



**BAYBURT İLİ UMUMİ ÇEŞMELERİNİN SU
KALİTESİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ
İLE TAHMİNİ**

Gülhat KOÇYİĞİT

Yüksek Lisans Tezi

İnşaat Mühendisliği

Ana Bilim Dalı

Dr. Öğr. Üyesi Ruşen SINIR

2025

(Her Hakkı Saklıdır)

T.C.
BAYBURT ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İNŞAAT MÜHENDİSLİĞİ ANA BİLİM DALI
İNŞAAT MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS PROGRAMI

BAYBURT İLİ UMUMİ ÇEŞMELERİNİN SU KALİTESİNİN MAKİNE
ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNİ

YÜKSEK LİSANS TEZİ

Gülhat KOÇYİĞİT

Danışman: Dr. Öğr. Üyesi Ruşen SINIR

BAYBURT - 2025

KABUL VE ONAY

Dr. Öğr. Üyesi Ruşen SINIR danışmanlığında, Gülhat KOÇYİĞİT tarafından hazırlanan “Bayburt İli Umumi Çeşmelerinin Su Kalitesinin Makine Öğrenmesi Yöntemleri İle Tahmini ” başlıklı bu çalışma, 24.04.2025 tarihinde yapılan savunma sınavı sonucunda oy birliği ile başarılı bulunarak jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Başkan : Dr. Öğr. Üyesi Ruşen SINIR

Üye : Dr. Öğr. Üyesi Bilal TAYFUR

Üye : Dr. Öğr. Üyesi Muhammet YILMAZ

Bu tezin kabulü Lisansüstü Eğitim Enstitüsü Yönetim Kurulu’nun/...../20... tarih ve 20...../.....-..... sayılı kararıyla onaylanmıştır.

Doç. Dr. Bünyamin ALIM
Enstitü Müdürü

BEYANNAME

Bayburt Üniversitesi Lisansüstü Eğitim Enstitüsü tez yazım kılavuzuna göre Dr. Öğr. Üyesi Ruşen SINIR danışmanlığında hazırlamış olduğum “Bayburt İli Umumî Çeşmelerinin Su Kalitesinin Makine Öğrenmesi Yöntemleri ile Tahmini” başlıklı yüksek lisans tezimin bilimsel etik değerlere ve kurallara uygun, özgün bir çalışma olduğunu, aksinin tespit edilmesi halinde her türlü yasal yaptırımını kabul edeceğimi beyan ederim.

24.04.2025

Gülhat KOÇYİĞİT



ÖN SÖZ

Tez çalışmamda bilgi ve deneyimleri ile beni yönlendiren danışman hocam Dr. Öğr. Üyesi Ruşen SINIR'a teşekkürlerimi sunarım. Tez savunma sınavımda jüri üyesi olan Dr. Öğr. Üyesi Bilal TAYFUR ve Dr. Öğr. Üyesi Muhammet YILMAZ'a katkılarından dolayı teşekkürlerimi sunarım. Tez için gerekli verileri temin etmemde yardımcı olan ve tez sürecindeki manevi desteklerinden dolayı Hakan DAĞLAR'a teşekkür ederim. Ayrıca verileri paylaşan Bayburt İl Sağlık Müdürlüğü ve Bayburt Halk Sağlığı Labaratuvarı çalışanlarına teşekkür ederim. Her daim desteklerini üzerimde hissettiğim aileme, özellikle çalışmamın devamını sağlamam konusundaki konuşmalarıyla beni motive eden anneme teşekkür ederim.

Gülhat KOÇYİĞİT



ÖZET

YÜKSEK LİSANS TEZİ

BAYBURT İLİ UMUMİ ÇEŞMELERİNİN SU KALİTESİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNİ

Gülhat KOÇYİĞİT

Bayburt Üniversitesi

Lisansüstü Eğitim Enstitüsü

İnşaat Mühendisliği Ana Bilim Dalı

Tez Danışmanı: Dr. Öğr. Üyesi Ruşen SINIR

Bayburt-2025, Sayfa: 69

Mavi altın olarak da anılan suyun içeriği, canlıların yaşamlarını, hayat kalitelerini ve sağlıklarını büyük oranda etkilemektedir. Suyun içerisinde bulunan her bir maddenin canlı hayatı üzerinde iyi veya kötü etkisi bulunmaktadır. Suyun içerisinde yer alan maddeler kadar madde miktarı da önemlidir. Bu bakımdan Dünya Sağlık Örgütü (WHO) su kalitesine ilişkin çalışmalar yaparak içerisinde bulunan maddeler ve miktarı ile ilgili kılavuz hazırlamıştır. Türkiye’de de bu yönde çalışmalar yapılmış 2005 yılında İnsani Amaçlı Sular Hakkında Yönetmelik ile standartlarını belirlemiştir. Teknolojinin gelişmesiyle birlikte yapay zekâ su kalitesi çalışmalarında da kullanılmaya başlanmış ve makine öğrenme yöntemleriyle su kalitesi tahminlerinde bulunulmuştur. Bu çalışmada, Bayburt ilinde yer alan şebeke sularının mikrobiyolojik ve kimyasal su analiz verileri ile Bayburt ili umumî çeşmelerine ait mikrobiyolojik su analiz verileri kullanılarak makine öğrenme yöntemleriyle su kalitesi tahmininde bulunulmuştur. İlk olarak veriler üzerinde ön işleme uygulanarak veriler makine öğrenimine hazır hale getirilmiş, daha sonra makine öğrenimi yöntemlerinden lojistik regresyon, knn, SVM, naive bayes, karar ağacı, rastgele orman makine öğrenme yöntemleri ile su kalitesi tahmin çalışmaları yapılmıştır. Bayburt şebeke suları analizi verileri kullanılarak yapılan makine öğrenme yöntemlerinde en uygun yöntemin karar ağacı yöntemi olduğu, Bayburt umumî çeşme suları analizi verileri kullanılarak yapılan makine öğrenme yöntemlerinde ise naive bayes, rastgele orman ve karar ağacı yöntemlerine ait sonuçların aynı olması dolayısıyla her üç yöntemin de kullanılabilir olduğu sonucuna varılmıştır. Su kalitesi tahmin çalışmalarında specificity değerinin önemli olduğu ve model değerlendirmesinde kullanılması gerektiği sonucuna varılmıştır.

