bursa'da faaliyet gösteren bir OSB'deki atıksu arıtma tesisinden örnek alınmıştır. OSB arıtma tesisine 36 firmadan konsantre asit/alkali atıksu, konsantre yağlı atıksu ve sürekli yıkama suyu gelmektedir. Dekantör ünitesinden çamur örneği alınmıştır.

Ölçülmesi hedeflenen PAH'lar, Amerika Birleşik Devletleri Çevre Koruma Ajansı (United States Environmental Protection Agency, US-EPA) tarafından yayınlanan öncelikli kirleticiler listesinde yer alan 16 PAH türünden 12 tanesidir. Uçuculuk özellikleri nispeten fazla olan 4 türün (Naphthalene (Nap), Acenaphthylene (Acy), Acenaphthene (Ace), Fluorene (Flu)) tespit edilmesi oldukça güçtür. Nitekim sunulan çalışmada bu türlerin ölçümü için çalışmalar yapılmış ancak geri kazanım değerleri <%10 çıktığı için rapor edilmemiştir. Ölçülen ve rapor edilen türler sırasıyla fenantren (Phe), antrasen (Ant), florenten, (Fl), piren (Pyr), benzo[a]antrasen (BaA), krisen (Chr), benzo[b]florenten (BbF), benzo[k]florenten (BkF), benzo[a]piren (BaP), indeno[1,2,3-c,d]piren (InP), dibenzo[a,h]antrasen (DahA) ve benzo[g,h,i]-perilen (BghiP) bileşikleridir.

2.1.2. UV düzeneği ve PAH giderim uygulamaları

Deneylerde Karaca (2013) tarafından özel olarak tasarlanan düzenek kullanılmıştır. Düzeneğin üstüne iki santimetre aralıklarla 254 nm'de 24 W gücünde üç Ultraviyole-C (UV-C) lamba (Philips TUV G8T5) yerleştirilmiştir. Paslanmaz çelikten imal edilen düzenekte, PAH ile her türlü girişimin önlenmesi için organik içeriği olmayan malzemeler kullanılmıştır. Düzenek boyutları 45 x 30 x 55 (en x boy x yükseklik) cm şeklinde olup, düzeneğe girişte konumlandırılan PUF kolonu ile düzenek içine alınan hava PAH'lardan arındırılmıştır. Böylelikle, havadaki PAH'ların çamura geçmesi engellenmiştir. Örnekler, cam petri kabına serilip UV kaynağından 18 cm mesafede yer alan rafa yerleştirilmiştir. Düzeneğe giren hava içeride 6 saniye kalmış olup çamurun sürekli temiz hava ile temas etmesi sağlanmıştır. İç ortamdaki nem ve sıcaklık takibi HOBOS-Thb M002 marka sensör ile yapılmış ve veriler H21-002 HOBOS veri derleyicide toplanmıştır. Düzeneğe ait detaylı bilgiler Karaca (2013) ve Karaca ve Tasdemir (2013) çalışmalarında verilmiştir. UV, UV-TiO₂ ve UV-DEA gibi çeşitli PAH giderim uygulamaları için deney süresi 24 saat olarak seçilmiştir. Fotokatalizör olarak TiO₂'nin nano kristal toz formu Degussa P25, (yüzey alanı ≈50 m²/g) ve reaksiyonları aktive etmek üzere fotosensitizör olarak Merck (8.03010.2500) marka DEA (GC alanı ≥% 99, yoğunluk:0,7 g/cm³ (20°C'de), pH:13) kullanılmıştır.

2.1.3. PAH ekstraksiyon, temizleme ve belirleme

İlk olarak, arıtma çamurlarının başlangıçtaki PAH konsantrasyonları belirlenmiştir. Bunun için 5 g çamur örneğine hacimce 1/1 oranında Diklorometan/Petrol eteri (DCM/PE)'den oluşan 20 mL'lik solvent karışımı ve 1 mL verim standardı ilave edilerek ultrasonik banyoda ekstrakte edilmiştir. Herbir çamur örneği düzenekte PAH giderim uygulamalarına tabi tutulmuş 24 saat sonunda düzenekten çıkan örneklerle de aynı ekstraksiyon işlemi uygulanmıştır. Sonrasında

solvent değişimi, hacim azaltma ve fraksiyonlarına ayırma işlemleri gerçekleştirilmiştir. Ekstraksiyona ilişkin detaylar ve PAH ön analiz aşamaları Karaca (2013), Karaca ve Tasdemir (2013) ve Salihoğlu ve diğ. (2010) çalışmalarındaki gibidir. PAH konsantrasyonları, üç eksenli detektörlü (MSD) bir Agilent 5975C atıl XL kütle seçici ile donatılmış bir Agilent 7890 Model Gaz Kromatografisi (GC) kullanılarak ölçülmüştür. MSD, optimum hassasiyet için seçilen iyon izleme (SIM) modunda çalıştırılmış PAH analizinden önce GC-MS sonuçlarının doğrusallığını belirlemek için yedi konsantrasyon seviyesinde (0,01; 0,1; 0,5; 1,25; 2,5; 5 ve 10 g/mL) kalibrasyonlar gerçekleştirilmiştir. Kalibrasyon eğrisinin tüm seviyelerinde, her tür için r^2 değeri 0,99'dan büyüktür. Her 24 örnekte bir orta nokta kalibrasyon standardının analizi yapılarak doğrulama işlemi gerçekleştirilmiş ve cihazın performansı değerlendirilmiştir. Bir uL'lik bir enjeksiyon için ölçülebilir PAH miktarı 0,1 pg olarak belirlenmiştir.

2.1.4. Kalite güvenilirliği / Kalite kontrol

Her bir örneğe (çamur ve şahit örnekleri) 4 ng/mL Standard Mix A PAH verim standardı ilave edilmiştir. Mix A PAH standardı Phenanthrene-d10, Chrysene-d12 ve Perylene-d12 içermekte olup her bir PAH için için ortalama verim değerleri sırasıyla %30, %50 ve %85 olarak hesaplanmıştır. Bu çalışmada, üç tekrarlı deneyler sonucunda elde edilen ortalama PAH giderim verimleri rapor edilmiştir. Toplam örnek sayısının %10'u kadar şahit örneği hazırlanarak deneysel çalışmalar sırasında herhangi bir kirlenme olup olmadığı tespit edilmiştir. Çamur örneklerine uygulanan tüm PAH ön analiz işlemleri şahitlere de uygulanmıştır. GC-MS, verim standartlarına göre kalibre edilmiş, her bir PAH bileşiği için gözlenebilir sınır değerleri (GSD'ler) şahit ortalamaları +3 standart sapma formülasyonu ile hesaplanmıştır. Kalite güvenilirliği/kalite kontrol uygulamalarına ait diğer detaylar Karaca (2013) ve Karaca ve Tasdemir (2013) çalışmalarındaki gibidir.

2.2. Veri Madenciliği Aşamaları

Veri madenciliği, büyük veritabanlarından yararlı bilgileri otomatik olarak keşfetme sürecidir. Veritabanlarında bilgi keşfinin ayrılmaz bir parçası olan veri madenciliği teknikleri ile veri ön işlemeden sonuçların son işlemesine kadar bir dizi dönüştürme adımı ile ham veriler yararlı bilgilere dönüştürülür. Veri madenciliği süreci beş adımdan oluşmaktadır (Fayyad ve diğ, 1996; Akpınar, 2014). Bilgi keşfinin uygulama alanı ve amacı belirlendikten sonra bir hedef veri seçimi veya oluşturulması ile veri madenciliği süreci başlar. Veri temizleme ve ön işleme adımı veri gürültüden arındırılır ve eksik veri alanları çeşitli stratejilerle giderilerek veri kalitesi artırılır. Dönüştürme aşamasında veri, kullanılacak algoritmalara uygun formata dönüştürülür. Sonraki aşamada veri madenciliği yöntemleri kullanılarak verinin barındırdığı gizli bilgiler ortaya çıkarılır. Son aşamada elde edilen bilgiler yorumlanır; eğer elde edilen bilgiler yeterli değilse önceki aşamalar tekrar edilir (Fayyad ve diğ, 1996; Akpınar, 2014).

Bu çalışmada PAH giderim uygulamaları aşamasında gerçekleştirilen deneylerden elde edilen veriler kullanılarak, PAH gideriminin mevcut deney koşullarına göre tahmin edilmesi ile giderim sağlayan deney koşullarının önceden tespit edilmesinde en başarılı olan tahmin yöntemlerinin belirlenmesi amaçlanmıştır. Bu amaçla veri seçimi gerçekleştirilmiş ve bölüm 2.3'de anlatılan veri hazırlama ve ön işleme adımları gerçekleştirilmiş ve veri dönüştürme ile verinin veri madenciliği yöntemlerinde kullanılabilir hale getirilmesi sağlanmıştır. Bölüm 2.4'de giderim tahmininde kullanılan veri madenciliği yöntemleri detaylandırılmış ve son olarak elde edilen sonuçlar değerlendirilmiştir. Bu şekilde veri madenciliği adımları tamamlanmıştır.

2.3. Veri Hazırlama ve Ön işleme

Bu çalışmada dört farklı tesisten çamur örnekleri alınmış, PAH giderim uygulamaları yapılmış ve giderim yüzdeleri hesaplanarak kaydedilmiştir. Tüm kirleticileri içeren ana veri

kümesinde, ön işleme (preprocessing) aşamasından önce 1752 örnek bulunmaktadır. Ön işlemden ve bilinmeyen örneklerin ortadan kaldırılmasından sonra, veri kümesi 1296 örnek içermektedir. Ana veri kümesinin öznitelikleri ve özellikleri Tablo 1’de özetlenmiştir.

