

**T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**KENTSEL ATIKSU ARITMA TESİSİ BİYOLOJİK
OKSİJEN İHTİYACININ (BOİ₅) MAKİNA ÖĞRENMESİ
YÖNTEMLERİ İLE TAHMİN EDİLMESİ**

YÜKSEK LİSANS TEZİ

Ercan SELVİ

Enstitü Anabilim Dalı : ÇEVRE MÜHENDİSLİĞİ

Tez Danışmanı : Dr. Öğr. Üyesi Beytullah EREN

Mayıs 2019

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

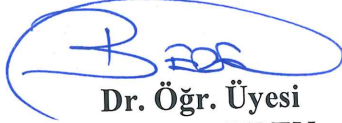
**KENTSEL ATIKSU ARITMA TESİSİ BİYOLOJİK
OKSİJEN İHTİYACININ (BOİ₅) MAKİNA ÖĞRENMESİ
YÖNTEMLERİ İLE TAHMİN EDİLMESİ**

YÜKSEK LİSANS TEZİ

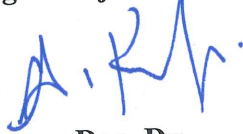
Ercan SELVİ

Enstitü Anabilim Dalı : ÇEVRE MÜHENDİSLİĞİ

Bu tez 29.05.2019 tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.



Dr. Öğr. Üyesi
Beytullah EREN
Jüri Başkanı



Doç. Dr.
Ahmet ÇELEBİ
Üye



Doç. Dr.
Ömer Hulusi DEDE
Üye

BEYAN

Tez içindeki tüm verilerin akademik kurallar çerçevesinde tarafımdan elde edildiğini, görsel ve yazılı tüm bilgi ve sonuçların akademik ve etik kurallara uygun şekilde sunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, tezde yer alan verilerin bu üniversite veya başka bir üniversitede herhangi bir tez çalışmasında kullanılmadığını beyan ederim.

Ercan SELVİ

29.05.2019

TEŞEKKÜR

Yüksek lisans eğitimim boyunca önemli bilgi ve tecrübelerinden yararlandığım, her konuda bilgi ve desteğini almış olduğum, araştırmanın planlanmasından modellemesine kadar tüm aşamalarında yardımlarını esirgemeyen, teşvik eden ve aynı titizlikte beni yönlendiren değerli danışman hocam Dr. Öğr. Üyesi Beytullah EREN'e ve sonuçların modellenmesi konusunda bilgi ve deneyimlerini benimle paylaşan değerli hocam Arş. Gör. Dr. Caner ERDEN'e teşekkürlerimi sunarım.

Birlikte geçirdiğimiz ömür boyunca maddi ve manevi desteklerini hissettiğim, sevgisini ve ilgisini benden hiç esirgemeyen; özellikle çok kıymetli eşim Ebru SELVİ'ye çok teşekkür ederim.

Ayrıca tez yazım süresinde teknik desteğini esirgemeyen Halil UZ ve Bilal SONSUZ'a ve yakın çalışma arkadaşlarımdan Dinçer ÖNER, Koray EREN ve Cafer BAYRAM'a teşekkür ederim.

İÇİNDEKİLER

TEŞEKKÜR	i
İÇİNDEKİLER	ii
SİMGELER VE KISALTMALAR LİSTESİ	iv
ŞEKİLLER LİSTESİ	v
TABLOLAR LİSTESİ	vi
ÖZET	vii
SUMMARY	viii
BÖLÜM 1.	
GİRİŞ	1
1.1. Çalışmanın Anlam ve Önemi	1
1.2. Çalışmanın Amaç ve Kapsamı.....	5
BÖLÜM 2.	
MATERYAL VE YÖNTEM	9
2.1. Makine Öğrenmesi.....	9
2.1.1. Denetimsiz öğrenme (Unsupervised Learning).....	11
2.1.1.1. Kümeleme (Clustering).....	11
2.1.2. Denetimli öğrenme (Supervised Learning)	11
2.1.2.1. Sınıflandırma yöntemi.....	12
2.1.2.2. Regresyon yöntemi	12
2.1.3. Takviyeli öğrenme	14
2.2. Çalışma Kapsamında Modellerde Kullanılan Yalancı Kod.....	14
BÖLÜM 3.	
ARAŞTIRMA BULGULARI	15

3.1. Veri Seti	15
3.2. Modelleme Sonuçları.....	16
3.3. Deney ve Model Sonuçlarının İstatiksel Karşılaştırması.....	20
3.3.1. Kök ortalama kare hata (root mean square error (RMSE))	20
3.3.2. Ortalama mutlak hata (mean absolute error (MAE)).....	21
3.3.3. Ortalama mutlak yüzde hata (mean absolute percentage error (MAPE))	21
BÖLÜM 4.	
TARTIŞMA VE SONUÇ	23
KAYNAKLAR	25
ÖZGEÇMİŞ	27

SİMGELER VE KISALTMALAR LİSTESİ

AKM	: Askıda Katı Madde
ANFIS	: Adaftif Ağ Temelli Bulanık Çıkarım Sistemi
ASM1	: Aktif Çamur Modelleri
BOİ ₅	: Biyolojik Oksijen İhtiyacı (5 gün)
IWA	: Uluslararası Su Birliği
KOİ	: Kimyasal Oksijen İhtiyacı
MLR	: Çoklu Lineer Regresyon
MLCM	: Makine Öğrenmesi Maliyet Modellemesi
NH ₄	: Amonyum
NO ₃	: Nitrat
NO ₂	: Nitrit
N _{org}	: Organik Azot
NSE	: Nash-Sutcliffe Verimliliği
PCA	: Temel Bileşenler Analizi
SGD	: Olasılıksal Dereceli Azalma
SVI	: Çamur Hacim İndeksi
SVM	: Destek Vektör Makinesi
T	: Sıcaklık
TP	: Toplam Fosfor
TKM	: Toplam Katı Madde
TN	: Toplam Azot
YSA	: Yapay Sinir Ağları
Q	: Debi

ŞEKİLLER LİSTESİ

Şekil 2.1. Makine öğrenmesi algoritmaları sınıflandırılma şeması.....	10
Şekil 3.1. Öğrenme algoritmaları için eğitim ve test sonuçlarına ait saçılım grafiği	17
Şekil 3.2. Eğitim seti için öğrenme algoritmalarına göre deney ve model sonuçlarının karşılaştırılması.....	18
Şekil 3.3. Test seti için öğrenme algoritmalarına göre deney ve model sonuçlarının karşılaştırılması.....	19

TABLÖLAR LİSTESİ

Tablo 3.1. Atıksu arıtma tesisi deneysel veri setine ait tanımlayıcı istatistikler (n=366).....	15
Tablo 3.3. Öğrenme algoritmaları modellerinden elde edilen performans sonuçları	22

ÖZET

Anahtar kelimeler: SVM öğrenme algoritması, SGD Regresör öğrenme algoritması, Pasif Agresif Regresör öğrenme algoritması, BOİ₅, atıksu arıtma

Temiz su kaynaklarının azalmasının bir sonucu olarak atıksuların arıtılması önem arz etmektedir. Atıksu arıtma tesislerinin verimli bir şekilde işletilmesi için laboratuvar ve sahadan alınan verilere göre arıtma proseslerine müdahale edilmesi gerekmektedir. Müdahalenin geç veya eksik olması durumu söz konusu olduğunda hem işletme verimi azalmakta hem de maliyet açısından maddi kayıplar ortaya çıkabilmektedir. Atıksu arıtma tesislerinin işletilmesinde laboratuvar analizleri arasında bulunan Biyolojik Oksijen İhtiyacı (BOİ₅) parametresinin analiz sonucu en erken 5 gün içerisinde sonuçlanmaktadır. Arıtma tesislerinin işletilmesi sırasında sürekli olarak analiz edilmesi gereken parametlerin Dolayısı ile atıksu arıtma tesisinin daha güvenli, hızlı, verimli ve düşük maliyetle işletilmesi ve kontrol altına alınabilmesi için tesise ait geçmiş verilerden faydalanılarak bazı parametrelerin tahmin edilmesi önem arz etmektedir. Bu amaçla çeşitli istatistiksel yöntemler ve modelleme teknikleri kullanılmaktadır.

Bu çalışmada makine öğrenmesi algoritmalarından 3 farklı denetimli öğrenme algoritması kullanılarak BOİ₅ parametresi tahmin edilmiştir. Oluşturulan modellerde girdi parametreleri olarak Debi(Q), Sıcaklık(T), pH, iletkenlik, Kimyasal Oksijen İhtiyacı (KOİ), Askıda Katı Madde (AKM) ve çıktı parametresi olarak Biyolojik Oksijen İhtiyacı (BOİ₅) kullanılmıştır. Modellemede kullanılan ham verilerin büyüklükleri birbirinden farklı olduğundan dolayı verileri aynı ölçeğe getirmek için normalizasyon işlemi yapılmıştır. Veriler normalize edildikten sonra veri setinin %90'ı eğitimde %10'u ise test setinde kullanılacak şekilde ikiye ayrılmıştır. Geliştirilen her bir model için iterasyon sayısı 100000 iterasyon olacak şekilde ayarlanmıştır. BOİ₅ tahmininde eğitim ve test verileri için elde edilen regresyon katsayıları (R^2) sırasıyla SVM öğrenme algoritması için, 0,9681; 0,9666, SGD Regresör öğrenme algoritması için, 0,5598;0,7061 ve Pasif Agresif Regresör öğrenme algoritması için ise 0,9840;0,9808 olarak belirlenmiştir. Buradan anlaşılacağı üzere geliştirilen modellerden hem SVM öğrenme algoritması hemde Pasif Agresif Regresör öğrenme algoritması BOİ₅ tahmini için çok iyi sonuçlar vermiştir ve tahmin amaçlı olarak arıtma tesislerinde kullanılabileceği anlaşılmıştır.