Anahtar Kelimeler: Su kalitesi, k-en yakın komşu, destek vektör makineleri, naive bayes, rastgele orman, karar ağacı.

ABSTRACT

M. SC. THESIS

DRINKING WATER QUALITY ESTIMATION OF THE FOUNTAINS IN BAYBURT WITH MACHINE LEARNING METHODS

Gülhat KOÇYİĞİT

Bayburt University

Institute of Graduate Studies

Department of Civil Engineering

Thesis Advisor: Assist. Prof. Dr. Ruşen SINIR

Bayburt-2025, Pages: 69

Water, often referred to as "blue gold," plays a critical role in influencing the survival, quality of life, and health of living organisms. The presence and quantity of substances within water can have either beneficial or detrimental effects on life. In this regard, the World Health Organization (WHO) has conducted studies on water quality, developing guidelines that specify the permissible substances and their concentrations. In Turkey, similar efforts have been undertaken, with standards established through the 2005 Regulation on Water Intended for Human Consumption. With advancements in technology, artificial intelligence has increasingly been employed in water quality studies, utilizing machine learning techniques to predict water quality. This study focuses on predicting water quality in Bayburt province using machine learning methods, based on microbiological and chemical water analysis data from the city's mains water supply, as well as microbiological water analysis data from public fountains in Bayburt. Initially, the data underwent preprocessing to ensure compatibility with machine learning applications. Subsequently, water quality predictions were performed using various machine learning methods, including logistic regression, k nearest neighbors (knn), support vector machines (svm), naive bayes, decision trees, and random forests. The results indicated that, for the mains water analysis data of Bayburt, the decision tree method was the most suitable. In contrast, for the public fountain water analysis data, naive bayes, random forest, and decision tree methods yielded identical results, suggesting that all three methods are equally viable. The study further concluded that specificity is a crucial metric in water quality prediction studies and should be incorporated into model evaluation processes.

Keywords: Water quality, k-nearest neighbors, support vector machines, naive bayes, random forest, decision tree.

İÇİNDEKİLER

KABUL VE ONAY	II
BEYANNAME	III
ÖN SÖZ.....	IV
ÖZET	V
ABSTRACT	VI
İÇİNDEKİLER	VII
TABLOLAR.....	X
ŞEKİLLER.....	XI
KISALTMALAR	XII
SİMGELER.....	XII
GİRİŞ	1
Araştırmanın Konusu ve Problemi	2
Araştırmanın Amacı.....	2
Araştırmanın Önemi ve Gerekçesi.....	2
Araştırmanın Sınırlılıkları	3
Varsayımlar	3
BİRİNCİ BÖLÜM	4
1. KURAMSAL ÇERÇEVE.....	4
1.1. SU KALİTESİ.....	11
1.1.1. Suda Bulunan Mikrobiyolojik Parametreler	12
1.1.1.1. <i>Escherichia coli</i>	12
1.1.1.2. Koliform.....	12
1.1.1.3. Enterokok	12
1.1.2. Suda Bulunan Kimyasal Parametreler	13
1.1.2.1. Antimon.....	13
1.1.2.2. Arsenik	13
1.1.2.3. Benzen.....	13
1.1.2.4. Bor	13
1.1.2.5. Bromat.....	13
1.1.2.6. Kadmiyum.....	13
1.1.2.7. Krom	13
1.1.2.8. Bakır.....	14
1.1.2.9. Kurşun.....	14
1.1.2.10. Cıva.....	14
1.1.2.11. Nikel.....	14
1.1.2.12. Nitrat	14
1.1.2.13. Siyanür	14
1.1.2.14. Selenyum.....	15
1.1.2.15. Akrilamid	15
1.1.2.16. Benzo (a) Piren	15
1.1.2.17. 1,2-Dikloretan.....	15
1.1.2.18. Nitrit.....	15
1.1.2.19. Florür.....	15