Ana veri kümesinde 10 farklı öznitelik bulunmaktadır. Veri madenciliği yöntemlerinin uygulanabilmesi ve başarılı olabilmesi için bu özniteliklerin ilgili formatlara dönüştürülmesi gerekmektedir. Veri madenciliği süreci aşamalarından biri olan veri dönüştürmede, verinin yöntemlere uygun hale getirilmesi için belli bir fonksiyona göre farklı değer ve/veya ölçeklere dönüşüm gerçekleştirilebilir (Akpınar, 2014). Veri dönüştürmenin gerekliliğini vurgulamak için PAH miktarı ve sıcaklık nümerik değerleri ele alınabilir. Örneğin PAH miktarı [0 ng/g katı madde; 2097,7 ng/g katı madde] aralığında değişirken sıcaklık [0°C; 60°C] aralığında değişmektedir. Veri madenciliği yöntemleri bu öznitelikleri değerlendirirken PAH miktarındaki büyük değerler, sıcaklıktaki düşük değerlere baskın gelebilir ve sıcaklığın gerçekteki etkisi sonuçlara yansımayaabilir. Özniteliklerin etkilerini başarılı bir şekilde sonuçlara dâhil etmek amacıyla özniteliklerin yapısına uygun dönüşümlerin gerçekleştirilmesi gerekmektedir. Örnek olarak, PAH miktarı ondalıklı bir yapıya sahip herhangi bir doğal sayı olabilir, dolayısıyla PAH miktarı [0,1] aralığına normalize edilebilir. Diğer taraftan sıcaklık değerleri için seviyeler önceden belirlenmiştir ve bunlar dışında herhangi bir doğal sayı olamaz. Bu sebeple sıcaklık için önceden belirlenen nümerik değerler yerine kategoriler kullanılabilir. Bu çalışmada gerçekleştirilen veri dönüştürme işlemleri Tablo 1’deki özniteliklere uygulanmış ve dönüşüm sonrası özniteliklerin özellikleri veri tipi ve seviyeler sütunlarında belirtilmiştir.

Tablo 1. Ana veri kümesi öznitelikleri ve özellikleri

Öznitelik	Tanım	Ham veri özelliği	Veri tipi	Seviyeler
Çamur tipi	Çamurun alındığı tesis	4 kategori: Kentsel, Otomotiv, Gıda, OSB.	Kategorik	4 kategori: Kentsel, Otomotiv, Gıda, OSB.
PAH bileşikleri	PAH bileşikleri	12 kategori: Phe, Ant, Fl, Pyr, BaA, Chr, BbF, BkF, BaP, InP, DahA ve BghiP.	Kategorik	12 kategori: Phe, Ant, Fl, Pyr, BaA, Chr, BbF, BkF, BaP, InP, DahA ve BghiP.
PAH miktarı (ng/g katı madde)	Her çamur tipindeki PAH miktarı	Nümerik değerler: [0 ng/g katı madde; 2097,7 ng/g katı madde].	Nümerik	Normalize edilmiş nümerik değerler: [0; 1].
Kimyasal tipi	PAH gideriminde kullanılan kimyasal	3 kategori: YOK, TiO ₂ , ve DEA.	Kategorik	3 kategori: YOK, TiO ₂ , ve DEA.
Kimyasal dozu (%)	Kullanılan kimyasalın dozu	Nümerik değerler: %0; %0,1; %0,5; %5, %10 ve %20.	Kategorik	6 kategori: 1, 2, 3, 4, 5 ve 6.
UV	UV açık veya kapalı	Açık, Kapalı.	İkili	Açık, Kapalı.
Termostat	Termostat kullanımının olup olmadığı	Var, Yok	İkili	Var, Yok
Düzenek	Düzenek kullanımının olup olmadığı	Var, Yok*	İkili	Var, Yok
Sıcaklık(°C)	Deneylerin yapıldığı sıcaklık	Nümerik değerler: [0°C; 60 °C].	Kategorik	6 kategori: 1, 2, 3, 4, 5, ve 6.
Giderim	Her bir uygulama için PAH giderim yüzdesi	Nümerik değerler:[-%100; %100]	İkili	Var, Yok

*Düzenek yok: PAH giderim uygulamalarının düzenekte değil açık hava koşullarında yapıldığını ifade eder.

Çamur tipi özniteliği, örneklerin toplandığı dört farklı tesisi gösterir. Bu çamur örnekleri, çok sayıda PAH bileşiklerini ve diğer organik kirleticileri içermektedir; ancak bu çalışmada US-EPA tarafından önceliklendirilen 12 PAH bileşiği dikkate alınmıştır. PAH miktarı, çamur örneklerindeki her bir PAH bileşiğinin konsantrasyonunu ifade eder. PAH miktarı (ng/g katı madde) niteliği [0 ng/g katı madde; 2097,7 ng/g katı madde] arasında olduğu için bu nitelik [0;1] aralığına normalize edilmiştir. Kimyasal türü (TiO₂, DEA) ve kimyasal dozu, PAH giderimi için kullanılan kimyasalları ve bunların oranını göstermektedir. Kimyasal dozun orijinal değerleri (%0; %0,1; %0,5; %5; %10 ve %20) sırasıyla (1, 2, 3, 4, 5 ve 6)'ya dönüştürülmüştür. Deneyler termostatlı veya termostatsız olarak tanımlanmış ve farklı sıcaklıklar kullanılmıştır. Orijinal

sıcaklık [0°C; 60°C] arasında değişmektedir. Bu nitelik de altı aralıkta gruplandırılmıştır; 1 (0°C-10°C), 2 (11°C-20°C), 3 (21°C-30°C), 4 (31°C-40°C), 5 (41°C-50°C) ve 6 (51°C-60°C). Deneyler giderim koşullarının farklı kombinasyonları ile gerçekleştirilmiştir. Deneylerin sonucunda PAH giderim yüzdeleri hesaplanmıştır. Bu yüzdeler [%-100; %100] arasında değişmektedir. Negatif değerler giderim olmadığını göstermiştir. Çıktı özneliği olarak giderim yüzdeleri yerine PAH gideriminin olup olmadığı dikkate alınmıştır. Farklı giderim uygulamalarının PAH'lar üzerindeki etkisini araştırmak için, ana veri kümesi PAH halka sayısına göre gruplandırılmıştır. Ana veri kümesi Tablo 2'deki gibi, 3 halkalı PAH'lara (Phe ve Ant), 4 halkalı PAH'lara (Fl, Pyr, BaA ve Chr) ve 5+6 halkalı (5 halkalılar ve 6 halkalılar toplamı) PAH'lara (BbF, BkF, BaP, InP, DahA ve BghiP) ayrılmıştır. Ek olarak, ana veri kümesi Tablo 3'deki gibi Phe, Ant, Fl, Pyr, BaA ve Chr olmak üzere altı bileşik dikkate alınarak bölünmüş ve her bir PAH için yeni bir veri kümesi oluşturulmuştur. Bu tablolarda veri kümesi no, veri kümesinin adı ve veri kümesi ile ilgili genel bilgiler verilmiştir. Tablo 2'de PAH bileşikleri sütununda ilgili veri kümesinde yer alan PAH bileşikleri belirtilmiştir. Tablo 3'deki veri kümeleri spesifik PAH bileşikleri için oluşturulduğundan veri kümesinin adında PAH bileşiği belirtilmiştir. Ayrıca Tablo 2 ve Tablo 3'de bulunan giderim çıktısı sütunu "giderim var" ve "giderim yok" kategorilerine ait olan örnek adetlerini gösterirken örneklem adedi sütunu toplam örnek adedini belirtmektedir. Giderim çıktısı sütununda örneklerin kategorilere göre dağılımının dengesiz olduğu görülebilir.

Tablo 2. Çoklu PAH'larla veri kümeleri özellikleri

Veri kümesi no.	Adı	Tanım	PAH bileşikleri	Giderim çıktısı	Örneklem adedi
1	TVO	Toplam veri orijinal (TVO): Orijinal boyut toplam veri kümesi.	12 tür: Phe, Ant, Fl, Pyr, BaA, Chr, BbF, BkF, BaP, InP, DahA, BghiP.	VAR:916; YOK:380.	1296
2	TVS	Toplam veri SMOTE (TVS): SMOTE uygulanmış toplam veri kümesi.	12 tür: Phe, Ant, Fl, Pyr, BaA, Chr, BbF, BkF, BaP, InP, DahA, BghiP.	VAR:916; YOK:1140.	2056
3	TVR	Toplam veri rassal (TVR): Rassal aşırı-örnekleme uygulanmış toplam veri kümesi.	12 tür: Phe, Ant, Fl, Pyr, BaA, Chr, BbF, BkF, BaP, InP, DahA, BghiP.	VAR:1832; YOK:760.	2592
4	3RO	Orijinal boyut 3-halkalı PAH kümesi	2 tür: Phe ve Ant	VAR:234; YOK:58.	292
5	3RS	SMOTE uygulanmış 3-halkalı PAH kümesi	2 tür: Phe ve Ant	VAR:234; YOK:174.	408
6	3RR	Rassal aşırı-örnekleme uygulanmış 3-halkalı PAH kümesi	2 tür: Phe ve Ant	VAR:468; YOK:116.	584
7	4RO	Orijinal boyut 4-halkalı PAH kümesi	4 tür: Fl, Pyr, BaA ve Chr	VAR:322; YOK:161.	483
8	4RS	SMOTE uygulanmış 4-halkalı PAH kümesi	4 tür: Fl, Pyr, BaA ve Chr	VAR:322; YOK:483.	805
9	4RR	Rassal aşırı-örnekleme uygulanmış 4-halkalı PAH kümesi	4 tür: Fl, Pyr, BaA ve Chr	VAR:644; YOK:322.	966
10	5RO	Orijinal boyut 5+6 halkalı PAH kümesi	6 tür: BbF, BkF, BaP, InP, DahA, ve BghiP	VAR:360; YOK:161.	521
11	5RS	SMOTE uygulanmış 5+6 halkalı PAH kümesi	6 tür: BbF, BkF, BaP, InP, DahA, ve BghiP	VAR:360; YOK:483.	843
12	5RR	Rassal aşırı-örnekleme uygulanmış 5+6 halkalı PAH kümesi	6 tür: BbF, BkF, BaP, InP, DahA, ve BghiP	VAR:720; YOK:322.	1042