ESTIMATION OF THE BIOLOGICAL OXYGEN NEED (BOD₅) OF URBAN WASTEWATER TREATMENT PLANT WITH MACHINE LEARNING METHODS

SUMMARY

Keywords: Support vector machine (SVM) learning algorithm, stochastic gradient descent (SGD) learning algorithm, passive aggressive regression learning algorithm, BOD₅, wastewater treatment

Treatment of wastewater is important as the clean water resources are getting decreased. In order to operate wastewater treatment plants efficiently, treatment processes should be intervened according to the data obtained from the laboratory and the field. In case of late or incomplete intervention, both operational efficiency and cost losses can occur. Biological Oxygen Demand (BOD₅) parameter, which is among the laboratory analyzes in the operation of wastewater treatment plants, results in 5 days. Therefore, it is important to estimate some parameters by utilizing the historical data of the wastewater treatment plant for safer, faster, efficient and low-cost operation and control. For this purpose, various statistical methods and modeling techniques are used.

In this study, BOD₅ parameter was estimated by using 3 different controlled learning algorithms from machine learning algorithms. Flow (Q), Temperature (T), pH, conductivity, Chemical Oxygen Demand (COD), Suspended Solids (SS) were used as input parameters, Biological Oxygen Demand (BOD₅) was used as an output parameter. Since the size of the raw data used in the modeling was different from each other, normalization was performed to bring the data to the same scale. After normalizing, the data was divided into two, using 90% of the data set in education and 10% in the test set. The number of iterations for each model developed is set to 100000 iterations. The regression coefficients (R^2) obtained for training and test data in the estimation of BOD₅ were found to be for SVM learning algorithm 0.9681; 0.9666, for the SGD Regressor learning algorithm 0.5598; 0.7061, and for the Passive Aggressive Regressor learning algorithm was determined as 0.9840; 0.9808. As can be seen from the developed models, both the SVM learning algorithm and the Passive Aggressive Regressor learning algorithm have given very good results for BOD₅ estimation and it was understood that it could be used in treatment plants for estimation purposes.

BÖLÜM 1. GİRİŞ

1.1. Çalışmanın Anlam ve Önemi

Çevre kirliliği tüm canlıların sağlığını ve yaşamını etkilediği gibi cansız cisimlerinde yapısına zarar vermektedir. Doğada bulunan hava, su ve toprağın kirlenmesi ile bütün canlı ve cansızlar varlıklar bu durumdan etkilenmektedir.

İnsan ve doğa ayrılmaz bir bütün olarak birbiri ile sürekli bir etkileşim içindedir. Son yıllarda hızla gelişen sanayi ve teknoloji ile birlikte bu alanda insan gücüne gereksinim artmış ve köyden kente göç hızla artmıştır. Ekonominin gelişmesi ile birlikte kentlerde insan nüfusu hızlı bir artış göstermiş ve insanların yaşam standartları da hızla artmaya başlamıştır. Yaşam standartlarının artması ve kent nüfusunun plansız bir şekilde büyümesi ile insan, etkileşim halinde olduğu doğaya olumsuz etkiler yansıtarak çevresel sorunların gün geçtikçe artmasına sebep olmuştur. Çevresel sorunları kaynağında çözmek ve gidermek sorunların büyümeden azaltılmasından büyük önem taşımaktadır. Hava kirliliği, su kirliliği, toprak kirliliği, gürültü kirliliği, ışık kirliliği en önemli çevre kirliliği türlerini oluşturmaktadır.

Su; insan yaşamı için yaşamsal öneme sahip olup aynı zamanda insanların sağlıklı bir yaşam sürebilmeleri ve ülkelerin kalkınması için temel bir ihtiyaçtır. Günümüzde su kalitesinin hemen her ülkede bozulmuş olması ve su kıtlığının giderek artması, suyun çok önemli bir problem haline gelmesine sebep olmuştur [1]. İyi su kalitesi, insan sağlığı, sosyal ve ekonomik kalkınma ve ekosistem için çok önemlidir. Nüfus arttıkça ve sanayiler tarafından doğal ortam bozuldukça, kaliteli ve yeterli su kaynaklarının olmasını sağlamak zorlaşmaktadır. Bu yüzden çözümün kaynağında yapılması gerekmektedir.

Su, tatlı su çıkarma, içme suyu arıtma, içme suyu dağıtım, kullanım, toplama ve atıksu arıtma, atıksu geri kazanma, geri kazanım suyu kullanımı gibi su döngüsünün her yerinde dikkatli bir şekilde yönetilmelidir. Kaliteli ve yeterli su kaynaklarının mevcudiyeti, atık suyu nasıl yönettiğimiz ile doğrudan bağlantılıdır. Artan kent nüfusu ve sanayiyle birlikte atık su miktarı ve suyun kirlilik yükü gün geçtikçe artmaktadır.

Atık su, evlerde, ticari binalarda ve endüstriyel tesislerde kullanıma uygun olmayan atıkların sıvı ve suyla taşınarak, kanalizasyon sistemine girebilecek herhangi bir yeraltı suyu sızması, yüzey suyu ve yağmur suyu girişi ile birleşimidir. Atık su, patojenler, kimyasallar ve diğer toksinler de dahil olmak üzere içeriğinde birçok kirletici barındırmaktadır. Çevre ve insan sağlığının korunması için atıksuların alıcı ortama verilmeden önce arıtılması gerekmektedir.

Atıksu arıtımında amaç, artık kullanıma uygun olmayan kirleticilerle dolu atıksuyu fiziksel, biyolojik ve kimyasal yollar ile içerdiği kirletici miktarını kabul edilebilir (ilgili mevzuatların izin verdiği) seviyeye getirerek doğaya vereceği zararı önlemek ve su döngüsüne yeniden kazandırmaktır [2].

Atıksular özelliklerine göre 3 grupta sınıflandırılabilir;

- Fiziksel Özellikleri: Renk, koku, iletkenlik, tuzluluk, bulanıklık, sıcaklık ve katı maddeler,
- Kimyasal Özellikleri: Organik, inorganik, metaller,
- Biyolojik Özellikleri: Bakteriler, virüsler, mantarlar ve patojenler olarak söylenebilir.

Ülkemizde ki su kaynaklarının korunması yönünde 1983 yılında yürürlüğe giren 2872 sayılı Çevre Kanunu ile hukuki düzenlemeler yapılmıştır. Çevre Kanunu'na dayanarak 2004 yılında Su kirliliği Kontrol Yönetmeliği yürürlüğe girmiş ve yayınlanmıştır. Söz konusu yönetmeliğin ilk maddesinde atıksuların arıtılmasının amacı, “ülkenin yeraltı

ve yerüstü su kaynakları potansiyelinin korunması ve en iyi bir biçimde kullanımının sağlanması için, su kirlenmesinin önlenmesini sürdürülebilir kalkınma hedefleriyle uyumlu bir şekilde gerçekleştirmek üzere gerekli olan hukuki ve teknik esasları belirlemektir.” şeklinde açıklanmıştır [3].

2006 yılında yürürlüğe girmiş olan Kentsel Atıksu Arıtma Yönetmeliğinin ilk maddesinde amacı, “kentsel atıksuların toplanması, arıtılması ve deşarjı ile belirli endüstriyel sektörlerden kaynaklanan atıksu deşarjının olumsuz etkilerine karşı çevreyi korumaktır. Bu Yönetmelik, kanalizasyon sistemlerine boşaltılan kentsel ve belirli endüstriyel atıksuların toplanması, arıtılması ve deşarjı, atıksu deşarjının izlenmesi, raporlanması ve denetlenmesi ile ilgili teknik ve idari esasları kapsar.” şeklinde açıklanmıştır [4].

Türkiye’de evsel atıksuların arıtılması amacıyla 2014 yılı sonu itibariyle toplam 597 atık su arıtma tesisi bulunmaktadır [1]. Bu tesislerin sadece 77 adedi ileri arıtmadır. Sadece arıtma tesisi yapmak ve kanunları oluşturmak atıksuların arıtıldıktan sonra deşarj edildikleri alıcı ortamları (göl, deniz, nehir vb.) korumak için tek başına yeterli olmamaktadır. Atıksu arıtma tesislerinin performanslı işletilmesi bu noktada büyük önem arz etmektedir. Atıksu arıtma tesislerinin işletilmesinde ilgili mevzuatların gerektirdiği atıksu arıtma tesisi giriş ve çıkış parametreler aşağıda açıklanmıştır.

Toplam katı madde (TKM): Atıksu içerisinden ızgaralanarak alınan katı maddeler uzaklaştırılarak, atıksu numunesinin 103-105 derecede buharlaştırıldıktan sonra geriye kalan maddelerdir. Bunlar metal tuzları, inorganik tuzlar, çözünmez tuzlar, çözünür tuzlar vb. Toplam katı madde önemli bir parametre olmasının sebebi, toplam çözünmüş katıların konsantrasyonu, suda yaşayan organizmaların hücrelerinde su dengesini etkilemektedir.

Biyolojik oksijen ihtiyacı (BOİ₅): Biyokimyasal oksijen ihtiyacı, mikroorganizmaların atıksu içersinde ki organik maddeleri (organik-BOİ) ve inorganik maddeleri (nitrifikasyon için BOİ) oksitlemeleri için gerek duydukları oksijen miktarıdır. Buna karbonlu maddelerin oksijen ihtiyacı denir. İki çeşit BOİ vardır.

- a. Organiklere ait BOİ (heterotrofik BOİ): Heterotrofik mikroorganizmaların organik maddeleri karanlık ortamda 5 günde ve 20 C°'de sıcaklıkta oksitlemeleri için gerekli oksijen miktarıdır. BOİ₅ ile gösterilir.
- b. Nitrifikasyon için BOİ (ototrofik BOİ): Ototrofik olan nitrifikasyon bakterileri amonyumu nitrata oksitlemek için ihtiyaç duyduğu oksijen miktarına nitrifikasyon için BOİ veya ototrofik BOİ adı verilir.