1.1.2.20. Klor ve klorlu bileşikler	15
1.1.2.21. Pestisitler	15
1.2. MAKİNE ÖĞRENMESİ	16
1.2.1. Makine Öğrenmesi Terminolojisi.....	16
1.2.1.1. Veri kümesi	16
1.2.1.2. Öznitelikler	16
1.2.1.3. Etiketler	16
1.2.1.4. Veri ön işleme.....	17
1.2.1.5. Veri bölme.....	17
1.2.1.6. Eğitim.....	17
1.2.1.7. Test.....	17
1.2.1.8. Model	17
1.2.1.9. Aşırı uyum.....	17
1.2.1.10. Uyum sağlayamama.....	17
1.2.1.11. Hiperparametre	17
1.2.1.12. P-value	18
1.2.1.13. Regularizasyon	18
1.2.1.14. Öğrenme algoritmaları	18
1.2.2. Gözetimli Öğrenme.....	18
1.2.2.1. K-en yakın komşu.....	19
1.2.2.2. Karar ağaçları	19
1.2.2.3. Rastgele orman	22
1.2.2.4. Lojistik regresyon	22
1.2.2.5. Destek vektör makineleri	23
1.2.2.6. Naive Bayes.....	23
1.2.2.7. Lineer regresyon	24
1.2.3. Gözetimsiz Öğrenme	26
1.2.3.1. K-ortalamlar kümeleme.....	26
1.2.3.2. Hiyerarşik kümeleme	27
1.2.3.3. Gürültülü uygulamaların yoğunluğa dayalı mekânsal kümelenmesi (DBSCAN) ...	27
1.2.4. Pekiştirmeli Öğrenme	28
1.2.5. Karmaşıklık Matrisi	28
1.3. TREND VE MEVSİMSELLİK ANALİZİ.....	30
1.3.1. Mevsimsellik Analizi ve Zaman Serisi Ayırıştırma	30
İKİNCİ BÖLÜM.....	32
2. MATERYAL VE YÖNTEM.....	32
2.1. BAYBURT ŞEBEKE SULARI ANALİZ VERİLERİ İLE MAKİNE ÖĞRENMESİ YÖNTEMLERİ	32
2.1.1. Lojistik Regresyon Yöntemi	33
2.1.2. Knn Yöntemi	33
2.1.3. SVM Yöntemi	34
2.1.4. Naive Bayes Yöntemi	35
2.1.5. Karar Ağacı Yöntemi.....	36
2.1.6. Rastgele Orman Yöntemi.....	36
2.2. BAYBURT İLİ UMUMÎ ÇEŞME SULARI ANALİZ VERİLERİ İLE MAKİNE ÖĞRENMESİ YÖNTEMLERİ.....	37

2.2.1. Lojistik Regresyon Yöntemi	42
2.2.2. Knn Yöntemi	42
2.2.3. SVM Yöntemi	42
2.2.4. Naive Bayes Yöntemi	43
2.2.5. Karar Ağacı Yöntemi.....	43
2.2.6. Rastgele Orman Yöntemi.....	44
2.3. BAYBURT İLİ UMUMÎ ÇEŞMELERİ TREND VE MEVSİMSELLİK ANALİZİ....	44
SONUÇ VE ÖNERİ.....	57
KAYNAKÇA.....	60
EKLER.....	64
Ek 1-a: Bayburt İli Şebeke Sularının Kimyasal Parametrelerine Ait Korelasyon Tablosu	64
Ek 1-b: Bayburt İli Şebeke Sularının Kimyasal Parametrelerine Ait Korelasyon Tablosu	65
Ek 1-c: Bayburt İli Şebeke Sularının Kimyasal Parametrelerine Ait Korelasyon Tablosu	66
Ek 2-a: Bayburt İli Şebeke Sularına Ait Makine Öğrenme Yöntemleri Sonuçları.....	67
Ek 2-b: Bayburt İli Çeşme Sularına Ait Makine Öğrenme Yöntemleri Sonuçları.....	68
ÖZ GEÇMİŞ	69

TABLÖLER

Tablo 1: Karmaşıklık Matrisi	29
Tablo 2: Lojistik Regresyon Karmaşıklık Matrisi	33
Tablo 3: Knn Karmaşıklık Matrisi -Minkowski/Euclidean Mesafe Hesabı	34
Tablo 4: Knn Karmaşıklık Matrisi- Manhattan Mesafe Hesabı	34
Tablo 5: SVM Karmaşıklık Matrisi-Linear Metrik	34
Tablo 6: SVM Karmaşıklık Matrisi-Polinomial Metrik.....	35
Tablo 7: SVM Karmaşıklık Matrisi-RBF Metrik	35
Tablo 8: SVM Karmaşıklık Matrisi-Sigmoid Metrik	35
Tablo 9: Naive Bayes Karmaşıklık Matrisi	36
Tablo 10: Karar Ağacı Karmaşıklık Matrisi-Entropi.....	36
Tablo 11: Karar Ağacı Karmaşıklık Matrisi-Gini.....	36
Tablo 12: Rastgele Orman Karmaşıklık Matrisi-Gini.....	37
Tablo 13: Rastgele Orman Karmaşıklık Matrisi-Entropi.....	37
Tablo 14: Umumî Çeşmeler Lojistik Regresyon Karmaşıklık Matrisi	42
Tablo 15: Umumî Çeşmeler Knn Yöntemi Karmaşıklık Matrisi	42
Tablo 16: Umumî Çeşmeler SVM Yöntemi Karmaşıklık Matrisi (Linear, RBF, Sigmoid)...	43
Tablo 17: Umumî Çeşmeler SVM Yöntemi Karmaşıklık Matrisi (Polinomial)	43
Tablo 18: Umumî Çeşmeler Naive Bayes Karmaşıklık Matrisi	43
Tablo 19: Umumî Çeşmeler Karar ağacı Yöntemi Karmaşıklık Matrisi	44
Tablo 20: Umumî Çeşmeler Rastgele Orman Yöntemi Karmaşıklık Matrisi	44
Tablo 21: Trend ve Mevsimsellik Analiz Sonuçları	45

