T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

KARIŞIK ENDÜSTRİYEL ATIKSU KARAKTERİZASYONUNUN ZAMANSAL DEĞİŞİMİNİN BELİRLENMESİ ve KİRLETİCİLERİN İSTATİSTİKSEL DEĞERLENDİRİLMESİ

Yasemin ÇELİK

DANIŞMAN
Prof. Dr.Serdar AYDIN

II. DANIŞMAN
Doç. Dr. Elif TÜZÜN

Çevre Mühendisliği Anabilim Dalı

Çevre Mühendisliği, Tezli Yüksek Lisans Programı

Şubat, 2026

TEZ KABUL VE ONAYI

Yasemin ÇELİK tarafından, **Prof. Dr. Serdar AYDIN** danışmanlığında hazırlanan ***" KARIŞIK ENDÜSTRİYEL ATIKSU KARAKTERİZASYONUNUN ZAMANSAL DEĞİŞİMİNİN BELİRLENMESİ ve KİRLETİCİLERİN İSTATİSTİKSEL DEĞERLENDİRİLMESİ"*** başlıklı bu çalışma, jürimiz tarafından **03/02/2026** tarihinde yapılan sınav sonucunda **oy birliği** ile başarılı bulunarak **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Tez Jürisi

		İmza	**Sonuç**
DANIŞMAN	Prof. Dr. Serdar AYDIN İstanbul Üniversitesi - Cerrahpaşa Çevre Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Doç. Dr. Selcan KARAKUŞ İstanbul Üniversitesi - Cerrahpaşa Fiziko Kimya Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Dr. Öğr. Üyesi Fatih ÖZBAŞ Fatih Sultan Mehmet Vakıf Üniversitesi Kültür Varlıklarını Koruma ve Onarım Bölümü		☒ Kabul ☐ Ret

Aileme ithaf ediyorum…

BÜTÇE DESTEKLERİ

KARIŞIK ENDÜSTRİYEL ATIKSU KARAKTERİZASYONUNUN ZAMANSAL DEĞİŞİMİNİN BELİRLENMESİ ve KİRLETİCİLERİN İSTATİSTİKSEL DEĞERLENDİRİLMESİ

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.

TEŞEKKÜR

Tez çalışmam ve meslek hayatım boyunca bilgi ve tecrübelerini paylaşan, her türlü konuda desteğini esirgemeyen çok değerli danışman hocam Sayın Prof. Dr. Serdar AYDIN'a ve bu çalışma sayesinde tanışma fırsatı bulduğum ve beni güler yüzü ile karşılayarak bilgi ve deneyimlerini paylaşan değerli danışmanım Sayın Doç. Dr. Elif TÜZÜN' e sonsuz saygı ve teşekkürlerimi sunarım.

Çalışmamı tamamlamamda bilgi ve birikimleri ile tezime katkıda bulunan Sayın Doç. Dr. Selcan KARAKUŞ ve Dr. Öğr. Üyesi Canan Hazal AKARSU'ya; Jüri Üyesi Dr. Fatih ÖZBAŞ'a ve lisanstan itibaren bilgi ve birikimlerini paylaşarak her zaman motive eden değerli hocalarım Sayın Prof. Dr. Emine ELMASLAR ÖZBAŞ ve Sayın Prof. Dr. Hüseyin Kurtuluş ÖZCAN'a; lisansüstü eğitim sürecim boyunca destek sağlayan Testmer Ölçüm ve Test Hizmetleri yönetici ve çalışanlarına; her vazgeçmemde beni cesaretlendiren ve her zaman yanımda olan can dostlarım Özlem ATILGAN ve İbrahim AKÇAVLI'ya; meslektaşım, değerli arkadaşım Dr. Ömer Yasin BALIK'a ve tüm arkadaşlarıma desteklerinden dolayı teşekkür ederim.

Tüm eğitim hayatım ve yaşamım boyunca hiçbir zaman maddi ve manevi desteklerini benden esirgemeyen, her zaman bana inanan ve güvenen, sevgileri ile bana güç veren ve benimle aynı heyecanlarla her zaman yanımda olan canım aileme teşekkür ederim.

Şubat 2026 **Yasemin ÇELİK**

İÇİNDEKİLER

Sayfa No

TEZ KABUL VE ONAYI ..... **ii**
BEYAN ..... **iii**
BÜTÇE DESTEKLERİ ..... **v**
TEŞEKKÜR ..... **vi**
İÇİNDEKİLER ..... **vii**
ŞEKİL LİSTESİ ..... **x**
TABLO LİSTESİ ..... **xii**
SİMGE VE KISALTMA LİSTESİ ..... **xiii**
ÖZET ..... **xvi**
ABSTRACT ..... **xviii**
1. GİRİŞ **1**
2. KAVRAMSAL ÇERÇEVE **3**
2.1. ORGANİZE SANAYİ BÖLGELERİ VE ATIKSU YÖNETİMİ 3
2.2. KARIŞIK ENDÜSTRİYEL ATIKSU ANALİZİ ve KARAKTERİZASYONU 6
2.3. ATIKSU VERİLERİNDE İSTATİSTİKSEL DEĞERLENDİRME 9
2.3.1. Zamansal Ve Çok Değişkenli Verilerde İstatistiksel Analiz Yaklaşımları 12
2.3.2. Atıksu Verilerinde Kullanılan Korelasyon Denklemleri; Pearson, Spearman ve Kendall Yaklaşımları 13
2.4. ENDÜSTRİYEL ATIKSULARIN KARAKTERİZASYONUNDA YAPAY ZEKÂ ve MAKİNE ÖĞRENMESİ YAKLAŞIMLARI 20
2.4.1. Yapay Sinir Ağı (ANN) 23
2.4.2. Makine Öğrenmesi (ML) 24
2.4.3. Arama Algoritması (SA) 26
3. YÖNTEM **34**
3.1. ARAŞTIRMA AMAÇLARI VE ARAŞTIRMA SORULARI 34
3.2. MATERYAL ve METOT 35
3.2.1. Veri Setinin Oluşturulması 35
3.2.2. Veri Setinin Değerlendirilmesi 37

3.2.3. KOİ Hariç İhlal Etiketi ve Hedef Değişken Tanımı 38
3.2.4. Zamansal Özellik Çıkarımı 39
3.2.5. Özellik Seti ve Eksik Değerlerin İşlenmesi 41
3.3. EĞİTİM-TEST AYRIMI VE MODELLEME YAKLAŞIMI 42
3.3.1. Firma Bazlı Zaman Sıralı 80/20 Ayrımı 42
3.3.2. Modeller ve Hiperparametreler 42
3.3.3. Değerlendirme Metrikleri 43
3.3.4. Sektör Bazlı Modelleme ve Özellik Önemi Analizi 44
4. BULGULAR 47
4.1. GLOBAL MODEL PERFORMANSI (ARAŞTIRMA SORUSU 1) 47
4.2. GLOBAL ÖLÇEKTE LİMİT AŞIMI RİSKİNİ AÇIKLAYAN TEMEL PARAMETRELER (ARAŞTIRMA SORUSU 2) 48
4.2.1. İhlal Geçmişi: Kümülatif İhlal Oranı ve Lag Etiketler 50
4.2.2. AKM'ye İlişkin Kısa ve Uzun Dönemli Yük Göstergeleri 50
4.2.3. Yağ Gres, pH ve Toplam Azotun Uzun Dönemli Etkisi 51
4.2.4. Global Modelin Ortaya Koyduğu Genel Desen 52
4.3. SEKTÖR BAZLI MODELLER VE LİMİT AŞIMI RİSKİNİ AÇIKLAYAN TEMEL PARAMETRELERİN SEKTÖREL FARKLILAŞMASI (ARAŞTIRMA SORUSU 3) 53
4.3.1. Sektör Bazlı Veri Kapsamı ve Örneklem Büyüklükleri 53
4.3.2. Sektör Bazlı Random Forest Performansı 54
4.3.3. Sektör Bazında Öne Çıkan Parametreler 55
4.3.4. Ortak Desenler: Sektörler Arasında Değişmeyen Parametreler 60
4.3.5. Sektöre Özgü Ayırt Edici Parametreler 60
5. TARTIŞMA 65
5.1. ARAŞTIRMA SORULARININ TARTIŞILMASI 65
5.2. OSB İÇİN PRATİK VE YÖNETSEL ÇIKARIMLAR 70
5.3. SINIRLILIKLAR VE GELECEKTEKİ ÇALIŞMALAR 72
6. SONUÇ VE ÖNERİLER 74
6.1. BAŞLICA SONUÇLAR 74
6.2. UYGULAMAYA YÖNELİK ÖNERİLER 75
6.3. GELECEK ÇALIŞMALAR İÇİN ÖNERİLER 75
KAYNAKLAR 77
EKLER 81
İNTİHAL RAPORU İLK SAYFASI 82
ETİK KURUL İZİN YAZISI 83

KURUM İZNİ YAZILARI **84**
ÖZGEÇMİŞ **85**

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.1. Türkiye OSB'lerin Bölgesel Dağılım Haritası, Ekim 2024 3

Şekil 2.2. Atıksu Arıtma Tesisi Yerleşim Planı 4

Şekil 2.3. Sektörlerden kaynaklanan atıksu kirletici parametrelerinin atıksuya etkisi 7

Şekil 2.4. Klasik İstatistiksel Değerlendirme Sürecini Gösteren Kavramsal Şema 11

Şekil 2.5. Pearson Korelasyon Katsayısının Grafik Görünümleri 14

Şekil 2.6. Değişkenler arası ilişkilerde monotonik fonksiyonların davranışı 15

Şekil 2.7. Kendall's Tau korelasyon katsayısının farklı veri yapıları altındaki uygulanabilirliğini gösteren örnek grafikler. 17

Şekil 2.8. Yapay Zekânın çalışma prensibi ve tasarımı 21

Şekil 2.9. Atıksu Arıtımında kullanılan YZ Modellerinin Sınıflandırılması (Wang ve ark., 2023) 22

Şekil 2.10. ANN modellerinin temel tasarımı a) ANN, b) RNN, c) CNN, d) FNN, e) DNN *Kaynak: Warg ve ark., 2023* 23

Şekil 2.11. ML modellerinin Şematik Diyagramı a) PCA, b) DT, c) SVM, d) PSO 25

Şekil 2.12. ML modellerinin Şematik Diyagramı a) RF, b) SOM, c) KNN, d) ANFIS 26

Şekil 2.13. SA modellerinin Şematik Gösterimi a) GA, b) GP 27

Şekil 3.1. Organize Sanayi Bölgesi Yerleşim Planı 35

Şekil 3.2. Organize Sanayi Bölgesindeki Firmaların Sektörel Dağılımı 36

Şekil 3.3. Bu Çalışmada Uygulanan Veri Temelli Modelleme Sürecinin Genel Akış Şeması 45

Şekil 4.1. RF Modelinden Elde Edilen En Yüksek Önem Skoruna Sahip İlk 15 Özellik 49

Şekil 4.2. SKKY 15.2 Limitleri için En Önemli 10 Özellik (RF) 56

Şekil 4.3. SKKY 6.1 Limitleri için En Önemli 10 Özellik (RF) 57

Şekil 4.4. SKKY 14.4 Limitleri için En Önemli 10 Özellik (RF) 58

Şekil 4.5. SKKY 14.5 Limitleri için En Önemli 10 Özellik (RF) 59
Şekil 4.6. SKKY 14.10 Limitleri için En Önemli 10 Özellik (RF) 60
Şekil 4.7. Seçili Sektörler için En Önemli 10 Özelliğin Önem Isı Haritası 63

TABLO LİSTESİ

Sayfa No

Tablo 2.1. Karışık Endüstriyel Atıksuların SKKY Yönetmeliği Kapsamındaki Parametreleri ve Sınır Değerleri (SKKY – Tablo 19).....6

Tablo 2.2. Atıksu Verilerinde Pearson, Spearman ve Kendall Korelasyon Katsayılarının Karşılaştırılması.....18

Tablo 2.3. Atıksu arıtımında yaygın olarak kullanılan Yapay Zekâ Modellerinin Kullanım Amaçları, Avantajları ve Dezavantajları 28

Tablo 3.1. Karışık Endüstriyel Atıksuların analizi için Kullanılan Standart Yöntemler.....37

Tablo 4.1. ML modellerinin test seti üzerindeki performans metrikleri 48

Tablo 4.2. En Yüksek Önem Skoruna Sahip İlk 15 Özellik.....48

Tablo 4.2 (Devam).....49

Tablo 4.3. Sektör bazlı Random Forest modelleri için train/test örneklem büyüklükleri 54

Tablo 4.4. Sektör bazlı Random Forest modellerinin doğrulama performansı 54

Tablo 4.5. SKKY 15.2 Limitleri için en önemli 10 özellik (Random Forest).....55

Tablo 4.6. SKKY 6.1 Limitleri için en önemli 10 özellik (Random Forest).....56

Tablo 4.7. SKKY 14.4 Limitleri için en önemli 10 özellik (Random Forest).....57

Tablo 4.8. SKKY 14.5 Limitleri için en önemli 10 özellik (Random Forest).....58

Tablo 4.9. SKKY 14.10 Limitleri için en önemli 10 özellik (Random Forest).....59

SİMGE VE KISALTMA LİSTESİ

Simgeler **Açıklama**

Cd : Kadmiyum
Cl^- : Klorür
CN^- : Siyanür
Cr : Krom
Cr^{+6} : Krom +6
Cu : Bakır
F^- : Florür
Fe : Demir
Hg : Civa
P : Fosfor
Pb : Kurşun
PFAS : Per- ve Polifloroalkil Maddeler
SO_4 : Sülfat
Zn : Çinko

Kısaltmalar **Açıklama**

ABD : Amerika Birleşik Devletleri
AKM : Askıda Katı Madde
ANFIS : Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi
ANN : Yapay Sinir Ağları
AS1 : Araştırma Sorusu 1
AS2 : Araştırma Sorusu 2
AS3 : Araştırma Sorusu 3
BOİ : Biyokimyasal Oksijen İhtiyacı
CA : Kümelenme Analizi
CCD : Merkezi Kompozit Tasarım
CNN : Evrişimsel Sinir Ağı
ÇO : Çözünmüş Oksijen

CV : Bilgisayar Görüsü
DA : Ayrım Analizi
DL : Derin Öğrenme
DNN : Derin Sinir Ağı
DT : Karar Ağacı
EN : Avrupa Normu
EPA : Amerika Birleşik Devletleri Çevre Koruma Ajansı
ExT : Aşırı Rastgele Ağaçlar Algoritması
FA : Faktör Analizi
F-BOD : Nihai Biyokimyasal Oksijen İhtiyacı
FNN : Bulanık Sinir Ağı
GA : Genetik Algoritma
GFFR : Genelleştirilmiş Bulanık Fonksiyon Regresyonu
GP : Genetik Programlama
ISO : Uluslararası Standardizasyon Örgütü
KNN : K-En Yakın Komşu Algoritması
KOİ : Kimyasal Oksijen İhtiyacı
LSD : En Küçük Anlamlı Fark Testi
ML : Makine Öğrenmesi
MLP : Çok Katmanlı Algılayıcı
MLP-PCA : Temel Bileşen Analizi ile Desteklenmiş Çok Katmanlı Algılayıcı
NLP : Doğal Dil İşleme
NN : Sinir Ağları
OSB : Organize Sanayi Bölgesi
PCA : Temel Bileşen Analizi
PR-AUC : Kesinlik–Duyarlılık Eğrisi Altında Kalan Alan
PSO : Parçacık Sürü Optimizasyonu
RBF : Radyal Tabanlı Fonksiyon Ağı
RF : Random Forest Algoritması
RNN : Tekrarlayan Sinir Ağı
ROC-AUC : Alıcı İşletim Karakteristiği Eğrisi Altında Kalan Alan
SA : Arama Algoritması
SHAP : Shapley Katkı Açıklamaları
SKKY : Su Kirliliği Kontrolü Yönetmeliği

SM : Standart Metot
SOM : Kendini Örgütleyen Harita
SVM : Destek Vektör Makineleri
TKN : Toplam Kjeldahl Azotu
TN : Toplam Azot
TP : Toplam Fosfor
TS : Toplam Katı Madde
UV-EP : UV-Elektro-Peroksone Prosesi
YZ : Yapay Zekâ
ZSF : Zehirlilik Katsayısı (Balıkbiyodeneyi)

ÖZET

YÜKSEK LİSANS TEZİ

KARIŞIK ENDÜSTRİYEL ATIKSU KARAKTERİZASYONUNUN ZAMANSAL DEĞİŞİMİNİN BELİRLENMESİ ve KİRLETİCİLERİN İSTATİSTİKSEL DEĞERLENDİRİLMESİ

Yasemin ÇELİK

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Çevre Mühendisliği Anabilim Dalı

Çevre Mühendisliği, Tezli Yüksek Lisans Programı

Danışman : Prof. Dr. Serdar AYDIN

II. Danışman : Doç. Dr. Elif TÜZÜN

Bu tez çalışmasında, Organize Sanayi Bölgeleri (OSB) bünyesinde oluşan karışık endüstriyel atıksuların zamansal ve çok değişkenli yapısı, çok yıllı izleme verileri kullanılarak analiz edilmiştir. Farklı sektörlerden kaynaklanan deşarjların ortak altyapı sistemlerinde birleşmesi sonucunda; KOİ, AKM, yağ-gres, besin elementleri, ağır metaller, renk ve pH gibi parametrelerin yüksek değişkenlik, doğrusal olmayan ilişkiler ve belirgin zamansal dalgalanmalar sergilediği belirlenmiştir. Çalışmada öncelikle tanımlayıcı istatistikler ile veri setinin dağılım özellikleri ve aykırı değer yapısı ortaya konulmuş; ardından Pearson, Spearman ve Kendall korelasyon analizleri ile parametreler arasındaki temel ilişkiler değerlendirilmiştir. Bu analizler, karışık endüstriyel atıksuların yalnızca klasik istatistiksel yöntemlerle tam olarak temsil edilemediğini göstermiştir.

Bu kapsamda, çalışmanın ana bulguları Random Forest makine öğrenmesi modeli kullanılarak elde edilmiştir. Model, doğrusal olmayan ilişkileri ve çok değişkenli etkileşimleri başarıyla yakalayarak, atıksu karakterinin tahmininde yüksek doğruluk sağlamıştır. Random Forest'ın sağladığı değişken önem analizi, bazı parametrelerin (özellikle organik yük ve askıda katı

madde ile ilişkili göstergelerin) atıksu karakteri üzerinde baskın rol oynadığını açık biçimde ortaya koymuş; böylece modelin yalnızca tahmin gücü değil, açıklanabilirliği de güçlendirilmiştir. Elde edilen sonuçlar, çevresel risklerin tekil parametre aşımlarından ziyade parametrelerin birlikte oluşturduğu etkileşimli yapı ile ilişkili olduğunu göstermektedir. Bu tez, istatistiksel analizler ile Random Forest gibi açıklanabilir makine öğrenmesi modellerinin birlikte kullanılmasının, farklı endüstri türlerinden kaynaklanan karışık atıksuların değişken yapısının anlaşılmasına katkı sağlayarak, sürdürülebilir endüstriyel su yönetimi politikalarının geliştirilmesine destek olmayı amaçlamaktadır.

Şubat 2026 , 104 sayfa.

Anahtar kelimeler: Karışık Endüstriyel Atıksu, Çok değişkenli istatistiksel analiz, Makine öğrenmesi, Korelasyon

ABSTRACT

M.Sc. THESIS

DETERMINATION of TEMPORAL VARIATION IN MIXED INDUSTRIAL WASTEWATER CHARACTERIZATION AND STATISTICAL ASSESSMENT OF POLLUTANTS

Yasemin ÇELİK

İstanbul University-Cerrahpaşa

Institute of Graduate Studies

Department of Environmental Engineering

Environmental Engineering, Master of Science Programme with Thesis

Supervisor : Prof. Dr. Serdar AYDIN

Co-Supervisor: Assoc. Prof. Dr.Elif TÜZÜN

In this thesis, the temporal and multivariate structure of mixed industrial wastewater generated within Organized Industrial Zones was analyzed using long-term monitoring data. As a result of discharges originating from different industrial sectors being combined within a shared infrastructure system, parameters such as chemical oxygen demand (COD), suspended solids (SS), oil and grease, nutrients, heavy metals, color, and pH were found to exhibit high variability, nonlinear relationships, and pronounced temporal fluctuations. The study first employed descriptive statistical analyses to characterize the distributional properties of the dataset and identify outlier structures; subsequently, Pearson, Spearman, and Kendall correlation analyses were applied to evaluate the fundamental relationships among parameters. These analyses demonstrated that mixed industrial wastewater cannot be fully characterized using classical statistical methods alone.

Within this framework, the main findings of the study were obtained using the Random Forest machine learning model. The model successfully captured nonlinear relationships and multivariate interactions, providing high predictive accuracy in representing wastewater characteristics. Variable importance analysis derived from the Random Forest model clearly revealed that certain parameters—particularly indicators associated with organic load and suspended solids—play a dominant role in defining wastewater behavior, thereby enhancing not only the model's predictive performance but also its interpretability. The results indicate that environmental risks are more strongly associated with the interactive behavior of multiple parameters rather than isolated exceedances of individual parameters. This thesis aims to support the development of sustainable industrial water management policies by contributing to a better understanding of the variable structure of mixed industrial wastewater originating from diverse industrial sectors through the integrated use of statistical analyses and explainable machine learning models such as Random Forest.

February 2026, 104 pages.

Keywords: Mixed industrial wastewater, Multivariate statistical analysis, Machine learning, Random forest, Correlation

1. GİRİŞ

Sanayileşmenin hızlanmasıyla birlikte endüstriyel faaliyetlerden kaynaklanan çevresel baskılar, özellikle su kaynakları üzerinde belirgin etkiler oluşturmaktadır. Bu etkilerin kontrol altına alınabilmesi amacıyla geliştirilen Organize Sanayi Bölgeleri (OSB), sanayi üretiminin planlı bir çerçevede yürütülmesini, ortak altyapı olanaklarının kullanılmasını ve çevresel etkilerin merkezi bir yönetim anlayışıyla denetlenmesini hedefleyen yapılar olarak öne çıkmaktadır. OSB bünyesinde faaliyet gösteren farklı sektörlerin ortak altyapı sistemlerine bağlı olması, özellikle atıksu yönetimini çevresel sürdürülebilirlik açısından kritik bir karar alanı hâline getirmektedir.

OSB'lerde oluşan atıksular, farklı üretim süreçlerinden kaynaklanan kirleticilerin aynı sistem içerisinde birleşmesi sonucu, fiziksel ve kimyasal açıdan yüksek değişkenlik gösteren karışık endüstriyel atıksu karakteri sergilemektedir. Bu tür atıksular; organik madde yükleri, askıda katı maddeler, besin elementleri, ağır metaller ve renk gibi çok sayıda parametreyi aynı anda içermekte ve zamansal olarak önemli dalgalanmalar gösterebilmektedir. Dolayısıyla, karışık endüstriyel atıksuların yalnızca anlık ölçüm sonuçları veya tekil parametreler üzerinden değerlendirilmesi, gerçek çevresel risklerin ortaya konulmasında çoğu zaman yetersiz kalmaktadır.

Mevcut uygulamalarda, atıksu yönetimi büyük ölçüde mevzuat sınır değerlerine uyumun izlenmesine dayanmaktadır. Ancak bu yaklaşım, parametreler arasındaki etkileşimleri, zamana bağlı değişim eğilimlerini ve olası risklerin erken aşamada tespit edilmesini sınırlı düzeyde ele alabilmektedir. Oysa karışık endüstriyel atıksuların çevresel etkisi, çoğunlukla kirletici parametrelerin birlikte oluşturduğu kümülatif yük ve bu parametrelerin zamansal davranışları üzerinden şekillenmektedir. Bu durum, atıksu verilerinin daha bütüncül ve veri temelli yaklaşımlarla değerlendirilmesini gerekli kılmaktadır.

Bu bağlamda, istatistiksel analizler endüstriyel atıksu verilerinin yorumlanmasında temel bir araç olarak öne çıkmaktadır. Tanımlayıcı istatistikler, verilerin dağılım yapısını ve değişkenlik düzeyini ortaya koyarken; korelasyon ve zamansal analizler, parametreler arasındaki ilişkilerin ve eğilimlerin anlaşılmasına olanak sağlamaktadır. Bununla birlikte, karışık endüstriyel atıksu

verilerinin çoğunlukla doğrusal olmayan, aykırı değerler içeren ve zamansal bağımlılık gösteren bir yapıya sahip olması, klasik istatistiksel yaklaşımların tek başına yeterli olmamasına neden olmaktadır (Bahna, 2009).

Son yıllarda yapay zekâ ve makine öğrenmesi temelli yöntemler, çok değişkenli ve karmaşık çevresel veri setlerinin analizinde önemli bir alternatif ve tamamlayıcı yaklaşım olarak kullanılmaya başlanmıştır (Wang ve ark., 2023). Bu yöntemler, büyük hacimli izleme verilerinden örüntülerin ortaya çıkarılması, kirletici davranışlarının tahmin edilmesi ve risk temelli yönetim stratejilerinin geliştirilmesi açısından önemli avantajlar sunmaktadır. Özellikle OSB ölçeğinde uzun dönemli izleme verilerinin bulunması, bu tür veri odaklı yaklaşımların uygulanabilirliğini artırmaktadır.

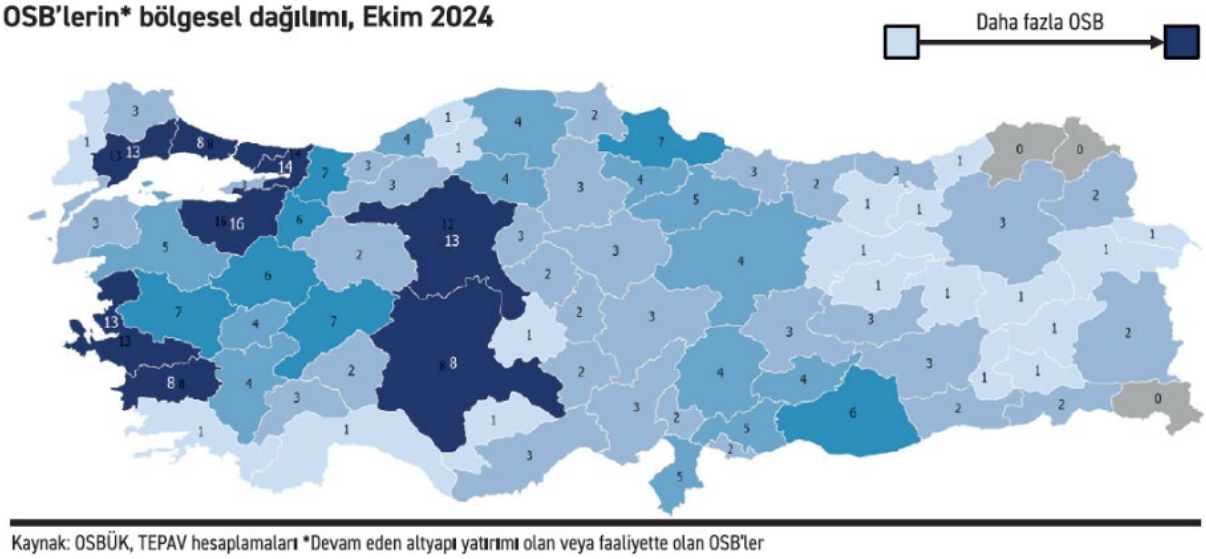
Bu tez çalışması kapsamında, Organize Sanayi Bölgeleri kaynaklı karışık endüstriyel atıksuların karakterizasyonu; mevzuat çerçevesi, istatistiksel değerlendirme yöntemleri ve yapay zekâ destekli analiz yaklaşımları birlikte ele alınmıştır. Çalışmanın temel amacı, uzun dönemli atıksu izleme verileri üzerinden parametrelerin zamansal ve çok değişkenli davranışlarını ortaya koymak, klasik ve ileri analiz yöntemlerinin birlikte kullanımını değerlendirmek ve elde edilen bulgular ışığında daha etkin ve öngörüye dayalı atıksu yönetim stratejilerine katkı sağlamaktır.

2. KAVRAMSAL ÇERÇEVE

2.1. ORGANİZE SANAYİ BÖLGELERİ VE ATIKSU YÖNETİMİ

Organize Sanayi Bölgeleri (OSB), sanayi faaliyetlerinin plansız ve dağınık biçimde gelişmesinin çevresel, ekonomik ve mekânsal açıdan çeşitli sorunlara yol açması sonucunda ortaya çıkan, planlı sanayi yerleşim alanlarıdır. Sanayi Devrimi sonrasında özellikle büyük şehirlerde kontrolsüz biçimde konumlanan sanayi tesisleri; yetersiz altyapı, artan çevre kirliliği, iş sağlığı ve güvenliği açısından riskler meydana getirmiştir. Bu olumsuzluklar, sanayi faaliyetlerinin belirli bölgelerde toplanmasını ve söz konusu alanların merkezi bir planlama yaklaşımıyla yönetilmesini zorunlu kılmıştır.

Bu kapsamda OSB kavramı, sanayi faaliyetlerinin belirli bir plan dâhilinde, ortak altyapı olanaklarından yararlanarak ve çevresel etkileri kontrol altında tutacak şekilde yürütülmesini amaçlayan bir yaklaşım olarak gelişmiştir. Dünya genelinde ilk uygulamalar, 20. yüzyılın başlarında ABD ve Batı Avrupa ülkelerinde "industrial park" veya "industrial estate" modelleri ile karşımıza çıkmaktadır (Özden, 2016).



Şekil 2.1. Türkiye OSB'lerin Bölgesel Dağılım Haritası, Ekim 2024

Kaynak: Güneş Sak, Ekonomi Gazetesi Köşe yazısından alınmıştır.