Veri kümelerindeki çıktı sınıfları arasındaki dengesizlikleri gidermek ve tahmin kalitesini iyileştirmek için aşırı-örnekleme yöntemleri uygulanmış (rastgele aşırı-örnekleme ve SMOTE) ve yeni veri kümeleri oluşturmuştur. Tablo 2 ve Tablo 3'de üretilen 30 veri kümesinin özellikleri detaylandırılmıştır. Tablo 3'deki veri kümeleri spesifik PAH'lar için üretildiğinden bu veri

kümelere ile yapılan tahminlerde PAH bileşiklerindeki özneliliği sütununda ilgili veri kümesindeki spesifik PAH bileşiminin adı yer alacaktır. Veri kümesindeki bütün örnekler için tekrar eden bu özneliliğin sonuçlar üzerinde bir etkisi olmayacağı için kullanılmamıştır. Aşırı-örnekleme yöntemleri sonraki bölümde detaylandırılmıştır.

Tablo 3. Spesifik PAH'larla veri kümeleri özellikleri

Veri kümesi no.	Adı	Tanım	Giderim çıktısı	Örnekleme adedi	Veri kümesi no.	Adı	Tanım	Giderim çıktısı	Örnekleme adedi
13	PheO	Orijinal boyut Phe veri kümesi	VAR:131; YOK:15.	146	22	PyrO	Orijinal boyut Pyr veri kümesi	VAR:83; YOK:63.	146
14	PheS	SMOTE uygulanmış Phe veri kümesi	VAR:131; YOK:45.	176	23	PyrS	SMOTE uygulanmış Pyr veri kümesi	VAR:83; YOK:189.	272
15	PheR	Rassal aşırı-örnekleme uygulanmış Phe veri kümesi	VAR:262; YOK:30.	292	24	PyrR	Rassal aşırı-örnekleme uygulanmış Pyr veri kümesi	VAR:166; YOK:126.	292
16	AntO	Orijinal boyut Ant veri kümesi	VAR:103; YOK:43.	146	25	BaAO	Orijinal boyut BaA veri kümesi	VAR:82; YOK:26.	108
17	AntS	SMOTE uygulanmış Ant veri kümesi	VAR:103; YOK:129.	232	26	BaAS	SMOTE uygulanmış BaA veri kümesi	VAR:82; YOK:78.	160
18	AntR	Rassal aşırı-örnekleme uygulanmış Ant veri kümesi	VAR:206; YOK:86.	292	27	BaAR	Rassal aşırı-örnekleme uygulanmış BaA veri kümesi	VAR:164; YOK:52.	216
19	FIO	Orijinal boyut Fl veri kümesi	VAR:97; YOK:11.	108	28	ChrO	Orijinal boyut Chr veri kümesi	VAR:60; YOK:61.	121
20	FIS	SMOTE uygulanmış Fl veri kümesi	VAR:97; YOK:33.	130	29	ChrS	SMOTE uygulanmış Chr veri kümesi	VAR:60; YOK:183.	241
21	FIR	Rassal aşırı-örnekleme uygulanmış Fl veri kümesi	VAR:194; YOK:22.	216	30	ChrR	Rassal aşırı-örnekleme uygulanmış Chr veri kümesi	VAR:120; YOK:122.	242

2.4. Veri Aşırı-Örnekleme Yöntemleri

Veri kümelerinde çıktı sütunu iki kategoriden oluşmaktadır. Tablo 2 ve Tablo 3 incelendiğinde veri kümelerinde çıktı sınıflarının dengeli dağılmadığı görülmektedir. Bu durumda az sayıda var olan sınıfı tahmin etmek güçleşmektedir. Veri madenciliği alanında, nadir olayların tahmin edilmesi, olayların seyrekliği ve gelişigüzeliliği nedeniyle zordur; ancak, nadir olayların yanlış sınıflandırılması da ağır maliyetlere neden olabilmektedir (Haixiang et al., 2017). Çıktılar arasındaki dengesizliği gidermek için veri kümesinin boyutunun artırılması modellere ekstra bir iş yükü getirirse de tahmin kalitesini iyileştiren alternatif bir çözüm yöntemidir ve araştırmacılar tarafından dengesiz veri kümelerinin analizinde başarıyla kullanılmaktadır (Japkowicz ve Stephen, 2002). Aşırı-örnekleme yaklaşımı (*oversampling*) ile çıktılar arasındaki dengesiz dağılım derecesini azaltmak için az görülen azınlık sınıfı örneklerinin sayısı çoğaltılarak veya yerine üretilerek artırılır (Nguwi ve Cho, 2010).

Aşırı-örnekleme yaklaşımları konusunda pek çok çalışma yapılmış olmasına rağmen (Malhotra ve Kamal, 2019; Jo ve Japkowicz, 2004; Bunkhumpornpat ve diğ., 2009; Bunkhumpornpat ve diğ., 2012; Nekooimehr ve Lai-Yuen, 2016; Maldonado ve diğ., 2019), en sık kullanılan yöntemler rassal aşırı-örnekleme (Wu ve Chang, 2003) ve sentetik azınlık aşırı-örnekleme (the synthetic minority over-sampling-SMOTE) tekniğidir (Chawla ve diğ., 2002). Bu çalışmada da literatürde sıklıkla kullanılan rassal aşırı-örnekleme ve SMOTE teknikleri ile çıktı sınıflarındaki dengesizlikler giderilmeye çalışılmıştır.

Rassal aşırı-örnekleme: Bu yöntem ile az sayıda var olan (azınlık) sınıf adedince ve baskın sınıf adedince rassal örnekler oluşturularak orijinal veride değiştirme yapılır (Wu ve Chang, 2003). Sunulan çalışmada ise yer değiştirme yerine, rassal aşırı-örnekleme ile veri kümelerindeki örnek adedi kadar örnekler üretilmiş ve sonrasında orijinal veri ile birleştirilerek orijinal veri boyutu iki katına çıkarılmıştır. Ayrıca tüm veri kümelerinde örnekler rassal olarak karıştırılmıştır.

Sentetik azınlık aşırı-örnekleme (SMOTE): Literatürde sıklıkla kullanılan yöntemlerden biri de sentetik azınlık aşırı-örneklemedir (SMOTE) (Chawla ve diğ., 2002). Bu yöntemde azınlık sınıfı kadar örnek sentetik olarak üretilir ve rassal aşırı-örneklemedeki yer değiştirme yerine üretilen veriler orijinal veriye eklenir ve azınlık sınıfı örnek adetleri artırılmış olur (Chawla ve diğ., 2002). Bu çalışmadaki veri kümelerine SMOTE yöntemi uygulanmış ve veri kümelerindeki azınlık sınıfı örnek sayısı artırılmıştır. SMOTE örneklerinin yüzdesi WEKA'da %200 olarak seçilerek veri kümelerindeki azınlık sınıfı örnek sayısı üç katına çıkarılmıştır. Sonrasında yine veri kümesindeki örnekler rassal olarak karıştırılmıştır.

Sonuç olarak orijinal boyuttaki veri kümelerindeki örnekler rassal aşırı-örnekleme ve SMOTE yöntemleri ile artırılarak çıktı sınıfları arasındaki dengesizlik giderilmiş ve tahmin kalitesi iyileştirilmiştir. Elde edilen 30 veri kümesi sonraki bölümde detaylandırılan veri madenciliği sınıflandırma yöntemleri ile kullanılarak PAH giderimleri tahmin edilmiştir.

2.5. Veri madenciliği sınıflandırma yöntemleri

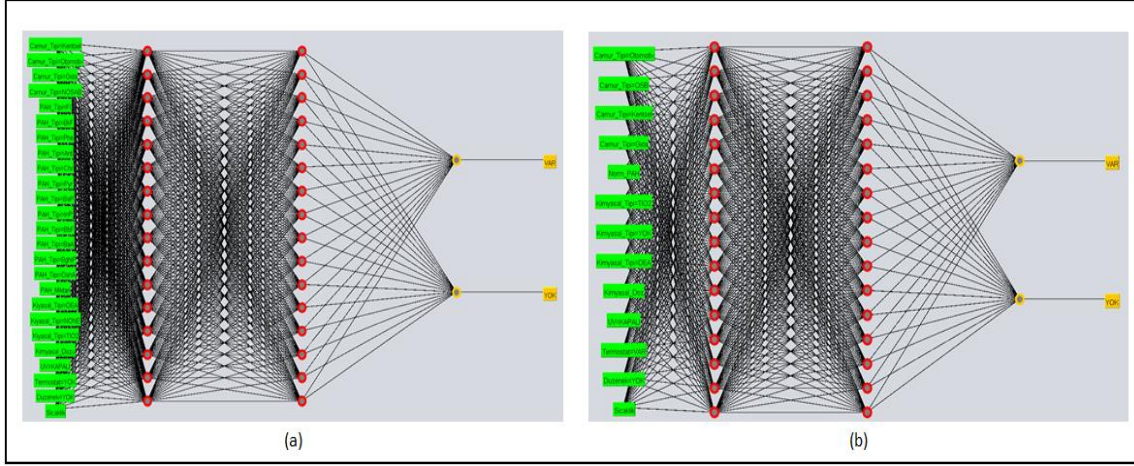
Bu çalışmada PAH giderimlerinin tahmini için beş sınıflandırma yöntemi kullanılmıştır. Tüm sınıflandırma yöntemleri açık kaynaklı veri madenciliği yazılımı WEKA (WEKA, 1993) ile kullanılmış ve 10-kat çapraz doğrulama (*10-fold cross validation*) tekniğine göre değerlendirilmiştir. 10-kat çapraz doğrulama yönteminde benzerlik ölçütüne göre, veri kümesi 10 küçük kümeye bölünür ve doğrulamanın her adımında dokuz küme modeli eğitmek için kullanılır ve onuncu küme ile model test edilir. Bu işlem bütün kümeler test kısmında kullanılacak şekilde 10 defa tekrarlanır. Sınıflandırma yöntemleri ve değerlendirme parametreleri alt başlıklarda detaylıca açıklanmıştır.