ŞEKİLLER

Şekil 1: Knn Yöntemi.....	19
Şekil 2: Karar Ağacı Yöntemi.....	20
Şekil 3: Logistik Regresyonda Kullanılan Sigmoid Fonksiyonu.....	22
Şekil 4: Destek Vektör Makinesi Yöntemi	23
Şekil 5: Naive Bayes Yöntemi	24
Şekil 6: Lineer Regresyon Yöntemi	24
Şekil 7: K-Ortalamlar Kümeleme Yöntemi	26
Şekil 8: Elbow Methodu (Dirsek Yöntemi).....	27
Şekil 9: DBSCAN Yöntemi	28
Şekil 10: Gordon Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları	38
Şekil 11: Kışla Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları.....	39
Şekil 12: Tarım Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları.....	39
Şekil 13: Aşağı Narkazan Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları	40
Şekil 14: Yukarı Narkazan Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları.....	41
Şekil 15: Kışla Çeşmesi <i>E. coli</i> Trend Analizi.....	45
Şekil 16: Kışla Çeşmesi <i>E. coli</i> Otokorelasyon Grafiği	46
Şekil 17: Kışla Çeşmesi Koliform Trend Analizi	46
Şekil 18: Kışla Çeşmesi Koliform Otokorelasyon Grafiği	47
Şekil 19: Aşağı Narkazan Çeşmesi <i>E. coli</i> Trend Analizi	48
Şekil 20: Aşağı Narkazan Çeşmesi <i>E. coli</i> Otokorelasyon Grafiği.....	48
Şekil 21: Aşağı Narkazan Çeşmesi Koliform Trend Analizi.....	49
Şekil 22: Aşağı Narkazan Çeşmesi Koliform Otokorelasyon Grafiği.....	49
Şekil 23: Gordon Çeşmesi <i>E. coli</i> Trend Analizi.....	50
Şekil 24: Gordon Çeşmesi <i>E. coli</i> Otokorelasyon Grafiği.....	50
Şekil 25: Gordon Çeşmesi Koliform Trend Analizi.....	51
Şekil 26: Gordon Çeşmesi Koliform Otokorelasyon Grafiği	51
Şekil 27: Tarım Çeşmesi <i>E. coli</i> Trend Analizi	52
Şekil 28: Tarım Çeşmesi <i>E. coli</i> Otokorelasyon Grafiği	52
Şekil 29: Tarım Çeşmesi Koliform Trend Analizi	53
Şekil 30: Tarım Çeşmesi Koliform Otokorelasyon Grafiği.....	53
Şekil 31: Yukarı Narkazan Çeşmesi <i>E. coli</i> Trend Analizi	54
Şekil 32: Yukarı Narkazan Çeşmesi Koliform Trend Analizi	54

KISALTMALAR

ARIMA	: Otoregresif Entegre Hareketli Ortalama
<i>E. coli</i>	: <i>Escherichia coli</i>
Knn	: K-En Yakın Komşu
LSTM	: Uzun Kısa Vadeli Bellek
MAE	: Ortalama Mutlak Hata (Mean Absolute Error)
MLP	: Çok Katmanlı Yapay Sinir Ağı (Multi-Layer Perceptron)
MSE	: Ortalama Kare Hata (Mean Squared Error)
RMSE	: Ortalama Kare Hatanın Karekökü (Root Mean Squared Error)
RF	: Rastgele Orman (Random Forest)
RBF	: Radyal Tabanlı Fonksiyon (Radial Basis Function)
ROC	: Alıcı İşletim Karakteristiği (Receiver Operating Characteristic)
SVC	: Destek Vektör Sınıflandırıcısı (Support Vector Classifier)
SVM	: Destek Vektör Makineleri (Support Vector Machine)
SVR	: Destek Vektör Regresyonu (Support Vector Regression)
WHO	: World Health Organization
YSA	: Yapay Sinir Ağları

SİMGELER

Bq	: bekerel (becquerel)
l	: litre
mg	: miligram
MSV	: milisievert
NH₃	: amonyak
NO₂	: nitrit
NO₃	: nitrat
µg	: mikrogram
µS	: mikrosiemens

GİRİŞ

Su, evrende büyük bir öneme sahiptir. Yaklaşık olarak 4 milyar yıl önce suya kavuşan gezegenimizin (Wilde vd., 2001) yine yaklaşık olarak üçte ikisi sularla kaplıdır. Gezegenimizi diğer gezegenlerden ayıran ve yaşanabilir kılan da suyun varlığıdır. Evrenimiz hidrolojik döngü ile yaşam bulmaktadır. Su, okyanuslar ve denizlerden atmosfere, atmosferden yağış vasıtasıyla yeryüzüne ulaşır ve böylece toprak, bitkiler... vs. tüm canlılar suya kavuşur. Yağışlarla yeryüzüne ulaşan su, yüzeysel akışlar ile denizler ve okyanuslara yeniden kavuşur ve döngü bu şekilde devam ederken su da kendini tazelemiş olur.

Dünya'nın %70'i suyla kaplıdır. Bu suyun %97,5'ini tuzlu olan okyanuslar ve denizler oluşturmakta iken geriye kalan %2,5'inin büyük bir kısmını kutuplarda buz örtüsü ve diğer kısmını da yeraltı suları olacak şekilde tatlı sular oluşturmaktadır. Erişebilir tatlı sular, dünyadaki suların %1'inden daha azdır (Yılmaz, 2015). Türkiye 501 milyar m³/yıl ortalama yağış miktarına sahiptir. Bu yağışın yarısından fazlası buharlaşarak atmosfere karışırken 69 milyar m³'ü yeraltına sızmaktadır. Türkiye, 112 milyar m³ kullanılabilir su potansiyeline sahiptir. Türkiye, 112 milyar m³ kullanılabilir toplam su potansiyelinin %72'sini tarımda, %12'sini sanayide, %16'sını içme ve kullanma suyu olarak tüketmektedir (Bulut, 2020).

Su eşittir bolluk demektir. Tarımda kullanılarak gıda güvenliğini, sanayide ve enerji üretiminde kullanılarak ekonomiyi büyük oranda etkileyen su, insanların temel ihtiyacı olan bir bileşiktir.