Türkiye'de Organize Sanayi Bölgeleri uygulamalarının temelleri, 1960'lı yıllarda yürürlüğe giren planlı kalkınma politikaları çerçevesinde başlamıştır. Bu dönemde OSB'ler ile; sanayinin ülke genelinde dengeli dağılımını sağlamak, şehir merkezlerindeki çevresel baskıyı azaltmak ve küçük ve orta ölçekli sanayi kuruluşlarını desteklemek planlanmıştır (Özden, 2016). Zaman içerisinde OSB'ler, yalnızca üretim alanları olmayıp; merkezi atıksu arıtma tesisleri, çevre yönetim birimleri ve ortak denetim mekanizmaları ile çevresel sürdürülebilirliğin sağlandığı sanayi alanları hâline gelmiştir. Şekil 2.1.' de Türkiye'de kurulu OSB sayılarının bölgesel dağılımı sunulmaktadır. Bu haritaya göre OSB sayıları; Marmara, Ege ve İç Anadolu Bölgelerinde yoğunlaştığı görülmektedir. Renk yoğunluğu, iller bazında OSB sayısının arttığını göstermektedir.

OSB'lerin hukuki ve kurumsal yapısı, 2000 yılında yürürlüğe giren 4562 sayılı Organize Sanayi Bölgeleri Kanunu ile belirlenmiştir. Bu Kanun ile OSB'lerin kuruluş, yönetim, işletme ve çevresel yükümlülükleri açık şekilde belirlenmekte; altyapı hizmetlerinin planlanması ve çevre mevzuatına uyum yükümlülükleri yasal çerçevesi tanımlanmaktadır (Özden, 2016). Böylece OSB'ler, sanayi faaliyetlerinin çevreye olan etkilerinin kontrol altına alınmasında önemli bir yapıyı oluşturmaktadır.

OSB bünyesinde faaliyet gösteren sanayi tesislerinden kaynaklanan atıksular, OSB'lerin kuruluş amaçları doğrultusunda planlanan ve Şekil 2.2.'de benzer yerleşim planı verilen merkezi atıksu arıtma tesislerinde arıtılarak alıcı ortama deşarj edilmektedir.



Şekil 2.2. Atıksu Arıtma Tesisi Yerleşim Planı

OSB içerisinde oluşturulan kanalizasyon sistemi aracılığıyla, farklı sektörlerden kaynaklanan ve bileşimleri bakımından yüksek çeşitlilik gösteren atıksular tek bir altyapı altında toplanmaktadır. Bu durum, merkezi atıksu arıtma tesislerinde çok sayıda kirletici türünü içeren karışık endüstriyel atıksuların yönetimini gerekli kılmakta; arıtma proseslerinin tasarımı,

işletilmesi ve izlenmesi açısından bütüncül yaklaşımların uygulanmasını zorunlu hale getirmektedir. Bu bağlamda, atıksu yönetimi hem çevresel koruma hem de sanayi sürekliliği açısından stratejik bir karar alanını oluşturmaktadır.

Endüstriyel üretim süreçlerinin çeşitliliği; atıksuların karakteristik yapısının sektöre, üretim tekniğine ve işletme koşullarına bağlı olarak farklılaşmasına neden olmaktadır. Bu durum, özellikle karışık endüstriyel atıksuların yönetimini zorlaştırmakta ve tek tip değerlendirme yaklaşımlarının yetersiz kalmasına yol açmaktadır. Bu nedenle, endüstriyel atıksuların çevresel etkilerinin etkin biçimde kontrol altına alınabilmesi için yalnızca anlık ölçümlere dayalı değerlendirmeler yerine, daha kapsamlı ve bütüncül yönetim stratejilerine ihtiyaç duyulmaktadır.

Klasik atıksu yönetimi uygulamaları çoğunlukla geriye dönük değerlendirmelere ve mevzuat limitlerinin aşılıp aşılmadığının kontrolüne dayanmaktadır. Ancak bu yaklaşım, ihlallerin nedenlerini ve tekrar etme eğilimlerini ortaya koymada sınırlı kalmaktadır. Özellikle zamansal değişkenlik gösteren atıksu parametreleri için anlık ölçümlere dayalı kararlar, çevresel risklerin erken aşamada tespit edilmesini güçleştirmektedir.

Mevzuat ve izleme açısından değerlendirildiğinde, karışık endüstriyel atıksuların yönetimi ilave zorluklar içermektedir. Farklı sektörlere özgü deşarj limitlerinin aynı izleme sisteminde uygulanması, değerlendirme sürecinde belirsizlikler doğurabilmektedir. Bazı parametrelerde sık ve tekrarlayan limit aşımları gözlenirken, diğer parametrelerin uzun süre mevzuat sınırları içerisinde kalması, çevresel riskin eksik veya hatalı algılanmasına neden olabilmektedir. Oysa çevresel etki çoğu zaman, parametrelerin birlikte oluşturduğu toplam yük ve etkileşim üzerinden şekillenmektedir. Türkiye'de karışık endüstriyel atıksuların yönetiminde SKKY Tablo 19 ile verilen yasal limitler ve parametreler ile atıksu karakterizasyonu belirlenmektedir. Tablo 2.1.' de SKKY Tablo 19 kapsamındaki parametreler ve limit değerlere yer verilmektedir.

Endüstriyel atıksuların etkin yönetimi; mevzuata uyumun yanı sıra alıcı ortamların korunması, arıtma tesislerinin verimli işletilmesi ve uzun vadeli çevresel risklerin azaltılmasını da kapsamaktadır. Bu bağlamda, düzenli izleme verilerinin sistematik olarak değerlendirilmesi ve kirletici parametrelerin zaman içindeki davranışlarının analiz edilmesi, proaktif ve risk temelli yönetim yaklaşımlarının geliştirilmesi açısından büyük önem taşımaktadır. Özellikle çok yıllı izleme verilerine dayalı analizler, endüstriyel atıksuların değişken yapısının ortaya konulmasına ve daha etkin yönetim stratejilerinin oluşturulmasına olanak sağlamaktadır.

Tablo 2.1. Karışık Endüstriyel Atıksuların SKKY Yönetmeliği Kapsamındaki Parametreleri ve Sınır Değerleri (SKKY – Tablo 19)

TABLO 19: KARIŞIK ENDÜSTRİYEL ATIKSULARIN ALICI ORTAMA DEŞARJ STANDARTLARI (KÜÇÜK VE BÜYÜK ORGANİZE SANAYİ BÖLGELERİ VE SEKTÖR BELİRLEMESİ YAPILAMAYAN DİĞER SANAYİLER)

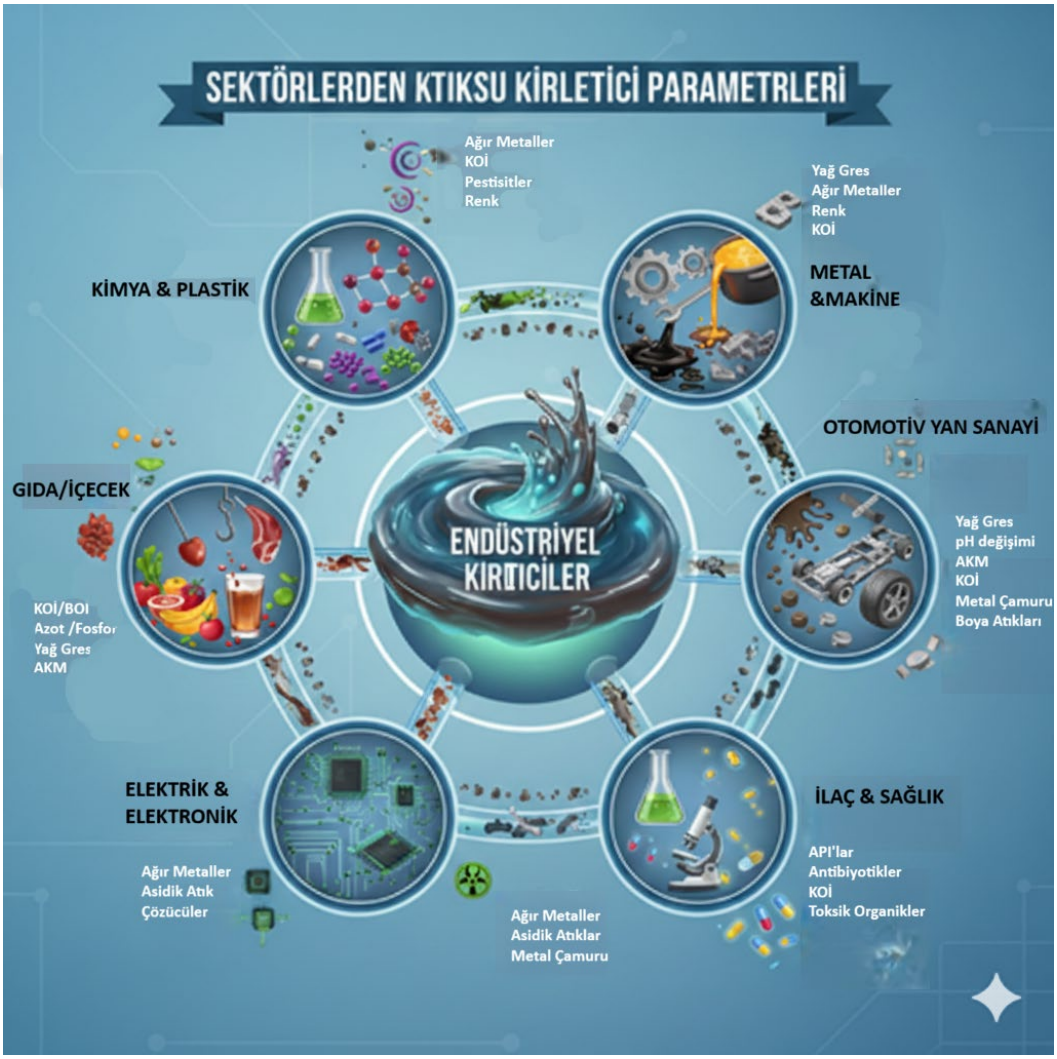
PARAMETRE	BİRİM	2 SAATLİK KOMPOZİT NUMUNE
Kimyasal Oksijen İhtiyacı (KOİ)	(mg/L)	250
Askıda Katı Madde (AKM)	(mg/L)	200
Yağ ve Gres	(mg/L)	20
Toplam Fosfor (P)	(mg/L)	2
Toplam Krom	(mg/L)	2
Krom (Cr^{+6})	(mg/L)	0.5
Kurşun (Pb)	(mg/L)	2
Toplam Siyanür (CN^{-})	(mg/L)	1
Kadmiyum (Cd)	(mg/L)	0.1
Demir (Fe)	(mg/L)	10
Florür (F^{-})	(mg/L)	15
Bakır (Cu)	(mg/L)	3
Çinko (Zn)	(mg/L)	5
Civa (Hg)	(mg/L)	0.05
Sülfat (SO_4^{-2})	(mg/L)	1500
Toplam Kjeldahl Azotu (TKN)	(mg/L)	20
Balık Biyodeneyi (ZSF)	-	10
Renk	(Pt-Co)	280
pH	-	6-9

2.2. KARIŞIK ENDÜSTRİYEL ATIKSU ANALİZİ ve KARAKTERİZASYONU

Karışık endüstriyel atıksular, farklı üretim sektörlerine ait proseslerden kaynaklanan ve fiziksel, kimyasal ve toksikolojik açıdan yüksek derecede değişkenlik gösteren atıksulardır. Bu atıksuların karakterizasyonu; kullanılan hammadde türleri, uygulanan proses teknolojileri, işletme koşulları ve üretim yoğunluğu gibi birçok etkene bağlı olarak değişkenlik göstermektedir. Bu nedenle, çevresel açıdan karmaşık bir yapıya sahip olan karışık endüstriyel atıksuların yönetiminde, tek bir sektöre özgü değerlendirme yaklaşımları yeterli olmamakta; daha kapsamlı analiz ve karakterizasyon çalışmalarına ihtiyaç duyulmaktadır.

Karışık endüstriyel atıksular; metal işleme, kimya, tekstil, gıda, boya, ilaç, kozmetik ve elektronik gibi farklı sektörlere ait deşarjların birleşmesiyle oluşmaktadır. Her sektörün kendine

özgü prosesleri ve kimyasal girdileri bulunması, atıksuların içerik ve konsantrasyon bakımından farklılaşmasına neden olmaktadır. Bu nedenle karışık endüstriyel atıksular, tek bir üretim prosesinin çıktısı olarak değil, çok sayıda prosesin eş zamanlı etkisiyle oluşan bileşik bir sistem olarak değerlendirilmekte ve analiz ile karakterizasyon çalışmaları bu çoklu yapıya göre yürütülmektedir. Şekil 2.3, farklı endüstriyel sektörlerden kaynaklanan atıksuların içerdiği başlıca kirletici parametrelerin karışarak karışık endüstriyel atıksu karakterinin oluşumunu sunmaktadır.



Şekil 2.3. Sektörlerden kaynaklanan atıksu kirletici parametrelerinin atıksuya etkisi

Atıksu karakterizasyonunda pH, askıda katı madde (AKM), kimyasal oksijen ihtiyacı (KOİ), yağ-gres, toplam azot ve fosfor, ağır metaller ve renk gibi parametreler temel göstergeler olarak değerlendirilmektedir. Bu parametreler yalnızca atıksuyun anlık kirlilik yükünü değil, aynı zamanda proses stabilitesini ve arıtılabilirlik potansiyelini de yansıtmaktadır. Karışık

endüstriyel atıksularda söz konusu parametrelerin geniş konsantrasyon aralıklarında ve zamana bağlı dalgalanmalar göstermesi, arıtma tesislerinin işletilmesi ve mevzuat uyumunun sağlanmasını zorlaştıran unsurlar olarak öne çıkmaktadır.

Karışık endüstriyel atıksuların analiz ve karakterizasyonu yalnızca mevzuat limitlerinin kontrolü amacıyla değil; kirletici kaynaklarının belirlenmesi, parametreler arasındaki ilişkilerin anlaşılması ve zaman içerisindeki değişimlerin izlenmesi ile yönetim stratejilerinin geliştirilmesi açısından kritik bir araçtır. Bu tür risk temelli değerlendirme, endüstriyel atıksu yönetiminde olası risklerin önceden öngörülmesine, arıtma süreçlerinin daha etkin planlanmasına ve çevresel etkilerin azaltılmasına önemli katkılar sağlamaktadır.

Endüstiriyel atıksuların karakterizasyonunu zorlaştıran bir unsur da, parametreler arasındaki karmaşık ve doğrusal olmayan etkileşimlerdir. Askıda katı madde konsantrasyonundaki artış, çoğu zaman yağ ve gres, renk veya bazı metal konsantrasyonlarında eş zamanlı değişimlere yol açabilmektedir (Long ve ark., 2018; Singh ve ark., 2004). Benzer şekilde pH değerindeki dalgalanmalar, metal iyonlarının çözünürlüğünü ve taşınımını etkileyerek atıksu kalitesini dolaylı yoldan değiştirebilmektedir. KOİ, atıksudaki toplam oksitlenebilir madde miktarını temsil eden bir parametre olup, birçok kirleticinin dolaylı göstergersi olarak değerlendirilmektedir (Nadiri ve ark., 2018). Yağ ve gres, yüzey aktif maddeler ve çözünmüş organik bileşikler KOİ konsantrasyonunun artmasına neden olmaktadır. Bu nedenle yağ ve gres konsantrasyonu ile KOİ arasında genellikle pozitif yönlü bir ilişki bulunmaktadır (Long ve ark., 2018). Toplam azot ve toplam fosfor parametreleri, özellikle gıda, kimya ve bazı metal yüzey işlemleri gibi sektörlerin katkı yaptığı karışık endüstriyel atıksularda öne çıkmaktadır. Çoğu zaman AKM ve organik yük konsantrasyonları ile bu iki parametre etkileşim halindedir (Giwa ve Hasan, 2017; Singh ve ark., 2004). Renk parametresi ise, özellikle tekstil, boya ve kimya sektörlerinden kaynaklanan atıksuların karışımında bir gösterge olmaktadır. Renk parametresinin de AKM ve KOİ ile pozitif yönlü bir ilişkisi bulunmaktadır (Katsoyiannis ve Samara, 2007). Bu ilişkiler çoğunlukla doğrusal bir yapı sergilemediğinden, parametrelerin birbirinden bağımsız olarak değerlendirilmesi gerçekçi sonuçlar vermemektedir.

Endüstriyel üretim süreçleri dinamik bir yapıya sahip olduğundan, atıksu bileşimi de zaman içerisinde sürekli olarak değişim göstermektedir. Bu bağlamda zamansal değişkenlik, endüstriyel atıksuların çevresel etkilerinin doğru şekilde izlenmesi ve yönetilmesinde temel bir araç olarak öne çıkmaktadır.

Zamansal değişkenlik, atıksu kalitesinin belirli bir periyot boyunca sabit kalmaması; aksine üretim koşulları, işletme uygulamaları ve dış etkenler doğrultusunda farklılaşması anlamına gelmektedir. Aynı tesisten farklı zamanlarda alınan numunelerde, kirletici parametrelerin hem konsantrasyon düzeyleri hem de birbirleriyle olan ilişkileri değişebilmektedir.

Zamansal değişkenliğin bir diğer boyutu, mevsimsel etkiler ile ilişkilidir. Özellikle açık alan depolama, soğutma suyu kullanımı veya hammadde temininin mevsimsel olduğu sektörlerde, atıksu bileşimi yılın farklı dönemlerinde belirgin farklılıklar gösterebilmektedir. Yağış rejimi, ortam sıcaklığı ve su kullanım alışkanlıklarındaki değişimler, atıksu parametrelerinin seyrelmesi veya yoğunlaşması üzerinde dolaylı etkilere sahip olabilmektedir.

Endüstriyel atıksularda zamansal değişkenlik yalnızca kısa vadeli dalgalanmalarla sınırlı değildir. Uzun vadede, proses değişiklikleri, teknoloji yatırımları, hammadde ikameleri ve çevresel mevzuata uyum kapsamında yapılan iyileştirmeler, atıksu karakterinin kademeli olarak dönüşmesine neden olabilmektedir. Bu tür yapısal değişimler, kirletici yüklerinde kalıcı artış veya azalışlar şeklinde ortaya çıkmaktadır.

Endüstriyel atıksularda zamansal değişkenlik kavramı; üretim dinamikleri, işletme uygulamaları, mevsimsel etkiler ve uzun dönemli yapısal değişimler ile birlikte ele alınmalıdır. Zamana bağlı veri setleri, kirletici parametrelerin davranış biçimlerini, artış ve azalış eğilimlerini ve olası risk dönemlerini ortaya koyma potansiyeline sahiptir. Bu yaklaşım, atıksu yönetiminde reaktif uygulamalardan ziyade proaktif ve öngörüye dayalı stratejilerin geliştirilmesine olanak tanımaktadır.

Karışık endüstriyel atıksuların tanımı ve karakterizasyonu; parametre çeşitliliği, zamansal değişkenlik, doğrusal olmayan etkileşimler ve kümülatif kirletici yük potansiyeli çerçevesinde ele alınmalıdır. Bu özellikler, söz konusu atıksuların karakterizasyonunda klasik değerlendirme yöntemlerinin sınırlarını ortaya koymakta ve veri temelli istatistiksel yaklaşımların önemini artırmaktadır.

2.3. ATIKSU VERİLERİNDE İSTATİSTİKSEL DEĞERLENDİRME

Endüstriyel atıksu izleme çalışmalarında elde edilen analiz sonuçları, atıksu kaynaklı çevre kirliliğinin değerlendirilmesinde temel veri kaynağını oluşturmaktadır. Ancak bu analizler, tek başına değerlendirildiğinde çoğu zaman sınırlı bir bilgi sunmakta; atıksuyun gerçek karakterini,

davranış biçimlerini ve potansiyel çevresel risklerini ortaya koymakta yetersiz kalmaktadır. Bu noktada istatistiksel değerlendirme, atıksu verilerinin anlamlandırılmasında ve değerlendirilebilir hale getirilmesinde kullanılarak izleme çalışmalarına katkı sunan bir araç olmaktadır.

Bir veri setinde istatistiksel değerlendirmenin temel amacı, elde edilen sayısal değerleri yalnızca bir sayı olarak ele almak yerine, bu değerlerin arkasındaki dağılımı, değişkenlik seviyesini ve birbirleri ile olan ilişki özelliklerini ortaya çıkarmaktır. Zamana bağlı değişkenlik gösteren verilerde istatistiksel analizler, bu değişkenliğin rastgele mi sistematik bir durum mu olduğunu ayırt etmeyi sağlamaktadır (Miran, 2021).

Tanımlayıcı istatistiksel değerlendirme, araştırma kapsamında elde edilen ham verilerin düzenlenmesi, özetlenmesi ve anlaşılır hâle getirilmesi açısından temel bir aşamadır. Bahna (2009)' a göre istatistik bilimi, tanımlayıcı ve çıkarımsal olmak üzere iki ana bileşene ayrılmakta; tanımlayıcı istatistiklerin temel amacının, verilerin merkezî eğilimini, dağılım yapısını ve değişkenliğini ortaya koymak olduğu vurgulamaktadır.

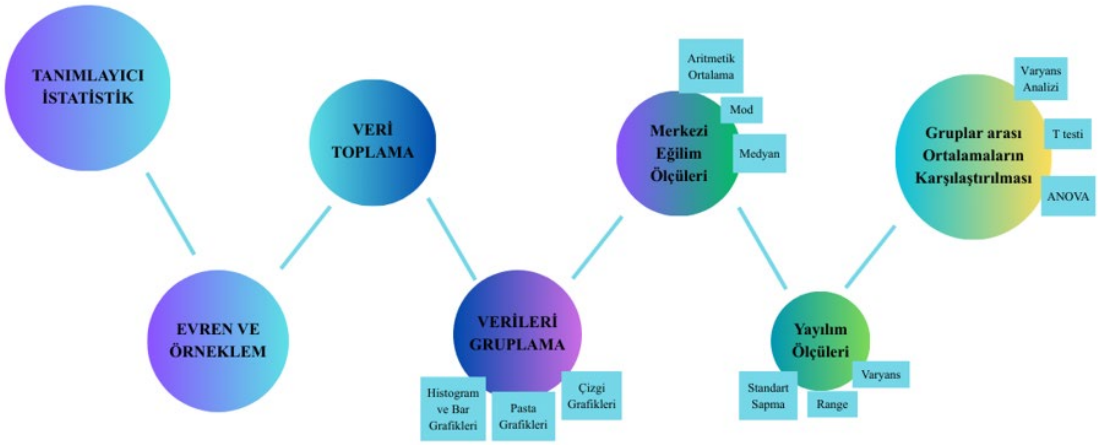
Tanımlayıcı istatistikler, herhangi bir hipotez testi yapılmadan önce veri setinin genel karakterini anlamaya imkân tanır. Ortalama, medyan, mod gibi merkezî eğilim ölçütleri, ölçümlerin tipik değerlerini yansıtırken; varyans, standart sapma, aralık ve yüzdelikler gibi yayılım ölçütleri, verilerin ne derece değişken olduğunu göstermektedir (Miran, 2021; Hanrahan, 2013; Bahna, 2009)

Bahna (2009), çarpık dağılıma sahip veya aykırı değer içeren veri setlerinde medyanın, aritmetik ortalamaya kıyasla daha temsil edici bir ölçüt olabileceğini açıklamaktadır.

Tanımlayıcı istatistiklerin bir diğer kritik işlevi, verilerin dağılım türünün belirlenmesidir. Normal veya normalden sapmış dağılımların tespiti, kullanılacak istatistiksel testlerin (parametrik veya parametrik olmayan) doğru seçilmesi açısından belirleyici rol oynamaktadır. Bu yönüyle tanımlayıcı istatistikler, yalnızca bir özetleme aracı değil, aynı zamanda analitik stratejinin şekillendirilmesini sağlayan bir ön değerlendirme mekanizması olarak da işlev görmektedir (Bahna, 2009).

Literatürde, tanımlayıcı istatistiklerin tablo ve grafikler aracılığıyla sunulmasının; veri kalitesinin kontrol edilmesi, olası ölçüm hatalarının veya aykırı değerlerin fark edilmesi ve

veriler arasındaki örüntülerin görsel olarak değerlendirilmesi için gerekli olduğu belirtilmektedir. Bu durum, tanımlayıcı istatistiklerin hem araştırmacı hem de okuyucu için verinin güvenilirliğini ve yorumlanabilirliğini artıran bir araç olduğunu göstermektedir (Bahna, 2009).



Şekil 2.4. Klasik İstatistiksel Değerlendirme Sürecini Gösteren Kavramsal Şema

Şekil 2.4'de Tanımlayıcı istatistiksel analiz sürecinin aşamalarını gösteren şematik yapı sunulmaktadır. Şekil, veri toplama aşamasından başlayarak evren ve örneklem kavramlarının tanımlanmasını, verilerin gruplandırılması ve grafiksel yöntemlerle görselleştirilmesini, merkezi eğilim ve yayılım ölçüleri aracılığıyla veri dağılımının karakterize edilmesini ve son olarak gruplar arası ortalamaların karşılaştırılmasına yönelik istatistiksel testleri kapsamaktadır.

Özellikle karışık endüstriyel atıksular söz konusu olduğunda, ortalama, medyan, minimum–maksimum değerler ve standart sapma gibi göstergeler farklı sektörlerden kaynaklanan kirletici yüklerin genel profilini ortaya koymakta ve veri setinin sistematik biçimde incelenmesine katkı sağlamaktadır. Bununla birlikte, yalnızca ortalama değerlere odaklanılması, atıksu karakterinin önemli yönlerinin gözden kaçmasına neden olabilmektedir (Hanrahan, 2013).

İstatistiksel değerlendirme, endüstriyel atıksu verilerinin yorumlanmasında temel bir aşama olup, parametreler arasındaki ilişkilerin nicel olarak ortaya konulmasını sağlar. Atıksularda ölçülen birçok parametre, benzer üretim süreçleri ve ortak kirletici kaynakları nedeniyle birlikte değişim gösterebilmekte; bu durum, atıksu karakterinin yalnızca tekil parametreler üzerinden değerlendirilmesini yetersiz kılmaktadır. Korelasyon analizleri gibi istatistiksel yöntemler, bu

birlikte değişim davranışlarını belirleyerek kirletici kaynakların ve proses temelli etkileşimlerin anlaşılmasına katkı sunmaktadır (Helsel & Hirsch, 2002).

İstatistiksel değerlendirmenin bir diğer önemli boyutu, atıksu verilerindeki zamansal değişimlerin incelenmesidir. Uzun dönemli izleme verileri, atıksu karakterinin zaman içerisinde nasıl değiştiğini ortaya koymakta; belirli parametrelerde gözlenen artış veya azalış eğilimleri, üretim koşullarındaki değişimler ya da uygulanan iyileştirme çalışmalarının bir yansıması olabilmektedir. Bu tür zamansal eğilimlerin belirlenmesi, yalnızca mevcut durumun değerlendirilmesi açısından değil, geleceğe yönelik risklerin öngörülmesi bakımından da önem taşımaktadır (Helsel & Hirsch, 2002).

Sonuç olarak tanımlayıcı istatistiksel değerlendirme, araştırma verilerinin yapısını anlamak, uygun analiz yöntemlerini belirlemek ve bulguları şeffaf ve sistematik bir biçimde sunmak için vazgeçilmez bir adımdır. Bu yönüyle, özellikle çevresel ve endüstriyel atıksu gibi çok değişkenli ve heterojen veri setlerinde, daha ileri istatistiksel veya modelleme çalışmalarının sağlam bir temele oturtulmasını sağlamaktadır (Bahna, 2009).

2.3.1. Zamansal Ve Çok Değişkenli Verilerde İstatistiksel Analiz Yaklaşımları

Endüstriyel atıksu verileri, çoğu zaman zamansal bağımlılık, mevsimsel değişkenlik ve çok sayıda fizikokimyasal parametrenin birlikte değişimi nedeniyle karmaşık bir yapı sergilemektedir. Bu tür veri setlerinde, klasik tek değişkenli ve parametrik istatistiksel yaklaşımlar verinin gerçek davranışını yeterince yansıtamayabilmektedir. Atıksu ve su kalitesi çalışmalarında verilerin sıklıkla normal dağılımdan sapma gösterdiği, aykırı değerler içerdiği ve parametreler arasında doğrusal olmayan ilişkiler bulunduğu literatürde açıkça belirtilmektedir (Helsel ve Hirsch, 2002). Bu nedenle, zamansal ve çok değişkenli verilerin analizinde Spearman ve Kendall gibi parametrik olmayan korelasyon katsayıları, özellikle monoton fakat doğrusal olmayan ilişkilerin değerlendirilmesinde yaygın biçimde kullanılmaktadır (Helsel ve Hirsch, 2002; Bahna, 2009). Ayrıca çok değişkenli korelasyon ve regresyon temelli analizlerin, atıksu arıtma süreçlerinde ölçülen parametreler arasındaki etkileşimleri açıklamada ve süreç performansının zamana bağlı olarak değerlendirilmesinde etkili sonuçlar sunduğu gösterilmiştir (Giwa ve Hasan, 2017).

Verilerin zamansal analizleri, atıksu parametrelerinin zaman boyunca nasıl bir değişim gösterdiğini ortaya koymayı amaçlayan temel değerlendirme yöntemleridir. Bu analizler, ölçülen değerlerdeki değişimlerin rastgele mi yoksa sistematik mi gerçekleştiğini ayırt

edilmesine olanak sağlamaktadır. Özellikle uzun süreli izleme çalışmalarında, zamana bağlı değişimlerin incelenmesi; atıksu karakteristiğinin belirlenmesi, meydana gelen kalıcı dönüşümleri, proses değişikliklerini ve işletme kaynaklı etkileri değerlendirmede belirleyici rol oynamaktadır.

Çok değişkenli analizlerde ise inceleme odağı, parametrelerin tek başlarına gösterdikleri davranışlar haricinde, birbirleriyle ilişkileri durumunda nasıl değiştikleri üzerine yoğunlaşmaktadır. Bu yaklaşım, parametreler arasındaki ilişkilerin nicel olarak ortaya konulmasını sağlarken, elde edilen sonuçların çevresel süreçler ve arıtma sistemi dinamikleri ile birlikte değerlendirilmesini gerektirerek atıksu verilerinin bütüncül ve gerçekçi biçimde yorumlanmasına olanak tanımaktadır.