BOİ kullanılan en yaygın parametre olmasına karşılık, ölçüm sonucunun uzun zaman alması, zor ayrıışan organik maddelerin beş günde tüketilememesi ve ortamda bakterilerin aktivitesini inhibe edici madde bulunması durumunda çözünmüş oksijen değerini gerekenden düşük okunması gibi nedenlerden dolayı dezavantajlıdır.

Kimyasal oksijen ihtiyacı (KOİ): Kimyasal Oksijen İhtiyacı (KOİ), su numunesinde bulunan organik maddenin kuvvetli kimyasal oksidasyon maddeleri yardımı ile CO₂, H₂O ve amonyuma oksidasyonu sırasında harcanan oksijen miktarını ifade eder. KOİ, mg/L olarak ifade edilir. BOİ ile sadece mikroorganizmalar tarafından kullanılabilen organik maddeler ölçülebilir iken, KOİ tüm organik maddeler ölçülebilir. Dolayısıyla, bir suyun biyolojik olarak arıtılıp arıtilamayacağına BOİ/KOİ oranına bakılarak karar verilir.

Toplam azot (TN): Sularda bulunan azot formları organik, amonyak, nitrit, nitrat ve azot gazıdır. Azot, proteinlerin sentezi için temel yapı taşı olduğundan, atıksuların biyolojik yollarla tasfiyesinde azot konsantrasyonunu bilmeye ihtiyaç vardır. Analitik olarak organik azot ve amonyak birlikte tayin edilerek “toplam azot” ya da daha doğru bir terimle “Kjeldahl azotu” olarak isimlendirilir.

Azot bileşiklerinin yüzey suları üzerinde çeşitli etkileri vardır;

- Organik Azot (N_{org}): Oksijeni kuvvetli bir biçimde tüketir.
- Amonyum (NH₄): Oksijeni tüketir. pH > 8'de, balıklar için zehirlidir.
- Nitrat (NO₃): Ötrifikasyona neden olur.

- Nitrit (NO_2): Balıklar için çok zehirlidir.

Toplam fosfor (TP): Fosfor, mikroorganizmaların büyümesi için gerekli temel elementlerden birisi olup, su ortamında birincil üretkenliği kısıtlayıcı bir nutrienttir. Fosforun büyümeyi kısıtlayıcı nutrient olması sebebiyle ham veya arıtılmış atıksuların, zirai drenajların veya bazı endüstriyel atıksuların su ortamlarına deşarjı, fotosentetik, akuatik mikro ve makro organizmaların istenmeyen miktarlarda gelişmelerine yol açar [5]. Fosfor, alıcı ortamlarda ötröfikasyona sebep olup, deşarj edilmeden önce mutlaka giderilmesi gerekmektedir. Fosfor giderimi kimyasal ya da biyolojik yollarla gerçekleştirilebilir. Klasik atıksu arıtma tesislerinde sadece %10-25 fosfat arıtımı gerçekleştirilebilir.

1.2. Çalışmanın Amaç ve Kapsamı

Ülkemizde son yıllarda temiz su kaynaklarının azalmasının bir sonucu olarak temiz su kaynaklarının korunması büyük önem arz etmektedir. Yetkili kuruluşlar, atıksu arıtma tesislerinin daha iyi işletilmeleri ve kontrol altına alınması konularına odaklanmıştır. Atıksu arıtma tesisinin hizmet verdiği bölgenin yaşam standardına bağlı olarak, atıksu arıtma tesisine gelen atıksu karakterinde ve saatlik debilerinde de ciddi değişimler görülebilmektedir. Bu sebeptir ki her arıtma tesisi farklı proses ve müdahaleler ile çalıştırılmaktadır.

Tesis işletilmesinde prosesden sorumlu olacak kişi tesis içerisindeki tüm proseslere hakim olmalı, laboratuvar dan ve enstrümanlardan alınan işletme parametrelerinin sayısal değerlerini iyi bilmeli ve kontrol/otomasyon açısından takibini çok iyi yapabilmelidir. Laboratuvar dan ve sahadan alınan verilere göre tesise müdahale edilmesi gerekmektedir. Müdahalenin geç veya eksik olması durumu söz konusu olduğunda hem işletme verimi azalmakta hem de maliyet açısından maddi kayıplar ortaya çıkabilmektedir. Özellikle giriş debisi ve su kalitesindeki salınımlar dengeleme tanklarının bulunmadığı durumlarda tesis performansını önemli ölçüde etkilemektedir. Gün içerisinde tesise gelen su kalitesi, endüstriyel tesislerin kaçak olarak deşarj ettiği atıksular tarafından bozulmaktadır. Bu gibi kaçak deşarj edilen sulara erken müdahale

edilmesi gerekmektedir. Bu ve bunun gibi bir sürü etken atıksu arıtma tesis prosesinin verimli şekilde kontrol edilmesini zorlaştırmaktadır. Kritik işletme durumlarında karar verme durumu geciktikçe tesis performansı düşmektedir. Atıksu arıtma tesisi işletilmesinde tesisin çeşitli noktalarından belirli zamanlarda alınan numuneler laboratuvarda analiz edilmekte ve proseslere analiz sonuçlarına göre müdahale edilmektedir. Laboratuvar analizleri arasında bulunan Biyolojik Oksijen İhtiyacı (BOI_5) parametresinin ölçüm sonucu en erken 5 gün içerisinde sonuç verebilmektedir. BOI_5 için yaklaşık bir sonuç almak için bakılan KOI parametre sonucu ise en erken 3 saat içerisinde sonuç verebilmektedir. Fakat KOI parametre ölçümü için gerekli olan kimyasallar maliyetli ve tehlikelidir. Ayrıca tesis içerisinde bulunan su kalite ölçüm cihazları ve debimetre gibi enstrümanların arıza yapma durumlarında sistemin düzgün işletilememesi söz konusudur. Bu nedenledir ki atıksu arıtma tesisinin daha güvenli, hızlı ve daha az maliyetli bir şekilde işletilmesi ve kontrol altına alınabilmesi için tesisin geçmişine ait analiz verileri ile geliştirilecek modellerin kullanılması ile bazı parametelerin tahmin edilmesi oldukça yararlı olacaktır.

Atık su arıtma tesislerinde geçmiş verilerden faydalanılarak bazı parametrelerin tahmin edilmesi için çeşitli istatistiksel yöntemler ve modelleme teknikleri kullanılmıştır. Yapay sinir ağları (YSA), adaptif ağ temelli bulanık çıkarım sistemi (ANFIS), çoklu istatistiksel yöntemler (temel bileşenler analizi, PCA ve çoklu lineer regresyon, MLR), Uluslararası Su Birliği (IWA) tarafından geliştirilen aktif çamur prosesi modelleri (ASM), Genetik algoritma, genetik programlama ve gen ifadeli programlama vb. yöntemler kullanılmıştır [6] [7] [8] [9] [10] [11] [12] [13] [14] [15].

Yapay Sinir Ağı (YSA) tekniği yüksek doğruluk, yetkinlik ve mühendislikteki umut vaat eden uygulamaları sayesinde proses performansının daha iyi tahmin edilmesi için kullanılabilir. Şimdiye kadar, mühendislik alanında YSA'nın birçok uygulaması yapılmıştır.

Deepnarain ve arkadaşları (2019) Güney Afrika Gauteng'de bulunan biyolojik arıtma sistemi, aerobik (%29), anoksik (%29) ve aerobik (%42) bölgeleri içeren 3 aşamalı bir Phoredox proses olarak çalıştırılan atıksu arıtma tesisleri 2 yıl boyunca aylık olarak

izlenerek çıkış atık su karakteristikleri ile çalışma koşulları (girdi olarak) ve çamur hacim endeksi (SVI) arasındaki ilişkiyi çıktı olarak tanımlamak için ana bileşen analizi ve regresyon ağacı modeli kullanılmıştır. Filamentli mikroorganizmaların çoğalmasını etkileyen çevresel faktörleri belirlemek için bir sınıflandırma ağacı modeli kullanılmıştır. Floresan in situ hibridizasyon analizi ile, Mikrotris parvicella, Thiothrix I & II ve Eikelboom Tipleri 0041, 0092 ve 021 N türlerini tanımlandı. Çamur hacim endeksi (SVI)'nin, çamur tutma süresindeki bir artışla arttığı, ancak çözünebilir KOİ ile negatif korelasyon gösterdiği bulundu. Microthrix parvicella'nın baskınlığı, 15.5 °C'nin altındaki sıcaklık düşüşüyle amonyum-azot gözlemlendi ve bu da kış ve ilkbahar mevsimlerinde SVI'da bir artışa neden oldu. Thiothrix'in büyümesi, kolayca biyobozunur KOİ'nin ve besin türleri arasındaki dengesiz orana bağlanabilir. Filament tip 0092, yüksek SVI'a katkıda bulundu ve gıda / mikroorganizma oranının 0.08 l/d'nin altına düşmesiyle galip geldi [16].

Torregrossa ve arkadaşları (2018), atıksularda yüksek performanslı enerji maliyeti modelleri üretmek için makine öğrenmesi kullanmıştır. Kuzeybatı Avrupa'da bulunan 317 atıksu arıtma tesisi veri tabanı kullanılarak, enerji maliyeti modellemesinde en önemli değişkenler belirlenmiştir. Ayrıca bu çalışma ise ilk defa enerji fiyatı model parametresi olarak kullanılmış ve önemi değerlendirilmiştir. Atık su arıtma tesislerinin proses parametreleri ve enerji tüketimi arasındaki ilişkileri tanımlamak için bu makalede, geleneksel yaklaşımların bir evrimi olan makine öğrenmesine dayanan Makine Öğrenmesi Maliyet Modellemesi (MLCM) metodolojisini kullanmaktadır. MLCM algoritmaları ile model performans göstergelerinin genellikle literatürde bulunanlardan daha iyi olduğu gösterilmiştir [17].