2.5.1. Çok Katmanlı Algılayıcı (ÇKA) Ağı

Kökleri 1940'lara kadar uzanan (McCulloch ve Pitts, 1943; Rosenblatt, 1958) yapay sinir ağları (YSA), çevre mühendisliği de dahil olmak üzere sayısız alanda kullanılmaktadır (Huang ve diğ., 2017; Olawoyin, 2016; Öztürk ve diğ., 2020; Sinan, 2010; Sui ve diğ., 2016). Girdiler çeşitli algoritmik fonksiyonlar ile işlenerek istenilen çıktıya dönüştürülür ve eğitim aşamasında ağda kullanılan fonksiyonların katsayıları elde edilir. Daha sonra görünmeyen olaylar (çıkıtlar) bu katsayılarla tahmin edilir. Temel işlevlerden biri olan geri yayılım ile bilgi ters yönde iletilir (Haykin, 2009; Maynard, 2020). Popüler YSA modellerinden biri olan Çok Katmanlı Algılayıcı (ÇKA), genellikle sınıflandırma problemlerinde kullanılır. ÇKA, girdi katmanı, gizli katman(lar) ve çıktı katmanı olarak adlandırılan üç katmana sahiptir.

Bu çalışmada farklı veri kümeleri için tasarlanan ÇKA ağları, çoklu PAH'larla veri kümelerinde Şekil 1(a)'daki gibi ve spesifik PAH'lı veri kümelerinde Şekil 1(b)'deki gibidir. Tablo 2'de özellikleri verilen çoklu PAH'ları içeren veri kümelerinde girdi katmanı düğümleri Tablo 1'deki öznitelikleri kapsamaktadır. Örnek olarak Şekil 1'de TVO (toplam veri orijinal boyut) veri kümesi için tasarlanan ÇKA ağı verilmiştir. Girdi katmanında yer alan PAH bileşikler özniteliği için veri kümesinde yer alan tüm PAH'lar için ayrı düğüm oluşturulmuştur. Tablo 2'de bulunan ve farklı PAH bileşiklerini barındıran diğer veri kümelerinde bu kısım farklılık gösterecektir. Tablo 3'deki ise spesifik PAH'lar için veri kümeleri oluşturulduğundan Şekil 1(b)'deki gibi girdi katmanında PAH bileşikler özniteliği bulunmamaktadır. Tüm veri kümeleri için iki gizli katman ve her birinde 16 düğüm vardır. Çıktı katmanında kategorik çıktılarda her çıktı için bir düğüm tanımlandığı için giderimin olup olmadığını gösteren iki düğüm vardır. Öğrenme oranı (η) güncellenecek ağırlıkların sayısının belirlenmesinde kullanılır. Momentum

parametresi (α) ise sinir ağındaki ağırlıkların güncellenmesinde kullanılır ve bu değer ağıın başarısını doğrudan etkiler. Bu çalışmada $\eta = 0,3$ ve $\alpha = 0,2$ alınmıştır (Haykin, 2009). Devir sayısı verilerin kaç defa işleneceğini belirtir. Diğer adıyla iterasyon sayısı (*epochs*) olan bu değer ÇKA'nın performansını doğrudan etkiler ancak ağıın başarısı ile doğru orantılı olduğu söylenemez. Fazla olması ağıın başarısını arttırabilir ancak gereğinden fazla olması ağıın eğitim süresini uzattığı gibi tahminlerin kalitesini de olumsuz etkileyebilir (Aydemir, 2018). Dolayısıyla en iyi değerde seçilmesi önemlidir. Yapılan analizler sonucu bu çalışma için iterasyon sayısı 1000 olarak belirlenmiştir. Tüm veri kümeleri için aynı parametreler kullanılmıştır.



Şekil 1:

(a)TVO veri kümesinde giderim tahmini için tasarlanan ÇKA ağı,(b) Spesifik PAH'lı veri kümelerinde giderim tahmini için tasarlanan ÇKA ağı

2.5.2. K-en yakın komşu (k-NN)

Literatürdeki en eski ve popüler sınıflandırma yöntemlerinden biri de k-en yakın komşu (k-NN) yöntemidir (Cover ve Hart, 1967). K-NN uygulamasında, etiketlenmemiş her örnek eğitim kümesindeki k-en yakın komşularının çoğunluk etiketine göre sınıflandırılır. K-NN yöntemi basit olmasına rağmen geçmiş verilerle başarılı bir şekilde birleştirildiğinde, genellikle iyi bilinen sınıflandırma yöntemlerindeki sonuçlara yakın sonuçlar elde eder. Bu çalışmada, k-NN algoritması için WEKA'daki *IBk* kütüphanesi kullanılmıştır (Aha ve diğ., 1991). Toplu tahmini gerçekleştirmek için örneklerin bir araya getirilip parti oluşturulması aşamasında parti büyüklüğü (*batch size*) 100 olarak kabul edilmiştir. Kullanılacak komşu sayısını belirten *k* değeri 1 alınmıştır (Aha ve diğ., 1991). Doğrusal en yakın komşu arama algoritması kullanılmış ve diğer tüm parametreler WEKA'da varsayılan değerlere ayarlanmıştır.

2.5.3. C4.5 karar ağacı (C4.5)

Karar ağaçları, sınıflandırma ve regresyon problemlerinde en çok kullanılan denetimli öğrenme algoritmalarından biridir. Karar ağacı, bir dizi karar kuralını uygulayarak büyük bir veri kümesini daha küçük kümelere böler. Anlaşılması ve yorumlanması kolay olmasına rağmen, verileri iyi açıklamayan karmaşık ağaçlar üretilebilir ve ağaç dallanması takip edilemeyebilir. Öte yandan, C4.5 algoritması, daha basit ağaçlar sağlayan mantıksal kurallara dayalı sınıflandırıcılar oluşturmak için bir sistem içerir (Quinlan, 1993). C4.5, bir karar ağacı oluşturmak için uyumlu sınıflandırma kurallarını kullanır ve bu kurallara budama sürecinde başvurulur. Budama işlemi, daha az önemli kuralları ele alır ve yeniden sınıflandırmayı olumsuz etkilemiyorlarsa bunları kaldırır. Her bir kural basitleştirildiği ve ayrı ayrı değerlendirildiği için, ortaya çıkan kural kümesi

orijinal ağaçtan daha basit ve daha anlaşılır olabilir. Bu çalışmada WEKA'daki *J48* algoritması adı verilen budanmış C4.5 karar ağacı kullanılmıştır. Budama için güven faktörü 0,25 ve parti büyüklüğü 100 alınmıştır (Quinlan, 1993).

2.5.4. Rastgele orman (RO)

Rastgele orman (*random forest*) algoritması, çeşitli karar ağaçları oluşturan ve bunları bir orman olarak birleştiren bir toplu öğrenme algoritmasıdır. Karar ağaçlarından elde edilen sonuçlar değerlendirilir ve nihai sonuç çoğunluk oyuyla belirlenir. Ormandaki tüm ağaçlar aynı dağılıma sahiptir; ancak her ağaç, çeşitliliği sağlayan ve daha iyi bir algoritma oluşturan rastgele bir özellik seçimi kullanılarak üretilir (Han ve diğ., 2012). Literatürdeki çalışmalar, RO modellerinin dengesiz veri kümelerinden öğrenmede başarılı olabileceğini göstermiştir (Khoshgoftaar ve diğ., 2007; Malhotra ve Kamal, 2019). RO yöntemi, varsayılan WEKA parametreleriyle kullanılmıştır (Breiman, 2001). İterasyon sayısı 100 olarak kabul edilmiş, ağacın maksimum derinliği sınırsız olarak alınmış ve rastgele seçilen özneliklerin sayısı sifira ayarlanmıştır (Breiman, 2001).