Literatürde yer alan; çözünmüş oksijen dağılımının azot giderimi üzerindeki etkilerini inceleyen çalışmada, amonyum ve toplam azot giderimi arasında güçlü ve istatistiksel olarak anlamlı bir korelasyon belirlenmiş; farklı havalandırma konfigürasyonlarının arıtma performansı üzerinde anlamlı etkiler oluşturduğu istatistiksel testlerle ortaya konulmuştur. Çalışmada, farklı havalandırma konfigürasyonları altında elde edilen veriler, Pearson korelasyon analizi, tek yönlü varyans analizi (ANOVA) ve LSD çoklu karşılaştırma testi kullanılarak değerlendirilmiştir (Daibes ve Alhaddad, 2026).

2.3.2. Atıksu Verilerinde Kullanılan Korelasyon Denklemleri; Pearson, Spearman ve Kendall Yaklaşımları

Zamansal ve çok değişkenli yapıya sahip endüstriyel atıksu verilerinin analizinde; parametreler arasındaki ilişkilerin nicel olarak ortaya konulması önemli bir yer tutmaktadır. Atıksu karakterizasyonunda ölçülen her bir parametre, çoğu zaman diğer parametrelerden bağımsız davranmamakta; aksine ortak proses kaynakları, benzer işletme koşulları veya çevresel etkenler nedeniyle birlikte değişim gösterebilmektedir. Bu nedenle, parametreler arasındaki ilişkilerin sistematik biçimde değerlendirilmesi, çok değişkenli istatistiksel analizlerin temel adımlarından biri olarak öne çıkmaktadır.

Korelasyon analizleri, iki değişken arasındaki ilişkinin yönünü ve gücünü nicel olarak ifade etmeyi amaçlayan istatistiksel araçlardır. Endüstriyel atıksu verilerinde korelasyon analizlerinin kullanımı, belirli parametrelerin birlikte artma veya azalma eğilimlerinin belirlenmesine, potansiyel ortak kirletici kaynaklarının anlaşılmasına ve daha ileri analizler için değişken seçiminin desteklenmesine katkı sağlamaktadır. Ancak atıksu verilerinin zamansal bağımlılık,

dağılım düzensizlikleri ve aykırı değerler içerebilmesi, kullanılacak korelasyon yönteminin dikkatle seçilmesini gerekli kılmaktadır.

Bu kapsamda en yaygın kullanılan korelasyon testleri arasında Pearson, Spearman ve Kendall korelasyon yaklaşımları yer almaktadır. Her bir yaklaşım, veri yapısına ve analiz amacına bağlı olarak farklı avantajlar sunmakta ve zamansal–çok değişkenli analizlerde tamamlayıcı roller üstlenmektedir.

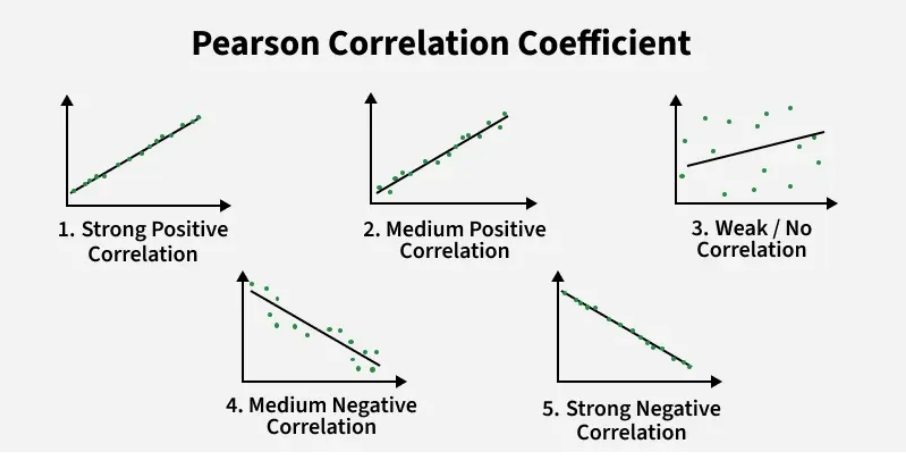
Pearson korelasyon katsayısı, iki sürekli değişken arasındaki doğrusal ilişkinin yönünü ve gücünü ölçen parametrik bir istatistiktir. Katsayı -1 ile +1 arasında değer alır; +1 tam pozitif doğrusal ilişkiyi, -1 tam negative doğrusal ilişkiyi, 0 ise doğrusal ilişki olmadığını ifade eder (Helsel ve Hirsch, 2002). Değişkenlerin sürekli ve yaklaşık normal dağılım gösterdiği, ilişkilerin doğrusal olduğu ve aykırı değerlere karşı duyarlılık gösteren Pearson Korelasyonu, atıksu parametreleri arasında birlikte artış veya azalış eğilimlerinin nicel olarak değerlendirilmesinde kullanılabilmektedir. Pearson Korelasyonunun matematiksel ifadesi, r, aşağıdaki formülde verilmiştir.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2}} \qquad (2\text{-}1)$$

$r: Pearson\ Korelasyon\ Katsayısı$

$x_i: x.değişken$ örnek değeri $\qquad y_i: y.değişken$ örnek değeri

$\bar{x}: x\ değişkenlerinin\ ortalaması \qquad \bar{y}: y\ değişkenlerinin\ ortalaması$



Şekil 2.5. Pearson Korelasyon Katsayısının Grafik Görünümleri

Şekil 2.5'da, iki değişken arasındaki doğrusal ilişkinin yönü ve gücü örnek dağılımlar üzerinden sunulmaktadır. Pozitif yönde güçlü ve orta düzey korelasyonlar, değişkenlerin birlikte artma eğilimini; zayıf veya ilişkisiz durum, belirgin bir doğrusal desenin bulunmadığını göstermektedir. Negatif yönde orta ve güçlü korelasyonlar ise bir değişken artarken diğerinin azaldığı ters yönlü doğrusal ilişkiyi ifade etmektedir.

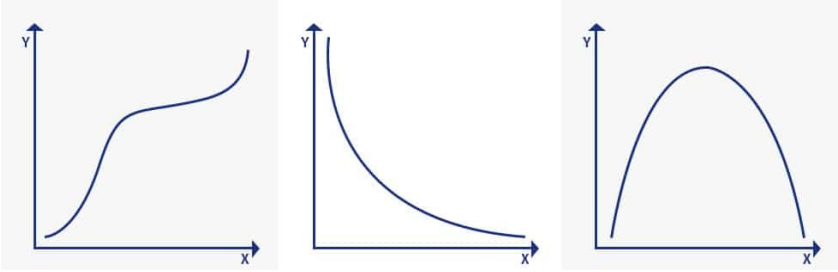
Örneğin, askıda katı madde ile kimyasal oksijen ihtiyacı arasındaki ilişki, doğrusal bir yapı sergilediğinde Pearson korelasyonu anlamlı bilgiler sunabilmektedir. Ancak atıksu verilerinde sıklıkla gözlenen aykırı değerler ve doğrusal olmayan ilişkiler, bu yöntemin tek başına yeterli olmasını çoğu zaman engellemektedir (Bahna, 2009).

Spearman korelasyon katsayısı, iki değişken arasındaki monoton ilişkinin gücünü ölçen parametrik olmayan (non-parametrik) bir istatistiktir. Hesaplama, ham değerler yerine sıralar (ranklar) üzerinden yapılır (Helsel ve Hirsch, 2002). Bu yaklaşım, değişkenlerin normal dağılım göstermesini gerektirmemekte ve doğrusal olmayan ancak aynı yönde ilerleyen ilişkileri ortaya koyabilmektedir (Bahna, 2009). Spearman korelasyon katsayısı aşağıdaki formül ile hesaplanmaktadır.

$$\rho = 1 - \frac{6 \sum d_i^2}{n\,(n^2-1)} \tag{2-2}$$

$\rho : Spearman\ Katsayısı$
$d_i : iki\ değişken\ arasındaki\ fark$
$n : gözlem\ sayısı$

Spearman katsayısının temelini oluşturan monoton fonksiyon kavramı, bağımsız değişken artışına bağlı olarak asla azalmayan veya asla artmayan fonksiyonu tanımlamaktadır.



a **b** **c**

Şekil 2.6. Değişkenler arası ilişkilerde monotonik fonksiyonların davranışı

Şekil 2.6. a'da değişkenler arasında monotonik artan ancak doğrusal olmayan bir ilişki, b'de monotonik azalan bir ilişki; c'de artış ve azalışın birlikte gözlendiği parabolik bir ilişki görseli sunulmaktadır.

Monotonik ilişki, Pearson katsayısında kullanılan doğrusal ilişkiye kıyasla daha az kısıtlayıcıdır. Zamansal atıksu verilerinde, parametrelerin belirli dönemlerde birlikte artma veya birlikte azalma eğilimi göstermesi durumunda, Spearman korelasyonu Pearson'a kıyasla daha esnek ve dayanıklı bir değerlendirme sunmaktadır.

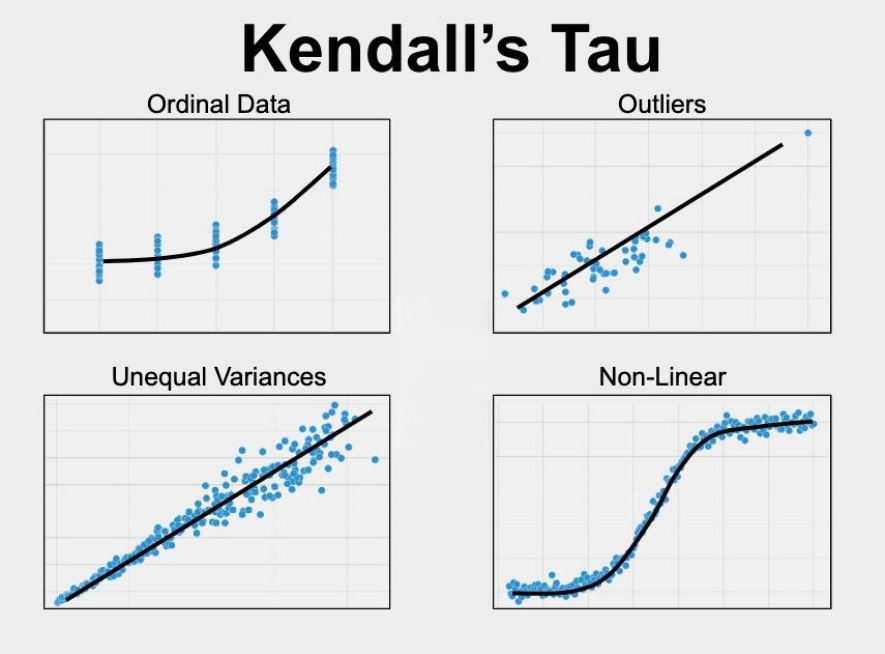
Özellikle mevsimsel etkiler, işletme kaynaklı dalgalanmalar ve ani konsantrasyon değişimlerinin söz konusu olduğu veri setlerinde, Spearman yaklaşımı ilişkilerin genel yönünü anlamak açısından önemli bir araçtır.

Kendall tau korelasyon katsayısı ise, iki değişken arasındaki ilişkiyi uyumlu ve uyumsuz gözlem çiftlerinin oranına dayalı olarak değerlendiren parametrik olmayan bir korelasyon ölçüsüdür (Helsel ve Hirsch, 2002). Bu yaklaşım, özellikle küçük örneklem büyüklüklerinde ve bağıl sıralamaların önemli olduğu durumlarda daha güvenilir sonuçlar sunabilmektedir. Gözlemler arası ilişkinin yönü ve gücünü tahmin etmeye yardımcı bir yorumlamaya olanak sağlamaktadır. Kendall tau korelasyon katsayısı aşağıdaki formül ile hesaplanmaktadır.

$$\tau = \frac{C-D}{\frac{1}{2}n(n-1)} \tag{2-3}$$

τ: $Kendall\ tau\ korelasyon\ katsayısı$
C: $Uyumlu\ çift\ sayısı$
D: $Uyumsuz\ çift\ sayısı$
n: $Gözlem\ sayısı$

Zamansal atıksu verilerinde, bir parametredeki artış eğiliminin başka bir parametredeki artış eğilimiyle ne ölçüde örtüştüğünün değerlendirilmesinde Kendall korelasyonu güçlü bir tamamlayıcı araç olarak öne çıkmaktadır.



Şekil 2.7. Kendall's Tau korelasyon katsayısının farklı veri yapıları altındaki

Şekil 2.7'de, sırasal (ordinal) veriler, aykırı değer içeren dağılımlar, değişen varyans yapısı ve doğrusal olmayan ilişkiler ayrı paneller halinde sunulmaktadır. Kendall's Tau katsayısı, değişkenler arasındaki ilişkiyi uyumlu ve uyumsuz gözlem çiftlerinin oranı üzerinden tanımlamakta olup, özellikle aykırı değerlere ve doğrusal olmayan yapılara karşı Pearson korelasyonuna kıyasla daha dayanıklı bir ölçüt sunmaktadır. Bu özellikleri sayesinde Kendall's Tau, zamansal bağımlılık ve düzensiz dağılım karakteri sergileyen endüstriyel atıksu verilerinde, parametreler arasındaki istatistiksel ilişkilerin güvenilir biçimde değerlendirilmesinde yaygın olarak tercih edilmektedir (Helsel ve Hirsch, 2002; Bahna, 2009).

Helsel ve Hirsch (2002), çevresel verilerin çoğunlukla normal dağılmadığını, aykırı değerler içerdiğini ve monoton ilişkiler sergilediğini vurgulamaktadır. Bu nedenle Pearson korelasyonu yalnızca varsayımların sağlandığı durumlarda önerilirken, Spearman ve Kendall korelasyon katsayıları, çevresel veri analizlerinde daha güvenilir ve dayanıklı alternatifler olarak öne çıkmaktadır. Helsel ve Hirsch (2002) verilerine göre derlenmiş Tablo 2.2 ile Atıksu verilerinde Pearson, Spearman ve Kendall Korelasyon Katsayılarının karşılaştırması verilmektedir.

Tablo 2.2. Atıksu Verilerinde Pearson, Spearman ve Kendall Korelasyon Katsayılarının Karşılaştırılması

Kaynak: Helsel ve Hirsch (2002) çalışmasından uyarlanmıştır.

Kriter	**Pearson (r)**	**Spearman (ρ)**	**Kendall (τ)**
Temel yaklaşım	Doğrusal ilişkinin gücünü ölçer	Monoton ilişkinin gücünü ölçer	Sıra uyumuna dayalı ilişkiyi ölçer
Parametrik yapı	Parametrik	Parametrik olmayan	Parametrik olmayan
Kullanılan veri tipi	Sürekli, nicel veriler	Sürekli veya ordinal veriler	Sürekli veya ordinal veriler
Normal dağılım varsayımı	Gerekli	Gerekli değil	Gerekli değil
Aykırı değerlere duyarlılık	Yüksek	Orta	Düşük
Örneklem büyüklüğüne duyarlılık	Orta	Düşük	Düşük (küçük örneklemler için uygun)
Bağlanmış sıralar (ties)	Sorun yaratabilir	Kısmen etkilenir	Daha dayanıklıdır
İlişki türü	Doğrusal	Doğrusal olmayan fakat yönlü	Doğrusal olmayan fakat yönlü
KOİ – AKM ilişkisi	Doğrusal yapı varsa uygundur; askıda katı madde kaynaklı KOİ artışını iyi yansıtır	Aykırı değerlerin yoğun olduğu işletmelerde daha güvenilirdir	Düzensiz üretim koşullarında daha istikrarlı sonuç verir
TN – TP ilişkisi	Zayıf veya doğrusal olmayan ilişkilerde yanıltıcı olabilir	Besin tuzu kaynaklı ortak değişimleri iyi yakalar	Kaynak farklılıklarının baskın olduğu durumlarda tercih edilir
Endüstriyel atıksular için uygunluk	Homojen ve stabil işletmelerde uygundur	Karışık ve değişken atıksular için uygundur	Karışık OSB atıksularında yüksek güvenilirlik
Yorumlanabilirlik	Yüksek ve yaygın	Yüksek	Daha muhafazakâr, akademik açıdan güçlü

Tablo 2.2.'de, endüstriyel atıksu karakterizasyonunda yaygın olarak kullanılan Pearson, Spearman ve Kendall korelasyon katsayıları; KOİ–AKM ve TN–TP gibi temel parametre çiftleri özelinde karşılaştırılmıştır. Pearson korelasyon katsayısı, doğrusal ilişki varsayımının sağlandığı ve veri dağılımının homojen olduğu durumlarda güçlü sonuçlar sunarken, karışık endüstriyel atıksularda sıkça gözlenen aykırı değerler ve doğrusal olmayan ilişkiler nedeniyle Spearman ve Kendall korelasyon katsayılarının daha güvenilir olduğu değerlendirilmektedir. Özellikle Kendall tau katsayısı, küçük örneklem büyüklükleri ve bağlanmış sıraların yoğun olduğu izleme verilerinde daha istikrarlı bir ilişki ölçütü sağlamaktadır (Bahna, 2009).

Tablo 2.2.' e göre zamansal ve çok değişkenli analizlerde bu üç korelasyon yaklaşımının birlikte kullanılması, parametreler arasındaki ilişkilerin farklı boyutlardan değerlendirilmesine olanak tanımaktadır. Pearson korelasyonu, güçlü doğrusal ilişkileri ortaya koyarken; Spearman ve Kendall yaklaşımları, doğrusal olmayan ancak tutarlı birlikte değişim örüntülerinin belirlenmesini mümkün kılmaktadır. Bu çok yönlü değerlendirme, atıksu karakterinin yalnızca tek bir istatistiksel ölçüt üzerinden değil, farklı varsayımlar altında test edilmesini sağlayarak analiz güvenilirliğini artırmaktadır.

Sonuç olarak, Pearson, Spearman ve Kendall korelasyon testleri; zamansal ve çok değişkenli atıksu verilerinin istatistiksel olarak anlamlandırılmasında birbirini tamamlayan araçlar olarak değerlendirilmelidir. Bu testlerin birlikte ve bilinçli biçimde kullanılması, atıksu karakterizasyonunun daha sağlam, tutarlı ve savunulabilir bir istatistiksel çerçeve içerisinde ele alınmasına olanak tanımaktadır.

Literatürde; su kalitesi değerlendirmelerinde parametreler arasındaki ilişkilerin kullanılması; kirlenme kaynaklarının anlaşılması ve izleme stratejilerinin geliştirilmesi açısından önemli bir araçtır. Bu kapsamda Agarwal ve Agarwal (2013), Hindistan'daki Kosi Nehri üzerinde gerçekleştirdikleri çalışmada çözünmüş oksijen (ÇO), biyokimyasal oksijen ihtiyacı (BOİ), kimyasal oksijen ihtiyacı (KOİ), toplam katı madde (TS) ve alkalinite parametreleri arasındaki doğrusal ilişkileri Pearson korelasyon ve regresyon analizleri ile incelemiştir. Elde edilen sonuçlar, KOİ ile BOİ arasında güçlü pozitif bir ilişki ($r \approx 0,87$) bulunduğunu; ÇO'nun ise hem BOİ hem de KOİ ile güçlü negatif korelasyonlar gösterdiğini ortaya koymuştur. Bu durum, organik ve inorganik kirletici yükündeki artışın çözünmüş oksijen seviyelerini düşürdüğünü ve su ekosistemi üzerinde baskı oluşturduğunu göstermektedir. Çalışmada ayrıca alkalinite ile toplam katı madde arasında anlamlı bir ilişki tespit edilmiş; regresyon denklemleri yardımıyla bazı parametrelerin diğerleri kullanılarak tahmin edilebileceği gösterilmiştir. Ölçülen ve regresyon ile tahmin edilen değerler arasındaki farkın düşük olması, doğrusal modellerin su kalitesi değerlendirmelerinde pratik ve maliyet etkin bir araç olarak kullanılabileceğini ortaya koymaktadır. Bununla birlikte, bu yaklaşımın tekil su kütleleri ve sınırlı parametre setleri için geçerli olduğu; karmaşık ve çok kaynaklı endüstriyel atıksu sistemlerinde zamansal değişkenlik ve çoklu etkileşimlerin daha ileri istatistiksel ve veri temelli yöntemlerle ele alınması gerektiğine değinilmektedir.

Özellikle su ve atıksu kalitesi çalışmalarında klasik istatistiksel yöntemler; verinin yapısını anlamak, temel ilişkileri ortaya koymak ve izleme ağlarını optimize etmek amacıyla yaygın biçimde kullanılmaktadır. Tanımlayıcı istatistikler (ortalama, standart sapma, varyans), parametrik ve parametrik olmayan korelasyon analizleri (Pearson, Spearman), varyans analizleri (ANOVA) ve regresyon modelleri bu yaklaşımın temel araçlarını oluşturmaktadır (Bahna, 2009).

Çok değişkenli istatistiksel yöntemler ise yüksek boyutlu çevresel veri setlerinin yorumlanmasında kritik rol oynamaktadır. Temel Bileşen Analizi (PCA) ve Faktör Analizi (FA), değişkenler arasındaki ortak varyansı açıklayarak veri indirgeme ve kirletici kaynaklarının tanımlanmasını mümkün kılmaktadır. Kümeleme Analizi (CA), benzer su kalitesi özelliklerine sahip izleme noktalarının sınıflandırılmasında; Ayrım Analizi (DA) ise sınırlı sayıda parametre ile mekânsal ve zamansal farklılıkların ayırt edilmesinde etkili sonuçlar sunmaktadır (Singh ve ark., 2004). Singh ve arkadaşlarının çalışması, DA'nın izleme parametrelerinin azaltılmasında yüksek doğruluk sağladığını göstermektedir.

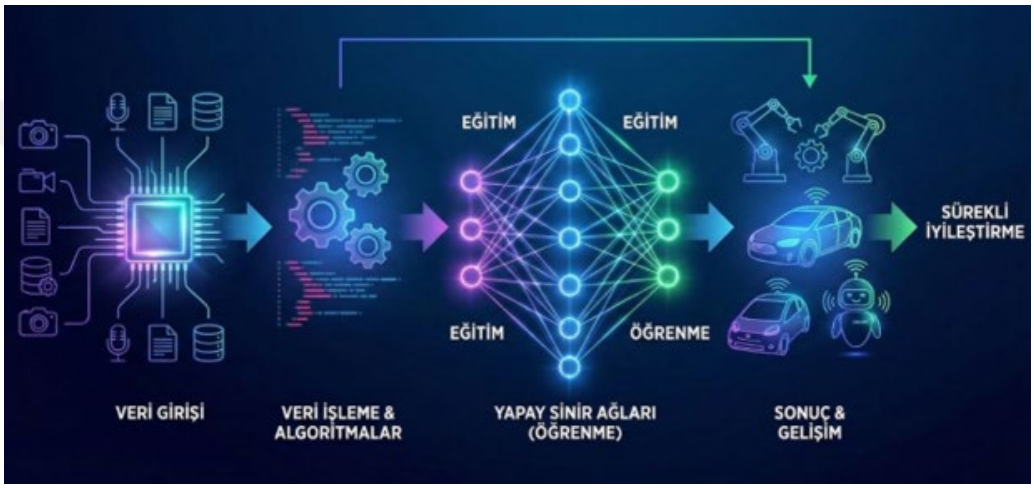
Korelasyon analizleri, özellikle pH, KOİ, BOİ, AKM, TN, TP ve ağır metaller arasındaki ilişkilerin ortaya konulmasında temel bir araçtır. Verma ve arkadaşlarının, 2025, nehir ve yüzey suları üzerinde yaptığı güncel çalışmalarda Pearson korelasyonu, hidrojeokimyasal süreçlerin ve mevsimsel etkilerin değerlendirilmesinde yaygın biçimde kullanıldığını göstermektedir. Bununla birlikte çalışma, klasik istatistiksel yöntemlerin doğrusal varsayımlara dayanması ve karmaşık, doğrusal olmayan süreçleri sınırlı biçimde temsil edebilmesi önemli bir kısıt olarak değerlendirilmektedir.

2.4. ENDÜSTRİYEL ATIKSULARIN KARAKTERİZASYONUNDA YAPAY ZEKÂ ve MAKİNE ÖĞRENMESİ YAKLAŞIMLARI

Endüstriyel atıksuların karmaşık ve doğrusal olmayan bir karakter sergilemesi, atıksu kalitesinin yalnızca belirli parametreler üzerinden ve anlık ölçümlere dayalı yaklaşımlarla değerlendirilmesini yetersiz kılmakta; çok değişkenli, zamansal bağımlılığa sahip ve yüksek boyutlu veri setlerinin birlikte ele alınmasını zorunlu hale getirmektedir (Singh ve ark., 2004; Verma ve ark, 2025).

Bu noktada, yapay zekâ ve makine öğrenmesi temelli veri odaklı yaklaşımlar, çok değişkenli ve zamansal bağımlılığa sahip atıksu sistemlerinin analizinde tamamlayıcı ve ileri düzey bir

çözüm sunmaktadır. Söz konusu yöntemler, büyük ve geniş spekturumlu çevresel verilerden örüntü çıkarma, gelecek çevre stratejileri için tahminler üretme ve çevre riskleri açısından erken uyarı sistemlerinin geliştirilmesinde avantajlar sağlamaktadır. Atıksu verilerinin düzenli ve kontrollü şekilde toplanması, bu tür veri esaslı dijital yaklaşımların uygulanabilirliğini artırmakta; özellikle çevre güvenliği açısından doğacak risklerin anlaşılmasına fırsat tanımaktadır. Literatürde, özellikle derin öğrenme ve ağaç tabanlı algoritmaların, endüstriyel atıksu ve arıtma süreçlerinin modellenmesinde yüksek tahmin başarımı sunduğu ve karar destek sistemlerinin geliştirilmesine katkı sağladığı bildirilmektedir (Khalil et al., 2023).



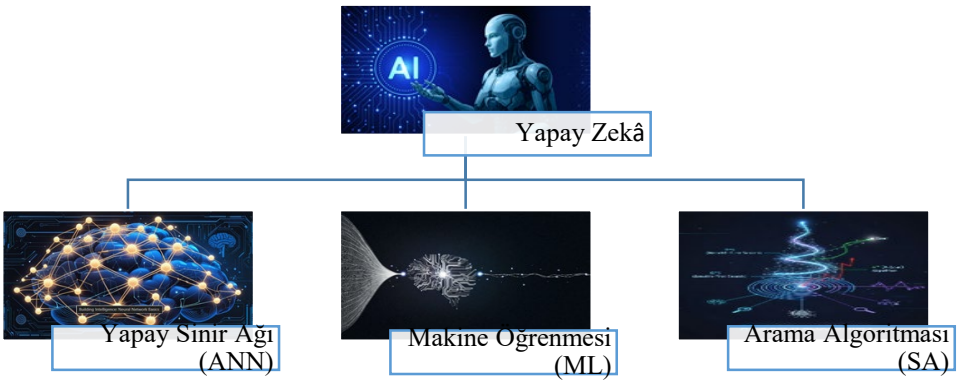
Şekil 2.8. Yapay Zekânın çalışma prensibi ve tasarımı

Yapay zekâ (YZ), bilgisayar sistemlerinin veri ve algoritmalar aracılığıyla öğrenme, akıl yürütme, çıkarım yapma ve karar verme gibi insan zekâsına özgü yetenekleri taklit edebildiği disiplinler arası bir alandır. Yapay zekâ sistemleri, çevresinden elde ettiği bilgileri analiz ederek deneyimlerden öğrenebilmekte ve bu öğrenme sürecini gelecekteki tahmin ve karar mekanizmalarında kullanabilmektedir (Russell, S., Norvig, P., 2009). Bu yüzyılın en etkili icatlarından biri olan YZ, hızla gelişmekte ve doğal dil işleme (NLP), bilgisayar görüsü (CV) ve otonom sürüş gibi birçok alanda uygulanmaktadır (Wang ve ark., 2023).

Şekil 2.8' da YZ'nin çalışma prensibi ve tasarımı verilmektedir. Yapay zekâ uygulamaları, yüksek hacimli ve karmaşık veri setlerinin algoritmik yaklaşımlar kullanılarak çözümlenmesi esasına dayanmaktadır. Bu yaklaşımda temel amaç, veriler içerisinde doğrudan gözlemlenemeyen ilişkilerin ve tekrar eden yapıların belirlenmesidir. Süreç, ölçüm veya gözlem yoluyla elde edilen ham verinin işlenerek analize uygun bir form kazanmasıyla başlamakta ve sistemin veri temelli algı oluşturmasına zemin hazırlamaktadır. Öğrenme

aşamasında, sistem önceden tanımlanmış kurallara bağlı kalmaksızın, kendisine sunulan çok sayıda örnek üzerinden istatistiksel ve matematiksel çıkarımlar üretmektedir. Geleneksel yazılım sistemlerinde karar mekanizmaları insan tarafından ayrıntılı biçimde kurgulanırken, yapay zekâ tabanlı yapılarda algoritmalar veri içerisindeki benzerlikleri ve ayrışmaları otomatik olarak belirlemektedir. Model, her yinelemede elde edilen çıktıları hedef değerlerle karşılaştırarak parametrelerini güncellemekte ve böylece tahmin performansını aşamalı olarak geliştirmektedir. Eğitim tamamlandığında sistem, daha önce karşılaşılmamış veri örnekleri karşısında da tutarlı ve güvenilir yorumlar üretebilecek bir genelleme kapasitesine ulaşmaktadır. Bu sayede, gerçek ortamdan sağlanan anlık girdiler hızlı biçimde değerlendirilmekte ve karar destek süreçlerine doğrudan katkı sağlayan çıktılar elde edilmektedir. Sonuç olarak, yapay zekâ sistemleri durağan veri işleme yaklaşımlarının ötesine geçerek öğrenme yeteneğine sahip, esnek ve uyarlanabilir bir yapı sunmaktadır (IWT Dijital Medya, 2025).

Wang ve ark., 2023 yılındaki atıksu arıtımında YZ uygulamaları ile ilgili incelemesinde, 2000'li yıllardan itibaren YZ'nin atıksu ile ilgili alanlarda yapılan yayın sayıları her geçen gün arttığı, özellikle 2015 ve sonrasında büyük bir artış yaptığı görülmektedir. Çalışmada, atıksu arıtımı konusunda kullanılan yapay zekâ modelleri üç ana kategoriye ayrılmaktadır. Bunlar Şekil 2.9.'da verildiği üzere Yapay Sinir Ağı (ANN), Makine Öğrenmesi (ML) ve Arama Algoritması (SA) olarak sınıflandırılmaktadır.