Kore'deki bir atık su arıtma tesisinde 1 günlük aralıklı T-N atık su konsantrasyonunu tahmin etmek için iki makine öğrenme modeli oluşturmaktır- yapay sinir ağları (YSA) ve destek vektör makineleri (SVM'ler). Günlük su kalitesi verileri ve meteorolojik veriler kullanılarak her iki modelin de performansı, belirleme katsayısı (R^2), Nash-Sutcliffe verimliliği (NSE), nispi verimlilik kriterlerine (d_{rel}) göre değerlendirildi. Ek olarak, Latin-Hypercube bir kerede bir faktör (LH-OAT) ve bir model arama algoritması sırasıyla duyarlılık analizine ve model parametresi optimizasyonuna

uygulandı. Sonuçlar, her iki modelinde, T-N atık su konsantrasyonunun 1 günlük aralık tahminine etkili bir şekilde uygulanabileceğini gösterdi. SVM modeli, eğitim aşamasında daha yüksek bir tahmin doğruluğu ve onaylama aşamasında da benzer sonuç göstermiştir. Bununla birlikte, duyarlılık analizi, YSA modelinin, 1 günlük aralıklı T-N konsantrasyon tahmini için, T-N konsantrasyonu ile entegre gıda atıkları ve atık su arıtımı modellemesi için girdi değerleri arasındaki ilişki göz önüne alınarak, daha üstün bir model olduğunu göstermiştir. Bu çalışma, entegre gıda atıklarının su kalitesinin ve atık su arıtma prosesinin erken tahmini için verimli ve sağlam doğrusal olmayan zaman serisi modelleme yöntemini önermiştir [18].

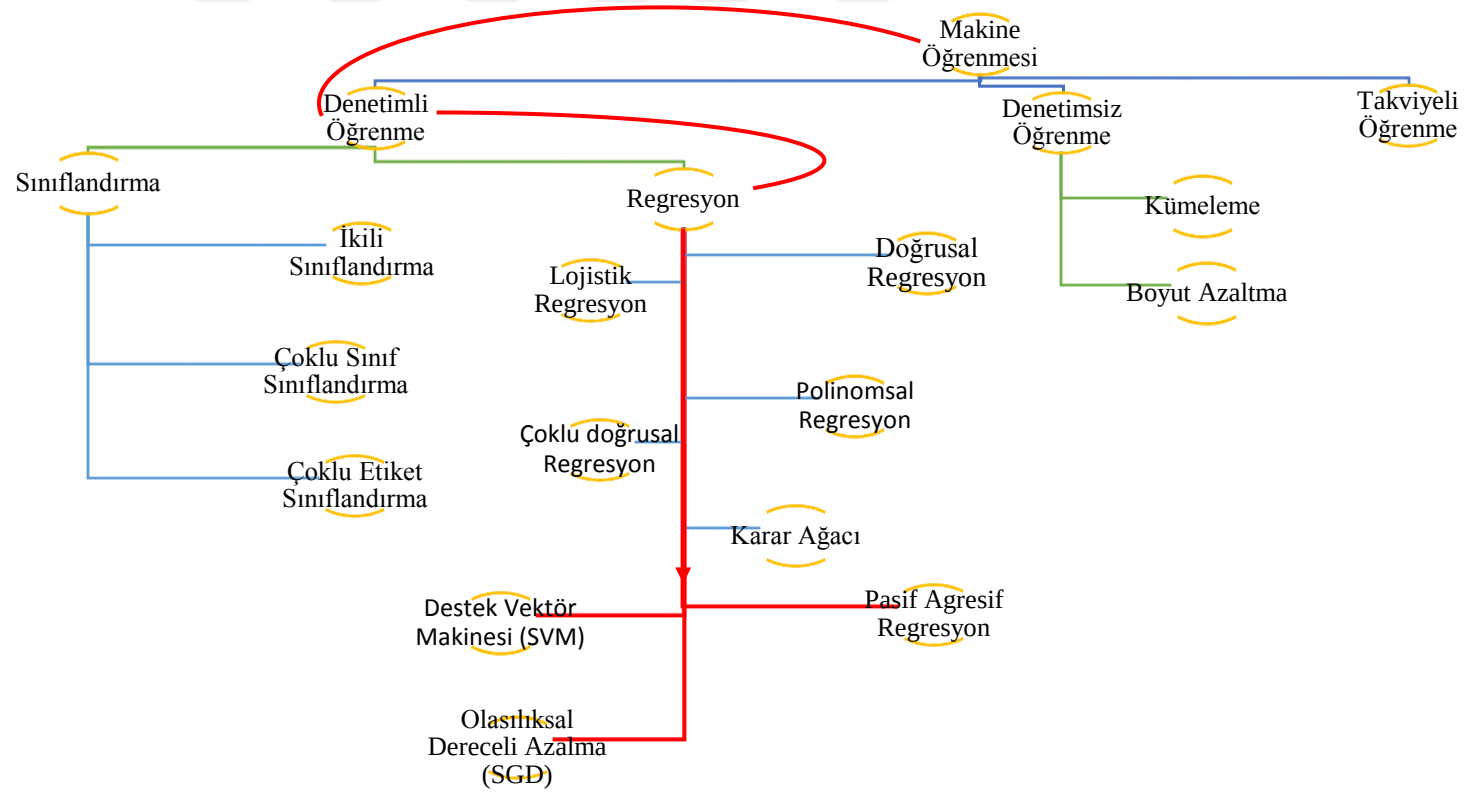
Su kalitesi, doğal çevre ve insan hayatı üzerindeki güçlü etkileri nedeniyle rezervuar yönetiminde en kritik konulardan biridir. Bu çalışma, rezervuarlarda sıkça kullanılan bir su kalitesi ölçümü olan Carlson's Trophic State Index'i tahmin etmek için bir makine öğrenme yaklaşımı oluşturmaktadır. Tayvan'da 20 rezervuardaki istasyonlardan on yıldan uzun süredir (1995-2016) toplanan veriler modelleme sistemine girdi olarak önceden işlendi. Dört iyi bilinen yapay zekâ tekniği, yapay sinir ağları (YSA), destek vektör makineleri, sınıflandırma ve regresyon ağaçları ve doğrusal regresyon, başlangıç ve topluluk senaryolarında analiz etmek için kullanıldı. Tahmini performansı değerlendirmek ve onu iki kurucu senaryodakilerle karşılaştırmak için bir meta-regresyon modelini bütünleştiren kullanıcı dostu bir arayüz geliştirilmiştir. Kapsamlı bir karşılaştırma, YSA'nın modelinin kademe yöntemine dayanarak, diğer tekli modellerden ve hibrit üstsezgisel regresyon modelinden daha doğru olduğunu göstermiştir. Hem tahminin doğruluğu hem de uygulamanın etkinliğinin su yönetimi çalışmalarının planlanmasında uygulayıcıları desteklediği düşünülmektedir. Buna göre, bu çalışma su kalitesi değerlendirmesinde potansiyel kullanım için yeni bir yaklaşım sunmaktadır [19].

Bu tez çalışması kapsamında yukarıda bahsedilen çalışmalardan farklı olarak bir atıksu arıtım tesisine ait geçmiş yıllardaki veriler kullanılarak 3 farklı makine öğrenme algoritması ile BOİ₅ parametresinin tahmin edilmesi amaçlanmıştır.

BÖLÜM 2. MATERYAL VE YÖNTEM

2.1. Makine Öğrenmesi

Makine Öğrenmesi, bilgisayarların insanlar gibi öğrenmelerini ve davranmalarını ve özerk bir şekilde, gözlem ve gerçek dünya etkileşimleri şeklinde veri ve bilgi besleyerek öğrenmelerini zaman içinde geliştirmelerini sağlayan bilimdir. Makine öğrenmesi ile geçmiş veriler dayanarak bilgisayara tecrübe kazandırmak yoluyla olayların analizi ve bunların gelecekte alabileceği durumları hakkında tahminleri yaptırılabilir. Kısaca bilgisayarın öğretilen bilgileri ve tecrübeleri analiz ederek gelecekte oluşacak benzeri olaylar hakkında kararlar verebilmesi ve problemlere çözümler üretebilmesi olarak tanımlanabilir [20]. Örnek verileri veya geçmiş deneyimleri kullanarak bir performans kriteri optimize edilebilir. Makine Öğrenmesi yöntemleri verinin yapısına göre Şekil 2.1.'de gösterilmektedir.



Şekil 2.1. Makine öğrenmesi algoritmaları sınıflandırılma şeması

2.1.1. Denetimsiz Öğrenme (Unsupervised Learning)

Denetimsiz Öğrenme, verilerdeki kalıpları bulmak için bir Makine Öğrenme teknikleri sınıfıdır. Denetlenmeyen algoritmaya verilen veriler etiketlenmemiştir, yani sadece girdi değişkenlerine (x) karşılık gelen çıktı değişkenleri verilmez. Denetimsiz öğrenmede, verilerdeki ilginç yapıları keşfetmek için algoritmalar kendilerine bırakılmıştır.

2.1.1.1. Kümeleme (Clustering)

Kümeleme, veri noktalarının gruplandırılmasını içeren bir makine öğrenme tekniğidir. Bir dizi veri noktası göz önüne alındığında, her veri noktasını belirli bir gruba sınıflandırmak için bir kümeleme algoritması kullanabiliriz. Teoride, aynı gruptaki veri noktalarının benzer özelliklere ve / veya özelliklere sahip olması gerekirken, farklı gruplardaki veri noktalarının birbirine benzemeyen özelliklere ve / veya özelliklere sahip olması gerekir. Kümeleme, denetimsiz bir öğrenme yöntemidir ve birçok alanda kullanılan istatistiksel verilerin analizi ve tespiti için yaygın bir tekniktir. Açık bir ifadeyle, amaç benzer özelliklere sahip grupları ayırmak ve kümeler oluşturaktır.