Canlıların yaşamlarını sürdürebilmeleri için gerekli olan suyun kalitesi, insan sağlığı açısından önem arz etmektedir. Yaşayan her canlının temiz, sağlıklı suya erişim hakkı vardır. Dünyada birçok insan sağlıklı suya ulaşamadığı için ölümle karşı karşıya gelmekte ve hatta bulaşıcı hastalıklar sonucunda ölmektedir. Dünya Sağlık Örgütü (WHO), kirlenmiş su ve kötü sanitasyonun, dizanteri, hepatit A, kolera, çocuk felci, ishal ve tifo gibi bulaşıcı hastalıklar ile bağlantılı olduğunu; her yıl yaklaşık 1 milyon insanın güvenli olmayan içme suyu tükettiği, kötü sanitasyon ve el hijyeninin iyi olmayışı sonucu ishalden öldüğünü tahmin etmektedir (World Health Organization, 2023). Bu veriler insan sağlığı açısından su kalitesinin ne kadar önemli olduğunu ve sağlıklı içme suları kullanılmadığında ne kadar ciddi sonuçlar doğurduğunu bizlere göstermektedir. Dünya Sağlık Örgütü bu konuya el atmış, yapmış olduğu çalışmalarla uluslararası su kalitesi standartları oluşturmuş ve kılavuzlar hazırlamıştır. Birçok ülke, bu standartlar çerçevesinde sanitasyon teknolojisi ile içme ve kullanma sularını iyileştirerek tüketiciye sunmaktadır. Türkiye'de de Sağlık Bakanlığı'na su kalitesi çalışmaları yapılmış "İnsani Tüketim Amaçlı Sular Hakkında Yönetmelik" ile de ilgili kurum ve

kuruluşlarca içme ve kullanma sularının tüketiciye sağlıklı bir şekilde ulaştırılması sağlanmıştır. Yönetmelik, belirli aralıklarla içme ve kullanma sularının tahlil edilmesini zorunlu kılmıştır. Bu tahlillerde suyun içerisinde insan sağlığına aykırı maddelerin, parazitlerin, mikroorganizmaların var olup olmadığı ve ilgili yönetmelikteki Ek-1’de yer alan kriterlere uygunluğu tespit edilerek rapor edilmektedir.

Son yıllarda teknoloji hızla gelişmiş ve yapay zekâ ile yeni bir ivme kazanmıştır. Birçok alanda kullanılan yapay zekâ, hidrolojide de yerini bulmuş ve çeşitli alanlarında kullanılmaya başlanmıştır. Yapay zekâ ve makine öğrenme yöntemleri geliştirilerek yüksek doğruluk oranları ile sonuçlanan su kalitesi tahmin çalışmaları yapılmıştır.

Araştırmanın Konusu ve Problemi

Bu çalışmada Bayburt ili şebeke sularının yıllık mikrobiyolojik ve kimyasal su analizi rapor verileri ile Bayburt ilinde içme ve kullanma suyu olarak da tüketilen mahallî çeşmelere ait su analizi rapor verileri Bayburt İl Sağlık Müdürlüğü’nden temin edilerek makine öğrenme yöntemleri ile su kalitesi tahmin çalışmaları yapılmıştır. Birinci bölümde su kalitesi, makine öğrenme yöntemleri ile trend ve mevsimsellik yöntemlerine ilişkin bilgiler yer almaktadır. İkinci bölümde çalışmada kullanılan materyal ve uygulanan yöntemlerden bahsedilmiştir. Sonuç kısmında ise belirlenen en iyi yöntemler ve sonuçlarına değinilmiştir.

Bayburt ili çeşme suları analizlerinde özellikle *Escherichia coli* ve koliform bakterilerine rastlanmaktadır. İnsan sağlığını önemli ölçüde etkileyen bu bakterilerin içme sularında hangi dönemlerde arttığı ya da mevsimselliğin var olup olmadığı araştırılmıştır.

Araştırmanın Amacı

Bayburt ili içme ve kullanma suyu analizlerinin makine öğrenme yöntemleri ile trend ve mevsimselliğin varlığının incelenerek; mevsimselliğin olması halinde gelecek dönemlere ait su kalitesi tahmininde bulunması ve önlem alınması amaçlanmıştır. Bayburt ilinde içme ve kullanma suyu olarak umumî çeşme sularını kullanan hanelerin olması, umumî çeşmelerin su kalitesinin analiz edilmesi ve halka duyurulması konusunda önem taşımaktadır. Bu çalışmanın amacı, umumî çeşme sularını kullanan hane halkının su kalitesi tahmini hakkında bilgilendirilmesi ile halkın daha güvenilir olan umumî çeşme sularına yönlendirilmesidir.

Araştırmanın Önemi ve Gerekçesi

Su, insan sağlığı açısından büyük bir öneme sahip olan vazgeçilmez bir kaynaktır. Ancak sadece suyun varlığı yeterli olmayıp suyun temiz ve sağlıklı olması da önemlidir. Bayburt ilinde içme ve kullanma sularında *Escherichia coli* ve koliform bakterilerine sık

rastlanması, bu bakterilerin içme suyundaki varlığının halk sağlığını tehdit etmesi, halkın da içme ve kullanma suyu olarak arıtma cihazı ve hazır su dışında umumî çeşme sularını da tercih etmesi; potansiyel sağlık tehditlerinin önüne geçilmesi için düzenli olarak çeşme suyu analizleri yaparak medya, sosyal medya vb. aracılığı ile halkı bilgilendirme gerekliliğini ortaya çıkarmıştır. Bu risklerin zamanında fark edilerek önlem alınması toplum sağlığı açısından kritik önemdedir.