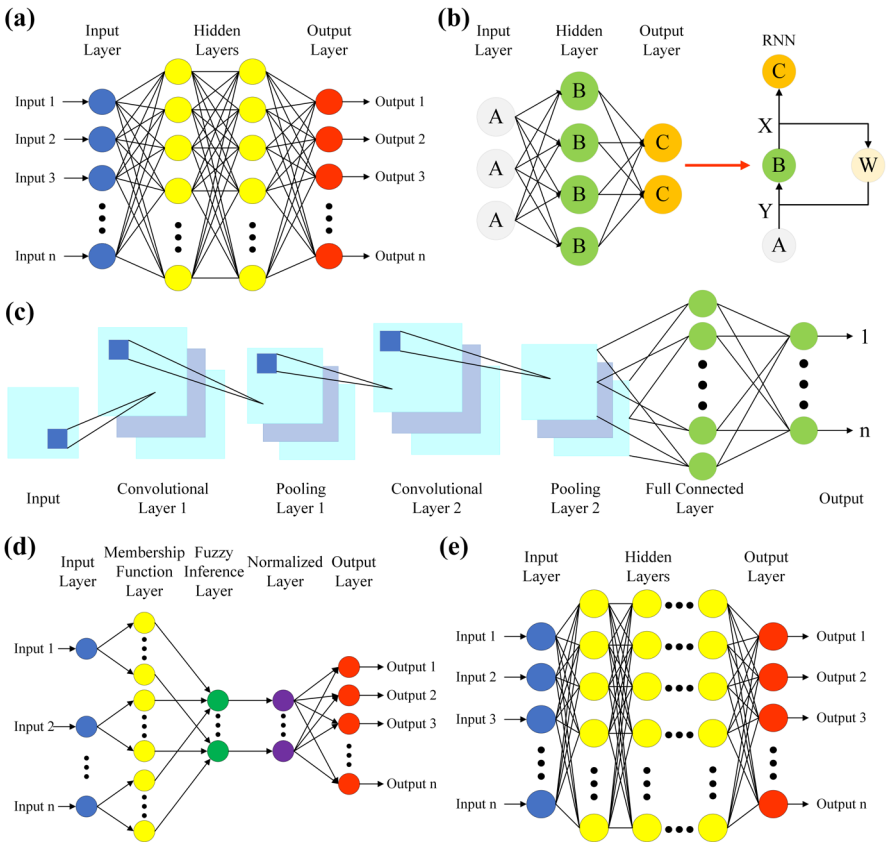
Şekil 2.9. Atıksu Arıtımında kullanılan YZ Modellerinin Sınıflandırılması

Kaynak: Wang ve ark., 2023

2.4.1. Yapay Sinir Ağı (ANN)

Yapay Sinir Ağları, biyolojik sinir sisteminden taklit edilerek geliştirilen ve giriş–çıkış değişkenleri arasındaki doğrusal olmayan ilişkileri öğrenebilen modellerdir. ANN yapıları genellikle bir giriş katmanı, bir veya birden fazla gizli katman ve bir çıkış katmanından oluşmakta; öğrenme süreci, nöronlar arasındaki bağlantı ağırlıklarının hata fonksiyonlarını minimize edecek şekilde optimize edilmesiyle gerçekleştirilmektedir. Atıksu arıtımı uygulamalarında ANN modelleri; su kalitesi parametrelerinin tahmini, proses performansının modellenmesi ve arıtma veriminin öngörülmesi gibi alanlarda yaygın olarak kullanılmaktadır (Wang ve ark., 2023).

Şekil 2.10.' da ANN ve alt modellerin temel öğrenme tasarımları şematize edilmiştir. Bu şema girdi bilgilerin nasıl sonuç çıktılarına dönüştüğünün özet gösterimini vermektedir.



Şekil 2.10. ANN modellerinin temel tasarımı a) ANN, b) RNN, c) CNN, d) FNN, e) DNN
Kaynak: Warg ve ark., 2023

Atıksu verilerinin değerlendirilmesinde kullanılan yapay sinir ağı alt modelleri ise 4 gruba ayrılmaktadır. Bunlar;

Tekrarlayan Sinir Ağı (RNN): Zaman bağımlı verileri işleme yeteneğine sahip yapılardır. Wang ve Ark. (2023)'e göre özellikle ardışık ölçümlerin bulunduğu atıksu izleme verilerinin modellemesinde kullanılmaktadır.

Evrişimsel Sinir Ağı (CNN): Özellikle görüntü ve spektral veri analizlerinde güçlü temsil yeteneğine sahip yapılardır. Sensör verilerinin, mikroskobik görüntülerin analizinde kullanılmakta olup karmaşık özelliklerin otomatik çıkarılmasına olanak sağlamaktadır.

Bulanık Sinir Ağı (FNN): Bulanık mantık ile yapay zekâ sinir ağlarının avantajlarını birleştirerek belirsizlik içeren sistemlerin modelleme yeteneğine sahiptir.

Derin Sinir Ağı (DNN): DNN'Ler çok sayıda gizli katman içeren ANN yapıları olup, yüksek boyutlu ve karmaşık veri setlerinden hiyerarşik özellikler öğrenebilme yeteneğine sahiptir. Atıksu değerlendimelerinde, özellike büyük veri setleri olduğunda yüksek tahmin doğruluğu sağlamaktadırlar (Wang ve ark., 2023, Singh ve ark, 2023).

2.4.2. Makine Öğrenmesi (ML)

Makine öğrenmesi, bilgisayar sistemlerinin açıkça programlanmaksızın verilerden öğrenmesini sağlayan algoritmalar bütünüdür. Atıksu arıtımı alanında ML modelleri; sınıflandırma, regresyon, kümeleme ve boyut indirgeme gibi amaçlarla kullanılmaktadır. En yaygın ML yaklaşımları arasında Temel Bileşen Analizi (PCA), Karar Ağaçları (DT), Destek Vektör Makineleri (SVM), Random Forest (RF), K-En Yakın Komşu (KNN), Kendini Örgütleyen Haritalar (SOM) ve Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS) yer almaktadır (Wang ve ark., 2023).

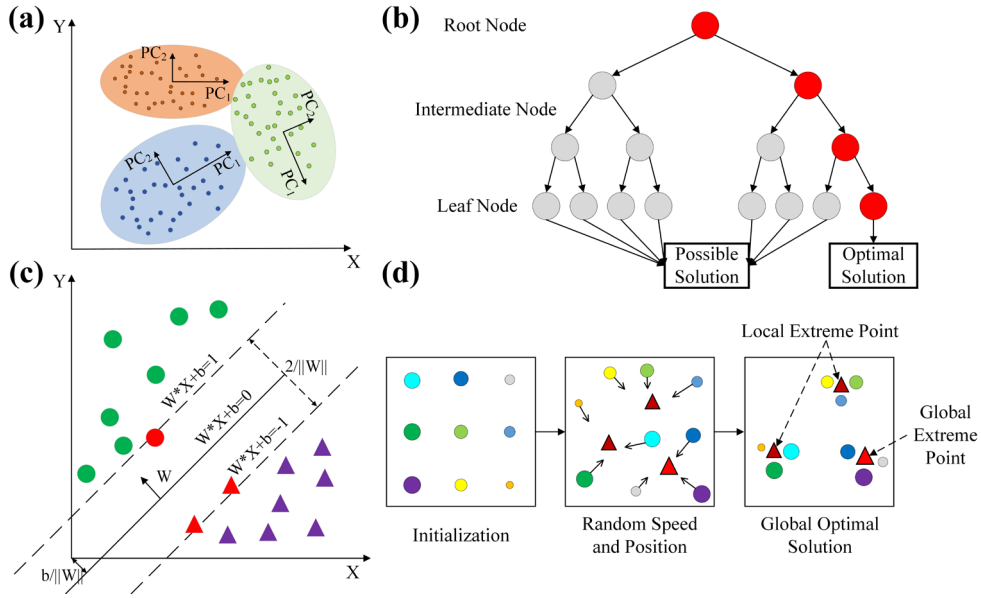
Temel Bileşen Analizi (PCA): PCA, yüksek boyutlu verileri, ortogonal bir dönüşümle orijinal veri bilgilerinin çoğunu koruyarak düşük boyutlu verilere dönüştürebilen, basit bir çok değişkenli istatistiksel makine öğrenme algoritmasıdır.

Karar Ağacı (DT): Sınıflandırma ve regresyon problemlerinin çözümünde kullanılan makine öğrenmesi temelli bir yöntemdir. Atıksu arıtımı uygulamalarında karar ağacı modelleri; su kalitesi sınıflandırması, proses performansının değerlendirilmesi, arıtma tesislerinin işletme

durumlarının ayrıştırılması ve kritik parametrelerin belirlenmesi gibi amaçlarla kullanılmaktadır (Wang ve ark., 2023).

Destek Vektör Makinesi (SVM): Karmaşık ve doğrusal olmayan veri setlerinde sınıflandırma ve regresyon problemlerini çözmede kullanılan bir algoritmadır. Atıksu arıtımı uygulamalarında SVM modelleri; su kalitesi parametrelerinin tahmini, arıtma veriminin öngörülmesi, işletme durumlarının sınıflandırılması ve proses performansının değerlendirilmesi gibi alanlarda yaygın olarak kullanılmaktadır (Wang ve ark., 2023).

Şekil 2.11. ve Şekil 2.12.'de makine öğrenmesi alt modellerinin çalışma prensiplerini gösteren şematik diyagram verilmektedir.



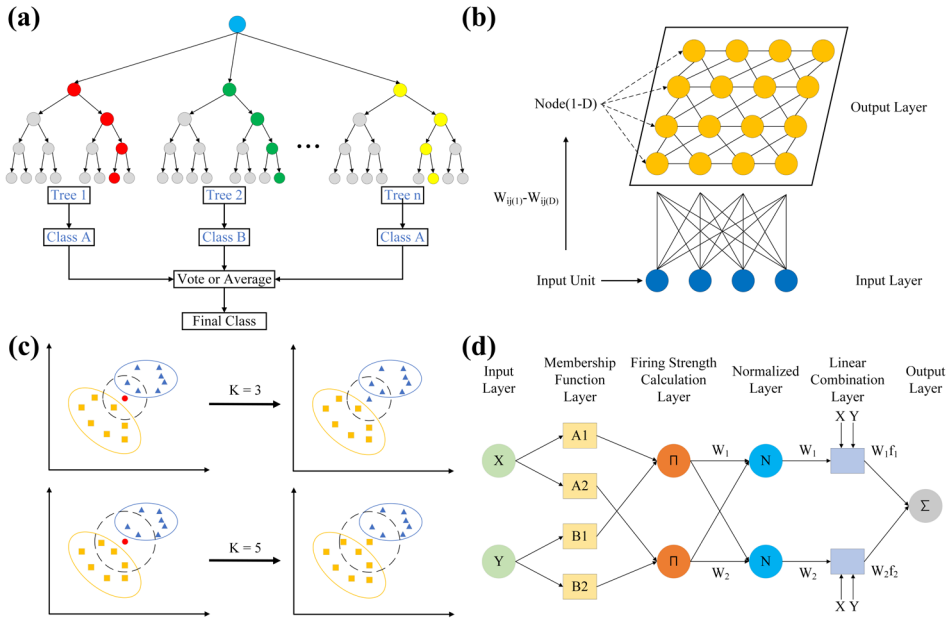
Şekil 2.11. ML modellerinin Şematik Diyagramı a) PCA, b) DT, c) SVM, d) PSO

Kaynak: Wang ve ark., 2023

Parçacık Sürü Optimizasyonu (PSO): Karmaşık ve çok boyutlu optimizasyon problemlerinin çözümü için geliştirilen, sürü zekâsı temelli bir algoritmadır (Wang ve ark., 2023).

Random Forest (RF): Birden fazla karar ağacı (DT) algoritmasını entegre çalıştırarak sınıflandırma ve regresyon problemlerinin çözümünü sağlayan algoritmadır. Wang ve arkadaşları (2023), RF modellerinin yüksek boyutlu ve karmaşık veri setlerinde kararlı performans sergilediğini vurgulamaktadır. Atıksu alanında parametre önem derecelerinin belirlenmesinde yaygın olarak kullanıldığı çalışmada belirtilmektedir.

SOM, KNN ve ANFIS alt modelleri makine öğrenmesinde sınıflandırma ve regresyon problemlerine çözüm için kullanılan algoritma türleridir.



Şekil 2.12. ML modellerinin Şematik Diyagramı a) RF, b) SOM, c) KNN, d) ANFIS

Kaynak: Wang ve ark., 2023

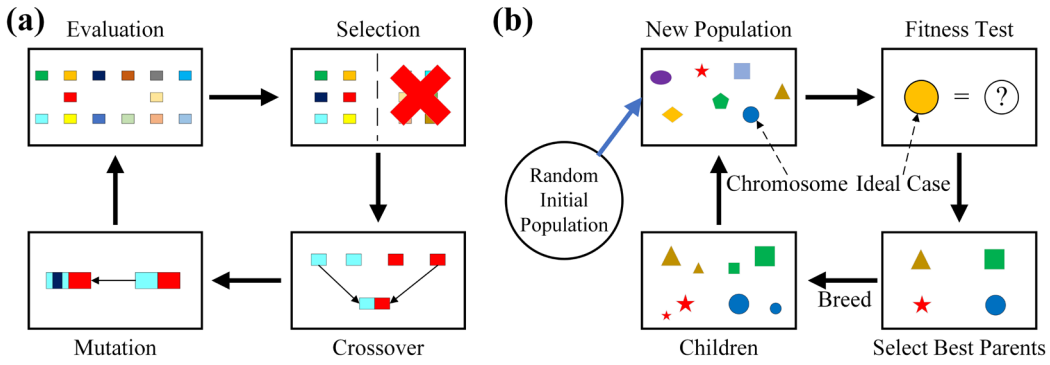
Makine öğrenmesi, çevresel sistemlerde çok sayıda değişken içeren ve doğrusal olmayan ilişkiler barındıran veri setlerinin modellenmesinde klasik yöntemlere güçlü bir alternatif sunmaktadır. ML modelleri; önceden tanımlanmış fiziksel denklemlere ihtiyaç duymadan, veri içerisindeki örüntüleri öğrenerek tahmin ve sınıflandırma yapabilmektedir.

2.4.3. Arama Algoritması (SA)

Sıralı veriler üzerinde bir bilginin aranmasında kullanılan yöntemlerin adıdır. Wnag ve ark. 2023, yapmış olduğu çalışmada, SA'nın Genetik Algoritma (GA) ve Genetik Programlama (GP) alt modellerinin atıksu uygulamalarında yaygın kullanıldığını belirtmişlerdir.

Genetik Algoritma (GA): Biyolojik evrim prensibini esas alan bir arama ve optimizasyon yöntemidir. Algoritmanın temel işleyişi; seçilim, çaprazlama ve mutasyon olmak üzere üç ana genetik operasyona dayanmaktadır.

Genetik Programlama (GP): Genetik Algoritma'nın bir uzantısı olup, klasik GA'dan farklı olarak çözüm olarak sabit parametreler yerine matematiksel ifadeler veya program yapıları üretmektedir.



Şekil 2.13. SA modellerinin Şematik Gösterimi a) GA, b) GP

Kaynak: Wang ve ark., 2023

Şekil 2.13'te arama algoritmalarının çalışma prensibini şematize eden görseller verilmektedir. Arama algoritmaları, atıksu arıtımında proses optimizasyonu, enerji ve maliyet etkinliğinin artırılması ve çok değişkenli karar problemlerinin çözümü açısından güçlü araçlar sunmaktadır. Ancak tek başlarına kullanıldıklarında hesaplama yükü ve yakınsama sorunları nedeniyle, literatürde çoğunlukla ANN, SVM ve RF gibi öğrenme tabanlı modellerle entegre edilerek hibrit yapılar oluşturulması önerilmektedir (Wang ve ark., 2023).

Yapay zekâ modellerinin kullanım amaçları, avantajları ve dezavantajları Tablo 2.3'te karşılaştırmalı olarak sunulmuştur. Wang ve ark., 2023, çalışmasına göre tablo incelendiğinde; yapay sinir ağı tabanlı modellerin (RNN, CNN, FNN ve DNN) özellikle doğrusal olmayan ve çok değişkenli veri yapılarının modellenmesinde yüksek tahmin doğruluğu sunduğu görülmektedir. Bu modeller, su kalitesi parametrelerinin tahmini ve proses performansının öngörülmesi gibi görevlerde etkin sonuçlar üretmekle birlikte, yüksek hesaplama maliyeti, eğitim sürecinin karmaşıklığı ve yorumlanabilirliğin sınırlı olması gibi dezavantajlara sahiptir. Buna karşılık, karar ağacı ve destek vektör makinesi gibi klasik makine öğrenmesi yöntemleri, daha sade model yapıları ve göreli olarak yüksek yorumlanabilirlikleri sayesinde özellikle karar destek ve parametre seçimi süreçlerinde öne çıkmaktadır. Arama ve optimizasyon algoritmaları (GA, PSO ve GP) ise doğrudan tahmin yapmaktan ziyade, arıtma proseslerinin işletme koşullarını iyileştirmeye yönelik güçlü araçlar olarak değerlendirilmektedir. Bu algoritmalar, çok boyutlu ve karmaşık çözüm uzaylarında küresel optimuma ulaşma potansiyeline sahip olmaları nedeniyle, enerji tüketiminin azaltılması ve giderim verimlerinin artırılması gibi hedefler açısından önemli avantajlar sunmaktadır. Ancak bu yöntemlerin hesaplama yükü ve yakınsama süreleri, tek başlarına kullanımını sınırlamakta; bu nedenle literatürde çoğunlukla

yapay sinir ağları ve makine öğrenmesi modelleriyle bütünleştirilmiş hibrit yapılar tercih edilmektedir. Bu durum, atıksu arıtımı gibi karmaşık sistemlerde tek bir yapay zekâ yaklaşımı yerine, problem yapısına uygun model kombinasyonlarının daha etkin ve güvenilir sonuçlar sunduğunu göstermektedir.

Tablo 2.3. Atıksu arıtımında yaygın olarak kullanılan Yapay Zekâ Modellerinin Kullanım Amaçları, Avantajları ve Dezavantajları

Kaynak: Wang ve ark., 2023

Yapay Zekâ Modeli	Kullanım Amacı	Avantajları	Dezavantajları
RNN	Regresyon	Zamansal serilerin modellenmesine uygun	Hesaplama maliyeti yüksek
	Sınıflandırma	Girdi uzunluğu açısından sınırlama yok	Eğitim süreci zor
	Tahmin		
CNN	Regresyon	Görüntü tabanlı veriler için uygun	Hesaplama maliyeti yüksek
	Sınıflandırma	Önemli özellikleri otomatik çıkarabilme	Eğitim süreci zor
	Segmentasyon		
FNN	Regresyon	Uygulaması ve yorumu görece kolay	Hesaplama maliyeti yüksek
	Sınıflandırma	Karmaşık doğrusal olmayan problemlere uygun	Model yapısı karmaşık
	Tahmin		
DNN	Regresyon	Yüksek doğruluk ve hızlı tahmin	Hesaplama maliyeti yüksek
	Sınıflandırma	Karmaşık doğrusal olmayan ilişkiler için uygun	Eğitim zor
	Tahmin		Aşırı uyum riski
PCA	Kümeleme	Basit ve uygulanması kolay	Bazı önemli bilgilerin kaybolma riski
	Boyut indirgeme		Gürültülü verilere duyarlı
DT	Regresyon	Kolay anlaşılır ve yorumlanabilir	Dengesiz veri setleri için uygun değil
	Sınıflandırma	Ön işleme gerek duymaz	Eğitim verimliliği düşük
	Optimizasyon		
SVM	Regresyon	Yüksek boyutlu problemlerde etkili	Hesaplama maliyeti yüksek
	Sınıflandırma	Doğrusal olmayan ayrımlarda başarılı	Büyük veri setlerine uygun değil
	Tahmin		

Tablo 2.3 (Devam)

Yapay Zekâ Modeli	Kullanım Amacı	Avantajları	Dezavantajları
PSO	Regresyon	Basit yapı	Başlangıç koşullarına duyarlı
	Sınıflandırma	Yüksek hesaplama verimliliği	Ayrık problemlere uygun değil
	Kümeleme	Geniş problem yelpazesinde uygulanabilir	
RF	Regresyon	Yüksek genelleme yeteneği	Yüksek doğruluk için çok sayıda ağaç gerekir
	Sınıflandırma	Yüksek boyutlu veri setleri için uygun	Hesaplama maliyeti yüksek
	Tahmin		
KNN	Regresyon	Basit ve kolay uygulanabilir	Hesaplama maliyeti yüksek
	Sınıflandırma	Doğrusal olmayan sınıflandırmalara uygun	Yüksek bellek kullanımı
SOM	Kümeleme	Yüksek boyutlu veriler için uygun	Hesaplama karmaşıklığı yüksek
	Boyut indirgeme	Veri görselleştirmeye elverişli	Eksik verilerle çalışmaya uygun değil
ANFIS	Regresyon	ANN ve bulanık mantığın avantajlarını birleştirir	Hesaplama maliyeti yüksek
	Sınıflandırma	Belirsiz verilerle çalışmaya uygundur	Uygun üyelik fonksiyonunun tanımlanması zor
	Tahmin		
GA	Regresyon	Karmaşık doğrusal olmayan problemler için uygun	Eğitim süreci zor
	Sınıflandırma	Çok amaçlı optimizasyona elverişli	Yerel arama yeteneği zayıf
	Optimizasyon		Yüksek boyutlarda uygun değil
GP	Regresyon	Karmaşık optimizasyon problemleri için uygun	Çok sayıda kontrol parametresi
	Sınıflandırma	Otomatik arama ile çözüm üretir	Yavaş yakınsama
	Optimizasyon		Yüksek boyutlarda uygun değil

2024 yılında yapılan bir çalışamada atıksu arıtma tesislerinin izlenmesinde su kalite parametrelerinin tahmini için Çok Katmanlı Algılayıcı (MLP), Genelleştirilmiş Bulanık Fonksiyon Regresyonu (GFFR), Temel Bileşen Analizi ile desteklenmiş Çok Katmanlı Algılayıcı (MLP-PCA) ve Radyal Tabanlı Fonksiyon Ağı (RBF) olmak üzere farklı makine

öğrenmesi modelini karşılaştırılmıştır (Afan ve ark., 2024). Bu ML tabanlı modeller iki senaryoda uygulanmıştır: ilki doğrudan giriş–çıkış verilerini kullanırken, ikincisi birincil ve ikincil çöktürücülerden elde edilen ara çıktıları içeren çok adımlı bir modelleme yaklaşımını kullanmaktadır. Sonuçlar, özellikle ikinci senaryoda iletkenlik tahmininde GFFR modelinin en iyi performansı gösterdiğini ortaya koymuştur. Diğer parametrelerin tahmininde yetersiz olmakla beraber, çok adımlı yaklaşım tüm modellerin tahmin başarısını belirgin biçimde iyileştirmiştir. Bu sonuçlar, su kalite tahmininde ara süreç verileri ile yapay zekâ modellerinin yüksek doğruluğa sahip olduğu ispatlamıştır.

Aasim ve arkadaşları, 2025, makine öğrenmesi modelleri, yalnızca tahmin doğruluğu açısından değil; özellik seçimi ve önem analizi yoluyla sistem davranışını açıklama potansiyeli açısından da önem taşıdığını yaptıkları çalışma ile göstermişlerdir. Özellikle SHAP (Shapley Additive exPlanations tekniği, makine öğrenmesi modeli tarafından yapılan tahminlerde her bir özelliğin tahmin düzeyine (daha yüksek veya daha düşük) olan katkısını incelemek amacıyla kullanılır) ve özellik önem sıralamaları, hangi parametrelerin süreç üzerinde baskın rol oynadığını ortaya koymaktadır. Biyokütle ve biyokömür temelli sistemlerde RF ve boosting tabanlı modellerin hem yüksek doğruluk hem de yorumlanabilirlik sunduğu gösterilmiştir.

Atıksu arıtma tesislerinde N_2O emisyonlarının modellenmesine yönelik yapılan bir çalışmada, kNN, AdaBoost ve DNN modellerinin mekanistik modellere kıyasla daha yüksek tahmin doğruluğu sağladığı; ayrıca model karmaşıklığı ve yorumlanabilirlik arasında denge kurulmasının operasyonel kararlar açısından kritik olduğu vurgulanmıştır (Khalil ve ark., 2023). Bu durum, YZ tabanlı yaklaşımların yalnızca "kara kutu" modeller olarak değil, süreç bilgisi ile entegre edilebilen araçlar olarak değerlendirilmesi gerektiğini göstermektedir.

Literatürdeki kapsamlı derleme ve bibliyometrik analizler, 2015 sonrası dönemde atıksu arıtımında YZ uygulamalarının hızla arttığını; özellikle besin maddeleri, ağır metaller ve adsorpsiyon süreçlerine odaklanıldığını ortaya koymaktadır (Baarimah ve ark., 2024). Ancak bu çalışmalar, veri ihtiyacının yüksek olması, model tekrarlanabilirliği ve fiziksel anlamlandırma eksikliği gibi sınırlamaların da altını çizmektedir (Wang ve ark., 2023).

Shokri ve arkadaşlarının 2025 yılında yaptıkları bir çalışmada; makine öğrenmesi temelli UV-elektro-peroksone (UV-EP) yönteminin petro-kimyasal atık sular üzerindeki etkilerini tahmin ve optimizasyonu amacıyla rapor edilmiştir. Reaksiyon süresi, ozon dozu, akım ve pH etkisi,

Merkezi Kompozit Tasarım (CCD) ve yapay sinir ağı (ANN) modeli aracılığıyla incelenmiştir. ANN modeli, deneysel verileri öğrenerek farklı işlem koşullarında KOI giderimini yüksek doğrulukla tahmin etmiş ve böylece optimum deneysel koşulların belirlenmesine imkan sağlamıştır. Bu şekilde, makine öğrenmesi çalışmada hem tahmin hem de süreç optimizasyonu amacıyla sunulmuştur. Bir başka çalışmada Hassnain ve arkadaşları (2025) makine öğrenmesi çalışmada bir uçucu yağ üretim tesisi atık su arıtma tesisi için nihai çıkış biyolojik oksijen talebinin (F-BOD) doğru tahmin edilmesini rapor etmiştir. Uzun zaman dilimindeki incelemeler ışığında ölçüm verileri kullanılarak farklı makine öğrenmesi modelleri değerlendirilmiştir. En yüksek doğrulukta Extra Trees (ExT) modeli bulunmuştur. Operasyonel ve fizikokimyasal parametrelere dayalı olarak F-BOD'nin gerçek zamanlı tahmin edilmesini ve geleneksel zaman alıcı gereksinimler ortadan kaldırmıştır. Ayrıca, sinir ağları (NN) ile gerçekleştirilen makine öğrenmesi tabanlı simülasyonlar, optimal işletme sınırlarının belirlenmesinde rol almıştır ve süreç kontrolü ile çevresel etkileri minimize edecek şekilde planlanmıştır. Özellik önem analizi ve Shapley Additive Explanations değerleri, F-BOD üzerinde en etkili değişkenleri ortaya koymuş ve makine öğrenmesinin sadece tahmin doğruluğunu artırmakla kalmayıp, karmaşık atıksu süreçlerinin anlaşılmasını da iyileştirdiğini göstermiştir.

Atıksu arıtma ve su kalitesi sistemleri, çok sayıda değişkenin doğrusal olmayan ve zamansal etkileşimi nedeniyle klasik modelleme yaklaşımlarıyla sınırlı biçimde temsil edilebilmektedir. Bu nedenle literatürde, yapay zekâ (YZ) ve makine öğrenmesi (ML) tabanlı yöntemlerin su kalitesi tahmini, proses izleme ve optimizasyon amaçlı kullanımı giderek artmaktadır (Ahmed ve ark., 2019). Kapsamlı derleme ve eğilim analizleri, özellikle yapay sinir ağları, destek vektör makineleri ve topluluk tabanlı algoritmaların KOİ, besin maddeleri ve emerging kirleticilerin giderim performansının modellenmesinde öne çıktığını ve proses optimizasyonu ile membran kirlenmesi kontrolünde etkin sonuçlar sunduğunu göstermektedir (Bhagat ve ark., 2020). Ayrıca literatür ve patent verilerine dayalı çalışmalar, çok amaçlı optimizasyon ve mikrobiyal süreçlerin modellenmesinin gelecek araştırma alanları arasında yer aldığını ortaya koymaktadır (Zhang ve ark., 2023; Zhao ve ark., 2020). Uygulamalı çalışmalarda ise makine öğrenmesi ile yanıt yüzey metodolojisinin birlikte kullanımı, fosfor gibi kritik kirleticilerin gideriminde hem tahmin doğruluğunu artırmakta hem de deneysel yükü azaltmaktadır (Han ve ark., 2025).

Özellikle son 10 yılda, atık su kalitesi değerlendirme ve izleme araştırmaları yapay zekâ destekli yöntemlere doğru evrildi. Bunlar arasında optik sensör üretimi, gerçek zamanlı izleme tekniklerinin ilerlemesi, optimizasyon çalışmaları ve yapay zekâ ve makine öğreniminin

entegrasyonlu modellerin çalışılması, hızlı patojen tespiti için nanoyapı esaslı sensör alanındaki gelişmeler ve çok parametreli su kalitesi testleri için Ardışık Enjeksiyon Analizi gibi otomatik ve çoklu yöntemlerin uygulanması bulunmaktadır. Yaygın olan yöntemler arasında spektroskopik incelemeler (Ruffino ve ark, 2020), elektrokimyasal sensörler (Sfragano ve ark, 2025), mikrobial sensörler (Yao ve ark, 2025), biyolojik analizler (Kandar ve ark., 2025), çok değişkenli istatistiksel veri analizleri (Santos ve ark., 2024) ve giderek artan bir şekilde IoT tabanlı gerçek zamanlı sensör ağları (Chavhan ve ark., 2025; Kubiak ve ark., 2025; Parthasarathy ve ark. 2025) bulunmaktadır. Veri türleri, su kirleticilerin laboratuvar tabanlı ölçümlerinden (örneğin, KOİ, BOİ, AKM) gerçek zamanlı spektral verilere, sensör verilerine ve otomatik izleme sistemleri aracılığıyla oluşturulan kapsamlı veri akışlarına kadar çeşitlenmektedir (Castano-Trias ve ark., 2023).