2.5.5. Torbalama

Torbalama (*Bagging*), homojen sınıflandırıcılardan oluşan bir topluluk oluşturmaya odaklanan popüler bir toplu öğrenme algoritmasıdır ve bir diğer adı önyükleme toplamasıdır (*bootstrap aggregation*) (Breiman, 1996). Bir sınıflandırıcıyı eğitmek için, n örnekle orijinal eğitim kümesi, yer değiştirme yöntemiyle m örnekle ($m < n$) yeni alt kümeleri oluşturmak için kullanılır. Bir alt kümedeki örnekler başka bir alt küme için seçilebilir ve bu nedenle bazı örnekler birden fazla alt kümede görünürken diğerleri hiçbir zaman görünmeyebilir (Han ve diğ., 2012). Daha sonra, her eğitim alt kümesi bir sınıflandırıcıyı (Torba) eğitmek için kullanılır. Her bir örnek için sınıflandırıcıların sınıflandırma sonuçları değerlendirilir ve sınıflandırıcıların çoğunun üzerinde fikir birliğine sahip olduğu bir sınıf seçilir. Torbalamanın başarısı, sınıflandırıcıların çeşitliliğine ve performanslarına, torba sayısına (sınıflandırıcılar) ve dengeli eğitim kümelerine bağlıdır. Bu çalışmada WEKA'daki *Bagging* kütüphanesi torbalama yöntemi için kullanılmıştır. Parti büyüklüğü 100, torba boyutu %100 ve sınıflandırıcı olarak da budanmış C4.5 karar ağacı seçilmiştir (Quinlan, 1993).

2.6. Değerlendirme Parametreleri

Sınıflandırma yöntemlerinin performanslarını değerlendirmek için çeşitli parametreler kullanılmıştır. Bu parametreler Tablo 4'te verilen karışıklık matrisi (*confusion matrix*) yardımıyla hesaplanmıştır. Karışıklık matrisi, her sınıf için doğru ve yanlış tahmin edilen örneklem adetlerini dört hücrede göstermektedir. Doğru Pozitif (DP), doğru tahmin edilen pozitif örneklem sayısını; Doğru Negatif (DN), doğru tahmin edilen negatif örneklem sayısını; Yanlış Pozitif (YP), yanlış tahmin edilen pozitif örneklem sayısını ve Yanlış Negatif (YN), yanlış tahmin edilen negatif örneklem sayısını ifade eder.

Tablo 4. İki sınıflı karışıklık matrisi

		Tahmin edilen sonuç	
		Pozitif tahmin	Negatif tahmin
Gerçek sonuç	Gerçek pozitif	Doğru Pozitif (DP)	Yanlış Negatif (YN)
	Gerçek negatif	Yanlış Pozitif (YP)	Doğru Negatif (DN)

Sınıflandırma yöntemlerinin performans değerlendirmesinde, doğruluk (*accuracy*), kesinlik⁺ (*precision*⁺), duyarlılık (*sensitivity/recall*), belirleyicilik (*specificity*), F-ölçütü (*F-measure*) ve AUC (*Area Under the ROC Curve*) parametreleri kullanılmıştır (Bradley, 1997). Alıcı işlem

karakteristikleri eğrisi (ROC- Receiver Operating Characteristic Curve) DP oranına karşın YP oranını gösterir. Bir diğer ifade ile doğru sınıflandırılmış azınlık sınıfı ile doğru sınıflandırılmış baskın sınıfın oranını verir. Eşitlik (1) – Eşitlik (6)’da verilen karşılaştırma parametreleri literatürde dengesiz verilerin sınıflandırılmasında sıklıkla kullanılan metriklerdir (Haixiang et al., 2017).

$$Kesinlik^+ = \frac{DP}{DP+YP} \quad (1)$$

$$Duyarluluk = \frac{DP}{DP+YN} \quad (2)$$

$$Belirleyicilik = \frac{DN}{DN+YP} \quad (3)$$

$$Doğruluk = \frac{DP+DN}{DP+YP+YN+DN} \quad (4)$$

$$AUC = [x = Duyarluluk, y = 1 - Belirleyicilik] \quad (5)$$

$$F - ölçütü = \frac{2 * Kesinlik^+ * Duyarluluk}{Kesinlik^+ + Duyarluluk} \quad (6)$$

Sonraki bölümde sınıflandırma yöntemleri 30 veri kümesi için çalıştırılmış ve performansları bahsedilen metriklerle değerlendirilmiştir.

3. BULGULAR

Tüm veri kümelerinde sınıflandırma yöntemleri ile giderim tahminleri yapılmış ve performans parametreleri hesaplanmıştır (Tablo EK-1–Tablo EK-4; erişim linki: <http://dx.doi.org/10.17632/jsx8yv2b5d.1>). Ayrıca tüm veri kümeleri için yöntemlerin ortalama performansları hesaplanmış ve Tablo 5’de verilmiştir. Doğruluk parametresi dikkate alındığında ÇKA (%86,615) ve C4.5 (%87,131) yöntemlerinin performansları diğer yöntemlere göre kötüdür. Bununla birlikte, azınlık sınıfını tahmin etmede belirleyici olan ortalama AUC değerine göre C4.5 (0,829) ve k-NN (0,853) diğer yöntemlere göre düşük performans göstermiştir. Bütün veri kümeleri için yüksek AUC ve % doğruluk değerine sahip yöntem RO’dur. RO sonrasında iyi performans gösteren yöntemler Torbalama ve ÇKA’dır.

Tablo 5. Sınıflandırma yöntemlerinin bütün veri kümelerinde ortalama performansları

Yöntemler	Doğruluk(%)	Duyarluluk	Kesinlik ⁺	Belirleyicilik	F-Ölçütü	AUC
ÇKA	86,615	0,866	0,866	0,773	0,866	0,884
K-NN	88,348	0,883	0,883	0,807	0,883	0,853
C4.5	87,131	0,871	0,866	0,734	0,867	0,829
RO	89,290	0,893	0,891	0,811	0,892	0,920
Torbalama	87,805	0,878	0,875	0,749	0,875	0,898

Aşırı-örnekleme yaklaşımının sınıflandırma yöntemleri üzerindeki etkisini görmek amacıyla yöntemlerin performansları veri kümelerinin özelliklerine göre de değerlendirilmiştir. Bu amaçla veri kümeleri üç grupta toplanmıştır: Orijinal veri kümeleri (orijinal), SMOTE ile iyileştirilmiş veri kümeleri (SMOTE) ve rassal aşırı-örnekleme ile iyileştirilmiş veri kümeleri (rassal aşırı-örnekleme). Tablo 6’da yöntemlerin üç veri grubunda % doğruluk ortalamaları ve Tablo 7’de yöntemlerin AUC ortalamaları hesaplanmıştır. İki tablo bir arada değerlendirildiğinde tüm yöntemlerin orijinal veri kümelerinde en düşük performans sergilediği görülmektedir. Dengesiz çıktı sınıflarının olduğu orijinal veri kümelerinde ortalama doğruluk %82,938’dir. Bu değer

SMOTE ile veriler iyileştirildiğinde %5,136 artarak %88,074'e çıkmıştır. En iyi tahmin sonuçları rassal aşırı-örnekleme ile iyileştirilen veri kümelerinde elde edilmiş ve tahmin kalitesi %9,564 oranında artarak %92,502 doğrulukla giderim tahmin edilmiştir.

Tablo 6. Orijinal ve iyileştirilmiş veri kümelerinde sınıflandırma yöntemlerinin % doğruluk ortalama değerleri

Veri Kümeleri	Yöntemler					
	ÇKA	K-NN	C4.5	RO	Torbalama	Ortalama
Orijinal	81,615	81,995	83,676	83,539	83,864	82,938
SMOTE	86,953	87,461	88,315	88,601	89,04	88,074
Rassal Aşırı-Örnekleme	91,275	95,588	89,403	95,73	90,512	92,502
Ortalama	86,615	88,348	87,131	89,29	87,805	

Tablo 7. Orijinal ve iyileştirilmiş veri kümelerinde sınıflandırma yöntemlerinin AUC ortalama değerleri

Veri Kümeleri	Yöntemler					
	ÇKA	K-NN	C4.5	RO	Torbalama	Ortalama
Orijinal	0,817	0,752	0,737	0,845	0,828	0,796
SMOTE	0,909	0,862	0,884	0,941	0,928	0,905
Rassal Aşırı-Örnekleme	0,926	0,944	0,866	0,974	0,939	0,93
Ortalama	0,884	0,853	0,829	0,92	0,898	

Azınlık sınıfının tahmin kalitesini gösteren önemli metriklerden AUC değerinin veri kümeleri ve yöntemlere göre ortalaması incelendiğinde orijinal verilerde ortalama AUC 0,796 iken bu değer SMOTE uygulanmış veri kümelerinde 0,905 ve rassal aşırı-örnekleme yöntemi ile iyileştirilmiş veri kümelerinde 0,930'a çıkmıştır. Rassal aşırı-örnekleme yönteminin azınlık sınıfının tahmin kalitesini artırdığı görülmüş ve bu sebeple tahmin yöntemlerinin performansları rassal aşırı-örnekleme ile iyileştirilmiş veri kümeleri özelinde tekrar karşılaştırılmıştır.

Tablo 8'de sınıflandırma yöntemleri rassal aşırı-örnekleme ile iyileştirilmiş veri kümelerindeki tahmin performanslarının ortalamaları alınarak karşılaştırılmıştır. Bu veri kümelerinde en başarılı tahminler RO yöntemi ile elde edilmiştir. RO yöntemi bütün parametrelerde en yüksek değerleri elde etmiş ve ortalama azınlık sınıfını en iyi tahmin eden yöntem olmuştur. Sonrasında k-NN yöntemi tüm parametrelerde diğer yöntemlerden daha yüksek değerler alarak ikinci en iyi yöntem olmuştur. Ortalama % doğruluk dikkate alındığında başarılı yöntemler sırasıyla şu şekildedir: RO (%95,730) > k-NN (%95,588) > ÇKA (%91,275) > Torbalama (%90,512) > C4.5 (%89,403). Dengesiz çıktının olduğu veri kümelerinde azınlık sınıfının tahmin kalitesini gösteren ortalama AUC değerlerine göre yöntemlerin başarı sıralaması ise şu şekildedir: RO (0,974) > k-NN (0,944) > Torbalama (0,939) > ÇKA (0,926) > C4.5 (0,866). AUC değerinin bire yakın olması azınlık sınıfının da baskın sınıf kadar iyi tahmin edildiğini gösterir. Bu parametreler ve diğer parametreler dikkate alındığında bu çalışmada giderim tahmininde en başarılı iki yöntemin RO ve k-NN yöntemleri olduğu söylenebilir. Sonrasında %doğruluk değerine göre üçüncü sırada yer alan ÇKA ağı, azınlık sınıfı tahmininde torbalama yöntemi kadar başarılı olamamıştır. C4.5 karar ağacı yöntemi ise diğer yöntemlerin gerisinde kalmıştır.