Araştırmanın Sınırlılıkları

Çalışmada Bayburt ili şebeke sularının mikrobiyolojik ve kimyasal analiz sonuçlarına ait veri kümesi ile Bayburt ili ile umumî çeşme çeşme sularının mikrobiyolojik analiz sonuçlarına ait veri kümesi kullanılmıştır. Bayburt ili şebeke sularına ait veri kümesinde eksik veriler yer almaktadır. Umumî çeşme sularına ait veriler ise sadece mikrobiyolojik analizlerle sınırlıdır. Makine öğrenme yöntemleri kullanılarak su kalitesi tahmin çalışmalarının daha iyi yürütülebilmesi için düzenli aralıklarla yapılan kimyasal ve mikrobiyolojik analizlere ihtiyaç vardır. Ayrıca izleme noktalarının yani çeşmelerin kimlik numaralandırmalarının olmaması, aynı çeşmeye ait farklı isimlendirmelerle analizlerin sonuçlandırılması çalışmayı zorlaştırmıştır. Umumî çeşmelerde trend ve mevsimselliğin olmaması nedeniyle geçmiş zaman verileri ile su kalitesi tahmin çalışmaları yapılmış, gerçek zaman tabanlı tahmin çalışmalarının yapılamadığı için erken uyarı sistemi geliştirilememiştir.

Varsayımlar

Eksik verilerin ortalama yöntemi ile doldurulmasının, kullanılan algoritma modellerinin doğruluğunu ve güvenilirliğini artırdığı, kullanılan makine öğrenmesi algoritmalarının su kalitesi tahmini için uygun olduğu kabul edilmiştir.

BİRİNCİ BÖLÜM

1. KURAMSAL ÇERÇEVE

Dijital teknolojiler, 20. yüzyılın ortalarından itibaren insan yaşamını köklü bir şekilde dönüştürmüştür. İlk elektronik bilgisayar olarak kabul edilen ENIAC'ın 1945 yılında geliştirilmesi, dijital dünyanın başlangıç noktası olarak görülür (Ceruzzi, 2003). 1960'lı yıllarda internetin ortaya çıkışı ve 1990'lı yıllarda World Wide Web'in (www) insanlığa sunulması, dijital teknolojilerde yeni bir dönemi başlatmıştır (Berners-Lee, 1999). Web 1.0 olarak adlandırılan ilk dönemde, kullanıcılar yalnızca statik içeriklere erişebilirken, 2000'li yıllarda web 2.0 ile etkileşimli, dinamik bir yapı ortaya çıkmıştır. Bu dönemde sosyal medya platformları, e-ticaret, internet bankacılığı, e-okul ve e-fatura gibi hizmetler, kullanıcıların hem tüketici hem de içerik üreticisi olmasına olanak tanımıştır (O'Reilly, 2005). Kablosuz ağlar ve mobil teknolojiler, bilgiye mekânsal sınırlamalar olmaksızın erişimi mümkün kılmıştır (Castells, 2009). Web 3.0 ile birlikte akıllı nesneler, robotik sistemler ve yapay zekâ, dijital dünyayı daha özerk ve bağlantılı bir evreye taşımıştır (Fuchs, 2017).

Yapay zekâ, modern bir inovasyon gibi görünse de temelleri 20. yüzyılın ortalarına dayanır. Alan Mathison Turing'in 1950 yılında yayımladığı "Computing Machinery and Intelligence" başlıklı makalesi, düşünen makineler fikrini ortaya atarak yapay zekânın teorik temelini oluşturmuştur (Turing, 1950). Yapay zekâ terimi, ilk kez 1956'daki Dartmouth Konferansı'nda resmi olarak kullanılmış ve bu dönem, yapay zekânın doğuşu olarak kabul edilmiştir (McCarthy vd., 1955). 1960'lı yıllarda semantik ağlar ve uzman sistemler geliştirilmiş, 1980'li yıllarda ise makine öğrenmesi, veri analizi ve örüntü tanıma için önemli bir araç haline gelmiştir (Russell & Norvig, 2020). 1990'lı yıllarda yapay sinir ağları (YSA) ve derin öğrenme alanındaki ilerlemeler, yapay zekânın daha karmaşık problemleri çözmesini sağlamıştır (LeCun vd., 2015). 2000'li yıllarda internetin yaygınlaşması, büyük veri (big data) ve bulut bilişim teknolojileri, yapay zekâ uygulamalarının ölçeklenmesini hızlandırmıştır (Goodfellow vd., 2016). 2010'lu yıllarda ise otonom araçlar, sesli yanıt sistemleri ve akıllı robotlar gibi yenilikler, yapay zekânın günlük yaşamda somut bir yer edinmesini sağlamıştır (Brynjolfsson & McAfee, 2014).

Dijital teknolojilerin evrimi, ENIAC'tan Web 3.0'a ve yapay zekâyâ uzanan bir süreçte, insan yaşamını köklü bir şekilde dönüştürmüştür. Yapay zekâ, Turing'in düşünen makineler vizyonundan günümüzün otonom sistemlerine kadar uzanan bir gelişim göstermiş ve makine öğrenmesi, bu alandaki yeniliklerin temel taşı haline gelmiştir. Gelecekteki çalışmalar, hibrit yapay zekâ modelleri, büyük veri entegrasyonu ve gerçek zamanlı izleme sistemleriyle bu

alanda daha ileri yenilikler sağlayabilir. Türkiye'nin Ulusal Yapay Zekâ Stratejisi, bu teknolojilerin sosyo-ekonomik kalkınmadaki rolünü güçlendirmeyi hedeflemektedir (T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi, 2021).