Literatürde, Makine öğrenmesi ve yapay sinir ağları (ANN), yinelemeli sinir ağları (RNN), bulanık sinir ağları (FNN) gibi derin öğrenme (DL) modelleri kirlilik tespiti, süreç modelleme, deneysel optimizasyon ve otomatik kontrol alanlarında yüksek tahmin potansiyeli ile atıksu sistemlerinin karmaşık ve doğrusal olmayan davranışlarının yönetilmesini sağlamaktadır. Bu özellikler, mikroplastikler, per- ve polifloroalkil maddeler (PFAS), ağır metaller ve ilaçlar gibi kirleticilerin ele alınmasında özellikle kıymetlidir. Bununla beraber işletme maliyetlerinin azaltılması ve etkili ve yüksek arıtma performansı ile katkı sağlamaktadır. Yapay zekânın geleneksel sistemlerinin sınırlılıklarını aşmadaki dönüştürücü potansiyelini eleştirmekle beraber gelişen özellikleri ile teknolojik değerlendirmeler amaçlamaktadır.

Son yıllarda atıksu arıtma ve su kalitesi izleme çalışmalarında yapay zekâ (YZ) ve makine öğrenmesi (ML) tabanlı yaklaşımların kullanımı belirgin biçimde artmıştır. Wang ve ark. (2023), atıksu arıtma sistemlerinde YZ ve ML modellerinin performans tahmini, süreç izleme ve tasarım amaçlı kullanımını ele aldıkları derleme çalışmalarında; özellikle yapay sinir ağları (ANN), destek vektör makineleri (SVM), random forest (RF) ve tekrarlayan sinir ağlarının (RNN/LSTM) KOİ, BOİ, besin maddeleri ve askıda katı madde gibi parametrelerin tahmininde yüksek doğruluk sağladığını rapor etmiştir. Nehir suyu kalitesi modellemesine odaklanan bir diğer kapsamlı çalışmada ise, yapay zekâ temelli modellerin çok değişkenli ve doğrusal olmayan su kalitesi verilerinde geleneksel istatistiksel yöntemlere kıyasla daha başarılı sonuçlar ürettiği; özellikle zamansal bağımlılık içeren veri setlerinde ANN ve hibrit modellerin öne çıktığı vurgulanmıştır (Tiyasha ve ark., 2020). Benzer şekilde, yapay zekâ ve makine öğrenmesi destekli izleme ve tasarım yaklaşımlarını inceleyen güncel bir çalışmada, bu yöntemlerin

yalnızca tahmin doğruluğunu artırmakla kalmadığı, aynı zamanda erken uyarı sistemleri, proses optimizasyonu ve karar destek mekanizmaları açısından da önemli avantajlar sunduğu belirtilmiştir (Singh ve ark., 2023). Bu çalışmalar birlikte değerlendirildiğinde, YZ ve ML tabanlı modellerin karmaşık, zamansal ve çok parametreli atıksu sistemlerinin anlaşılması ve yönetilmesinde güçlü bir araç sunduğu; özellikle uzun dönemli izleme verilerinin değerlendirilmesinde literatürde önemli bir metodolojik dönüşümün yaşandığı görülmektedir.

Bu tez çalışması, karışık endüstrilerin bulunduğu Organize Sanayi Bölgesi (OSB)'ndeki atıksu izleme programından elde edilen veriler ışığında, ML temelli bir erken uyarı ve tahmin yaklaşımı geliştirmeyi hedeflemiştir. Tez kapsamında, deneysel parametreler (kimyasal oksijen ihtiyacı (KOİ) dışındaki) mevzuat limitlerinin aşılması riskinin öngörülebilirliği ele almış; bu sorunu belirleyen deneysel parametreler hem global (tüm firmaları kapsayan) hem de sektörel düzeyde araştırılmıştır. Böylece, atıksu izlenmesi ve gelecekteki farklı numuneler alınmadan önce durumunun tahmin edilebilmesi hedeflenmiştir.

3. YÖNTEM

Bu tez çalışmasında, karışık endüstrilerin bulunduğu Organize Sanayi Bölgesinde yer alan üretim tesislerinden kaynaklanan atıksu kirlilik yükü izlemesi programı ile elde edilen verilerin istatistiksel olarak değerlendirilmesi amacıyla makine öğrenmesi tabanlı yapay zekâ yaklaşımları ile proaktif bir atıksu yönetimi hedeflenmiştir.

Çalışma kapsamında kullanılan veri setlerinin kapsamı ve özellikleri, numune alma ve analiz süreçleri, uygulanan istatistiksel ve veri temelli analiz yöntemleri sistematik biçimde bu bölümde sunulmaktadır.

3.1. ARAŞTIRMA AMAÇLARI VE ARAŞTIRMA SORULARI

Bu çalışma, OSB'de yürütülen düzenli atıksu izleme programından elde edilen çok yıllı veri setini kullanarak, makine öğrenmesi (ML) temelli bir erken uyarı yaklaşımı geliştirmeyi amaçlamaktadır. Bu kapsamda hem global (tüm firmalar bazında) hem de sektörel düzeyde, deneysel parametreler (KOİ hariç) için mevzuat ihlali riskinin öngörülebilirliği ve bu riski belirleyen temel etmenler incelenmiştir.

Çalışma şu üç araştırma sorusuna odaklanmaktadır:

- **Araştırma Sorusu 1 (AS1):**

Endüstriyel nitelikli atıksu numunelerinde, bir sonraki numunenin KOİ dışındaki parametreler açısından "limit altında mı / üstünde mi" olacağı makine öğrenmesi ile ne kadar iyi tahmin edilebilir?

Başka bir ifadeyle: "Bugünkü atıksu profiline bakarak, bir sonraki numunede KOİ dışı mevzuat ihlali riski var mı?" sorusuna hangi doğruluk düzeyi ile cevap verilebilir?

- **Araştırma Sorusu 2 (AS2):**

KOİ hariç limit aşımı riskini en çok hangi parametreler ve parametre kombinasyonları açıklamaktadır?

Küresel (global) model düzeyinde, ihlal riskini taşıyan numunelerin ortaya çıkmasında hangi kimyasal/fiziksel parametreler ve hangi zamansal özetler (lag, hareketli ortalama, kümülatif ortalamalar, ihlal geçmişi vb.) belirleyicidir?

- **Araştırma Sorusu 3 (AS3):**

Sektör/altyük profillerine göre KOİ dışı limit aşımının "key driver"ları nasıl değişmektedir?

SKKY sektör grupları (örneğin 15.2, 6.1, 14.4, 14.5, 14.10) bazında kurulan modellerde, ihlal riskini açıklayan temel parametreler ve zamansal özellikler birbirinden farklılaşmakta mıdır?

Bu araştırma soruları doğrultusunda, 3.2. Bölümde veri ön-işleme ve modelleme yaklaşımı, 4. Bölümde ise global ve sektörel bulgular detaylı olarak sunulmaktadır.

Bu bölümde, OSB atıksu izleme verisinin AS1–AS3 kapsamındaki soruları yanıtlayabilecek şekilde nasıl yapılandırıldığı, yani KOİ hariç ihlal etiketinin nasıl tanımlandığı ve global ile sektörel makine öğrenmesi modellerinin nasıl kurulduğu ayrıntılı olarak sunulmaktadır.

3.2. MATERYAL ve METOT

3.2.1. Veri Setinin Oluşturulması

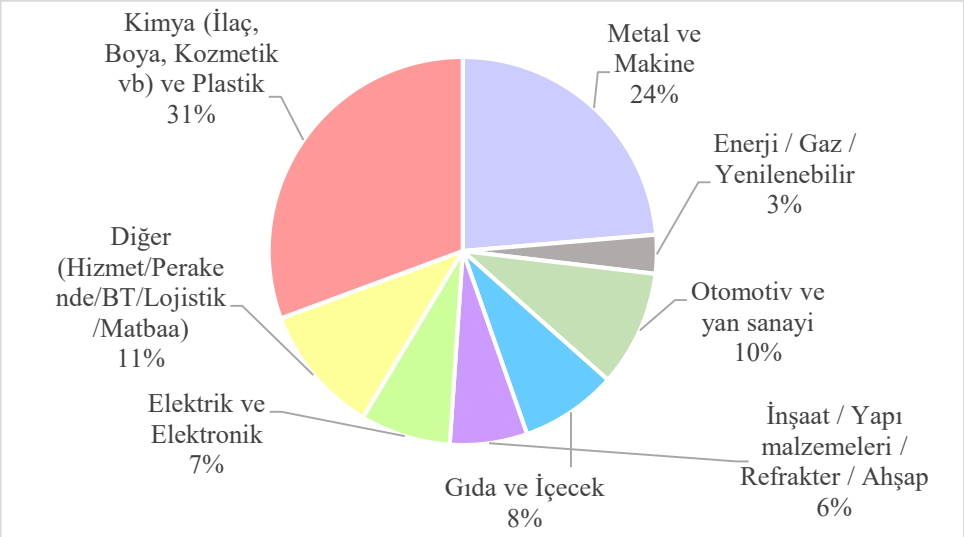
Bu çalışmada kullanılan veri seti, karışık endüstrilerden oluşan OSB'de bulunan farklı sektörlerde üretim yapan tesislerinden kaynaklı atıksuların analiz sonuçlarının derlenmesinden oluşmaktadır. OSB üretim tesislerinden numune alımı, yıl boyunca ayda iki kez olacak şekilde gerçekleştirilmiştir.



Şekil 3.1. Organize Sanayi Bölgesi Yerleşim Planı

Şekil 3.1'de numune alımı gerçekleştirilen OSB yerleşim planı sunulmaktadır. Şekilde lacivert alanlar sanayi alanlarını, turkuaz alan teknopark bölgelerini, gri bölümler idari, sosyal ve teknik alanları ve yeşil bölümler park alanlarını göstermektedir.

OSB'de bulunan 38 ayrı tesisten, her tesisin kendi sektörüne ait atıksu izleme parametrelerine göre tesis bazında sektörel veri setleri oluşturulmuştur. Verilerin sektörel dağılımı Şekil 3.2.'de verilen grafik ile sunulmuştur. Atıksu yükünün ağırlıklı olarak kimya (ilaç, boya, kozmetik vb.) ve plastik sektörlerinden kaynaklandığı, bunu metal ve makine sanayinin izlediği; enerji, otomotiv, gıda ve diğer sektörlerin ise daha düşük oranlarda katkı sağladığı görülmektedir.



Şekil 3.2. Organize Sanayi Bölgesindeki Firmaların Sektörel Dağılımı

Firmalardan alınan numunelerin hepsi 24 saatlik kompozit numune olarak TS EN ISO 5667-10 standardına göre alınmıştır. Numunelerin laboratuvardaki analiz çalışmalarında Tablo 3.1'de sunulan metotlar kullanılmıştır.

Tablo 3.1. Karışık Endüstriyel Atıksuların analizi için Kullanılan Standart Yöntemler

Parametre	Deney Metodu (Uluslararası Standartlar)	Analiz Tekniği / Notlar
pH	SM 4500 H^+ B	Elektrometrik Yöntem
Kimyasal Oksijen İhtiyacı (KOİ)	SM 5520 B	Açık Reflüks Titrimetrik Yöntem
Askıda Katı Madde (AKM)	SM 2540 D	Gravimetrik Yöntem
Yağ ve Gres	SM 5520 B	Gravimetrik Yöntem
Toplam Fosfor (TP)	SM 4500 P B ve SM 4500 P E	Persülfat Özütleme + Spektrofotometrik Yöntem
Toplam Kjeldahl Azotu (TKN)	SM 4500 N_{org} B	Makro Kjeldahl Yöntemi, Titrasyon
Krom+6 (Cr^{+6})	SM 3500 Cr B	Spektrofotometrik Yöntem
Kurşun (Pb)	EPA 200.7	ICP-OES Yöntemi
Kadmiyum (Cd)	EPA 200.7	ICP-OES Yöntemi
Demir (Fe)	EPA 200.7	ICP-OES Yöntemi
Bakır (Cu)	EPA 200.7	ICP-OES Yöntemi
Çinko (Zn)	EPA 200.7	ICP-OES Yöntemi
Civa (Hg)	EPA 200.7	ICP-OES Yöntemi
Toplam Krom (Cr)	EPA 200.7	ICP-OES Yöntemi
Toplam Siyanür (CN^-)	SM 4500 CN^- C ve SM 4500 CN^- E	Distilasyon + Spektrofotometrik Yöntem
Florür (F^-)	TS EN ISO 10304-1	Iyon Kromatografi Yöntemi
Sülfat (SO4)	SM 4500 $SO4^{2-}$ E	Türbidimetrik Yöntem
Balık Biodeneyi	SKKY Numune Alma ve Analiz Metotları Tebliğ Ek 1	Akut Toksisite Testi
Renk	SM 2120 C	Spektrofotometrik Yöntem

3.2.2. Veri Setinin Değerlendirilmesi

Bu çalışma, Organize Sanayi Bölgesi (OSB) içerisindeki endüstriyel tesislerden 2020–2025 yılları arasında alınan kanal numunelerine dayanmakta, ancak modelleme aşamasında yalnızca 2023–2025 dönemindeki gözlemler kullanılmaktadır. Böylece:

- Ölçüm altyapısı ve raporlama pratiklerinin görece oturduğu,
- Mevcut OSB atıksu izleme politikaları ve SKKY limitleri ile uyumlu,
- Zaman içinde rejim değişimine (proses değişiklikleri, yeni tesisler vb.) daha az maruz kalan bir alt dönem analiz edilmiştir.

Ham veri seti, her bir numune için aşağıdaki bilgileri içermektedir:

- Firma adı ve SKKY sektör kod(lar)ı (ör. “SKKY 14.4 Limitleri; SKKY 14.5 Limitleri”),
- Numune tarihi ve numune türü (kanal numunesi – endüstriyel deşarj),

- OSB yönetmeliği/SKKY Tablo 19 ve Tablo 25'e göre hesaplanmış parametre limitleri,
- Laboratuvarda ölçülen atıksu parametreleri:

    - AKM, Toplam Azot, Toplam Fosfor, pH,
    - Yağ-Gres, Fenol, Sülfat (SO_4^{2-}), Toplam Sülfür,
    - Toplam Çinko, Toplam Bakır, Toplam Nikel, Toplam Krom, Toplam Kurşun, Toplam Kadmiyum,
    - Klorür (Cl^-), Yüzey aktif madde, TKN,
    - Renk, Katran ve Petrol Kökenli Yağlar,
    - (KOİ de ölçülmekte, ancak bu çalışmanın hedef ve özellik setinden bilinçli olarak hariç tutulmuştur; bkz. 3.2.3).

Tez düzeyindeki ön işleme adımları sonucunda, tarihler standardize edilmiş, parametre adları normalize edilmiş ve her numune için parametre bazlı limit değerleri ve limit aşımı `(*_exceed)` bayrakları oluşturulmuştur. Bu haliyle elde edilen tablo `Tez_Verileri_preprocessed.xlsx` olarak kaydedilmiş ve bu çalışma için başlangıç noktası olarak kullanılmıştır.

3.2.3. KOİ Hariç İhlal Etiketi ve Hedef Değişken Tanımı

OSB verilerinin ilk taramasında, mevzuat ihlallerinin büyük çoğunluğunun KOİ limit aşımından kaynaklandığı görülmüştür. Bu durum, diğer parametrelerdeki ihlalleri görece "gölgeleyen" bir etki yaratmakta ve modelin özellikle ağır metal, yağ-gres, sülfür gibi kritik parametreleri ikincil önemde görme riskini doğurmaktadır.

Bu nedenle, çalışmada iki adımlı bir etiketleme yapılmıştır:

1. Her numune için, parametre bazlı limit aşımı bayrakları kullanılarak KOİ dışındaki tüm parametrelerdeki aşım durumu hesaplanmıştır.
    - `*_exceed` kolonları incelenmiş, `KOİ_exceed` hariç tüm exceed kolonları OR operatörü ile birleştirilmiştir.
2. Bu bileşik bayrak, yeni bir ikili etiket olarak tanımlanmıştır:
    - `non_compliant_excl_KOİ = 1` → Numune, KOİ dışındaki en az bir parametrede limit aşımı içermektedir.
    - `non_compliant_excl_KOİ = 0` → Tüm parametreler limit altındadır (KOİ ihlali olsa bile bu etikette dikkate alınmamaktadır).

Araştırma sorusu gereği, tahmin edilmek istenen büyüklük bir sonraki numunenin ihlal durumu olduğundan, her firma için tarih sırasına göre:

- `non_compliant_excl_KOİ_next` etiketi tanımlanmış,
- Her satır (t_k) için hedef değişken, aynı firmanın bir sonraki numunesindeki (t_{k+1}) `non_compliant_excl_KOİ` değeri olarak atanmıştır.

Son numune için bir "sonraki numune" olmadığı için, bu gözlemler modelleme setinden çıkarılmıştır.
2023–2025 dönemi ve KOİ hariç etiket tanımı ile birlikte, hedef değişken dağılımı daha dengeli bir hal almış**,** eğitim setinde yaklaşık:

- `non_compliant_excl_KOİ_next = 1` oranı ≈ %66,
- `non_compliant_excl_KOİ_next = 0` oranı ≈ %34

seviyesine gelmiştir. Böylece aşırı sınıf dengesizliğinin (ilk durumda ≈ %93 ihlal) getireceği kara-kutu davranışlar önemli ölçüde azaltılmıştır.

3.2.4. Zamansal Özellik Çıkarımı

Çalışmanın hedefi, **"**bugünkü atıksu profiline bakarak, bir sonraki numunede (t_{k+1}) KOİ dışı mevzuat ihlali riski var mı?**"** sorusuna cevap vermektir. Bu nedenle, her firma için zamana bağlı süreç aşağıdaki şekilde operasyonalize edilmiştir:

- Her firma için kayıtlar tarih bazında sıralanmış,
- Her satır, t_k anındaki ölçümler ve geçmiş numunelerin istatistikleri ile tanımlanmış,
- Hedef değişken ise aynı firmanın bir sonraki numunesi (t_{k+1}) için `non_compliant_excl_KOİ_next` etiketi olarak alınmıştır.

Bu amaçla, KOİ hariç seçilmiş parametreler için üç tür zamansal özellik üretilmiştir:

1. **Anlık (t_k) parametre değerleri**

AKM, Toplam Azot, Toplam Fosfor, pH, Yağ-Gres, Fenol, Sülfat (SO_4^{2-}), Toplam Sülfür, Toplam Çinko, Toplam Bakır, Toplam Nikel, Toplam Krom, Toplam Kurşun, Toplam Kadmiyum, Klorür (Cl^-), Yüzey aktif madde, TKN, Renk, Katran ve Petrol Kökenli Yağlar.

2. **Kısa dönem geçmiş özellikleri (lag + rolling)**

Her bir parametre için:

- **Lag özellikleri:**
  - $X_lag1 = X(t_{k-1})$,
  - $X_lag2 = X(t_{k-2})$,
  - $X_lag3 = X(t_{k-3})$.
- **Son üç numunenin hareketli ortalaması:**
  - $X_mean_last3 = mean\{ X(t_k), X(t_{k-1}), X(t_{k-2}) \}$

Bu yapı, hem en güncel numunenin kirlilik profilini, hem de hemen önceki birkaç numunedeki eğilimi (artış/azalış) modele yansıtmaktadır.

Ayrıca, KOİ hariç ihlal etiketi için de lag özellikleri oluşturulmuştur:

- `non_compliant_excl_KOİ_lag1`, `lag2`, `lag3`.

3. **Uzun dönem kümülatif özellikler**

Her firma için, numune sırası boyunca aşağıdaki kümülatif özetler hesaplanmıştır:

- **Toplam numune sayısı:**
  - `num_samples_so_far` (firmanın o ana kadar kaç kez ölçüldüğü),
- **KOİ hariç kümülatif ihlal sayısı ve oranı:**
  - `cum_violations_excl_KOİ`
  - `cum_violation_ratio_excl_KOİ = cum_violations_excl_KOİ / num_samples_so_far`,
- **Parametre bazlı kümülatif ortalamalar:**
  - $cum_mean_X = mean\{ X(t_1), \ldots, X(t_k) \}$ Tüm parametreler için hesaplanmıştır.

Bu tasarım sayesinde, model her satırda:

- Anlık kirlilik seviyesini (t_k)**,**
- Kısa dönemli salınımları (lag ve hareketli ortalama)**,**
- Uzun dönemli kirlilik profili ve ihlal geçmişini (kümülatif oranlar)

aynı anda görebilmektedir. İlk üç numune (yeterli lag/rolling bilgisi olmadığı için) ve son numune (t_{k+1} etiketi tanımlanamadığı için) modelleme setinden düşürülmüştür.

3.2.5. Özellik Seti ve Eksik Değerlerin İşlenmesi

Zamansal özellik çıkarımı sonrasında, her satır için:

- Kimlik ve meta veriler: `firm`, `skky_codes`, `date`,
- Etiketler: `non_compliant`, `non_compliant_excl_KOİ`, `non_compliant_excl_KOİ_next`,
- Limit ve limit aşım bayrakları: `*_limit`, `*_exceed`,
- Üretilmiş tüm sayısal özellikler (anlık, lag, rolling, kümülatif)

mevcut hale gelmiştir.

Model girişinde:

- Kimlik ve meta değişkenler (firma adı, SKKY kodları, tarih) ile,
- Limit ve exceed bayrakları,
- Etiket kolonları

özellik setinden dışlanmıştır. Geriye kalan sayısal kolonlar, modelleme özellik seti (X) olarak kullanılmıştır.

Eksik değerler için:

- Global modelde**,** her özellik için eğitim seti medyanı hesaplanmış,
- Eğitim setindeki eksikler, ilgili sütunun eğitim medyanı ile doldurulmuş,
- Data leakage'i engellemek için test seti için de aynı medyan vektörü kullanılmıştır.

Böylece, hem farklı firmalar arasında ölçülen/ölçülmeyen parametrelerin yarattığı boşluklar, hem de düzensiz numune alma pratiklerinden kaynaklanan eksiklikler, tutarlı ve denetimli bir şekilde ele alınmıştır.

3.3. EĞİTİM-TEST AYRIMI VE MODELLEME YAKLAŞIMI

3.3.1. Firma Bazlı Zaman Sıralı 80/20 Ayrımı

Zaman bağımlılığını ve firma içi korelasyonları koruyabilmek için, eğitim–test ayrımı firma bazlı ve kronolojik yapılmıştır:

- Her firma için numuneler tarih sırasına göre sıralanmış,
- Gözlem sayısı 5'in altında olan firmalar tamamen eğitim setine alınmış (testte çok az gözlem oluşmasını önlemek için),
- Diğer firmalarda:
    - İlk ~%80'lik bölüm eğitim**,**
    - Son ~%20'lik bölüm test seti olarak ayrılmıştır.

Bu strateji sayesinde:

- Model, geleceğe yönelik tahmin senaryosunu daha gerçekçi bir şekilde taklit etmekte (geçmiş numunelerle eğitilerek, daha sonraki numuneleri tahmin etmektedir),
- Aynı firmanın verilerinin train ve test setinde karışmasına izin verilmekte, ancak zaman sırası korunmaktadır; böylece hem genellenebilirlik hem de veri etkinliği sağlanmaktadır.

3.3.2. Modeller ve Hiperparametreler

Çalışmada dört farklı sınıflandırma modeli değerlendirilmiştir:

1. **Lojistik regresyon (baseline)**

    - Doğrusal karar sınırı sağlayan, yorumlanabilir bir temel model,
    - Sınıf dengesizliğini dengelemek üzere `class_weight = "balanced"` seçeneği kullanılmıştır.

2. **Random Forest**

    - 300 ağaçtan oluşan, maksimum derinliği sınırlandırılmamış,
    - `min_samples_split = 5`, `class_weight = "balanced"` ayarları ile,
    - Karmaşık, doğrusal olmayan ilişkileri yakalayabilen bir topluluk modeli.

3. **XGBoost**

    - 400 ağaç, max_depth = 5, learning_rate = 0.05,
    - subsample = 0.8, colsample_bytree = 0.8 ile düzenlileştirilmiş bir boosting yaklaşımı,
    - Sınıf ağırlığı varsayılan bırakılmış; dengesizlik sorununu veri ön-işleme seviyesinde (KOİ hariç etiket ve 2023–2025 filtrelemesi) hafifletmek hedeflenmiştir.

4. **LightGBM**

    - 400 ağaç, learning_rate = 0.05,
    - subsample = 0.8, colsample_bytree = 0.8,
    - class_weight = "balanced" seçeneği ile, hafif ve hızlı bir gradient boosting uygulaması.

Tüm modeller Python ortamında, scikit-learn, XGBoost ve LightGBM kütüphaneleri kullanılarak uygulanmıştır.

3.3.3. Değerlendirme Metrikleri

Sınıflar tam dengeli olmadığından ve asıl ilgi odağı ihlalli numuneleri (pozitif sınıf) yakalamak olduğundan, değerlendirmede aşağıdaki metrikler kullanılmıştır:

- Doğruluk (accuracy),
- Pozitif sınıf için:
    - Hassasiyet (precision_pos),
    - Duyarlılık / yakalama oranı (recall_pos),
    - F1 skoru (f1_pos),
- ROC-AUC (Receiver Operating Characteristic alanı),

- PR-AUC (Precision–Recall eğrisi alanı).

Özellikle PR-AUC ve pozitif sınıf F1 skoru, dengesiz sınıf yapısında modelin ihlalli numuneleri ne kadar istikrarlı ve güvenilir biçimde yakalayabildiğini değerlendirmek için temel alınmıştır.

3.3.4. Sektör Bazlı Modelleme ve Özellik Önemi Analizi

Araştırma sorusu 3 kapsamında, SKKY sektör gruplarına göre limit aşımı riskini açıklayan temel parametrelerin değişimi incelenmiştir. Bunun için:

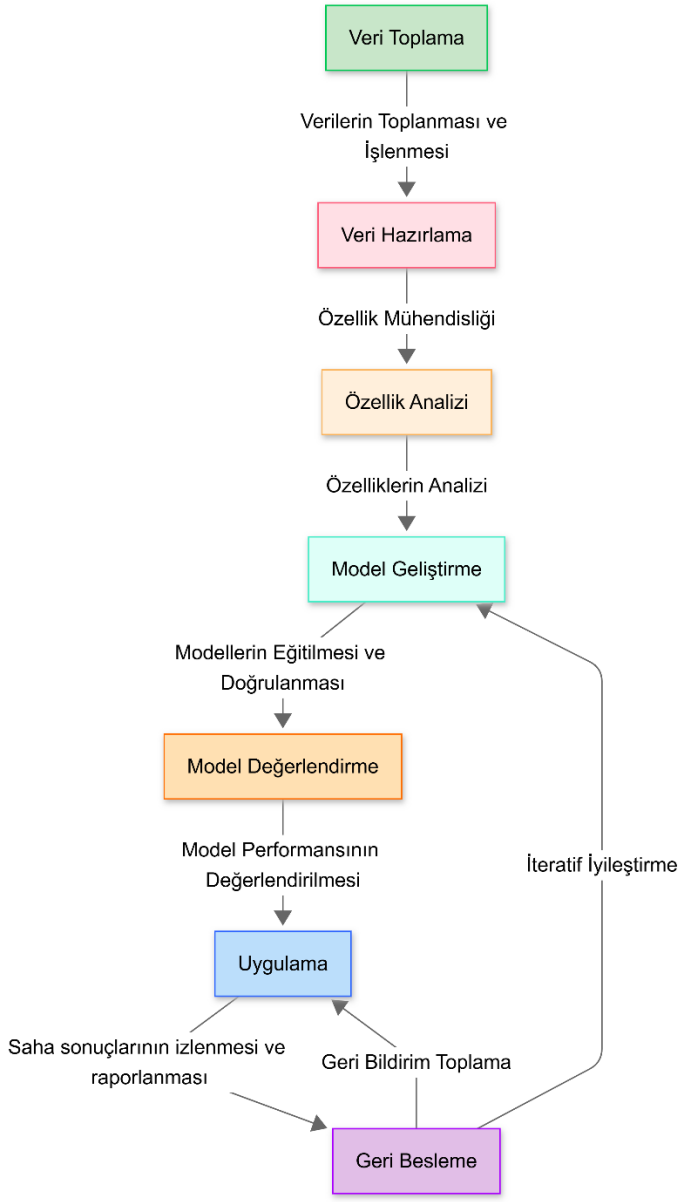
1. `skky_codes` alanında birden fazla sektör kodu olan firmalarda kodlar ayrıştırılmış ve satırlar explode edilerek,
    - Örneğin "`SKKY 14.4 Limitleri; SKKY 14.5 Limitleri`" kaydı iki ayrı satıra dönüştürülmüş,
    - Böylece aynı numune ilgili olduğu tüm sektör gruplarına dahil edilmiştir.
2. Her SKKY sektörü için:
    - Aynı özellik seti kullanılmış,
    - Yine firma bazlı zaman sıralı 80/20 eğitim–test ayrımı uygulanmış,
    - Sektör özelinde Random Forest modeli eğitilmiştir.
3. Yeterli veri olmayan/tek sınıfa yakın dağılıma sahip sektörler elenmiş; sonuçta:
    - SKKY 15.2,
    - SKKY 6.1,
    - SKKY 14.4,
    - SKKY 14.5,
    - SKKY 14.10

sektör grupları için anlamlı sonuçlar elde edilmiştir.

Global model ve sektör modelleri için özellik önemi analizleri Random Forest'ın önem skorları üzerinden üretilmiş ve hem global ölçekte hem de sektör bazında karşılaştırılmıştır.

Bu çalışmada izlenen veri temelli analiz ve modelleme yaklaşımı Şekil 3.3.'te şematik olarak sunulmuştur. Metodoloji, endüstriyel atıksu verilerinin toplanması ve ön işlenmesi ile başlamış; veri hazırlama ve özellik mühendisliği adımlarının ardından seçilen değişkenler analiz edilmiştir. Elde edilen özellikler kullanılarak modeller geliştirilmiş, eğitim ve doğrulama aşamalarından sonra performansları değerlendirilmiştir. Model çıktıları saha sonuçları ile

karşılaştırılarak izlenmiş ve elde edilen geri bildirimler doğrultusunda süreç iteratif olarak iyileştirilmiştir.