Sonuç olarak, veri iyileştirme yöntemlerinden rassal aşırı-örnekleme yöntemi ile tahmin yöntemlerinin tamamında performans artışı gerçekleşmiş ve RO ve k-NN yöntemleri ile en iyi performans değerleri elde edilmiştir. RO dışındaki yöntemlerin performansları verilerin dağılımlarından RO'ya kıyasla daha fazla etkilenmektedir. Örnek olarak k-NN yöntemi azınlık çıktı verilerinin az olduğu orijinal veri kümelerinde azınlık sınıfının tahmininde kötü performans

sergilerken rassal aşırı-örnekleme ile iyileştirilmiş veri kümelerinde en başarılı ikinci yöntem olmuştur.

Tablo 8. Sınıflandırma yöntemlerinin rassal aşırı-örnekleme ile iyileştirilmiş veri kümelerinde ortalama performansları

Yöntemler	Doğruluk(%)	Duyarlılık	Kesinlik ⁺	Belirleyicilik	F-Ölçütü	AUC
ÇKA	91,275	0,913	0,912	0,824	0,912	0,926
K-NN	95,588	0,956	0,956	0,917	0,956	0,944
C4.5	89,403	0,894	0,892	0,76	0,892	0,866
RO	95,730	0,957	0,958	0,917	0,957	0,974
Torbalama	90,512	0,905	0,903	0,776	0,903	0,939

4. DEĞERLENDİRME ve SONUÇ

Arıtma çamurlarındaki PAH'ların uygun teknolojilerle giderimi çevre ve halk sağlığı açısından büyük önem arz etmektedir. Bu bileşiklerin gideriminde çeşitli fiziksel, kimyasal ve/veya biyolojik yöntemler kullanılmakta olup fotoparçalanma uygulamaları bunlardan biridir. Fotoparçalanma uygulamalarında laboratuvar koşullarında (sıcaklık, katkı maddesi türü ve miktarı vb.) çok sayıda deney yapılmalı ve sonrasında tam ölçekli çalışmalara geçilmelidir. Deneylere başlamadan önce arıtılacak olan çamurda başlangıç PAH konsantrasyonuna göre UV teknolojisinin etkin bir şekilde uygulanıp uygulanamayacağı konusunda bir öngörüde/tahminde bulunmak ve buna göre deneysel çalışmalara başlamak yerinde bir yaklaşım olacaktır. Böylelikle, PAH'ların giderilemeyeceği tahmin edilen durumlar için UV uygulamaları yerine alternatif uygulamalara yönelmesi sağlanacak, bu sayede zaman ve maliyet kaybının önüne geçilmiş olacaktır. Süreci kısaltmak ve maliyetleri düşürmek için daha önce bu konuda yapılmış benzer deneysel çalışmalardan elde edilen veriler kullanılarak veri madenciliği yöntemleriyle çamurlardaki kirlenmelerin giderimleri için tahminlerde bulunmak mümkündür. Bu kapsamda çamurun karakterizasyonuna göre arıtma yöntemleriyle (UV uygulamaları vb.) PAH gideriminin gerçekleşip gerçekleşmeyeceği tahmin edilebilir. Sunulan çalışmada arıtılacak çamurun PAH kirlilik seviyesine göre fotoparçalanma uygulamalarıyla giderilip giderilemeyeceğinin veri madenciliği yöntemleriyle tahmin edilmesi amaçlanmıştır. Böylelikle deney yükü azaltılmış ve UV uygulamalarıyla PAH gideriminin olmayacağı tahmini durumunda alternatif yöntem arayışına girilmiş olacaktır.

Bu kapsamda, bir kentsel ve üç endüstriyel tesisten alınan arıtma çamur örneklerinde 12 PAH türüne ait konsantrasyonlar tespit edilmiştir. Sonraki aşamada farklı deneysel koşullarda UV ışığı ile fotoparçalanma uygulamaları gerçekleştirilmiş ve PAH bileşenlerinin her biri için giderim yüzdeleri hesaplanmıştır. Elde edilen veriler veri madenciliği çalışmalarında kullanılmıştır. Örneklem sayısının az olmasından dolayı tahmin kalitesini iyileştirmek için giderim yüzdesi ikili kategorik veriye dönüştürülmüş ve giderim olup olmadığı tahmin edilmiştir. Böylece çıktının numerik tahmini yerine kategorik tahmini sağlanmıştır. Veri kümelerinde benzer özellik gösteren PAH halka grupları ve arıtma çamurlarında US-EPA tarafından önceliklendirilen bazı PAH türleri ayrı ayrı ele alınmış ve her biri için bir veri kümesi oluşturulmuştur.

Dengesiz çıktıya sahip az örneklemler veri kümelerinde tahmin kalitesini iyileştirmek için rassal aşırı-örnekleme ve SMOTE yöntemleri ile veri kümelerindeki dengesizlikler giderilmiştir. Literatürde kategorik çıktılarda tahmininde sıklıkla kullanılan sınıflandırma yöntemleri belirlenmiştir. Bu yöntemlerle PAH türlerine göre ayrılmış orijinal, rassal aşırı-örnekleme ile iyileştirilmiş ve SMOTE ile iyileştirilmiş toplam 30 farklı veri kümesinde giderimler tahmin edilmiştir. Yöntemlerin performanslarının değerlendirilmesinde kategorik çıktılarda ve dengesiz veri kümelerinde kullanılan değerlendirme metrikleri dikkate alınmış ve yöntemler kesinlik⁺,

duyarlılık, belirleyicilik, %doğruluk, AUC ve F-ölçütü parametrelerine göre karşılaştırılmıştır. Bu çalışma ile literatürde ilk kez 12 PAH türünü barındıran dengesiz veri kümeleri aşırı örnekleme yöntemleriyle dengelenmiş ve çeşitli sınıflandırma yöntemi ile giderim tahminleri gerçekleştirilmiştir. Yapılan deneyler ve karşılaştırmalar sonrasında az sayıda ve dengesiz veride rassal aşırı-örnekleme yönteminin tahmin kalitesini %9,564 oranında arttırdığı görülmüştür. Bununla birlikte SMOTE ile %5,136 oranında tahminler iyileştirilmiştir. Tüm parametreler dikkate alındığında giderim tahmininde en başarılı iki yöntemin RO ve k-NN yöntemleri olduğu görülmüştür.

Sonuç olarak RO ve k-NN yöntemleri dengesiz veri kümelerinde giderim tahmininde başarılı bir şekilde kullanılabilir. Tasarlanan deneysel koşullar ve girdi PAH bileşikleri ile bu yöntemlerle daha deneyler yapılmadan giderimin olup olmayacağı tahmin edilebilir. Eğer giderimin olmayacağı tahmin edilirse deneysel koşulları değiştirilerek tahminler tekrarlanabilir. Böylece, daha az maliyet ile giderimi sağlayacak deneysel koşullar belirlenebilir. Ayrıca bu çalışmada önerilen yöntemler farklı matrikslerde (su ve toprak) farklı bileşiklerin (inorganik ve organik kirleticiler için) giderimlerinin tahmininde de kullanılabilir. Bu sınıflandırma yöntemleri ile daha az maliyet ve kaynakla giderim sağlanıp sağlanamayacağı girdi bileşenlerine göre tahmin edilebilir ve uzun soluklu deneysel çalışmaların azaltılması sağlanabilir.

ÇIKAR ÇATIŞMASI

Yazarlar, bilinen herhangi bir çıkar çatışması veya herhangi bir kurum/kuruluş ya da kişi ile ortak çıkar bulunmadığını onaylamaktadırlar.

YAZAR KATKISI

Burcu Çağlar Gençosman çalışmanın kavramsal ve tasarım süreçlerinin belirlenmesi ve yönetimi, verinin dönüştürülmesi, uygun yöntemlerle analizi ve sonuçların yorumlanması aşamalarında, Gizem Eker Şanlı çalışmanın kavramsal ve tasarım süreçlerinin belirlenmesi ve yönetimi, verinin toplanması ve sonuçların yorumlanması aşamalarında katkı sağlamışlardır.

KAYNAKLAR

1. Aha, D. W., Kibler, D. ve Albert, M. K. (1991) Instance-based learning algorithms, *Machine Learning*, 6(1), 37–66. <https://doi.org/10.1007/BF00153759>.
2. Akpınar, H. (2014) *Data: Veri Madenciliği Veri Analizi*, (1.baskı), Papatya Yayıncılık Eğitim, İstanbul.
3. Aydemir, S. (2018) *Weka ile Yapay Zeka*, (1.baskı), Seçkin Yayıncılık, Ankara.
4. Ballesteros-Gómez, A., Caballero-Casero, N., García-Fonseca, S., Lunar, L. ve Rubio, S. (2019) Multifunctional vesicular coacervates as engineered supramolecular solvents for wastewater treatment, *Chemosphere*, 223, 569–576. <https://doi.org/10.1016/j.chemosphere.2019.02.089>.
5. Bhagat, S. K., Tung, T. M. ve Yaseen, Z. M. (2020) Development of artificial intelligence for modeling wastewater heavy metal removal: State of the art, application assessment and possible future research, *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2019.119473>.
6. Blanchard, M., Teil, M. J., Ollivon, D., Legenti, L. ve Chevreuil, M. (2004) Polycyclic aromatic hydrocarbons and polychlorobiphenyls in wastewaters and sewage sludges from the Paris area (France), *Environmental Research*, 95(2), 184–197. <https://doi.org/10.1016/j.envres.2003.07.003>.