2.1.2. Denetimli öğrenme (Supervised Learning)

Bu sınıftaki algoritmalar, dışarıdan temin edilen örneklerden genel hipotez üretmesine neden olan ve daha sonra gelecek örnekler hakkında tahminler yapan algoritmalar arayışıdır. Denetimli sınıflandırma, akıllı sistemler tarafından en sık gerçekleştirilen görevlerden biridir.

Bu algoritma, belirli bir belirleyici kümesinden (bağımsız değişkenler) tahmin edilecek bir sonuç-hedef değişkeninden (veya bağımlı değişken) oluşur. Bu değişkenleri kullanarak, veri girişleri istenen çıktılarla eşleştiren bir işlev üretilmektedir. Eğitim süreci, model eğitim verilerinde istenen bir doğruluk seviyesine ulaşana kadar devam eder [21].

2.1.2.1. Sınıflandırma yöntemi

Sınıflandırma yöntemi için geliştirilen algoritmalar, verilerin belli özelliklerine göre sınıflandırılarak ayrılmasına dayanmaktadır. Eğer yöntem, hangi verinin, hangi koşullarda, hangi sınıfa ait olacağı bilgisi ile sınıflandırlara ayrılarak eğitilirse, yeni oluşturduğu veri setindeki verileri de eğitildiğine benzer biçimde sınıflandırabilir.

2.1.2.2. Regresyon yöntemi

Regresyon analizi yöntemi, istenilen en az iki veya daha fazla değişken arasındaki birbirleri ile olan ilişkilerin incelenmesini sağlayan güçlü bir istatistiksel yöntemdir. Pek çok regresyon analizi türü varken, hepsi bir veya daha fazla bağımsız değişkenin bağımlı değişken üzerindeki etkisini incelemektedir. Regresyon analizi, hangi değişkenlerin ilgi konusu üzerinde etkili olduğunu belirlemede güvenilir bir yöntemdir. Bir regresyon gerçekleştirme süreci, hangi faktörlerin en önemli olduğunu, hangi faktörlerin göz ardı edilebileceğini ve bu faktörlerin birbirlerini nasıl etkilediğini güvenle belirlenmesini sağlar. Bir regresyon analizi yapmak için, varsayılan bir ya da birkaç bağımsız değişkenden etkilendiği düşünülen bağımlı bir değişken tanımlanması gerekmektedir.

Regresyon modellerinde kullanılan başlıca teknikler şunlardır:

Liner Regresyon (Linear Regression): Liner regresyonda, bir değişken üzerindeki veriler ile ikinci değişkendeki değerleri tahmin edilmektedir. Tahmin edilen değişken, ölçüt değişkenidir ve y olarak adlandırılır. Tahminlerde temel alınan değişken, tahmin değişkenidir ve x olarak adlandırılır. Yalnızca bir tahmin değişkeni olduğunda, tahmin yöntemi çağrılır. Doğrusal regresyon, noktalar arasında en uygun düz çizgiyi bulmaktan oluşur. En uygun çizgiye regresyon çizgisi denir.

Lojistik Regresyon (Logistic Regression): Lojistik regresyon, farklı sınıflara gözlemler atamak için kullanılan ve sınıflandırma problemleri için kullanılan tahminsel bir analiz algoritmasıdır ve olasılık kavramına dayanır.

Karar Ağaçları (Decision Trees): Karar ağaçları, en basit ve en kullanışlı Makine Öğrenme yapılarından biridir. Karar Ağacı öğrenme algoritması, sınıflandırma ve regresyon problemini çözmek için eğitim verilerinden karar ağaçları oluşturur.

Destek Vektör Makinası (Support Vector Machine (SVM)): Destek vektörü makine algoritmasının amacı, bir N-boyutlu uzayda (N- özelliklerin sayısı) veri noktalarını belirgin bir şekilde sınıflandıran bir hiper düzlem bulmaktır. İki veri noktası sınıfını ayırmak için, seçilebilecek birçok olası hiper düzlem vardır. Amacımız her iki sınıfın veri noktaları arasındaki maksimum mesafeyi bulmak. Marj mesafesinin maksimize edilmesi, bazı veri takviyeleri sağlayarak gelecekteki veri noktalarının daha güvenle sınıflandırılmasını sağlar.

Polinom Regresyonu (Polynomial Regression): Veriler arası ilişki her zaman doğrusal değildir. Optimum ilişkiyi bulmak için bir eğri gereklidir. Polinom fonksiyonlarından örnekle bu yöntemde de bir terimin karesi veya küpü alınarak doğrusal olmayan bir model oluşturulmak istenebilir. Bu gibi durumlarda kullanılabilen bir algoritma modelidir. Polinom regresyon yöntemi, bağımlı ve bağımsız değişken arasındaki ilişkinin en iyi yaklaşımını sağlar [22].

Olasılıksal Dereceli Azalma (Stochastic Gradient Descent (SGD)): Metin sınıflandırmasında ve doğal dil işlemede sıklıkla karşılaşılan büyük ölçekli ve seyreltilmiş makine öğrenmesi problemlerine başarıyla uygulanmıştır. Verilerin seyrek olduğu göz önüne alındığında, bu modüldeki sınıflandırıcılar, 10^5 ' den fazla eğitim örneği ve 10^5 ' den fazla özellik içeren problemlere kolayca ölçeklenebilir.

Pasif Agresif Regresör (Passive Aggressive Algorithms): Pasif Agresif Algoritmalar ark Crammer tarafından geliştirilen sınıflandırma ve regresyon için çevrimiçi öğrenme algoritmalarıdır. Bu algoritma çok basittir ve performanslarının online Perceptron ve MIRA gibi diğer birçok alternatif yöntemden daha üstün olduğu kanıtlanmıştır.

2.1.3. Takviyeli öğrenme

Bu algoritmayı kullanarak, makine belirli kararlar vermek için eğitilmiştir. Makine, deneme yanılma kullanarak sürekli olarak kendini eğittiği bir ortama maruz kalır. Bu makine geçmiş deneyimlerden öğrenir ve doğru iş kararları vermek için mümkün olan en iyi bilgiyi yakalamaya çalışır [20].

2.2. Çalışma Kapsamında Modellerde Kullanılan Yalancı Kod

Atıksu arıtma tesisine ait verilerin modellenmesinde kullanılan yalancı kod aşağı belirtilmektedir;

Verileri hazırla

data = arıtma.csv

Parametreleri belirle Test oranı = 0.1, iterasyon_sayısı = 100, en_iyi_skor=0, i = 0

Test ve train verilerini belirle x_test, x_train, y_test, y_train

Öğrenme algoritmalarını oluştur SVR, SGD, PassiveAggressiveRegressor

For öğrenme algoritması in öğrenme algoritmaları listesi:

While i<iterasyon_sayısı:

i = i +1

clf = öğrenme algoritması

clf.fit Modelin öğrenmesini sağla

if **clf.score** > en_iyi_skor: Modelin skorunu hesapla

clf.save() Model daha iyi bir skora sahipse kaydet

en_iyi_skor = clf.score yeni skoru kaydet

end if

end while

end for

BÖLÜM 3. ARAŞTIRMA BULGULARI

Tez çalışması kapsamında bir biyolojik atıksu arıtma tesisinde geçmiş yıllara ait deneysel veriler kullanılarak makine öğrenmesi algoritmalarından 3 farklı denetimli öğrenme algoritması kullanılarak BOI_5 parametresinin tahmin edilmesi amaçlanmıştır. Bu amaçla bir program geliştirilerek problemin çözümüne ilişkin en uygun öğrenme algoritması belirlenmiştir.

3.1. Veri Seti

Çalışma kapsamında bir atıksu biyolojik arıtma tesisinde deneysel olarak ölçülen parametrelerin farklı makine öğrenme algoritmaları kullanılarak BOI_5 parametresinin tahmin edilmesi amaçlanmaktadır. Modellemede kullanılacak veri setine ait tanımlayıcı istatistikler Tablo 3.1.'de özetlenmiştir.

Tablo 3.1. Atıksu arıtma tesisi deneysel veri setine ait tanımlayıcı istatistikler (n=366)

Veri Seti	$\bar{X}$	σ	X_{min}	$X_{25\%}$	$X_{50\%}$	$X_{75\%}$	X_{maks}	BOI_5 ile Korelasyon
BOI_5 (mg/l)	186	123	76	155	186	215	333	1
Q ($m^3/gün$)	196629	117956	99528	187980	202621	210901	237215	0,0658
T ($^{\circ}C$)	23	16	13	21	23	25	28	-0,0300
pH	8	6	6	7	8	8	8	-0,2004
İletkenlik ($\mu S/cm$)	1296	874	570	1131	1208	1366	2670	-0,0352
KOI (mg/l)	313	199	101	243	307	367	834	0,3620
AKM (mg/l)	151	99	40	108	138	180	562	0,2373

Tablo 3.1.'den de görüldüğü gibi modellemede kullanılacak ham verilerin büyüklükleri birbirinden farklı olduğundan dolayı verileri modellemede kullanmak üzere aynı ölçeğe getirmek için normalizasyon işlemi yapılmıştır. Burada Min-Max

normalizasyon yöntemi kullanılmıştır. Bu yöntem verileri doğrusal olarak normalize eder.

Minimum; birverinin alabileceği en düşük değer iken, maksimum; verinin alabileceği en yüksek değeri ifade eder. Bir veriyi min-max yöntemi ile 0 ile 1 aralığına indirmek için denklem 3.1. kullanılmaktadır.