Yapay zekâ, günümüzde birçok alanda dönüştürücü bir rol oynamaktadır. Milli savunma, tıp, mühendislik, finans, e-ticaret, istatistik ve oyun endüstrisi gibi disiplinlerde yapay zekâ uygulamaları yaygındır (Russell & Norvig, 2020). Hidroloji gibi bilimsel disiplinlerde, yapay zekâ ve makine öğrenmesi, su kalitesi tahmini gibi kritik uygulamalarda yüksek doğruluk ve etkinlik sunmaktadır. Yapay zekâ, su kaynaklarının yönetimi, akış tahmini, sel modellemesi ve su kalitesi tahmini gibi konularda önemli katkılar sunmaktadır (Shen, 2018).

Su kalitesi tahmini, çevresel sürdürülebilirlik ve halk sağlığı açısından kritik bir öneme sahiptir. Bu bakımdan bu alanda makine öğrenme yöntemleri gibi yapay zekâ alanları kullanılarak birçok çalışma yapılmıştır. Hidroloji alanında, yapay zekâ ve makine öğrenme yöntemleri, su kalitesi parametrelerindeki (örneğin, çözünmüş oksijen, nitrat, fosfat, klorofil-a) trendleri analiz etmek ve gelecekteki değişimleri tahmin etmek için kullanılmaktadır. Yapay sinir ağları, rastgele orman ve derin öğrenme modelleri, bu alanda sıkça tercih edilen yöntemlerdir. Örneğin, Najah vd. (2013), Malezya'daki Johor Nehri'nde çözünmüş oksijen ve biyokimyasal oksijen ihtiyacını tahmin etmek için YSA kullanmış ve %89 doğruluk elde etmiştir. Benzer şekilde, Zhu vd. (2019), Çin'deki Taihu Gölü'nde toplam azot ve fosfor seviyelerini tahmin etmek için RF ve SVM modellerini uygulamış, RF modelinin %92 doğruluk sağladığını rapor etmiştir. Türkiye'de yapılan bir çalışmada Özçelik (2020), içme suyu kaynaklarındaki kimyasal parametrelerin tahmini için makine öğrenmesi modellerini kullanmış ve LSTM modelinin zaman serisi verilerinde yüksek performans gösterdiğini belirtmiştir. Derin öğrenme yöntemleri, özellikle LSTM, zaman serisi verilerindeki mevsimsel ve uzun vadeli trendleri modellemede geleneksel istatistiksel yöntemlere (Mann-Kendall testi) üstünlük sağlamaktadır (Kim vd., 2022).

Panigrahi vd. (2023), Hindistan Odisha ve Vietnam bölgesindeki yeraltı suyu kalitesini değerlendirmiş ve içilebilirlik tahminini yapmışlardır. Toplamda 15 kimyasal parametre kullanılarak yeraltı suyu kalitesi değerlendirilmiştir. Veri önışlemede eksik verilerin giderilmesi, aykırı değerlerin düzeltilmesi ve verilerin normalize edilmesi gibi işlemler yapılmıştır. Makine öğrenimi algoritmalarından rastgele orman, karar ağacı, destek vektör makineleri, naive bayes, AdaBoost, XGBoost, k-en yakın komşu makine öğrenme yöntemleri kullanılmıştır. Modelleme sonucunda Odisha bölgesinde %70, Vietnam bölgesinde %96 doğruluk oranı ile sonuçlanmıştır. K-en yakın komşu algoritmasında en yüksek accuracy değeri k değeri 2,3 olduğunda elde edilmiştir. En düşük hata oranı ise k değeri 10 seçildiğinde elde

edilmiştir. Destek vektör makineleri yönteminde Odisha en yüksek doğruluk değeri polinomial svm ile elde edilmiş, Vietnam en yüksek doğruluk değeri ise lineer svm, RBF svm ve polinomial svm ile elde edilmiştir. Değerlendirme metriklerinden accuracy, precision, recall, F1-Score ve AUC-ROC kullanılmıştır. En iyi performansı gösteren model, genellikle F1-Score ve AUC-ROC değerlerine göre belirlenmiştir. XGBoost ve polinomial svm en iyi doğruluk değerlerini vermiştir.

Turan (2023), kaggle sitesinden edinmiş olduğu 10 parametreye ait veriyi kullanarak, 15 adet makine öğrenme modeli üzerinde çalışmış, veri setindeki boş alanlar ortalama hesabı yaparak doldurmuştur. Accuracy, AUC-ROC ve F1-Score değerlendirme metriklerini kullanmış, çalışmada en iyi sonucu LGBM Classifier ile SVC modelleri vermiştir. Her üç değerlendirme metriğine bakılarak LGBM Classifier modelinde %0,92 ve SVC modelinde ise %0,91 doğruluk değerini elde etmiştir. Aynı veri seti üzerinde çalışan Nasir vd. (2022), destek vektör makinesi, rastgele orman, lojistik regresyon, karar ağacı, CATBoost, XGBoost ve çok katmanlı algılayıcı (MLP) sınıflandırıcıları ile çalışmış, CATBoost modelinin %94,51'lik bir oranla en doğru sınıflandırıcı olarak bulmuştur. İki çalışma karşılaştırıldığında artırma algoritmaları tahminlerde daha iyi sonuç vermiştir. Yine aynı verilerle çalışan Kaddoura (2022), boş olan verileri veri kümesinden çıkararak rastgele orman, gradyan artırmalı ağaçlar (gradient boosted trees), lojistik regresyon, XGBoost, karar ağacı, k-en yakın komşu, extra trees, yapay sinir ağları, destek vektör makineleri, lasso lars yöntemlerini kullanmış, karmaşıklık matrisini kullanarak edinmiş olduğu F1 -score ve AUC-ROC değerlerine göre destek vektör makinesi ve k-en yakın komşunun üstün olduğuna kanaat getirmiştir.