Şekil 3.3. Bu Çalışmada Uygulanan Veri Temelli Modelleme Sürecinin Genel Akış Şeması

Bu çalışmada, KOİ dışı mevzuat ihlali riskinin tahmin edilmesi amacıyla farklı varsayımlara ve öğrenme mekanizmalarına sınıflandırma modelleri kullanılmıştır. Tez çalışmasındaki atıksu verilerinin değerlendirilmesi amacıyla farklı modellerin seçimi, yorumlanabilirlik, karmaşık ve doğrusal olmayan ilişkileri yakalama ve değerlendirmesi sağlayacak şekilde yapılmıştır. Bu bağlamda, tez çalışmasında lojistik regresyon modeli temel (baseline) yaklaşım çalışılmıştır; elde edilen sonuçlar, daha gelişmiş topluluk ve boosting tabanlı yöntemlerle karşılaştırılmıştır.

Lojistik regresyon, katsayılar üzerinden parametre etkilerinin yorumlanabilmesi ve doğrusal karar sınırı sunması nedeniyle referans model olarak kullanılmıştır. Ayrıca, Random Forest modeli, çok sayıda karar ağacının bir araya getirilmesiyle oluşturulan bir topluluk yaklaşımı olarak, deneysel değişkenler arasındaki karmaşık ilişki ve doğrusal olmayan etkileşimleri modelleyebilmektedir. Gradient boosting tabanlı XGBoost ve LightGBM modelleri ise, ardışık ağaç yapıları aracılığıyla hata odaklı öğrenme gerçekleştirmesiyle beraber yüksek tahmin yeteneği ve esneklik sağlayan yöntemler olarak görülmektedir. Bu modeller, karmaşıklığın azaltılması ve örnekleme stratejileri sayesinde aşırı öğrenmeye karşı daha dayanıklı bir durum sağlamaktadır. Çalışmada kullanılan modeller, sınıf dengesizliği problemini dikkate alacak şekilde oluşturulmuş ve her bir algoritma, erken uyarı amacıyla yanlış negatiflerin minimize edilmesi hedefi doğrultusunda incelenmiştir. Böylece, farklı model sınıflarının mevzuat ihlali riskini öngörme konusundaki güçlü ve zayıf yönleri sistematik olarak araştırılmıştır.

4. BULGULAR

3. Bölümde tanımlanan veri yapısı ve modeller kullanılarak elde edilen bulgular, üç araştırma sorusu çerçevesinde sunulmaktadır. Önce global modelin bir sonraki numunedeki KOİ dışı ihlal durumunu ne ölçüde öngörebildiği (AS1) incelenmiş, ardından global ölçekte ihlal riskini açıklayan başlıca parametreler ve zamansal özellikler değerlendirilmiştir (AS2). Son olarak, SKKY sektör grupları için kurulan modeller aracılığıyla, limit aşımı riskini açıklayan temel parametrelerin sektörel farklılaşması analiz edilmiştir (AS3).

4.1. GLOBAL MODEL PERFORMANSI (ARAŞTIRMA SORUSU 1)

Global modelde, tüm firmalar bir arada ele alınarak 2023–2025 döneminde bir sonraki numunenin ihlal durumu tahmin edilmiştir. Eğitim setinde pozitif sınıf oranı ≈ %66 civarındadır; yani veri ılımlı düzeyde dengesiz olup, ihlalli numuneler ağır basmaktadır.

Dört modelin test seti üzerindeki performans metrikleri Tablo 4.1'de özetlenmiştir. Random Forest, hem en yüksek doğruluk (0,854) hem de en yüksek F1 skoru (0,884) değerlerini sağlamaktadır. Özellikle pozitif sınıf duyarlılığının 0,948 gibi çok yüksek bir seviyede olması, KOİ dışı ihlalli numuneleri yakalama konusunda modelin son derece başarılı olduğunu göstermektedir. OSB bağlamında, ihlalli numunelerin kaçırılmasının maliyeti yüksek olduğu için, bu yüksek recall düzeyi önemli bir avantajdır. XGBoost ve LightGBM modelleri de benzer ROC-AUC ve PR-AUC değerleri ile güçlü performans sergilemekte; doğruluk ve F1 skorları Random Forest'ın biraz altında kalmaktadır. Lojistik regresyon ise doğrusal bir temel model olarak, beklenenden daha iyi bir baz çizgi sunmakla birlikte, karmaşık ilişkileri yakalama gücü sınırlıdır. Bu sonuçlar, Araştırma Sorusu 1 kapsamında, "bugünkü atıksu profiline bakarak bir sonraki numunedeki KOİ dışı ihlal riskini öngörme" hedefinin, Random Forest tabanlı global model ile uygulamaya geçirilebilir bir doğruluk düzeyiyle gerçekleştirilebildiğini göstermektedir.

Tablo 4.1. ML modellerinin test seti üzerindeki performans metrikleri

Model	accuracy	precision_pos	recall_pos	f1_pos	roc_auc	pr_auc
Logistic Regression	0,817	0,848	0,839	0,843	0,876	0,880
Random Forest	0,854	0,829	0,948	0,884	0,877	0,884
XGBoost	0,834	0,820	0,919	0,867	0,868	0,877
LightGBM	0,841	0,827	0,922	0,872	0,872	0,880

4.2. GLOBAL ÖLÇEKTE LİMİT AŞIMI RİSKİNİ AÇIKLAYAN TEMEL PARAMETRELER (ARAŞTIRMA SORUSU 2)

Araştırma Sorusu 2 kapsamında, KOİ hariç mevzuat ihlal riskini açıklayan temel etmenleri belirlemek için, Random Forest global modelinden elde edilen özellik önemleri analiz edilmiştir. Random Forest, Tablo 4.1'de gösterildiği üzere, en yüksek doğruluk ve F1 skorlarını sağlaması nedeniyle, limit aşımı riskini açıklayan temel parametrelerin yorumlanmasında ana model olarak seçilmiştir. Tablo 4.2'de, toplam önem skorunun yaklaşık %65'ini açıklayan ve en yüksek önem skoruna sahip ilk 15 özellik listelenmektedir. Bu özellikler, büyük ölçüde üç ana grupta toplanmaktadır:

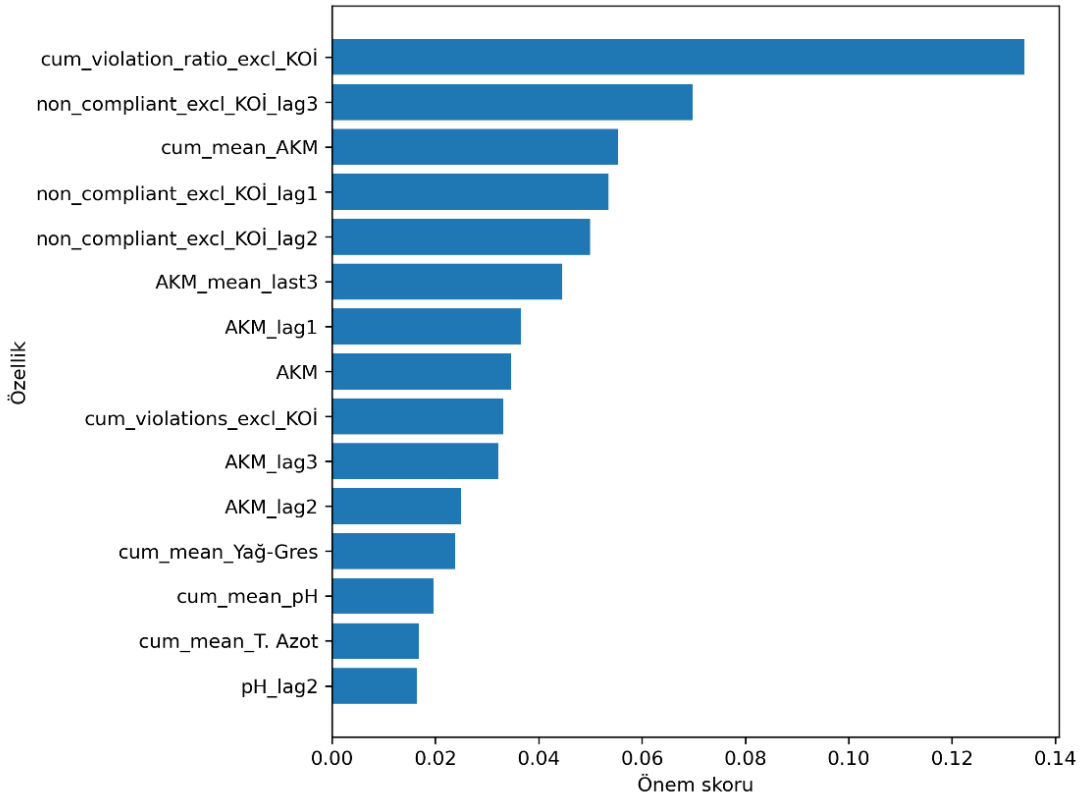
(i) ihlal geçmişini temsil eden etiket tabanlı değişkenler,
(ii) AKM'ye ilişkin kısa ve uzun dönemli yük göstergeleri ve
(iii) Yağ–Gres, pH ve toplam azota ilişkin kümülatif ortalamalar.

Tablo 4.2. En Yüksek Önem Skoruna Sahip İlk 15 Özellik

Önem Sırası	Özellik	Özelliğin Tanımı	Önem Skoru
1	cum_violation_ratio_excl_KOİ	İlgili firmada bugüne kadar alınan tüm numunelerde KOİ hariç parametrelerde yaşanan ihlal sayısının, toplam numune sayısına oranı	0,134
2	non_compliant_excl_KOİ_lag3	İlgili firmada üç önceki numunenin (t_{k-3}) KOİ hariç parametreler açısından ihlalli (1) veya uyumlu (0) olup olmadığı	0,070
3	cum_mean_AKM	İlgili firmada bugüne kadar alınan tüm numunelerde ölçülen AKM değerlerinin kümülatif ortalaması	0,055
4	non_compliant_excl_KOİ_lag1	Bir önceki numunenin (t_{k-1}) KOİ hariç parametreler açısından ihlalli (1) veya uyumlu (0) olması	0,054
5	non_compliant_excl_KOİ_lag2	İki önceki numunenin (t_{k-2}) KOİ hariç parametreler açısından ihlalli (1) veya uyumlu olması	0,050

Tablo 4.3 (Devam)

Önem Sırası	Özellik	Özelliğin Tanımı	Önem Skoru
6	AKM_mean_last3	Son üç numunede ölçülen AKM değerlerinin (t_k, t_{k-1}, t_{k-2}) hareketli ortalaması	0,045
7	AKM_lag1	Bir önceki numunedeki (t_{k-1}) AKM değeri	0,037
8	AKM	Tahmin edilen bir sonraki numuneye referansla, bugünkü numunenin (t_k) AKM değeri	0,035
9	cum_violations_excl_KOİ	İlgili firmada bugüne kadar KOİ hariç parametrelerde gerçekleşen toplam ihlal sayısı	0,033
10	AKM_lag3	Üç numune önceki (t_{k-3}) AKM değeri	0,032
11	AKM_lag2	İki numune önceki (t_{k-2}) AKM değeri	0,025
12	cum_mean_Yağ-Gres	İlgili firmada bugüne kadar ölçülen tüm Yağ–Gres değerlerinin kümülatif ortalaması	0,024
13	cum_mean_pH	İlgili firmada bugüne kadar ölçülen tüm pH değerlerinin kümülatif ortalaması	0,020
14	cum_mean_T. Azot	İlgili firmada bugüne kadar ölçülen tüm toplam azot (T. Azot) değerlerinin kümülatif ortalaması	0,017
15	pH_lag2	İki önceki numunedeki (t_{k-2}) pH değeri	0,016
Toplam			0,646



Şekil 4.1. RF Modelinden Elde Edilen En Yüksek Önem Skoruna Sahip İlk 15 Özellik

Şekil 4.1.'de RF modelde kullanılan değişkenlerde en yüksek önem skoruna ait ilk 15 özellik sunulmuştur. Grafik, özellikle KOİ hariç parametrelere ilişkin kümülatif ve gecikmeli uygunsuzluk göstergeleri ile AKM'ye bağlı değişkenlerin, ihlal davranışının açıklanmasında önemli rol oynadığını göstermektedir.

4.2.1. İhlal Geçmişi: Kümülatif İhlal Oranı ve Lag Etiketler

En yüksek önem skoruna sahip değişken, `cum_violation_ratio_excl_KOİ`'dir (önem skoru 0,134). Bu değişken, ilgili firmanın bugüne kadar aldığı tüm numuneler içinde KOİ dışı ihlalli numunelerin oranını temsil etmektedir. Başka bir ifadeyle, firmanın kronik KOİ dışı ihlal eğilimini sayısallaştırmaktadır. Random Forest modelinin bu değişkeni birinci sıraya yerleştirmesi, tarihsel ihlal oranının, gelecekteki ihlal riskini tahmin etmede tek başına en güçlü gösterge olduğunu ortaya koymaktadır.

Bu göstergeyi, ihlal geçmişine ilişkin diğer etiket tabanlı özellikler izlemektedir. `non_compliant_excl_KOİ_lag1`, `non_compliant_excl_KOİ_lag2` ve `non_compliant_excl_KOİ_lag3`, sırasıyla bir, iki ve üç önceki numunenin (t_{k-1}, t_{k-2}, t_{k-3}) KOİ dışı ihlalli (1) veya uyumlu (0) olup olmadığını göstermektedir. `cum_violations_excl_KOİ` ise firmanın bugüne kadarki toplam KOİ dışı ihlal sayısını temsil etmektedir (önem sırası 9; önem skoru 0,033).

Bu yapı, modelin ihlal riskini değerlendirirken, yalnızca son ölçümü değil, firmanın ihlal geçmişini hafızasında tuttuğunu göstermektedir. Özellikle son üç numunedeki ihlal durumları ile kümülatif ihlal oranı ve toplam ihlal sayısı birlikte değerlendirildiğinde**,** ihlal yapan firmanın ihlale devam etme eğilimi veri temelli olarak doğrulanmaktadır. Dolayısıyla, operasyonel açıdan bakıldığında, düzenli olarak ihlal yapan firmalar için daha sık numune alma, daha yakın izleme ve hedefli denetim planlaması yapılması gerektiğine işaret etmektedir.

4.2.2. AKM'ye İlişkin Kısa ve Uzun Dönemli Yük Göstergeleri

Tablo 4.2'deki ilk 15 özellik içinde, AKM'ye ilişkin değişkenlerin çok baskın bir şekilde yer aldığı görülmektedir:

- `cum_mean_AKM` (önem sırası 3; 0,055): Firmanın bugüne kadarki AKM kümülatif ortalamasını **t**emsil eder. Yani, AKM'nin yapısal olarak sürekli yüksek veya düşük seyredip seyretmediğini yansıtır.

- `AKM_mean_last3` (6. sıra; 0,045): Son üç numunedeki AKM değerlerinin (t_k, t_{k-1}, t_{k-2}) hareketli ortalamasıdır ve kısa dönemli yüklenmeyi temsil eder.
- `AKM` (8. sıra; 0,035): Bugünkü numunenin (t_k) anlık AKM değeridir.
- `AKM_lag1`, `AKM_lag2`, `AKM_lag3` (7., 11. ve 10. sıralar; önem skorları 0,037; 0,025; 0,032): Sırasıyla bir, iki ve üç önceki numunelerdeki AKM değerleridir.

Bu sonuçlar, KOİ dışı ihlal riskinin önemli bir bölümünün AKM dinamikleri ile yakından ilişkili olduğunu göstermektedir. Hem uzun dönemli AKM ortalaması, hem de son birkaç numunedeki AKM seviyeleri bir sonraki numunenin KOİ dışı ihlalli olup olmayacağını güçlü biçimde belirlemektedir.

Bu durum, sahadaki gözlemle de uyumludur: AKM'nin artması, çoğu zaman proses yüklenmesinin, yetersiz ön arıtmanın veya çökeltim performansındaki düşüşün bir göstergesidir ve birlikte seyrettiği diğer parametrelerde (örneğin yağ-gres, renk, toplam azot, sülfür, yüzey aktif maddeler) KOİ dışı limit aşımlarını tetikleyebilmektedir. Dolayısıyla global model, AKM'yi hem "yapısal kirletici yük" hem de "operasyonel stres göstergesi" olarak kullanmakta ve ihlal riskini buna göre artırmaktadır.

Bu bulgu, OSB özelinde şu tür uygulamalara temel sağlayabilir:

- AKM seviyesinin belirli eşiklerin üzerine çıktığı firmalarda numune alma sıklığının artırılması,
- AKM'de ani artışlar gözlenen firmalar için, proses işletme parametrelerinin ve ön arıtma sistemlerinin daha yakından izlenmesi,
- AKM–yağ-gres–renk gibi parametre kombinasyonlarına dayalı otomatik uyarı kuralları geliştirilmesi.

4.2.3. Yağ Gres, pH ve Toplam Azotun Uzun Dönemli Etkisi

İlk 15 özellik içinde, KOİ dışı ihlallerin kimyasal kompozisyon boyutunu temsil eden üç önemli kümülatif ortalama parametresi de yer almaktadır:

- `cum_mean_Yağ-Gres` (12. sıra; 0,024): Firmanın bugüne kadarki Yağ–Gres değerlerinin kümülatif ortalamasıdır**.** Uzun dönemli yağ-gres yükü yüksek olan firmalar, KOİ dışı ihlal riski açısından modelde daha riskli profiller olarak değerlendirilmektedir.

- `cum_mean_pH` (13. sıra; 0,020): Tüm numunelerdeki pH değerlerinin kümülatif ortalamasını temsil eder. Bu değişken, firmanın atıksu pH'ının genel olarak asidik, nötr veya alkali bir rejimde olup olmadığını gösterir.
- `cum_mean_T. Azot` (14. sıra; 0,017): Firmanın bugüne kadar ölçülen toplam azot (T. Azot) değerlerinin kümülatif ortalamasıdır ve azot içeriği yüksek prosesleri işaret eder.

Bunlara ek olarak:

- `pH_lag2` (15. sıra; 0,016) — iki önceki numunedeki pH değerini temsil eder — modelin pH'daki kısa dönemli sapmaları da dikkate aldığını göstermektedir.

Bu değişkenler birlikte değerlendirildiğinde, Random Forest global modeli KOİ dışı ihlal riskini yalnızca AKM ve ihlal geçmişi ile değil, aynı zamanda yağ–gres yükü, azot içeriği ve pH rejimi ile de ilişkilendirmektedir. Özellikle:

- Yağ–Gres kümülatif ortalaması yüksek olan firmalar,
- Toplam azotu uzun dönemli olarak yüksek seyreden tesisler,
- pH'sı sürekli olarak asidik veya bazik rejime yakın olan prosesler

KOİ dışı ihlal riski açısından daha kritik gruplar olarak öne çıkmaktadır.

4.2.4. Global Modelin Ortaya Koyduğu Genel Desen

Random Forest global modelinin ilk 15 özelliği, toplam önem skorunun yaklaşık %65'ini açıklamaktadır. Bu 15 özellik birlikte değerlendirildiğinde, KOİ dışı mevzuat ihlal riskinin şu üçlü yapı ile en iyi şekilde açıklandığı söylenebilir:

1. İhlal geçmişi:
    - Kümülatif ihlal oranı ve toplam ihlal sayısı,
    - Son üç numunenin ihlal/uyum durumu (lag etiketleri).
2. Kısa ve uzun dönem AKM dinamikleri:
    - AKM kümülatif ortalaması,
    - Son üç numunenin AKM ortalaması,
    - AKM'nin lag1–lag3 değerleri ve anlık (t_k) seviyesi.
3. Kimyasal kompozisyonun uzun dönemli bileşeni:
    - Yağ–Gres, pH ve toplam azotun kümülatif ortalamaları,

- pH'nin kısa dönem lag değeri (pH_lag2).

Bu yapı, Araştırma Sorusu 2 kapsamında, KOİ hariç limit aşımı riskinin yalnızca tek bir parametreye değil, ihlal geçmişi + AKM yükü + kimyasal kompozisyonun kombinasyonuna bağlı olduğunu göstermektedir. Dolayısıyla, OSB'de kurulacak olası bir erken uyarı sisteminde:

- İhlal geçmişi (özellikle kümülatif ihlal oranı),
- AKM seviyeleri ve AKM'nin kısa/uzun dönem davranışı,
- Firmaya özgü yağ–gres, pH ve azot profilleri

birlikte ele alınarak, firma bazlı risk skorlarının dinamik olarak güncellenmesi mümkündür.

4.3. SEKTÖR BAZLI MODELLER VE LİMİT AŞIMI RİSKİNİ AÇIKLAYAN TEMEL PARAMETRELERİN SEKTÖREL FARKLILAŞMASI (ARAŞTIRMA SORUSU 3)

Araştırma Sorusu 3 kapsamında, global Random Forest modelinin sektörel alt gruplara ayrıldığında nasıl davrandığı ve KOİ hariç mevzuat ihlal riskini açıklayan değişkenlerin sektör bazında nasıl farklılaştığı incelenmiştir. Bunun için, SKKY kodları kullanılarak her numune, bağlı olduğu sektör grubuna atanmış; birden fazla SKKY tablosuna giren firmaların gözlemleri ilgili tüm sektör gruplarına çoğaltılmıştır. Yeterli örneklem büyüklüğüne sahip sektörler için, global model ile aynı feature set ve aynı eğitim stratejisi kullanılarak sektör bazlı Random Forest modelleri kurulmuştur.

4.3.1. Sektör Bazlı Veri Kapsamı ve Örneklem Büyüklükleri

Öncelikle, 2023–2025 döneminde yeterli gözlem sayısına ulaşmış sektör grupları belirlenmiştir. Bu amaçla, her bir sektör için eğitim ve test setlerindeki gözlem sayıları hesaplanmış; hem eğitim hem de doğrulama için anlamlı istatistiksel çıkarıma imkân verecek şekilde en yüksek örneklem büyüklüğüne sahip beş sektör grubu seçilmiştir. Seçilen sektör grupları ve örneklem büyüklükleri Tablo 4.3'te özetlenmektedir.

Tablo 4.4. Sektör bazlı Random Forest modelleri için train/test örneklem büyüklükleri

sector_group	n_train	n_test
SKKY 14.10 Limitleri	213	54
SKKY 14.4 Limitleri	272	69
SKKY 14.5 Limitleri	232	59
SKKY 15.2 Limitleri	318	82
SKKY 6.1 Limitleri	232	59

4.3.2. Sektör Bazlı Random Forest Performansı

Seçilen beş sektör grubu için kurulan Random Forest modellerinin doğrulama performansları Tablo 4.4'te verilmiştir.

Tablo 4.5. Sektör bazlı Random Forest modellerinin doğrulama performansı

sector_group	accuracy	precision_pos	recall_pos	f1_pos	roc_auc	pr_auc
SKKY 14.10 Limitleri Kimya-Deterjan Üretimi	0,741	0,727	0,941	0,821	0,849	0,917
SKKY 14.4 Limitleri Kimya-Boya	0,826	0,804	0,978	0,882	0,758	0,827
SKKY 14.5 Limitleri Kimya-Boya Hammade	0,831	0,800	1,000	0,889	0,729	0,819
SKKY 15.2 Limitleri Metal-Metal hazırlama	0,890	0,855	1,000	0,922	0,923	0,947
SKKY 6.1 Limitleri Gıda-Meşrubat üretimi	0,763	0,763	1,000	0,865	0,398	0,745

Tablo 4.4'e göre:

- Doğruluk (accuracy) değerleri tüm sektörlerde 0,74–0,89 aralığındadır. En yüksek doğruluk SKKY 15.2 Limitleri sektöründe (0,890) elde edilmiştir.
- Pozitif sınıf için F1-skoru, yani "bir sonraki numunede ihlal olacak mı?" sorusunu ne kadar dengeli tahmin ettiğimiz, tüm sektörlerde 0,82'nin üzerinde, özellikle SKKY 15.2 ve SKKY 14.5 gruplarında sırasıyla 0,922 ve 0,889'dur.
- Recall_pos değerleri 0,94–1,00 aralığındadır; yani modeller, test setindeki ihlalli numunelerin çok büyük bir kısmını başarılı şekilde yakalamaktadır. SKKY 14.5, SKKY 15.2 ve SKKY 6.1 sektörlerinde pozitif sınıf için kaçırılan ihlal yoktur (recall_pos = 1,000).
- PR-AUC (Precision–Recall eğrisi altındaki alan) değerleri 0,745–0,947 arasında değişmekte olup, özellikle SKKY 15.2 (0,947) ve SKKY 14.10 (0,917) sektörlerinde

yüksek seviyededir. Bu metrik, ihlal sınıfının göreli azınlıkta olduğu durumlarda ROC-AUC'den daha açıklayıcıdır.

- ROC-AUC değerleri çoğu sektörde kabul edilebilir düzeydeyken (0,729–0,923), SKKY 6.1 sektöründe nispeten düşüktür (0,398). Bu durum, bu sektörde skor dağılımının daha dengesiz olduğunu veya karar eşiği sabitken sınıflar arası ayrımın ROC açısından daha zor olduğunu göstermektedir. Buna karşın, aynı sektör için F1-skorunun 0,865 ve PR-AUC'nin 0,745 olması**,** uygulama açısından ihlal yakalama performansının yine de tatmin edici olduğunu ortaya koymaktadır.

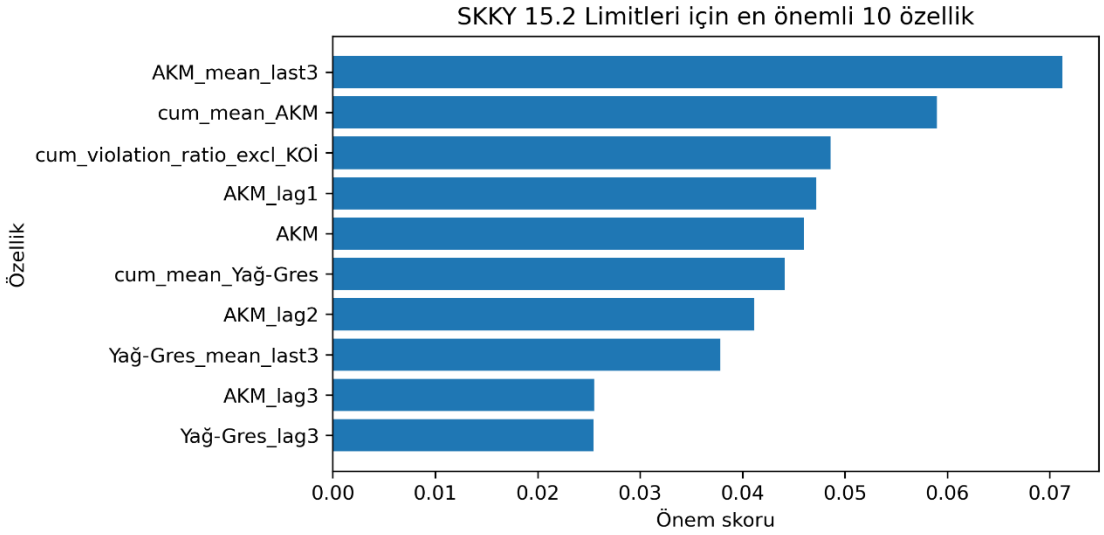
4.3.3. Sektör Bazında Öne Çıkan Parametreler

Tablo 4.5-4.9, veri setinde temsil gücü yüksek olan beş ana sektör grubu için Random Forest modelinin göreli önem skorlarına göre sıralanmış ilk 10 özelliğini sunmaktadır.

Tablo 4.6. SKKY 15.2 Limitleri için en önemli 10 özellik (Random Forest)

Sıra	Özellik	Önem Skoru
1	AKM_mean_last3	0.071
2	cum_mean_AKM	0.059
3	cum_violation_ratio_excl_KOİ	0.049
4	AKM_lag1	0.047
5	AKM	0.046
6	cum_mean_Yağ-Gres	0.044
7	AKM_lag2	0.041
8	Yağ-Gres_mean_last3	0.038
9	AKM_lag3	0.026
10	Yağ-Gres_lag3	0.025

Tablo 4.5. ve Şekil 4.2.'de verilen bulgular, model performansının ağırlıklı olarak AKM ve Yağ Gres parametrelerine ilişkin kümülatif, ortalama ve gecikmeli (lag) göstergeler tarafından belirlendiği ortaya koymaktadır.

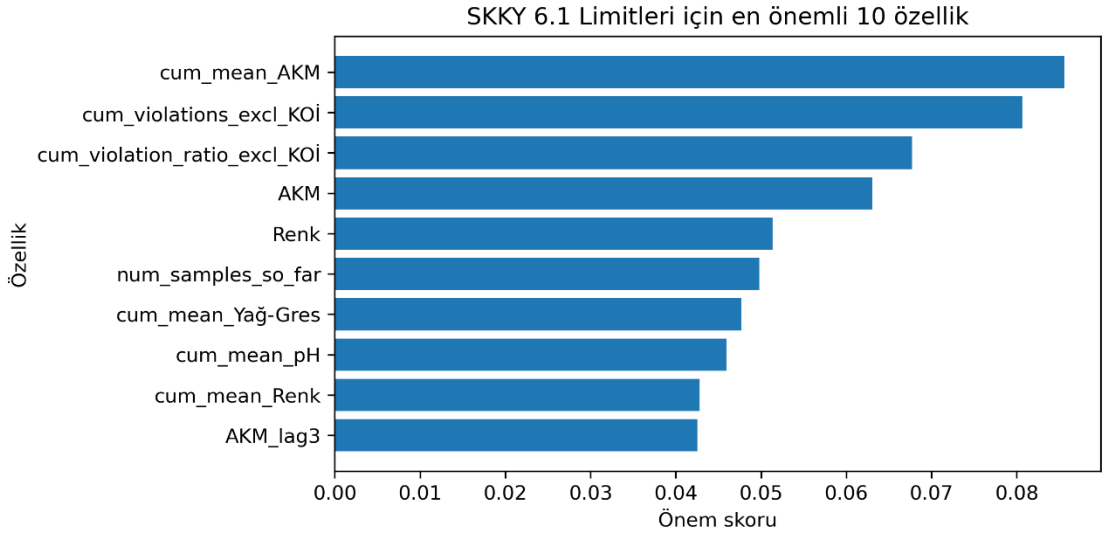


Şekil 4.2. SKKY 15.2 Limitleri için En Önemli 10 Özellik (RF)

Tablo 4.6. ve Şekil 4.3.'te, SKKY Tablo 6.1 deşarj limitleri esas alınarak oluşturulan modelde en yüksek öneme sahip ilk 10 değişkenin önem skorları sunulmaktadır. Bulgular, özellikle AKM ve KOİ hariç parametrelere ilişkin kümülatif ihlal göstergelerinin model çıktısı üzerinde belirleyici olduğunu ortaya koymaktadır.