7. Boztoprak, H., Özbay, Y., Güçlü, D., ve Küçükhemek, M. (2016) Prediction of sludge volume index bulking using image analysis and neural network at a full-scale activated sludge plant, *Desalination and Water Treatment*, 57(37), 17195–17205. <https://doi.org/10.1080/19443994.2015.1085909>.
8. Bradley, A. P. (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognition*, 30(7), 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2).
9. Breiman, L. (1996) Bagging predictors, *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1007/bf00058655>.
10. Breiman, L. (2001) Random Forests, *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
11. Bunkhumpornpat, C., Sinapiromsaran, K. ve Lursinsap, C. (2009) Safe-level-SMOTE: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem, *Lecture Notes in Computer Science*, 475–482. https://doi.org/10.1007/978-3-642-01307-2_43.
12. Bunkhumpornpat, C., Sinapiromsaran, K. ve Lursinsap, C. (2012) DBSMOTE: Density-based synthetic minority over-sampling technique, *Applied Intelligence*, 36(3), 664–684. <https://doi.org/10.1007/s10489-011-0287-y>.
13. Chawla, N. V., Bowyer, K. W., Hall, L. O. ve Kegelmeyer, W. P. (2002) SMOTE: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>.
14. Corominas, L., Garrido-Baserba, M., Villez, K., Olsson, G., Cortés, U. ve Poch, M. (2018) Transforming data into knowledge for improved wastewater treatment operation: A critical review of techniques, *Environmental Modelling and Software*, 106, 89–103. <https://doi.org/10.1016/j.envsoft.2017.11.023>.
15. Cover, T. M. ve Hart, P. E. (1967) Nearest Neighbor Pattern Classification, *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>.
16. Da Rocha, O. R. S., Dantas, R. F., Bezerra Duarte, M. M. M., Lima Duarte, M. M. ve da Silva, V. L. (2013) Solar photo-Fenton treatment of petroleum extraction wastewater, *Desalination and Water Treatment*, 51(28–30), 5785–5791. <https://doi.org/10.1080/19443994.2013.792136>.
17. Dong, D., Li, P., Li, X., Xu, C., Gong, D., Zhang, Y. ve Li, P. (2010) Photocatalytic degradation of phenanthrene and pyrene on soil surfaces in the presence of nanometer rutile TiO_2 under UV-irradiation, *Chemical Engineering Journal*, 158(3), 378–383. <https://doi.org/10.1016/j.cej.2009.12.046>.
18. Eker, G. ve Hatipoğlu, M. (2019) Effect of UV wavelength, temperature and photocatalyst on the removal of PAHs from industrial soil with photodegradation applications, *Environmental Technology*, 40(28), 3793–3803. <https://doi.org/10.1080/09593330.2018.1491635>.
19. Eker, G., Şengül, B. ve Cindoruk, S. S. (2019) Performance Evaluation of Diethylamine to the Removal of Polycyclic Aromatic Hydrocarbons (PAHs) from Polluted Soils with Sunlight, *Polycyclic Aromatic Compounds*, 1–13. <https://doi.org/10.1080/10406638.2019.1578809>.

20. European Communities, E. (2001) Pollutants in urban wastewater and sewage sludge, https://ec.europa.eu/environment/archives/waste/sludge/pdf/sludge_pollutants.pdf, Erişim Tarihi: 12.10.2020.
21. Fan, M., Hu, J., Cao, R., Ruan, W. ve Wei, X. (2018) A review on experimental design for pollutants removal in water treatment with the aid of artificial intelligence, *Chemosphere*, 200, 330-343. <https://doi.org/10.1016/j.chemosphere.2018.02.111>.
22. Farzin, S., Chianeh, F. N., Anaraki, M. V. ve Mahmoudian, F. (2020) Introducing a framework for modeling of drug electrochemical removal from wastewater based on data mining algorithms, scatter interpolation method, and multi criteria decision analysis (DID), *Journal of Cleaner Production*, 266, 122075. <https://doi.org/10.1016/j.jclepro.2020.122075>
23. Fayyad, U., Piatetsky-Shapiro, G. ve Smyth, P. (1996) From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37-37.
24. Filibeli, A. (1996) Arıtma Çamurlarının İşlenmesi, *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Yayınları*, 255.
25. Gaur, N., Narasimhulu, K. ve PydiSetty, Y. (2018) Recent advances in the bio-remediation of persistent organic pollutants and its effect on environment, *Journal of Cleaner Production*, 198, 1602–1631. <https://doi.org/10.1016/j.jclepro.2018.07.076>.
26. Ghasemi, N., Gbeddy, G., Egodawatta, P., Zare, F. ve Goonetilleke, A. (2019) Removal of polycyclic aromatic hydrocarbons from wastewater using dual-mode ultrasound system, *Water and Environment Journal*, 1-10. <https://doi.org/10.1111/wej.12540>.
27. Hadjimichael, A., Comas, J. ve Corominas, L. (2016) Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector, *AI Communications*, 29(6), 747–756. <https://doi.org/10.3233/AIC-160714>.
28. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H. ve Bing, G. (2017) Learning from class-imbalanced data: Review of methods and applications, *Expert Systems with Applications*, 73, 220–239. <https://doi.org/10.1016/j.eswa.2016.12.035>.
29. Han, J., Kamber, M. ve Pei, J. (2012) *Data Mining: Concepts and Techniques*, (3. Basım), USA: Morgan Kaufmann Publishers, Elsevier.
30. Haykin, S. S. (2009) *Neural Networks and Learning Machines* (3. Basım), Upper Saddle River, NJ, USA: Pearson Publishing.
31. Japkowicz, N. ve Stephen, S. (2002) The class imbalance problem: A systematic study, *Intelligent Data Analysis*, 6(5), 429–449. <https://doi.org/10.3233/ida-2002-6504>.
32. Jing, L., Chen, B. ve Zhang, B. (2014) Modeling of UV-induced photodegradation of naphthalene in marine oily wastewater by artificial neural networks. *Water, Air, and Soil Pollution*, 225(4), 1–14. <https://doi.org/10.1007/s11270-014-1906-0>.
33. Karaca, G. (2013) Arıtma çamurları ve Nilüfer Çayı sedimentindeki poliaromatik hidrokarbon (PAH) seviyelerinin belirlenmesi ve giderimlerinin araştırılması, Doktora Tezi, Fen Bilimleri Enstitüsü, Bursa Uludağ Üniversitesi, Bursa.
34. Karaca, G. ve Tasdemir, Y. (2013) Effects of Temperature and Photocatalysts on Removal of Polycyclic Aromatic Hydrocarbons (PAHs) from Automotive Industry Sludge, *Polycyclic Aromatic Compounds*, 33(4), 380–395. <https://doi.org/10.1080/10406638.2013.782880>.
35. Khoshgoftaar, T. M., Golawala, M. ve Van Hulse, J. (2007) An empirical study of learning from imbalanced data using random forest, *International Conference on Tools with Artificial Intelligence*, ICTAI, 2, 310–317. <https://doi.org/10.1109/ICTAI.2007.46>.