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

x' : Normalize veri

x_i : i. Girdi verisi

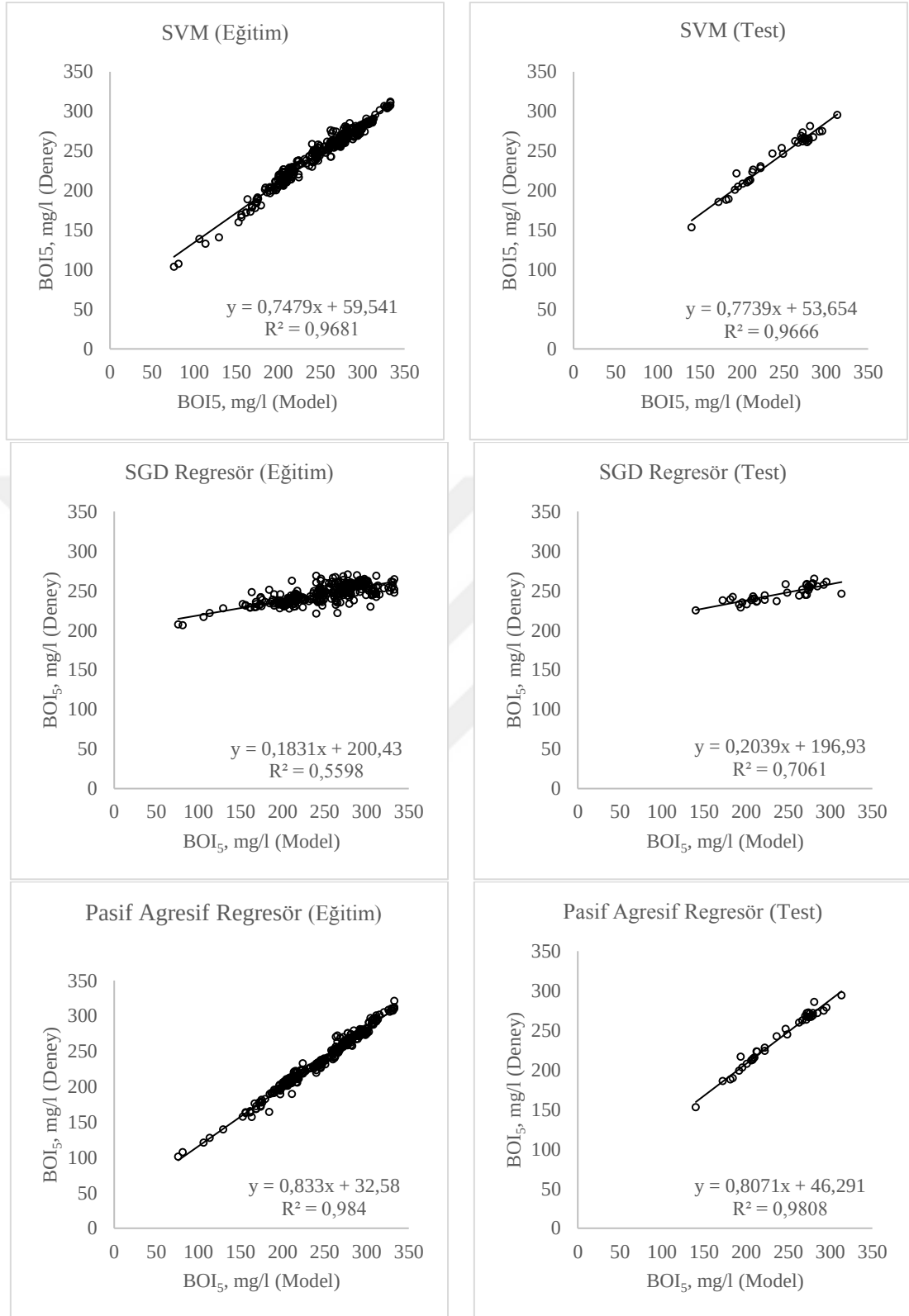
x_{min} : Girdi verisi içerisinde yer alan en küçük sayıyı

x_{max} : Girdi verisi içerisinde yer alan en büyük sayıyı ifade etmektedir.

Veriler normalize edildikten sonra veri setinin eğitim ve test seti olarak ikiye ayrılması gerekmektedir. Modelleme için geliştirilen programda kullanılan öğrenme algoritmaları problem için iterasyon sayısına göre optimum çözümü üretecek eğitim ve test setini kendisi seçmekte ve en optimum sonucu veren eğitim ve test setini kaydetmektedir. Geliştirilen program veri setinin %90'ı eğitimde %10'u ise test setinde kullanılacak şekilde 100000 iterasyon yapılacak şekilde ayarlanmıştır.

3.2. Modelleme Sonuçları

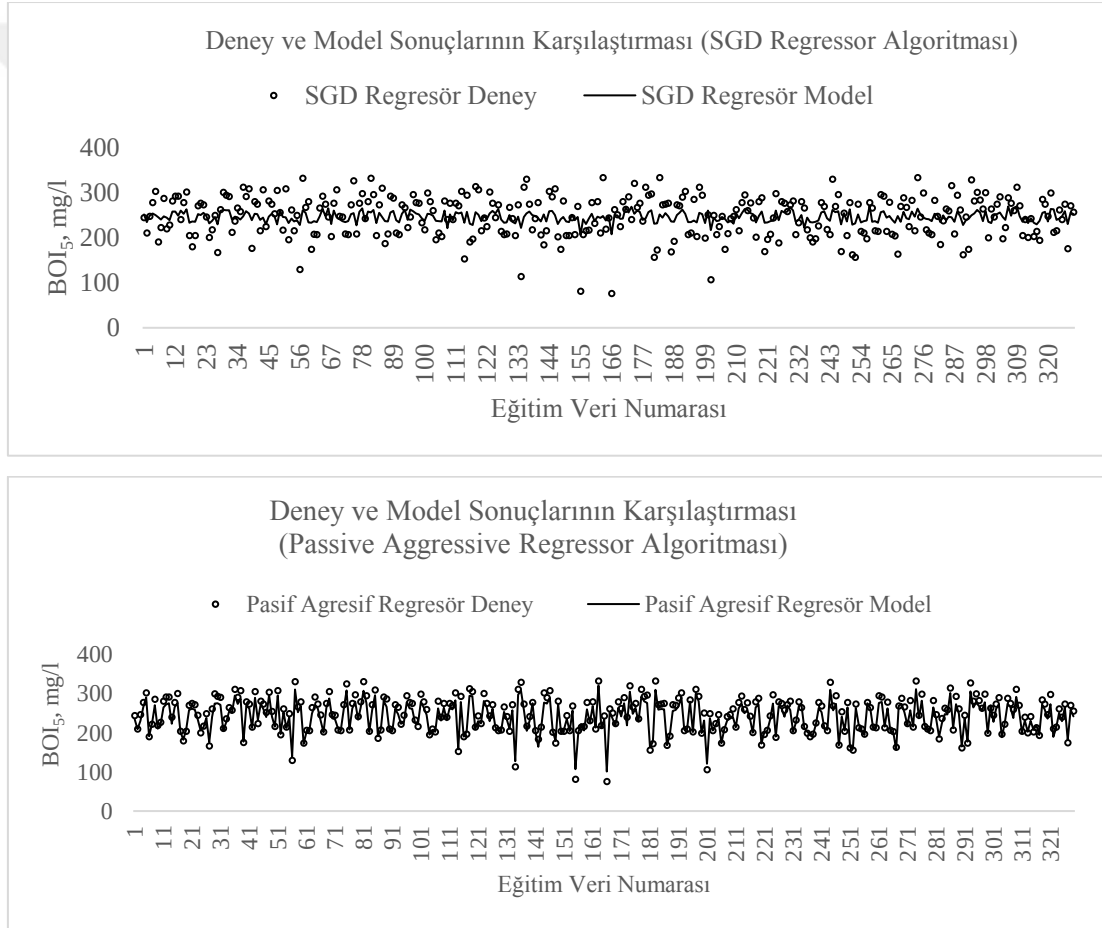
Biyolojik atıksu artıma tesisinde BOI_5 parametresinin tahmin edilmesi için 3 farklı öğrenme algoritması (Destek Vektör Makinesi (SVM), Olasılıksal Dereceli Azalma (SGD Regresör) ve Pasif Agresif Regresör) geliştirilen program aracılığı ile denemiştir. Her bir algoritma için 100000 iterasyon sonunda eğitim ve test setleri için elde edilen sonuçlar aşağıdaki Şekil 3.1.'de gösterilmektedir.



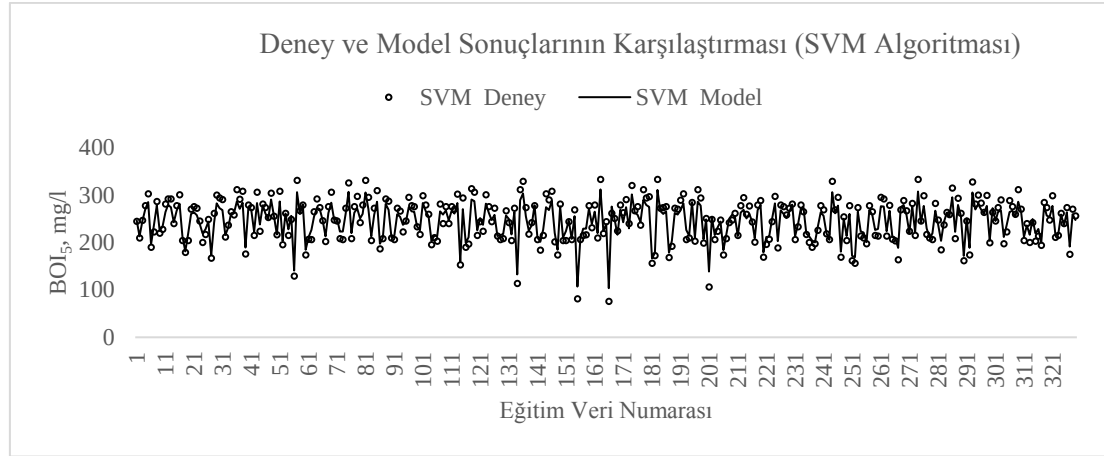
Şekil 3.1. Öğrenme algoritmaları için eğitim ve test sonuçlarına ait saçılım grafiği

Şekil 3.1.'de görüldüğü gibi BOI_5 tahmininde eğitim ve test verileri için elde edilen regresyon katsayıları (R^2) sırasıyla SVM öğrenme algoritması için, 0,9681;0,9666, SGD regresör öğrenme algoritması için, 0,5598;0,7061 ve Pasif Agresif Regresör öğrenme algoritması için ise 0,9840;0,9808 olarak belirlenmiştir. Buradan anlaşılacağı üzere hem SVM öğrenme algoritması hemde Pasif Agresif Regresör öğrenme algoritması BOI_5 tahmini için çok iyi sonuçlar vermiştir.