Tablo 4.7. SKKY 6.1 Limitleri için en önemli 10 özellik (Random Forest)

Sıra	Özellik	Önem Skoru
1	cum_mean_AKM	0.086
2	cum_violations_excl_KOİ	0.081
3	cum_violation_ratio_excl_KOİ	0.068
4	AKM	0.063
5	Renk	0.051
6	num_samples_so_far	0.050
7	cum_mean_Yağ-Gres	0.048
8	cum_mean_pH	0.046
9	cum_mean_Renk	0.043
10	AKM_lag3	0.043

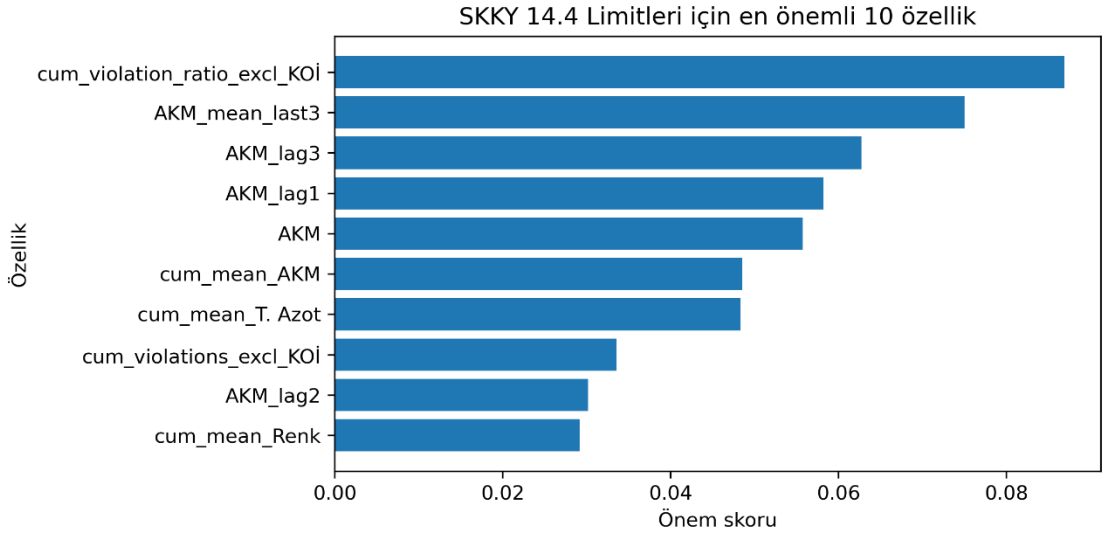


Şekil 4.3. SKKY 6.1 Limitleri için En Önemli 10 Özellik (RF)

Tablo 4.8. SKKY 14.4 Limitleri için en önemli 10 özellik (Random Forest)

Sıra	Özellik	Önem Skoru
1	cum_violation_ratio_excl_KOİ	0.087
2	AKM_mean_last3	0.075
3	AKM_lag3	0.063
4	AKM_lag1	0.058
5	AKM	0.056
6	cum_mean_AKM	0.049
7	cum_mean_T. Azot	0.048
8	cum_violations_excl_KOİ	0.034
9	AKM_lag2	0.030
10	cum_mean_Renk	0.029

Şekil 4.4. ve Tablo 4.7.'de, SKKY Tablo 14.4 deşarj limitleri esas alınarak oluşturulan modelde en yüksek öneme sahip ilk 10 değişkenin önem skorları sunulmaktadır. Bulgular, modelin karar mekanizmasının büyük ölçüde KOİ hariç parametrelere ilişkin kümülatif ihlal oranları ve AKM temelli zamansal göstergeler tarafından belirlendiğini göstermektedir.

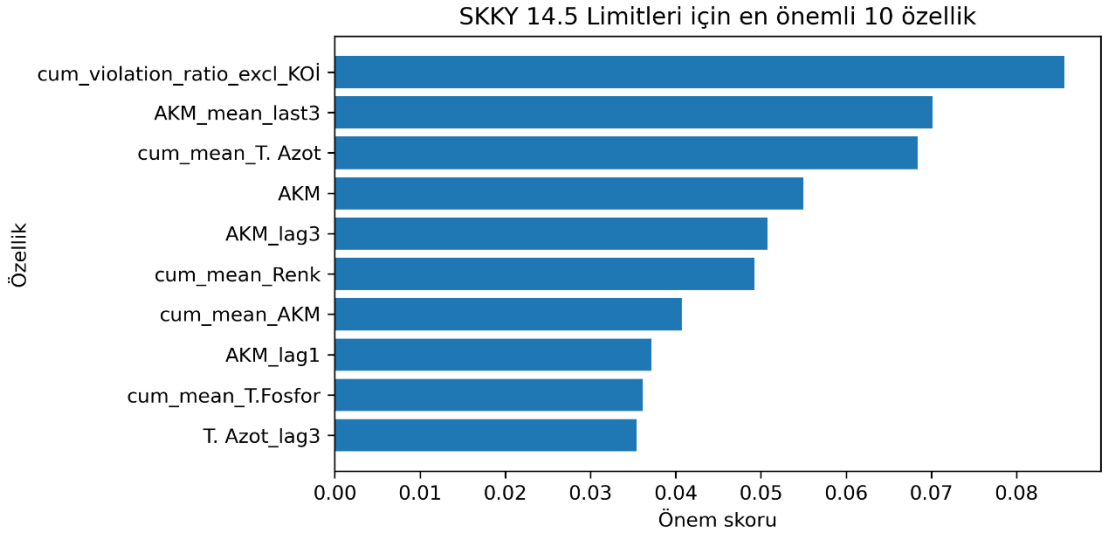


Şekil 4.4. SKKY 14.4 Limitleri için En Önemli 10 Özellik (RF)

Tablo 4.9. SKKY 14.5 Limitleri için en önemli 10 özellik (Random Forest)

Sıra	Özellik	Önem Skoru
1	cum_violation_ratio_excl_KOİ	0.086
2	AKM_mean_last3	0.070
3	cum_mean_T. Azot	0.068
4	AKM	0.055
5	AKM_lag3	0.051
6	cum_mean_Renk	0.049
7	cum_mean_AKM	0.041
8	AKM_lag1	0.037
9	cum_mean_T.Fosfor	0.036
10	T. Azot_lag3	0.035

Şekil 4.5. ve Tablo 4.8.' de, SKKY Tablo 14.5 deşarj limitleri esas alınarak oluşturulan modelde en yüksek öneme sahip ilk 10 değişkenin önem skorları sunulmaktadır. Bulgular, bu limit seti kapsamında KOİ hariç parametrelere ilişkin kümülatif ihlal oranlarının ve AKM ile azot-fosfor bileşenlerine ait zamansal göstergelerin model çıktısını belirgin biçimde etkilediğini ortaya koymaktadır.

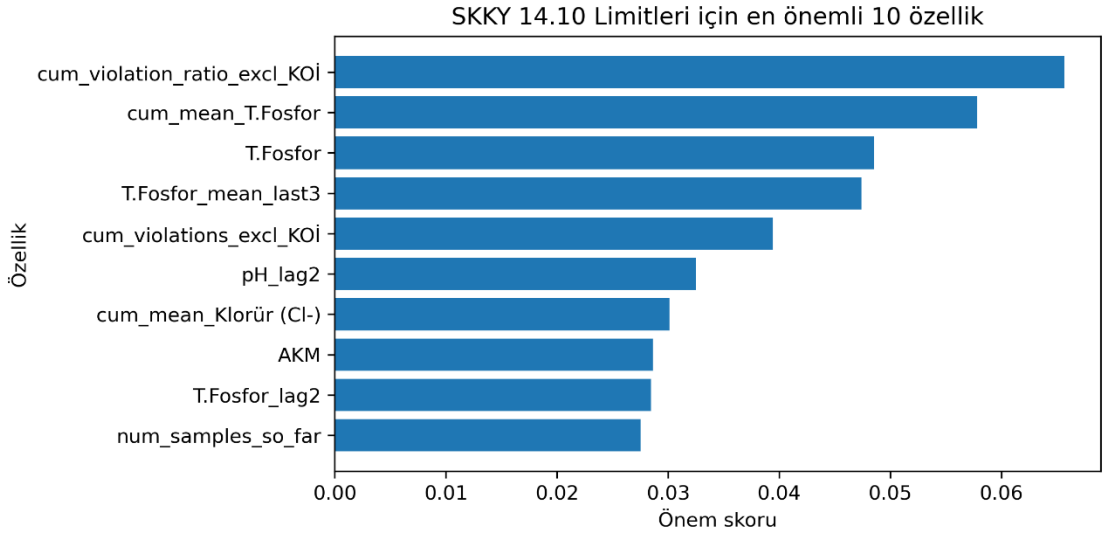


Şekil 4.5. SKKY 14.5 Limitleri için En Önemli 10 Özellik (RF)

Şekil 4.6. ve Tablo 4.9.' da, SKKY Tablo 14.10 deşarj limitleri esas alınarak geliştirilen modelde en yüksek öneme sahip ilk 10 değişkenin önem skorları sunulmaktadır. Bulgular, bu limit setinde fosfor parametresine ait kümülatif, anlık ve gecikmeli göstergelerin model çıktısı üzerinde baskın bir rol oynadığını açık biçimde ortaya koymaktadır.

Tablo 4.10. SKKY 14.10 Limitleri için en önemli 10 özellik (Random Forest)

Sıra	**Özellik**	**Önem Skoru**
1	cum_violation_ratio_excl_KOİ	0.066
2	cum_mean_T.Fosfor	0.058
3	T.Fosfor	0.049
4	T.Fosfor_mean_last3	0.047
5	cum_violations_excl_KOİ	0.039
6	pH_lag2	0.033
7	cum_mean_Klorür (Cl-)	0.030
8	AKM	0.029
9	T.Fosfor_lag2	0.028
10	num_samples_so_far	0.028



Şekil 4.6. SKKY 14.10 Limitleri için En Önemli 10 Özellik (RF)

4.3.4. Ortak Desenler: Sektörler Arasında Değişmeyen Parametreler

Tablolardan görüldüğü üzere, Random Forest ana modelinin sektörel versiyonları arasında bazı çok net ortak desenler ortaya çıkmaktadır:

- **Geçmiş ihlal öyküsü her sektörde kritik:** `cum_violation_ratio_excl_KOİ` ve `cum_violations_excl_KOİ` neredeyse tüm sektörlerde ilk 3–5 özellik arasında yer alıyor. Bu, firmanın ihlal geçmişinin, bir sonraki numunenin mevzuat üstü kalma riskini açıklamada sürekli ve güçlü bir sinyal taşıdığını gösteriyor.
- **AKM tüm sektörlerde baskın:** Hem anlık değer (`AKM`) hem de kısa dönemli dinamikler (`AKM_lag1`, `AKM_lag2`, `AKM_lag3`, `AKM_mean_last3`) hemen her sektörel modelin ilk sıralarında. Bu sonuç, AKM'nin hem SKKY limitleri hem de pratik saha gözlemleri açısından *endüstriyel atıksu kalitesinin ana taşıyıcı parametresi* olduğunu doğruluyor.
- **Uzun dönem ortalamalar kısa dönem lag'leri tamamlıyor:** `cum_mean_AKM`, `cum_mean_Yağ-Gres`, `cum_mean_Renk`, `cum_mean_T. Azot`, `cum_mean_T.Fosfor` gibi *kümülatif ortalamalar* da pek çok sektörde ilk 10'a giriyor. Yani model sadece son üç numunedeki sıçramalara değil, firmanın uzun dönem tipik kirlilik profiline de duyarlı.

4.3.5. Sektöre Özgü Ayırt Edici Parametreler

Bunun yanında, her sektör grubunda üst sıralara çıkan bazı parametreler sektörün proses yapısına uygun ayırt edici parametreler olarak öne çıkıyor:

- **SKKY 15.2 (örneğin metal işleme / makine vb.)**

Bu grupta ilk 10 özelliğin neredeyse tamamı AKM ve Yağ-Gres etrafında toplanmış durumda.

- `AKM_mean_last3`, `AKM_lag1-3`, `AKM`, `cum_mean_AKM`
- `cum_mean_Yağ-Gres`, `Yağ-Gres_mean_last3`, `Yağ-Gres_lag3`

Bu sonuç, proses kaynaklı *katı madde ve yağ-gres yükünün* kısa ve uzun dönem dinamiklerinin, bir sonraki numunenin uyum durumunu belirlemede merkezi rol oynadığını gösteriyor.

- **SKKY 6.1 (gıda/benzeri sektörler)**

Burada da AKM önde olmakla birlikte, liste daha çok parametreli bir yapıya sahip:

- `cum_mean_AKM`, `AKM` yanında
- `Renk`, `cum_mean_Renk`, `cum_mean_pH`, `cum_mean_Yağ-Gres` gibi parametreler üst sıralarda.

Bu durum, özellikle renk, organik yük ve pH dengesinin bir arada önemli olduğu gıda tipi atıksular için beklenen bir desen; model, ihlal riskini tek bir parametreye indirgemek yerine, birden çok kalite göstergesinin kümülatif davranışını dikkate alıyor.

- **SKKY 14.4 ve 14.5 (kimya/organik kimya vb.)**

Her iki grupta da:

- `cum_violation_ratio_excl_KOİ` + `cum_violations_excl_KOİ`
- AKM'nin lag ve rolling özellikleri (`AKM_mean_last3`, `AKM_lag1-3`, `AKM`)
- Ve özellikle azot ve fosfor yükü (`cum_mean_T. Azot`, `cum_mean_T.Fosfor`, `T. Azot_lag3`) en üst sıralarda.

Bu, kimyasal üretim/işleme süreçlerinde hem katı madde yükü hem de azot–fosfor bileşenlerinin mevzuat uyumsuzluğuna birlikte katkı verdiğini; modelin de bu çok bileşenli yapıyı yakalayabildiğini gösteriyor.

- **SKKY 14.10 (yüksek fosfor yükü içeren sektörler)**

Bu sektörde en çarpıcı fark, Toplam Fosfor'un üç farklı zaman perspektifinden ilk 4 sıraya yerleşmesi:

- `cum_mean_T.Fosfor` (uzun dönem ortalama),
- `T.Fosfor` (anlık değer),
- `T.Fosfor_mean_last3` (son 3 numunenin ortalaması),
- ayrıca `T.Fosfor_lag2` de ilk 10 içinde.

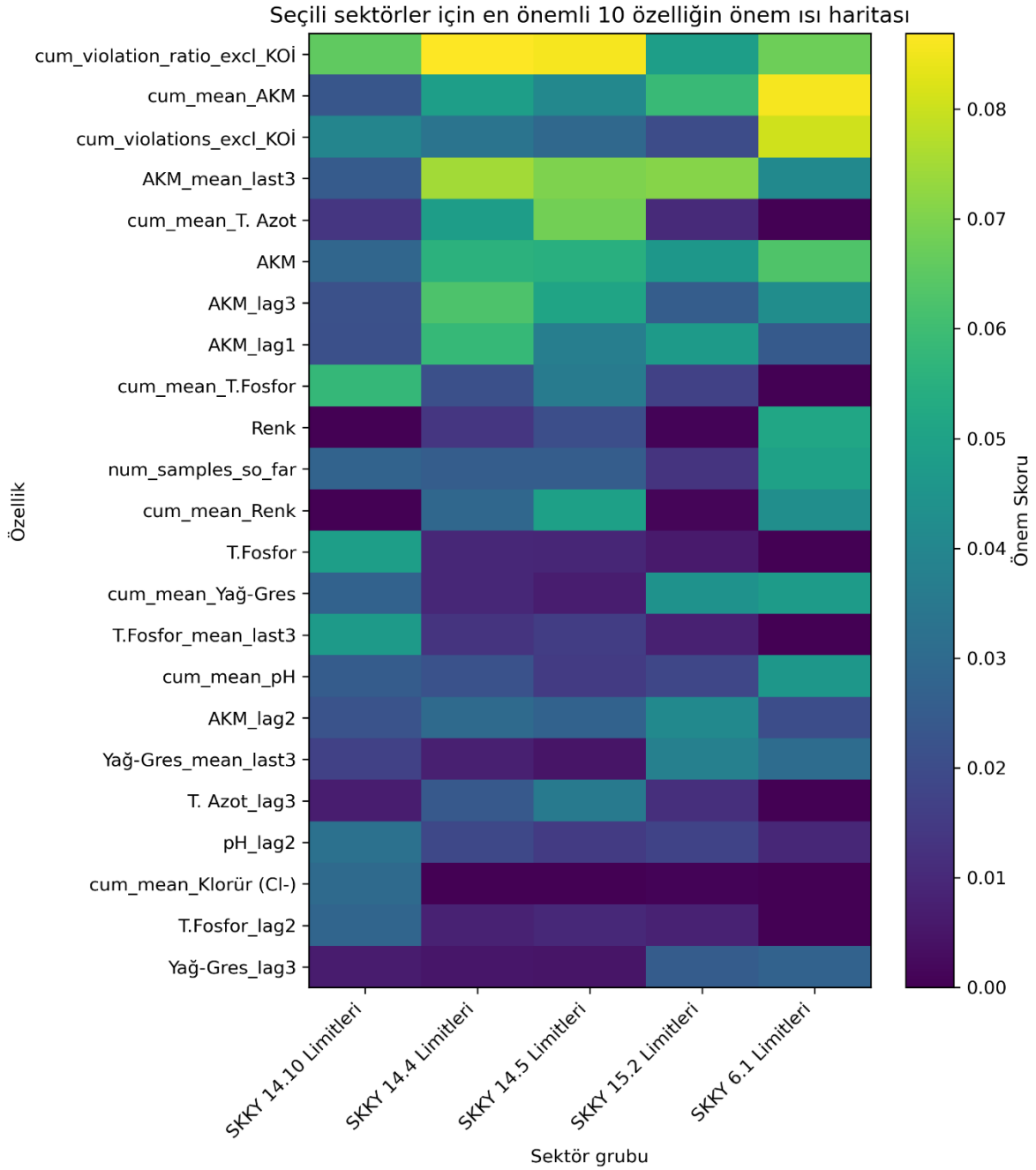
Buna ek olarak `pH_lag2` ve `cum_mean_Klorür (Cl-)` da anlamlı önem skorlarına sahip. Bu kombinasyon, bu sektör grubunda fosfor yükü + pH + klorür bileşiminin ihlal riskini belirleyen temel üçlü olduğunu düşündürüyor.

Şekil 4.7. seçili sektör grupları (SKKY 14.10, 14.4, 14.5, 15.2 ve 6.1) için geliştirilen RF modelde en yüksek öneme sahip 10 değişkenin özellik önem skorlarının karşılaştırmalı ısı haritasını göstermektedir. Bulgular, sektörler arasında ortak ve ayrışan belirleyicilerin açık biçimde ortaya çıktığını göstermektedir.

Isı haritasında en dikkat çekici bulgu, `cum_violation_ratio_excl_KOİ` değişkeninin tüm sektör gruplarında yüksek önem skorlarıyla öne çıkmasıdır. Bu durum, KOİ dışındaki parametrelere ait kümülatif ihlal oranlarının, sektörden bağımsız olarak deşarj limiti ihlalinin en güçlü göstergelerinden biri olduğunu ve sistemlerin belirgin bir geçmiş ihlal hafızası taşıdığını ortaya koymaktadır. Benzer şekilde `cum_mean_AKM` ve `AKM_mean_last3` değişkenlerinin özellikle SKKY 14.4, 14.5 ve 6.1 limitlerinde yüksek önem değerleri alması, AKM yükünün hem uzun hem de kısa vadeli birikiminin birçok sektörde ortak bir risk unsuru olduğunu göstermektedir.

Şekil 4.7 sektör bazlı farklılaşmalar yönünden incelendiğinde; SKKY 14.10 ve 14.5 limitlerinde `T. Fosfor` ve `cum_mean_T.Fosfor` değişkenlerinin belirgin biçimde öne çıktığı, bu sektörlerde fosfor kaynaklı yüklerin uyumsuzluk riskinde kritik rol oynadığı görülmektedir. Buna karşılık SKKY 15.2 limitlerinde `AKM_lag` değişkenleri ve `cum_mean_Yağ-Gres` parametresinin daha baskın hâle gelmesi, bu sektör grubunda partikül ve yağ-gres kaynaklı süreçlerin zamansal etkilerinin ön plana çıktığını göstermektedir. SKKY 6.1 limitlerinde ise `cum_mean_AKM`, `cum_violations_excl_KOİ` ve `num_samples_so_far` değişkenlerinin yüksek önem skorları,

uzun dönem izleme ve veri birikiminin bu sektörlerde karar mekanizmasını güçlendirdiğini ortaya koymaktadır.



Şekil 4.7. Seçili Sektörler için En Önemli 10 Özelliğin Önem Isı Haritası

pH, klorür ve renk gibi parametrelerin ise çoğu sektörde daha düşük fakat sektörlere özgü durumlarda anlamlı önem skorları aldığı; bu parametrelerin doğrudan belirleyici olmaktan ziyade destekleyici göstergeler olarak rol oynadığı anlaşılmaktadır.

Genel olarak ısı haritası, SKKY limitlerine uyumsuzluğun açıklanmasında kümülatif ihlal davranışı, AKM temelli zamansal yüklenme ve sektöre özgü baskın parametrelerin birlikte etkili olduğunu ortaya koymakta; çalışmada kullanılan gecikmeli ve kümülatif değişken temelli yöntemsel yaklaşımın farklı sektör grupları için tutarlı ve ayırt edici sonuçlar ürettiğini açık biçimde göstermektedir.

5. TARTIŞMA

Bu tezde, Organize Sanayi Bölgesi'nde (OSB) yürütülen düzenli atıksu izleme programından elde edilen çok yıllı veri seti kullanılarak, KOİ dışındaki parametreler için mevzuat ihlali riskini öngörebilen makine öğrenmesi (ML) tabanlı bir erken uyarı yaklaşımı geliştirilmiştir. Çalışma, üç temel araştırma sorusu (AS1–AS3) etrafında kurgulanmış; global (tüm firmaları kapsayan) ve sektörel (SKKY tablo bazlı) modeller birlikte ele alınmıştır. Bu bölümde, elde edilen bulgular bu üç araştırma sorusu çerçevesinde tartışılmakta, ardından OSB uygulaması ve metodolojik sınırlılıklar değerlendirilmektedir.

5.1. ARAŞTIRMA SORULARININ TARTIŞILMASI

AS1 – "Bir sonraki numunenin KOİ dışındaki parametreler açısından limit altında mı / üstünde mi olacağı ML ile ne kadar iyi tahmin edilebilir?"

Çalışmada, 2023–2025 dönemindeki endüstriyel kanal numuneleri için, t_k anındaki ölçümler ve zamansal özetler (lag, hareketli ortalama, kümülatif ortalamalar, ihlal geçmişi) kullanılarak t_{k+1} anındaki KOİ hariç mevzuat ihlali durumu (`non_compliant_excl_KOİ_next`) tahmin edilmiştir.

Global model sonuçları, özellikle Random Forest algoritmasının bu görev için güçlü bir performans sunduğunu göstermektedir. Test setinde:

- Doğruluk (accuracy) ≈ 0,85,
- Pozitif sınıf F1 skoru (ihlal sınıfı) ≈ 0,88,
- Pozitif sınıf duyarlılığı (recall_pos) ≈ 0,95,
- PR-AUC ≈ 0,88 düzeyindedir.

Bu metrikler, "bugünkü atıksu profiline bakarak, bir sonraki numunede KOİ dışı mevzuat ihlali riski var mı?" sorusuna Random Forest global modeli ile pratikte kullanılabilir bir doğruluk düzeyiyle cevap verilebileceğini göstermektedir. OSB bağlamında önemli olan "ihlal yapan numuneleri kaçırmamak" olduğundan, yüksek recall ve yüksek F1 skoru, modelin erken uyarı amaçlı kullanımını desteklemektedir.

Tüm sektör gruplarında KOİ hariç kümülatif ihlal oranı ve ihlal sayısı değişkenlerinin yüksek önem düzeyine sahip olması, mevzuat uyumsuzluğunun tek bir parametreye indirgenemeyeceğini açıkça göstermektedir. Gerçek zamanlı kontrol ve korelasyon temelli ozonlama çalışmaları da dahil olmak üzere birçok literatür çalışmasında, kümülatif yük göstergelerinin proses kontrolü ve erken uyarı sistemlerinde daha güvenilir sonuçlar verdiği bildirilmektedir (Chys ve ark., 2017). Bu bağlamda, çalışmadaki bulgular yalnızca istatistiksel değil, işletme ve yönetim açısından da uygulanabilir sonuçlar sunmaktadır.

Benzer şekilde, literatürde yapay zekâ ve makine öğrenmesi tabanlı atıksu çalışmalarında, tarihsel veri ve kümülatif göstergelerin model başarısını anlamlı ölçüde artırdığı bildirilmektedir. Özellikle RF ve ANN tabanlı modellerin, geçmiş yüklenme ve ihlal bilgilerini kullanarak karmaşık ve doğrusal olmayan ilişkileri yakalayabildiği belirtilmektedir (Zare Abyaneh, 2014). Bu yönüyle, tez bulguları literatürdeki veri-temelli modelleme yaklaşımıyla yüksek uyum göstermektedir.

AS2 – "KOİ hariç limit aşımı riskini en çok hangi parametreler ve parametre kombinasyonları açıklamaktadır?"

Global Random Forest modelinden elde edilen özellik önemleri, KOİ dışı ihlal riskinin üç ana bileşen etrafında şekillendiğini ortaya koymuştur:

1. İhlal geçmişi:

- `cum_violation_ratio_excl_KOİ` ve `cum_violations_excl_KOİ` gibi kümülatif ihlal göstergeleri ile
- `non_compliant_excl_KOİ_lag1, lag2, lag3` gibi son üç numunenin ihlal/uyum durumunu temsil eden etiketler, en yüksek önem skorlarına sahiptir.

Bu yapı, kronik ihlal eğiliminin gelecekteki ihlal riskini güçlü biçimde belirlediğini göstermektedir.

2. AKM'ye ilişkin kısa ve uzun dönemli yük göstergeleri:

- `cum_mean_AKM, AKM_mean_last3, AKM, AKM_lag1-3,`

global modelde ardışık biçimde üst sıralarda yer almakta; bu da AKM'nin hem yapısal kirlilik yükü hem de proses stres göstergesi olarak ihlal riskine doğrudan yansıdığını göstermektedir.

3. Kimyasal kompozisyonun uzun dönemli bileşeni:

- `cum_mean_Yağ-Gres`, `cum_mean_pH`, `cum_mean_T. Azot` ve belirli ölçüde `pH_lag2` gibi değişkenler, firmanın uzun dönem yağ–gres, pH ve azot profilinin KOİ dışı ihlal riskini etkilediğini ortaya koymaktadır.

Dolayısıyla global model, ihlal riskini tek bir parametre üzerinden değil, "ihlal geçmişi + AKM dinamiği + uzun dönem kimyasal kompozisyon" üçlüsünün bileşimi üzerinden açıklamaktadır. Bu çerçeve, OSB ölçeğinde risk sürücülerinin yorumlanması açısından anlamlı ve yorumlanabilir bir yapı sunmaktadır.

Random Forest model çıktılarında tüm sektör gruplarında AKM'nin anlık, gecikmeli ve kümülatif bileşenleriyle yüksek önem skorlarına sahip olması, askıda katı maddelerin endüstriyel atıksu sistemlerinde temel bir belirleyici olduğunu doğrulamaktadır. Bulgular, özellikle KOİ dışı ihlal davranışlarının AKM birikimi ile güçlü biçimde ilişkili olduğunu göstermektedir. Bu sonuç, atıksu arıtma performansının AKM ve ilişkili katı madde dinamikleriyle yakından bağlantılı olduğunu ortaya koyan önceki çalışmalarla örtüşmektedir. Literatürde, AKM'nin arıtma verimliliği ve mevzuat uyumu açısından kritik bir belirleyici olduğu, özellikle biyolojik arıtma süreçlerinde çökeltim performansı ve yüklenme rejimini yansıttığı vurgulanmaktadır (Bumbac ve ark., 2018).

Literatürde, AKM'nin KOİ, besin maddeleri ve bazı toksik bileşenlerle güçlü korelasyon gösterdiği; özellikle yüksek AKM yüklerinin biyolojik arıtma proseslerinde oksijen transferini ve çökelme verimini olumsuz etkilediği bildirilmektedir. Çok değişkenli istatistiksel tekniklerin uygulandığı nehir ve atıksu çalışmalarında da AKM ve organik yük parametrelerinin birlikte hareket ettiği ve kirlenme kaynaklarının ayırt edilmesinde kritik rol oynadığı gösterilmiştir (Singh, 2004). Bu bağlamda, çalışmada AKM'nin farklı zaman pencereleriyle ön plana çıkması, zamansal bağımlılığın ihlal riskinin açıklanmasında temel bir unsur olduğunu, literatürde tanımlanan yapısal ilişkileri veri temelli olarak doğruladığını desteklemektedir.

AS3 – "Sektör/altyük profillerine göre KOİ dışı limit aşımının 'key driver'ları nasıl değişmektedir?"