36. Kong, Q., Wu, H., Liu, L., Zhang, F., Preis, S., Zhu, S. ve Wei, C. (2018) Solubilization of polycyclic aromatic hydrocarbons (PAHs) with phenol in coking wastewater treatment system: Interaction and engineering significance, *Science of the Total Environment*, 628–629, 467–473. <https://doi.org/10.1016/j.scitotenv.2018.02.077>.
37. Lin, Y. J., Teng, L. S., Lee, A. ve Chen, Y. L. (2004) Effect of photosensitizer diethylamine on the photodegradation of polychlorinated biphenyls, *Chemosphere*, 55(6), 879–884. <https://doi.org/10.1016/j.chemosphere.2003.11.059>.
38. Maldonado, S., López, J. ve Vairetti, C. (2019) An alternative SMOTE oversampling strategy for high-dimensional datasets, *Applied Soft Computing Journal*, 76, 380–389. <https://doi.org/10.1016/j.asoc.2018.12.024>.
39. Malhotra, R. ve Kamal, S. (2019) An empirical study to investigate oversampling methods for improving software defect prediction using imbalanced data, *Neurocomputing*, 343, 120–140. <https://doi.org/10.1016/j.neucom.2018.04.090>.
40. Manoli, E. ve Samara, C. (2008). The removal of Polycyclic Aromatic Hydrocarbons in the wastewater treatment process: Experimental calculations and model predictions, *Environmental Pollution*, 151(3), 477–485. <https://doi.org/10.1016/j.envpol.2007.04.009>.
41. Maynard, M. (2020) *Neural Networks: Introduction to Artificial Neurons, Backpropagation and Multilayer Feedforward Neural Networks with Real-World Applications*, Bağımsız basım.
42. McCulloch, W. S. ve Pitts, W. (1943) A logical calculus of the ideas immanent in nervous activity, *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. <https://doi.org/10.1007/BF02478259>.
43. Mojiri, A., Zhou, J. L., Ohashi, A., Ozaki, N. ve Kindaichi, T. (2019) Comprehensive review of polycyclic aromatic hydrocarbons in water sources, their effects and treatments, *Science of the Total Environment*, 696, p. 133971. <https://doi.org/10.1016/j.scitotenv.2019.133971>.
44. Nekooimehr, I. ve Lai-Yuen, S. K. (2016) Adaptive semi-supervised weighted oversampling (A-SUWO) for imbalanced datasets, *Expert Systems with Applications*, 46, 405–416. <https://doi.org/10.1016/j.eswa.2015.10.031>.
45. Newhart, K. B., Holloway, R. W., Hering, A. S. ve Cath, T. Y. (2019) Data-driven performance analyses of wastewater treatment plants: A review, *Water Research*, 157, 498–513. <https://doi.org/10.1016/j.watres.2019.03.030>.
46. Nguwi, Y. Y. ve Cho, S. Y. (2010) An unsupervised self-organizing learning with support vector ranking for imbalanced datasets, *Expert Systems with Applications*, 37(12), 8303–8312. <https://doi.org/10.1016/j.eswa.2010.05.054>.
47. Ofman, P. ve Struk-Sokołowska, J. (2019) Artificial Neural Network (ANN) Approach to Modelling of Selected Nitrogen Forms Removal from Oily Wastewater in Anaerobic and Aerobic GSBP Process Phases, *Water*, 11(8), 1594. <https://doi.org/10.3390/w11081594>.
48. Oh, J. Y., Choi, S. D., Kwon, H. O. ve Lee, S. E. (2016) Leaching of polycyclic aromatic hydrocarbons (PAHs) from industrial wastewater sludge by ultrasonic treatment, *Ultrasonics Sonochemistry*, 33, 61–66. <https://doi.org/10.1016/j.ultsonch.2016.04.027>.
49. Öztürk, N., Şentürk, H., Gündoğdu, A., & Duran, C. (2020). İçme Suyu Arıtma Tesisi Atık Çamuru Üzerine Metilen Mavisi Adsorpsiyonu ve Yapay Sinir Ağları ile Modellenmesi, *Uludağ University Journal of The Faculty of Engineering*, 25(2), 1083–1104. <https://doi.org/10.17482/uumfd.674224>.

50. Picos-Benítez, A. R., López-Hincapié, J. D., Chávez-Ramírez, A. U. ve Rodríguez-García, A. (2017) Artificial intelligence based model for optimization of COD removal efficiency of an up-flow anaerobic sludge blanket reactor in the saline wastewater treatment, *Water Science and Technology*, 75(6), 1351–1361. <https://doi.org/10.2166/wst.2017.005>.
51. Quinlan, J. R. (1993) *C4.5: Programs for Machine Learning*,. San Mateo, California, USA: Morgan Kaufmann Publishers, Elsevier.
52. Rababah, A. ve Matsuzawa, S. (2002) Treatment system for solid matrix contaminated with fluoranthene. II - Recirculating photodegradation technique, *Chemosphere*, 46(1), 49–57. [https://doi.org/10.1016/S0045-6535\(01\)00090-X](https://doi.org/10.1016/S0045-6535(01)00090-X).
53. Ren, R. Y., Yang, L. H., Han, J. L., Cheng, H. Y., Ajibade, F. O., Guadie, A. ve Wang, A. J. (2020) Perylene pigment wastewater treatment by fenton-enhanced biological process, *Environmental Research*, 186, 109522. <https://doi.org/10.1016/j.envres.2020.109522>.
54. Rosenblatt, F. (1958) The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>.
55. Saber, A. N., Zhang, H., Cervantes-Avilés, P., Islam, A., Gao, Y., An, W. ve Yang, M. (2020) Emerging concerns of VOCs and SVOCs in coking wastewater treatment processes: Distribution profile, emission characteristics, and health risk assessment, *Environmental Pollution*, 265. <https://doi.org/10.1016/j.envpol.2020.114960>.
56. Salari, D., Daneshvar, N., Aghazadeh, F. ve Khataee, A. R. (2005) Application of artificial neural networks for modeling of the treatment of wastewater contaminated with methyl tert-butyl ether (MTBE) by UV/H₂O₂ process, *Journal of Hazardous Materials*, 125(1–3), 205–210. <https://doi.org/10.1016/j.jhazmat.2005.05.030>.
57. Salihoğlu, N. K., Salihoğlu, G., Taşdemir, Y., Cindoruk, S. S., Yolsal, D., Oğulmuş, R. ve Karaca, G. (2010). Comparison of polycyclic aromatic hydrocarbons levels in sludges from municipal and industrial wastewater treatment plants, *Archives of Environmental Contamination and Toxicology*, 58(3), 523–534. <https://doi.org/10.1007/s00244-009-9389-5>.
58. Salihoğlu, N. K., Karaca, G., Salihoğlu, G. ve Taşdemir, Y. (2012) Removal of polycyclic aromatic hydrocarbons from municipal sludge using UV light, *Desalination and Water Treatment*, 44(1–3), 324–333. <https://doi.org/10.1080/19443994.2012.691689>.
59. Sinan, R. K. (2010) Eysel Atıksu Arıtma Tesislerinde Ön Arıtım ve Biyolojik Arıtım Çıkış Parametrelerinin YSA ile Tahmini, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Selçuk Üniversitesi, Konya.
60. Sprovieri, M., Feo, M. L., Prevedello, L., Manta, D. S., Sammartino, S., Tamburrino, S. ve Marsella, E. (2007) Heavy metals, polycyclic aromatic hydrocarbons and polychlorinated biphenyls in surface sediments of the Naples harbour (southern Italy), *Chemosphere*, 67(5), 998–1009. <https://doi.org/10.1016/j.chemosphere.2006.10.055>.
61. Stevens, J. L., Northcott, G. L., Stern, G. A., Tomy, G. T. ve Jones, K. C. (2003) PAHs, PCBs, PCNs, organochlorine pesticides, synthetic musks, and polychlorinated n-alkanes in U.K. sewage sludge: Survey results and implications, *Environmental Science and Technology*, 37(3), 462–467. <https://doi.org/10.1021/es020161y>.
62. Syafiuddin, A. ve Boopathy, R. (2020) A review of polycyclic aromatic hydrocarbons and their substitutions in full-scale wastewater treatment plants, *Environmental Quality Management*, 1-17. <https://doi.org/10.1002/tqem.21694>.

63. Talebkeikhah, F., Rasam, S., Talebkeikhah, M., Torkashvand, M., Salimi, A. ve Moraveji, M. K. (2020) Investigation of effective processes parameters on lead (II) adsorption from wastewater by biochar in mild air oxidation pyrolysis process, *International Journal of Environmental Analytical Chemistry*, 1–21. <https://doi.org/10.1080/03067319.2020.1777291>.
64. Wang, J., Tian, Z., Huo, Y., Yang, M., Zheng, X. ve Zhang, Y. (2018) Monitoring of 943 organic micropollutants in wastewater from municipal wastewater treatment plants with secondary and advanced treatment processes, *Journal of Environmental Sciences (China)*, 67, 309–317. <https://doi.org/10.1016/j.jes.2017.09.014>.
65. Wang, Y. R., Sun, G. D., ve Jin, Q. (2020) Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network, *Applied Soft Computing Journal*, 92, 106333. <https://doi.org/10.1016/j.asoc.2020.106333>.
66. WEKA (1993) Waikato Environment for Knowledge Analysis, https://waikato.github.io/weka-wiki/downloading_weka/, Erişim Tarihi: 14.12.2020.
67. Wu, G. ve Chang, E. Y. (2003) Class-Boundary Alignment for Imbalanced Dataset Learning, *Proceedings of the ICML'03 Workshop on Learning from Imbalanced Data Sets*. Washington, USA.
68. Yaqub, M., Asif, H., Kim, S. ve Lee, W. (2020) Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency using a long short-term memory (LSTM) neural network, *Journal of Water Process Engineering*, 37, 101388. <https://doi.org/10.1016/j.jwpe.2020.101388>.
69. Yaqub, M. ve Lee, S. H. (2020) Micellar enhanced ultrafiltration (MEUF) of mercury-contaminated wastewater: Experimental and artificial neural network modeling, *Journal of Water Process Engineering*, 33, 101046. <https://doi.org/10.1016/j.jwpe.2019.101046>.
70. Ye, Z., Yang, J., Zhong, N., Tu, X., Jia, J. ve Wang, J. (2020) Tackling environmental challenges in pollution controls using artificial intelligence: A review, *Science of the Total Environment*, 699, 134279. <https://doi.org/10.1016/j.scitotenv.2019.134279>.
71. Zhang, L., Li, P., Gong, Z. ve Li, X. (2008) Photocatalytic degradation of polycyclic aromatic hydrocarbons on soil surfaces using TiO₂ under UV light, *Journal of Hazardous Materials*, 158(2–3), 478–484. <https://doi.org/10.1016/j.jhazmat.2008.01.119>.
72. Zhang, X., Yu, T., Li, X., Yao, J., Liu, W., Chang, S. ve Chen, Y. (2019) The fate and enhanced removal of polycyclic aromatic hydrocarbons in wastewater and sludge treatment system: A review, *Critical Reviews in Environmental Science and Technology*, 49(16), 1425–1475. <https://doi.org/10.1080/10643389.2019.1579619>.
73. Zhao, L., Dai, T., Qiao, Z., Sun, P., Hao, J. ve Yang, Y. (2020) Application of artificial intelligence to wastewater treatment: A bibliometric analysis and systematic review of technology, economy, management, and wastewater reuse, *Process Safety and Environmental Protection*, 133, 169–182. <https://doi.org/10.1016/j.psep.2019.11.014>.