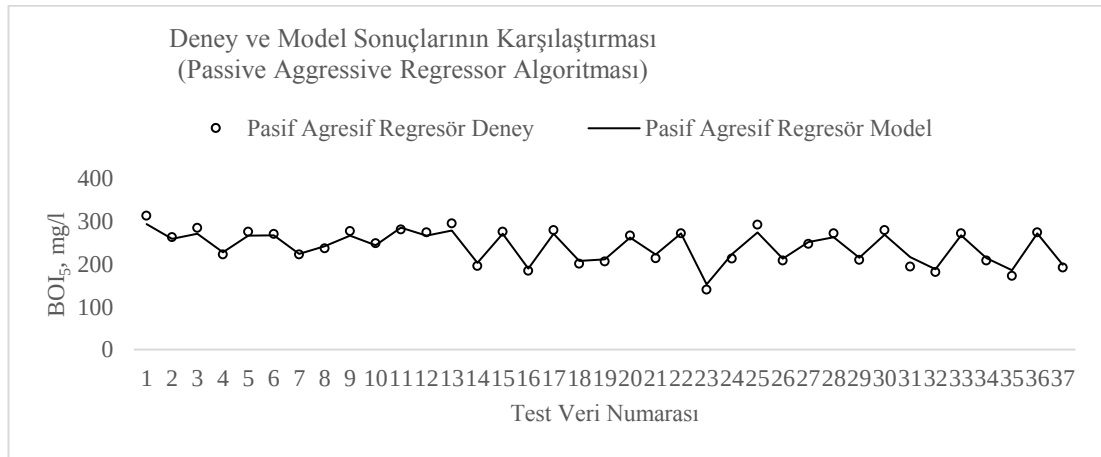
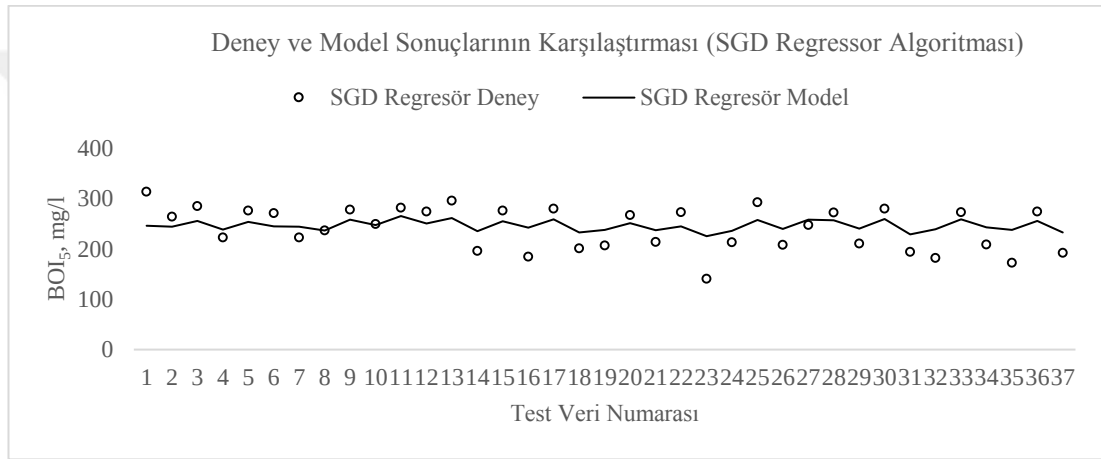
Aşağıda Şekil 3.2.'de eğitim ve test setine ait veriler için her bir öğrenme algoritmasından elde edilen deney ve model sonuçları karşılaştırılmıştır.



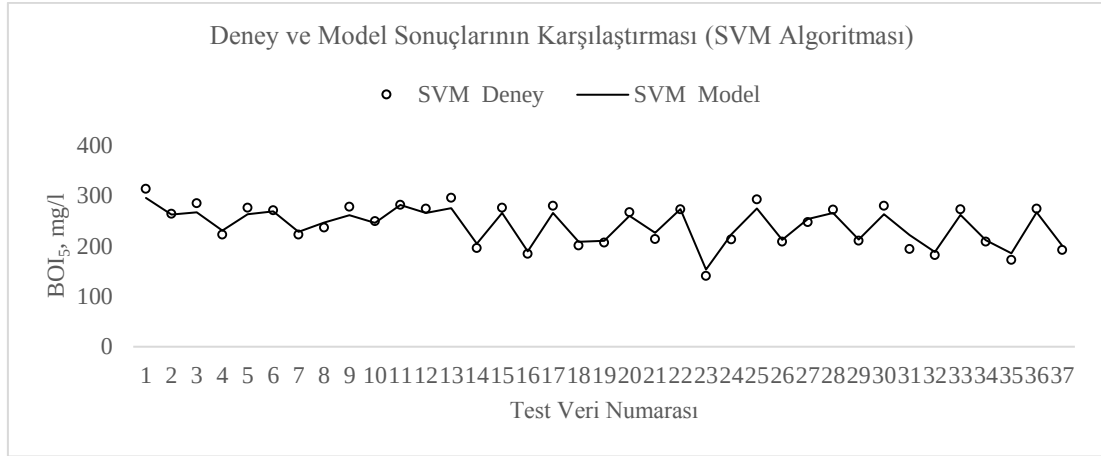
Şekil 3.2. Eğitim seti için öğrenme algoritmalarına göre deney ve model sonuçlarının karşılaştırılması



Şekil 3.2. (Devamı)



Şekil 3.3. Test seti için öğrenme algoritmalarına göre deney ve model sonuçlarının karşılaştırılması



Şekil 3.3. (Devamı)

3.3. Deney ve Model Sonuçlarının İstatiksel Karşılaştırması

Geliştirilen modelde bir makine öğrenme algoritmasının veri setine uygulanmasından sonraki işlem, kullanılan öğrenme algoritmasının verilerimize ne kadar uygun olduğunun belirlenmesidir. Farklı makine öğrenmesi algoritmalarını değerlendirmek ve uygulamak için farklı performans değerlendirme yöntemleri kullanılmaktadır. Çalışmadaki öğrenme algoritmalarının performansının değerlendirilmesinde kullanılan yöntemler aşağıda açıklanmıştır.

3.3.1. Kök ortalama kare hata (root mean square error (RMSE))

Modelinin tahmin ettiği sonuçlar ile gerçek veriler arasındaki uzaklığın bulunmasında kullanılan bir ölçüdür. RMSE değerinin 0 olması kullanılan modelinin hiç hata yapmadığı anlamına gelir. Kök ortalama kare hata hesaplamak için denklem 3.1. kullanılmaktadır.

$$RMSE = \sqrt{\frac{\sum_{j=1}^n (A_j - P_j)^2}{n}} \quad (3.1)$$

n : Veri Sayısı

A_j : Gerçek veri

P_j : Tahmin edilen veri

ifade etmektedir.

3.3.2. Ortalama mutlak hata (mean absolute error (MAE))

Modelinin tahmin ettiği veriye en iyi uyan çizgi arasındaki ortalama dikey mesafeyi hesaplayarak hatanın bulunmasında kullanılan bir ölçüdür. MAE değerinin sıfıra yakın olması kullanılan makine öğrenmesi modelinin daha iyi performans gösterdiği anlamına gelir. Ortalama mutlak hata hesaplamak için denklem 3.2. kullanılmaktadır.

$$MAE = \frac{1}{n} \sum_{j=1}^n |A_j - P_j| \quad (3.2)$$

n : Veri Sayısı

A_j : Gerçek veri

P_j : Tahmin edilen veri

ifade etmektedir.

3.3.3. Ortalama mutlak yüzde hata (mean absolute percentage error (MAPE))

Ortalama mutlak yüzde hata, modelin tahmin ettiği verilerin gerçek değerler ile olan yüzdelik hataların mutlak değerleri toplamalarının ortalamasını almaktadır. Gerçek değerler arasında sıfır değeri olmamalıdır.

Ortalama mutlak yüzde hata hesaplamak için denklem 3.3. kullanılmaktadır.

$$MAE = \frac{100}{n} \sum_{j=1}^n \frac{|A_j - P_j|}{A_j} \quad (3.3)$$

n : Veri Sayısı

A_j : Gerçek veri

P_j : Tahmin edilen veri

ifade etmektedir.

Çalışmada kullanılan öğrenme algoritmalarının tahmin performansları için hesaplanan performans sonuçları Tablo 3.2. 'de verilmiştir.

Tablo 3.2. Öğrenme algoritmaları modellerinden elde edilen performans sonuçları

Model	RMSE		MAE		MAPE		R ²	
	Eğitim	Test	Eğitim	Test	Eğitim	Test	Eğitim	Test
SVM	13,275	1,479	11,349	0,2432	0,0476	0,0013	0,9678	0,9988
SGD	37,712	34,235	31,064	29,324	0,1415	0,1354	0,5603	0,7076
PAR	12,465	9,2998	10,444	7,892	0,0414	0,0342	0,9836	0,9817

Tablo 3.2. 'ten de anlaşılabacağı üzere bu çalışmada BOİ₅ tahmini için kullanılan öğrenme algoritmalarından en iyi performansa sahip olan en düşük RMSE, MAE ve MAPE değerleri ve en yüksek R² değerine sahip Pasif Agresif Regressör (PAR) öğrenme algoritmasıdır.