SKKY sektör grupları için kurulan Random Forest modelleri, global modelde gözlenen genel desenin sektörel düzeyde hem korunduğunu hem de zenginleştiğini göstermektedir:

Ortak desenler:

- Hemen tüm sektörlerde `cum_violation_ratio_excl_KOİ` ve `cum_violations_excl_KOİ` ilk sıralarda yer almakta;
- AKM'nin anlık, lag ve hareketli ortalama özellikleri her sektörde baskın kalmaktadır;
- Parametrelerin kümülatif ortalamaları (örneğin `cum_mean_AKM`, `cum_mean_Yağ-Gres`, `cum_mean_Renk`, `cum_mean_T. Azot`, `cum_mean_T.Fosfor`) sektörel modellerde de güçlü önem skorlarına sahiptir.

Sektöre özgü farklılaşmalar:

- SKKY 15.2 grubunda AKM ve Yağ–Gres kombinasyonu belirleyici olup, askıda katı madde ve yağ yükü temel risk sürücüleri olarak öne çıkmaktadır.
- SKKY 6.1 grubunda AKM'ye ek olarak Renk, `cum_mean_Renk`, `cum_mean_pH` ve `cum_mean_Yağ-Gres` gibi parametreler daha çok ağırlık kazanmakta; bu da organik madde, renk ve pH dengesinin birlikte önemli olduğunu göstermektedir.
- SKKY 14.4 ve 14.5 gruplarında, AKM ile birlikte `cum_mean_T. Azot`, `cum_mean_T.Fosfor` ve `T. Azot_lag3` gibi azot–fosfor bileşenleri üst sıralara yükselmektedir.
- SKKY 14.10 grubunda ise fosfor yükü öne çıkmakta; `cum_mean_T.Fosfor`, `T.Fosfor`, `T.Fosfor_mean_last3`, `T.Fosfor_lag2` gibi değişkenler ilk sıralarda yer almakta, ayrıca `pH_lag2` ve `cum_mean_Klorür` gibi parametreler sektöre özgü kritik göstergeler olarak öne çıkmaktadır.

Sektör bazlı doğrulama sonuçları, beş ana sektörde doğruluk değerlerinin 0,74–0,89, pozitif sınıf F1 skorlarının ise 0,82–0,92 aralığında olduğunu göstermektedir. Özellikle SKKY 15.2 ve

14.5 gruplarında hem yüksek doğruluk hem de çok yüksek F1 skorları elde edilmiştir. Bu bulgular, KOİ dışı ihlal riskinin sektörel düzeyde de ML ile güvenilir biçimde tahmin edilebildiğini ve her sektör için farklı parametre kombinasyonlarının ön plana çıktığını doğrulamaktadır.

SKKY 14.4 ve 14.5 kapsamındaki sektörlerde, Toplam Azot ve Toplam Fosfor'un kümülatif ve gecikmeli bileşenlerinin AKM ile birlikte üst sıralarda yer alması, kimyasal ve organik kimya temelli üretim süreçlerinde çok bileşenli kirletici yapının hâkim olduğunu göstermektedir. RF model çıktıları, azot ve fosfor yüklerinin yalnızca eşik aşımı anlarında değil, zaman içinde biriken yükler aracılığıyla mevzuat uyumsuzluğunu tetiklediğini ortaya koymaktadır.

Literatürde, özellikle endüstriyel kaynaklı atıksularda azot–fosfor bileşenlerinin birlikte artmasının, alıcı ortamlarda ötrofikasyon riskini artırdığı ve biyolojik arıtma sistemlerinin stabilitesini zorladığı sıklıkla vurgulanmaktadır. Uzun dönemli izleme verilerine dayalı çalışmalar, besin maddelerinin gecikmeli etkilerinin çoğu zaman anlık ölçümlerle yakalanamadığını; bu nedenle kümülatif göstergelerin daha açıklayıcı olduğunu belirtmektedir (Ying ve ark., 2025). Bu yönüyle, çalışmada elde edilen bulgular, tek parametreli değerlendirmelerin yetersizliğini ve çok değişkenli, zamansal temelli modelleme yaklaşımlarının gerekliliğini desteklemektedir.

Tez çalışmasında; SKKY 14.10 limitlerine tabi sektörlerde Toplam Fosfor'un kümülatif, anlık ve gecikmeli göstergelerinin RF modelde baskın hâle gelmesi, bu sektörlerde fosfor yükünün ihlal riskinin temel belirleyicisi olduğunu göstermektedir. Buna pH ve klorür gibi parametrelerin eşlik etmesi, fosfor kaynaklı yüklerin çoğu zaman kimyasal denge ve iyonik yapı ile birlikte değerlendirilmesi gerektiğini ortaya koymaktadır. Bu bulgu, literatürde özellikle besin maddesi (N–P) yüklerinin arıtma performansı ve su kalite göstergeleri üzerindeki kritik rolünü vurgulayan çalışmalarla paralellik göstermektedir. Çok değişkenli istatistiksel ve yapay zekâ tabanlı analizlerde, azot ve fosforun birlikte ele alınmasının model açıklayıcılığını artırdığı belirtilmektedir (Kothari ve ark., 2021). Dolayısıyla, bu çalışmada gözlenen sektör bazlı ayrışmalar, literatürde tanımlanan proses–parametre ilişkilerinin saha verisiyle desteklenmiş bir yansımasıdır.

5.2. OSB İÇİN PRATİK VE YÖNETSEL ÇIKARIMLAR

Bu çalışma, OSB atıksu izleme programını "sadece geriye dönük değerlendirme yapan bir sistem" olmaktan çıkarıp, veri temelli bir erken uyarı ve risk yönetim araç kutusuna dönüştürme potansiyeline sahiptir. Elde edilen bulgular, uygulamaya dönük olarak şu başlıklar altında özetlenebilir:

1. ML tabanlı erken uyarı modülü:

- Random Forest global modeli, mevcut veri altyapısına entegre edilerek, her yeni numune girişinde bir sonraki numunenin KOİ dışı ihlal riskini olasılık skoru olarak üretebilir.
- Bu skorlar, risk eşiklerine göre "düşük–orta–yüksek" kategorilere ayrılarak numune alma sıklığı, sahaya gidilecek denetim programı ve ön arıtma iyileştirme önerileri için kullanılabilir.

2. Kümülatif ihlal oranına dayalı dinamik risk sınıflaması:

- `cum_violation_ratio_excl_KOİ`'nin global ve sektörel modellerde ısrarlı biçimde ilk sıralarda yer alması, bu göstergenin mevzuat uyumu açısından temel bir risk göstergesi olarak kullanılabileceğini göstermektedir.
- OSB'de, belirli bir eşik değerin üzerinde kümülatif ihlal oranına sahip firmalar için daha sık numune alma, detaylı proses analizi, eğitim ve yaptırım stratejileri kademeli olarak tanımlanabilir.

3. AKM ve eşlik eden parametrelere odaklı izleme stratejileri:

- AKM'nin hem global hem de sektörel modellerde baskın risk sürücüsü olması, AKM'ye yönelik çevrimiçi sensörler, daha sık laboratuvar analizi ve proses kontrolü mekanizmalarının önemini vurgulamaktadır.
- AKM–Yağ-Gres–Renk–T.Azot–T.Fosfor gibi parametre kombinasyonları, sektöre göre uyarlanmış "kritik parametre setleri" olarak tanımlanabilir.

4. Sektöre özgü rehberlik ve süreç iyileştirme:

- SKKY 14.10 için fosfor odaklı; SKKY 14.4/14.5 için azot–fosfor ağırlıklı; SKKY 6.1 için renk–pH–organik yük ağırlıklı; SKKY 15.2 için AKM–Yağ-Gres ağırlıklı rehber dokümanlar oluşturulabilir.

- Bu rehberler, firmaların kendi proseslerini gözden geçirmesinde, ön arıtma ekipmanı seçimi ve işletme parametrelerinin optimizasyonunda veri temelli referanslar sağlayabilir.

5. Atıksu izleme politikasının gözden geçirilmesi:

- Çalışma, sistematik olarak ihlalli numune üreten firmaların erken dönemde tespit edilebildiğini göstermektedir. Bu firmalar için, OSB yönetmeliğinde öngörülen 30 günlük sıklaştırılmış izleme gibi mekanizmalar, ML modelleriyle desteklenerek daha hedefli ve etkin hale getirilebilir.

Yapay zekâ temelli doğrudan ilişkili literatür, bu tür modellerin en önemli katkısının yalnızca yüksek tahmin doğruluğu değil, erken uyarı ve risk temelli karar destek kapasitesi olduğunu ortaya koymaktadır. Özellikle Random Forest gibi ağaç tabanlı yöntemlerin, değişken önem skorları aracılığıyla hangi parametrelerin ihlal riskini sürüklediğini açıklayabilmesi, bu modelleri çevresel denetim uygulamaları açısından güçlü kılmaktadır (Baarimah ve ark., 2024). Bu tezde elde edilen bulgular da, kümülatif ve gecikmeli göstergelerin ön plana çıkmasıyla, denetim ve izleme stratejilerinin reaktif yaklaşımlardan ziyade önleyici ve hedefli biçimde kurgulanması gerektiğini göstermektedir.

Yapay zekâ ve makine öğrenmesi temelli atıksu tahmin çalışmalarında, Random Forest, ANN ve bulanık mantık tabanlı komite modellerinin doğrusal yöntemlere kıyasla daha yüksek tahmin başarısı sunduğu gösterilmektedir (Hoang ve ark., 2026; Giwa ve Hasan, 2017). Bu açıdan gerçekleştirilen çalışma, kullanılan Random Forest modeli, değişkenler arası doğrusal olmayan ilişkileri ve zamansal bağımlılıkları yakalayarak çok parametreli ihlal dinamiklerini başarıyla açıklayabilmiştir. Özellikle kümülatif ve gecikmeli değişkenlerin ön plana çıkması, literatürde önerilen veri temelli modelleme yaklaşımlarının pratik bir uygulaması niteliğindedir.

Genel olarak bu çalışmada elde edilen bulgular, endüstriyel atıksularda mevzuat ihlallerinin tekil parametreler yerine, çok bileşenli ve zamansal olarak biriken kirletici yükler tarafından

belirlendiğini göstermektedir. Literatürle uyumlu biçimde, AKM, azot ve fosfor gibi temel parametrelerin gecikmeli ve kümülatif etkileri ön plana çıkmakta; bu durum, veri temelli ve makine öğrenmesi tabanlı yaklaşımların klasik istatistiksel yöntemlere kıyasla daha bütüncül bir değerlendirme sunduğunu ortaya koymaktadır. Bu yönüyle çalışma, hem metodolojik hem de uygulamaya dönük olarak literatüre özgün bir katkı sağlamaktadır.

5.3. SINIRLILIKLAR VE GELECEKTEKİ ÇALIŞMALAR

Bu tez çalışmasının bulguları güçlü olmakla birlikte, bazı sınırlılıklar ve gelecekteki araştırma alanları mevcuttur:

1. KOİ dışı odaklanma:

- Çalışma, bilinçli bir tercihle KOİ'yi etiketten ve özellik setinden hariç tutmuş; KOİ dışındaki parametrelerin ihlal dinamiklerine odaklanmıştır.
- Gelecekte, KOİ ve KOİ dışı parametrelerin birlikte ele alındığı çoklu çıktı (multi-label veya multi-task) modeller geliştirilerek, hem toplam organik yük hem de spesifik parametre bazlı ihlaller eş zamanlı tahmin edilebilir.

2. Zamansal modelleme mimarisi:

- Bu çalışmada zamansal bağımlılık, lag, hareketli ortalama ve kümülatif istatistikler üzerinden klasik sınıflandırma modellerine gömülü olarak temsil edilmiştir.
- İlerleyen çalışmalarda, yinelemeli sinir ağları (RNN, LSTM), 1D-CNN veya Transformer tabanlı dizisel modeller kullanılarak, daha karmaşık zamansal desenler ve dönemsel etkiler (sezonsallık, kampanya dönemleri, üretim değişiklikleri vb.) yakalanabilir.

3. Veri kapsamı ve genellenebilirlik:

- Modelleme, 2023–2025 dönemindeki veriler üzerine kurulmuş; bu sayede izleme rejimi ve proses yapısının görece sabit olduğu bir alt dönem analiz edilmiştir.

- Farklı yıllar, farklı organize sanayi bölgeleri ve farklı ülkelerdeki veri setleriyle yapılacak karşılaştırmalı çalışmalar, önerilen yaklaşımın genellenebilirliğini test etmek açısından değerli olacaktır.

4. Parametre çeşitliliği ve yardımcı değişkenler:

- Bu çalışmada firmanın üretim miktarı, debi, spesifik proses adımları, arıtma tesislerinin gerçek zamanlı işletme parametreleri (pompa hızları, kimyasal dozlar vb.) ve meteorolojik değişkenler modele dahil edilmemiştir.

- Gelecekte bu tür yardımcı değişkenlerin eklenmesiyle, model performansı daha da artırılabilir ve "neden–sonuç" ilişkileri daha derinlemesine analiz edilebilir.

5. Modelin saha entegrasyonu ve karar destek sistemleri:

- Tez kapsamında modeller çevrimdışı olarak değerlendirilmiştir. Bir sonraki adım, GOSB'nin mevcut veri tabanı ve izleme otomasyonuna entegre edilecek bir karar destek arayüzü (örneğin web tabanlı bir gösterge paneli) tasarlamaktır.

- Bu sayede, hem laboratuvar personeli hem de çevre birimi uzmanları, her numune için anlık risk skorlarını takip edebilir; sonuçlara göre hareket planlarını dinamik olarak güncelleyebilir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışması, yoğun sanayi yüküne sahip bir organize sanayi bölgesinde, rutin atıksu izleme verisinin yalnızca mevzuat uyumu kontrolü için değil, aynı zamanda geleceğe dönük risk tahmini ve erken uyarı amacıyla da kullanılabileceğini göstermiştir. KOİ dışı parametreler için geliştirilen global ve sektörel Random Forest modelleri, hem metodolojik hem de uygulamaya dönük düzeyde güçlü bir çerçeve sunmaktadır.

6.1. BAŞLICA SONUÇLAR

- Global Random Forest modeli, bir sonraki numunenin KOİ dışı mevzuat ihlali riskini yüksek doğruluk, yüksek pozitif sınıf F1 skoru ve yüksek recall ile tahmin edebilmektedir. Bu performans düzeyi, OSB ölçeğinde pratik bir erken uyarı aracı geliştirmek için yeterlidir.

- İhlal riski, tek bir parametreye indirgenmeyip "kümülatif ihlal geçmişi + AKM dinamiği + uzun dönem kimyasal kompozisyon" bileşimiyle açıklanmaktadır. Bu sonuç, ihlallerin hem kronik davranış örüntüleri hem de proses yüküyle yakından ilişkili olduğunu ortaya koymaktadır.

- Sektör bazlı modeller, global deseni büyük ölçüde korumakta; ancak fosfor (SKKY 14.10), azot–fosfor kombinasyonu (SKKY 14.4–14.5), AKM–Yağ-Gres (SKKY 15.2) ve Renk–pH–organik yük (SKKY 6.1) ekseninde sektöre özgü farklılaşmalar üretmektedir. Bu durum, risk sürücülerinin sektör/altyük profiline göre anlamlı biçimde değiştiğini göstermektedir.

- Çalışma, literatürde görece sınırlı sayıda ele alınan "organize sanayi bölgesi ölçeğinde, çok yıllı izleme verisiyle KOİ dışı ihlal riskinin tahmini ve risk sürücülerinin global ve sektörel düzeyde analizi" boşluğunu doldurmakta ve önerilen yaklaşımın benzer veri altyapısına sahip diğer OSB'lere ve kentsel atıksu yönetim sistemlerine uyarlanabilir olduğunu göstermektedir.

6.2. UYGULAMAYA YÖNELİK ÖNERİLER

1. **Erken uyarı modülünün hayata geçirilmesi:**

Geliştirilen global Random Forest modelinin, OSB'nin veri tabanı ve analiz yazılımlarına entegre edilerek her numune girişinde risk skoru üretmesi önerilmektedir. Bu skorlar, numune alma stratejilerinin, saha denetim planlarının ve ön arıtma önerilerinin dinamik biçimde güncellenmesinde kullanılabilir.

2. **Kümülatif ihlal oranına dayalı risk sınıflandırması:**

`cum_violation_ratio_excl_KOİ` göstergesi, OSB düzeyinde resmi bir risk sınıflandırma kriteri olarak tanımlanabilir. Eşik değerler belirlenerek, kademeli uyarı, eğitim ve yaptırım basamakları oluşturulmalıdır.

3. **Kritik parametre setlerinin tanımlanması:**

AKM başta olmak üzere, her SKKY sektör grubu için belirlenen kritik parametre kombinasyonları (örneğin AKM–Yağ-Gres–Renk–T.Azot–T.Fosfor) resmi izleme rehberlerine ve firma bilgilendirme dokümanlarına entegre edilmelidir.

4. **Sektöre özgü teknik rehberler:**

SKKY 14.10, 14.4, 14.5, 15.2 ve 6.1 grupları için, model bulgularına dayalı sektörel rehberler hazırlanmalı; bu rehberlerde proses iyileştirme, ön arıtma tasarımı ve işletme parametrelerine yönelik öneriler sunulmalıdır.

5. **Dijital karar destek sistemi geliştirilmesi:**

ML modelleri, bir gösterge paneli (dashboard) ile birlikte tasarlanmalı; OSB yönetimi ve laboratuvar personeli için kullanıcı dostu bir karar destek aracı olarak sunulmalıdır. Böylece hem mevzuat uyumu hem de çevresel performans daha şeffaf ve izlenebilir hale gelecektir.

6.3. GELECEK ÇALIŞMALAR İÇİN ÖNERİLER

- KOİ ve KOİ dışı parametreleri birlikte tahmin eden çoklu çıktı modellerinin (multi-label/multi-task) geliştirilmesi,

- Zamansal bağımlılığın derin öğrenme tabanlı mimarilerle (LSTM, 1D-CNN, Transformer) modellenmesi,
- Farklı OSB'lerden ve ülkelerden veri setleriyle çok merkezli ve karşılaştırmalı çalışmalar yürütülmesi,
- Üretim miktarı, debi, işletme parametreleri ve meteorolojik verilerin modele entegrasyonu,
- Geliştirilen modellerin gerçek zamanlı izleme sistemleri, sensör ağları ve SCADA altyapılarıyla bütünleştirilmesi,

bu tezin ortaya koyduğu çerçevenin hem bilimsel açıdan derinleştirilmesi hem de uygulama alanının genişletilmesi için öncelikli araştırma alanları olarak önerilmektedir.

KAYNAKLAR

Afan, H. A., Wan Mohtar, W. H. M., Khaleel, F., Kamel, A. H., Mansoor, S. S., Alsultani, R., Ahmed, A. N., Sherif, M., & El-Shafie, A. (2024). Data-driven water quality prediction for wastewater treatment plants. *Heliyon,* 10(18), e36940. https://doi.org/10.1016/j.heliyon.2024.e36940

Agarwal, M., & Agarwal, A. (2013). Linear regression and correlation analysis of water quality parameters: A case study of River Kosi at District Rampur, India. *International Journal of Innovative Research in Science, Engineering and Technology*, 2(12), 7273–7279. www.ijirset.com

Ahmed, A. N., Othman, F. B., Afan, H. A., Ibrahim, R. K., Fai, C. M., Hossain, M. S., Ehteram, M., & Elshafie, A. (2019). Artificial intelligence and water quality: From drinking water to wastewater treatment – A review. *Journal of Hydrology*, 578, 124084. https://doi.org/10.1016/j.jhydrol.2019.124084

Baarimah, A. O., Bazel, M. A., Alaloul, W. S., Alazaiza, M. Y. D., Al-Zghoul, T. M., Almuhaya, B., Khan, A., & Mushtaha, A. W. (2024). Artificial intelligence in wastewater treatment: Research trends and future perspectives through bibliometric analysis. *Case Studies in Chemical and Environmental Engineering,* 10, 100926. https://doi.org/10.1016/j.cscee.2024.100926

Bahna, S. L. (Ed.). (2009). Statistics for clinicians. Annals of Allergy*, Asthma & Immunology*, 103(4, Suppl. 1). https://doi.org/10.1016/S1081-1206(10)60166-1

Bhagat, S. K., Tung, T. M., & Yaseen, Z. M. (2020). Artificial intelligence as an upcoming technology in wastewater treatment: A comprehensive review. *Journal of Cleaner Production*, 250, 119473. https://doi.org/10.1016/j.jclepro.2019.119473

Bumbac, C., Manea, E. E., Tiron, O., & Badescu, V. R. (2018). Correlation between wastewater treatment performances and sludge characteristics. *Revista de Chimie (Bucharest*), 69(1), 10–13. http://www.revistadechimie.ro

Castaño-Trias M, Rodríguez-Mozaz S, Buttiglieri G. (2023). A decade of water monitoring in a Mediterranean region: Pharmaceutical prioritisation for an upgraded analytical methodology. *Environmental Nanotechnology Monitoring Management* 20:100850. https://doi.org/10.1016/j.enmm.2023.100850

Chavhan, N., Bhattad, R., Khot, S., Patil, S., Pawar, A., Pawar, T., et al. (2025). APAH: An autonomous IoT driven real-time monitoring system for industrial wastewater. *Digital Chemical Engineering*, 14, 100217. https://doi.org/10.1016/j.dche.2025.100217

Chys, M., Audenaert, W. T. M., Deniere, E., Mortier, S. T. F. C., Van Langenhove, H., Nopens, I., Demeestere, K., & Van Hulle, S. W. H. (2017). Surrogate-based correlation models in

view of real-time control of ozonation of secondary treated municipal wastewater. *Environmental Science & Technology*, 51(24), 14233–14243. https://doi.org/10.1021/acs.est.7b04905

Daibes, R. N., & Alhaddad, G. D. (2026). Analysis of dissolved oxygen distribution effects on nitrogen removal efficiency in oxidation ditch systems. *Scientific Reports*, 16, 2046. https://doi.org/10.1038/s41598-025-31648-0

Giwa, A., & Hasan, S. W. (2017). Statistical correlation analysis and energy consumption of electrically-enhanced membrane bioreactor for wastewater treatment. *Desalination and Water Treatment*, 68, 60–69. https://doi.org/10.5004/dwt.2017.20215

Han, Y., Zhu, L., Li, X., Cui, D., Li, J., Cen, K., & Chen, D. (2025). Integrated machine learning and response surface methodology for screening and optimization of high-performance Mg-modified bamboo biochar for phosphorus adsorption in water. *Industrial Crops and Products*, 231, 121213. https://doi.org/10.1016/j.indcrop.2025.121213

Hanrahan, G. (2013). *Çevre Kimyasında Temel Kavramlar: Problemler Ve Çözüm Notları* (İ. Toröz, Çev. Ed.). Nobel Yayıncılık. ISBN: 9786051336527

Hassnain, M., Lee, S. M. W., & Azhar, M. R. (2025). Soft sensing of biological oxygen demand in industrial wastewater using machine learning models. *Journal of Water Process Engineering*, 78, 108699. https://doi.org/10.1016/j.jwpe.2025.108699

Helsel, D. R., & Hirsch, R. M. (2002). *Statistical methods in water resources* (Techniques of Water-Resources Investigations Book 4, Chapter A3). U.S. Geological Survey.

Hoang, A. T., Chen, W.-H., Guerrero-Pérez, M. O., Rodríguez-Castellón, E., Lopez-Escalante, M. C., Nguyen, V. N., Paramasivam, P., Nguyen, X. P., & Truong, T. H. (2026). Turning waste into energy: Application of machine learning and explainable artificial intelligence for determining key factors in wastewater-to-hydrogen conversion. *Energy*, 344, 139934. https://doi.org/10.1016/j.energy.2026.139934

IWT Dijital Medya (2025). https://www.iwtdijitalmedya.com/blog/yapay-zeka-nedir [Erişim Tarihi: 14.01.2026]

Kandar B, Ghorui N, Datta C, Ghanta KC, Dutta S. (2025). Treatment of biologically treated synthetic refinery wastewater using Stenotrophomonas maltophilia sp. NITD 24: Experiment and analysis. *J Environmental Chemical Engineering* 13(3):116305. https://doi.org/10.1016/j.jece.2025.116305

Katsoyiannis, A., & Samara, C. (2007). The fate of dissolved organic carbon (DOC) in the wastewater treatment process and its importance in the removal of wastewater contaminants. *Environmental Science and Pollution Research*, 14(5), 284–292. https://doi.org/10.1065/espr2006.05.302

Khalil, M., AlSayed, A., Liu, Y., & Vanrolleghem, P. A. (2023). Machine learning for modeling N_2O emissions from wastewater treatment plants: Aligning model performance,

complexity, and interpretability. *Water Research*, 245, 120667. https://doi.org/10.1016/j.watres.2023.120667

Kothari, V., Vij, S., Sharma, S. K., & Gupta, N. (2021). Correlation of various water quality parameters and water quality index of districts of Uttarakhand. *Environmental and Sustainability Indicators*, 9, 100093. https://doi.org/10.1016/j.indic.2020.100093

Kubiak A, Lima S, Li Puma G. (2025). Adaptive bioinspired Light Emitting Diode-Internet of Things (LED-IoT) assisted solar packed bed photocatalytic reactor system for naproxen removal from real wastewater matrices. *Chemical Engineering Journal*; 522:167518. https://doi.org/10.1016/j.cej.2025.167518

Long, S., Zhao, L., Shi, T., Li, J., Yang, J., Liu, H., Mao, G., Qiao, Z., & Yang, Y. (2018). Pollution control and cost analysis of wastewater treatment at industrial parks in China. *Journal of Cleaner Production*, 172, 2435–2442. https://doi.org/10.1016/j.jclepro.2017.11.167

Miran, B. (2021). *Temel istatistik*. İzmir. ISBN: 975-93088-0-0

Parthasarathy J, Kumar SS, Sundararajan J. (2026). IoT-integrated carbon nanotube sensors for real-time monitoring of pesticide contamination in agricultural systems. *Microchemical Journal*. 220, 116675. https://doi.org/10.1016/j.microc.2025.116675

Ruffino B, Korshin G V., Zanetti M. (2020). Use of spectroscopic indicators for the monitoring of bromate generation in ozonated wastewater containing variable concentrations of bromide. *Water Research*, 182, 116009. https://doi.org/10.1016/j.watres.2020.116009

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education.

Sak, G. (2025), *Türkiye organize sanayi bölgeleri ile bugüne geldi*, Ekonomi Gazetesi » Köşe Yazıları, 10 Kasım 2025, https://www.ekonomigazetesi.com/kose-yazisi/turkiye-organize-sanayi-bolgeleri-ile-bugune-geldi-64183 [Erişim: 21.01.2026]

Santos AV, Amaral MCS, Oliveira SMAC. (2024). Artificial Neural Networks and Multivariate Statistical Process Control to improve ammonia removal on membrane bioreactors treating refinery wastewater. *Journal of Water Process Engineering*; 67:106126. https://doi.org/10.1016/j.jwpe.2024.106126

Sfragano PS, Emanuele A, Laschi S, Ferraro G, Fratini E, Palchetti I. (2025). Electrochemical sensing of perfluorooctanoic acid in wastewater: Characterization of a molecularly imprinted polymer-based sensor. *Biosensors and Bioelectronics*: X, 28, 100726. https://doi.org/10.1016/j.biosx.2025.100726

Shokri, A., Derakhshandeh, M., & Farhami, N. (2025). Using UV-Electro-Peroxone process for treating industrial wastewater: RSM and machine learning statistical analysis, energy consumption and kinetic studies. *Chemical Engineering Journal Advances*, 100886. https://doi.org/10.1016/j.ceja.2025.100886

Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Research*, 38(18), 3980–3992. https://doi.org/10.1016/j.watres.2004.06.011

Singh, N. K., Yadav, M., Singh, V., Padhiyar, H., Kumar, V., Bhatia, S. K., & Show, P.-L. (2023). Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. *Bioresource Technology*, 369, 128486. https://doi.org/10.1016/j.biortech.2022.128486

Tiyasha, Tran Minh Tung, & Yaseen, Z. M. (2020). A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology*, 585, 124670. https://doi.org/10.1016/j.jhydrol.2020.124670

Verma, P., Mehta, P., & Taloor, A. K. (2025). Water quality assessment of river Ujh in district Udhampur, J&K for drinking and agricultural purposes. *Discover Geoscience*, 3, 64. https://doi.org/10.1007/s44288-025-00158-x

Wang, Y., Cheng, Y., Liu, H., Guo, Q., Dai, C., Zhao, M., & Liu, D. (2023). A review on applications of artificial intelligence in wastewater treatment. *Sustainability*, 15(18), 13557. https://doi.org/10.3390/su151813557

Yao K, Liao C, Wang Z, Jiang E, Li N, Wang X. (2025). Can Microbial Electrochemical Sensors Measure Biochemical Oxygen Demand in Saline Wastewater*? Water & Ecology* 1(2):100009. https://doi.org/10.1016/j.wateco.2025.100009

Ying, S., Yao, P., Wang, Z., Luo, Y., Zhao, B., Guan, R., Cao, M., Xuan, M., Xu, R., He, Y., Zhang, H., & Ding, J. (2026). Research on water pollution monitoring and qualitative source identification in a typical coastal river network. *Environments,* 13(1), 1. https://doi.org/10.3390/environments13010001

Zare Abyaneh, H. (2014). Evaluation of multivariate linear regression and ANN in water quality prediction. *Journal of Environmental Health Science & Engineering*, 12, 40. https://doi.org/10.1186/2052-336X-12-40

Zhang, S., Jin, Y., Chen, W., Wang, J., Wang, Y., & Ren, H. (2023). Artificial intelligence in wastewater treatment: A data-driven analysis of status and trends. *Chemosphere*, 336, 139163. https://doi.org/10.1016/j.chemosphere.2023.139163

Zhao, X., Wang, Y., Ren, H., & Jin, Y. (2020). Artificial intelligence in wastewater treatment: Research trends and future perspectives. *Environmental Science and Pollution Research*, 27, 40652–40665. https://doi.org/10.1007/s11356-020-10186-4

Zou, L., Li, H., Wang, S., Zheng, K., Wang, Y., Du, G., & Li, J. (2019). Characteristic and correlation analysis of influent and energy consumption of wastewater treatment plants in Taihu Basin. *Frontiers of Environmental Science & Engineering*, 13(6), 83. https://doi.org/10.1007/s11783-019-1167-7