BÖLÜM 4. TARTIŞMA VE SONUÇ

Temiz su kaynaklarını koruma çalışmalarının başında atıksu arıtma tesisleri yapmak ve bu tesislerin performanslı bir şekilde işletilmesi büyük önem arz etmektedir. Atıksu arıtma tesislerinin verimli bir şekilde işletilmesi için proses ve laboratuvar verileri çok önemlidir. Tesise alınan atık suyun karakterinin önceden tespit edilmesi tesisi daha performanslı işletmemizi sağlayacaktır. Bu nedenle atıksu karakterini önceden tespit edebilmek için geçmiş verilerden faydalı olarak geliştirilmiş modeller kullanarak bazı atıksu parametrelerin tahmin edilmesi tesisin verimli bir şekilde işletilmesi açısından önemlidir.

Atıksu arıtma tesisinin işletilmesinde yönetmelikler ile belirlenmiş tesis çıkış suyu kalitesini ölçülmesi amacıyla bakılan parametreler BOI_5 , KOI , AKM, pH, Sıcaklık, TP ve TN'dur. BOI_5 ve KOI parametreleri biyolojik atıksu arıtma tesislerinde organik maddenin tespit edilmesinde kullanılan en yaygın ölçümlerden biridir. BOI_5 parametre ölçümünün uzun zaman alması (en az 5 gün), deney prosedürünün hassas olması ve KOI parametresinin ölçümü ise kimyasal sarf maliyeti ve zor bir deney olması sebebiyle işletme sırasında maliyet ve zamandan kazanmak için bu çalışmada biyolojik atıksu arıtma tesislerinin önemli işletme parametrelerinden biri olan BOI_5 tahmini için geçmiş yıllara ait verileri kullanarak 3 farklı öğrenme algoritmasının tahmin performansının belirlenmesi ve bu amaçla kullanılabilirliği araştırılmıştır.

Çalışmada biyolojik atıksu arıtma tesisine ait Debi (Q), Sıcaklık (T), pH, İletkenlik, KOI , Askıda katı madde (AKM) parametreleri girdi olarak kullanılmış ve BOI_5 parametresi 3 farklı makine öğrenme algoritması kullanarak tahmin edilmiş ve elde edilen sonuçlar çeşitli performans parametreleri açısından karşılaştırılmıştır. Eğitim ve test verileri ile elde edilen sonuçlar (Tablo 3.3) incelendiğinde BOI_5 tahmini için kullanılan öğrenme algoritmalarından en iyi performansa sahip olan en düşük RMSE

(E:12,465/T:9,2998) , MAE (E:10,444/T:7,892) ve MAPE (E:0,0414/T:0,0342) değerleri ve en yüksek R^2 (E:0,9836/T:0,9817) değerine sahip Pasif Agresif Regressör (PAR) öğrenme algoritması olduğu görülmüştür. Dolayısı ile Pasif Agresif Regressör (PAR) öğrenme algoritmasının atıksu arıtma tesislerinin işletilmesi için BOI_5 tahmininde etkili bir şekilde kullanılabileceği anlaşılmıştır. Bu sayede BOI_5 tahmini yapılarak tesislerin işletilmesinde hızlı ve deneysel çalışma gerekenimi düşeceğinden maliyetsiz bir sonuca ulaşılabilir. Aynı çalışma diğer çıkış parametreleri içinde denenerek tesis işletilmesinde daha hızlı sonuçlar alınabilir fakat sıcaklık ve pH gibi enstürman ile ölçülen parametrelerde çok kullanışlı olmayacaktır.



KAYNAKLAR

- [1] M. Çakmakci, B. Özkaya, M. S. Bilgili, Ö. Apaydın, ve S. İdris, “Atıksu Arıtma Eylem Planı (2014-2023)”, Çevre ve Şehircilik Bakanlığı, 2014.
- [2] “Wastewater Engineering”. <http://web.deu.edu.tr/atiksu/ana52/aryen3.html>. Erişim Tarihi: 09.04.2019.
- [3] Çevre ve Orman Bakanlığı, *Su Kirliliği Kontrolü Yönetmeliği*, c. 25687. 2004.
- [4] Çevre ve Orman Bakanlığı, *Kentsel Atıksu Arıtımı Yönetmeliği*, c. 26047. 2006.
- [5] Admin, “Fosfor Analizi”, *Yeşil Aşkı* <http://www.yesilaski.com/fosfor-analizi.html>. Erişim Tarihi: 01.05.2019
- [6] C. Gökhan, “Arıtma Proseslerinin Yapay Zeka Ve Çoklu İstatistiksel Yöntemler ile Modellenmesi”, Yüksek Lisans Tez, Süleyman Demirel Üniversitesi, 2006.
- [7] H. Moral, A. Aksoy, ve C. F. Gokcay, “Modeling of the activated sludge process by using artificial neural networks with automated architecture screening”, *Comput. Chem. Eng.*, c. 32, sy 10, ss. 2471-2478, Eki. 2008.
- [8] M. S. Bhatti, D. Kapoor, R. K. Kalia, A. S. Reddy, ve A. K. Thukral, “RSM and ANN modeling for electrocoagulation of copper from simulated wastewater: Multi objective optimization using genetic algorithm approach”, *Desalination*, c. 274, sy 1-3, ss. 74-80, Tem. 2011.
- [9] M. S. Nasr, M. A. E. Moustafa, H. A. E. Seif, ve G. El Kobrosy, “Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT”, *Alex. Eng. J.*, c. 51, sy 1, ss. 37-43, Mar. 2012.
- [10] M. Ay ve O. Kisi, “Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques”, *J. Hydrol.*, c. 511, ss. 279-289, Nis. 2014.

- [11] W. Wu, G. C. Dandy, ve H. R. Maier, "Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling", *Environ. Model. Softw.*, c. 54, ss. 108-127, Nis. 2014.
- [12] A. Salahi, T. Mohammadi, R. Mosayebi Behbahani, ve M. Hemmati, "Asymmetric polyethersulfone ultrafiltration membranes for oily wastewater treatment: Synthesis, characterization, ANFIS modeling, and performance", *J. Environ. Chem. Eng.*, c. 3, sy 1, ss. 170-178, Mar. 2015.
- [13] A. Giwa, S. Daer, I. Ahmed, P. R. Marpu, ve S. W. Hasan, "Experimental investigation and artificial neural networks ANNs modeling of electrically-enhanced membrane bioreactor for wastewater treatment", *J. Water Process Eng.*, c. 11, ss. 88-97, Haz. 2016.
- [14] X. Wu, Y. Yang, G. Wu, J. Mao, ve T. Zhou, "Simulation and optimization of a coking wastewater biological treatment process by activated sludge models (ASM)", *J. Environ. Manage.*, c. 165, ss. 235-242, Oca. 2016.
- [15] E. Hong, A. M. Yeneneh, T. K. Sen, H. M. Ang, ve A. Kayaalp, "ANFIS based Modelling of dewatering performance and polymer dose optimization in a wastewater treatment plant", *J. Environ. Chem. Eng.*, c. 6, sy 2, ss. 1957-1968, Nis. 2018.
- [16] N. Deepnarain vd., "Decision tree for identification and prediction of filamentous bulking at full-scale activated sludge wastewater treatment plant", *Process Saf. Environ. Prot.*, c. 126, ss. 25-34, Haz. 2019.
- [17] D. Torregrossa, U. Leopold, F. Hernández-Sancho, ve J. Hansen, "Machine learning for energy cost modelling in wastewater treatment plants", *J. Environ. Manage.*, c. 223, ss. 1061-1067, Eki. 2018.
- [18] H. Guo vd., "Prediction of effluent concentration in a wastewater treatment plant using machine learning models", *J. Environ. Sci.*, c. 32, ss. 90-101, Haz. 2015.
- [19] J.-S. Chou, C.-C. Ho, ve H.-S. Hoang, "Determining quality of water in reservoir using machine learning", *Ecol. Inform.*, c. 44, ss. 57-75, Mar. 2018.
- [20] Ö. Ünsal, "Mesleki Alan Seçimlerinin Makine Öğrenmesi Algoritması Kullanılarak Belirlenmesi", Yüksek Lisans Tez, Gazi Üniversitesi, 2011.

- [21] “Essentials of Machine Learning Algorithms (with Python and R Codes)”, *Analytics Vidhya*, Erişim Tarihi:08.05.2019.
- [22] G. Seif, “Machine Learning”, *Towards Data Science*.
<https://towardsdatascience.com/machine-learning/home>. Erişim Tarihi: 06.05.2019.



ÖZGEÇMİŞ

Ercan SELVİ, 16.02.1989'da Sakarya'da doğdu. İlk, orta ve lise eğitimini Sakarya'da tamamladı. 2006 yılında Akyazı Şehit Yüzbaşı Halil İbrahim Sert Lisesi'nden mezun oldu. 2007 yılında başladığı Sakarya Üniversitesi Çevre Mühendisliği Bölümü'nü 2011 yılında bitirdi. 2012 yılında askerlik eğitimin yedek subay olarak Şanlıurfa-Suruç'da karakol komutanı olarak tamamladı. 2012-2013 yılları arasında geri kazanım alanında faaliyet gösteren özel bir şirkette çalıştım. 2013 yılında Sakarya Üniversitesi Çevre Mühendisliği Bölümünde yüksek lisans eğitimine başladı. 2015 yılından beri su ve atıksu arıtma alanında faaliyet gösteren özel firmalarda çalışmaktayım.