Poudel vd. (2023), kaggle üzerinden edinmiş olduğu, 9 parametre ve 1 etiketten oluşan 3276 veriyi içeren veri kümesi ile çalışmış, eksik verileri ortalama hesabı ile doldurmuştur. Lojistik regresyon, yapay sinir ağları, rastgele orman ve k-en yakın komşu sınıflandırıcılarını kullanmış; en iyi modeli seçerken de değerlendirme metriklerinden accuracy, precision, recall F1-Score ve AUC (Area Under the Curve) kullanmıştır. Bunlar içerisinde en yüksek accuracy, recall, ve AUC scores değere sahip olan rasgele orman sınıflandırıcısını en iyi model olarak tespit etmiştir. Rastgele ormanın en iyi sınıflandırıcı olduğuna karar vermesinin bir diğer nedeninin de birden çok karar ağacına sahip olarak aykırı değerlere daha dirençli olmasını ileri sürmüştür.

Yurtsever ve Emeç (2023), kaggle platformu üzerinden edinmiş oldukları veri seti üzerinde çalışmışlardır. Veri önışlemede standardizasyon yapılmış, eksik veriler SVR (Support Vektör Regression) ile tahmin edilmeye çalışılmıştır. Eksik verilerin çıkarılması, ortalama yöntemi ile doldurulması ve SVR ile tahmin edilerek hazırlanmış olan veri kümeleri için

karmaşıklık matrisinden faydalananak accuracy, precision, recall and F1-Score metriklerini değerlendirme için kullanmışlardır. Çalışmada hem eksik verileri tamamlamada kullanılan üç yöntem hem de uygulanan algoritmalar karşılaştırılmıştır. En iyi sonucu eksik verileri regresyon yöntemi ile doldurma ve SXH algoritması uygulama ile elde etmişlerdir. Bu çalışma, içme suyu kalitesinin değerlendirilmesinde hibrit makine öğrenimi modellerinin etkinliğini göstermektedir. Özellikle, SVR ve XGBoost yöntemlerinin birleşimiyle oluşturulan SXH modelinin yüksek performansı, literatürdeki benzer çalışmalara kıyasla önemli bir katkı sağlamıştır.

Kuşçu ve Kılıç (2023), Porsuk çayına ait 2007-2017 yılları arasındaki verileri, veri kümesi olarak kullanmışlardır. Veri kümesinde 9 adet parametre bulunmaktadır. Verilerin ölçü birimleri farklı olduğundan veriler üzerinde normalizasyon yapılmıştır. Çoklu doğrusal regresyon, yapay sinir ağları, destek vektör makineleri, gradyan artırma ve extreme gradyan artırma algoritmaları kullanılmıştır. Değerlendirme metriği olarak MAE ve MSE kullanılmıştır. Değeri düşük olan algoritmanın performansı daha iyi olarak değerlendirilmektedir. Çalışmada en düşük değer XGBoost algoritmasında elde edilmiştir.

Abuzir ve Abuzir (2022), kaggle platformu üzerinden elde ettikleri 10 su kalite parametresinden oluşan veri kümesini kullanmışlardır. Algoritmaları uygulamadan önce veri ön işleme sürecinde, eksik ve aykırı veriler temizlenmiş, ardından min-max normalizasyon uygulanarak veriler [0,1] aralığına ölçeklendirilmiştir. Ayrıca, boyut indirgeme ve en etkili özelliklerin seçimi için Pearson korelasyon katsayısı ve Temel Bileşenler Analizi (PCA) kullanılmıştır. Karar ağacı, naive bayes ve MLP algoritmaları uygulanarak model performansları accuracy, ROC eğrisi ve sınıflandırma doğruluğu gibi metriklerle değerlendirilmiş; MLP algoritması en yüksek doğruluğu göstermiştir.

Ahmed vd. (2019), Pakistan'daki Rawal havzasından elde edilen verileri kullanarak dört parametre üzerinden makine öğrenimi modelleri uygulamışlardır. Kutu grafiği analizi kullanılarak aykırı değerler belirlenmiş ve veri setinden çıkarılmıştır. Veriler, min-max normalizasyon yöntemiyle [0,1] aralığına ölçeklendirilmiştir. Korelasyon analizi ve Temel Bileşenler Analizi (PCA) kullanılarak en etkili özellikler belirlenmiş ve veri boyutu azaltılmıştır. Su kalitesi tahmini için on beş adet makine öğrenme algoritması uygulanmıştır. Model performansları MAE, MSE, RMSE ve R-Squared metrikleri ile regresyon algoritmaları; accuracy, precision recall ve F1-Score metrikleri ile de sınıflandırma algoritmaları için değerlendirilmiş ve MLP algoritması en yüksek doğruluk oranını vermiştir.

Azroul vd. (2022), California Su Kaynakları departmanından aldığı gerçek veri kümesi ile çalışmıştır. Veri kümesinden modele etki eden verileri seçtikten sonra makine öğrenme

algoritmalarını uygulamıştır. Çoklu lineer regresyon, gradient boosting, lasso regresyon, SVM, karar ağacı, ve yapay sinir ağları algoritmaları ile çalışmıştır. Tahmin algoritmalarından çoklu lineer regresyon, gradient boosting, lasso regresyon arasından RMS, MSE ve R-Squared değerlendirme metrik sonuçlarına bakıldığında çoklu lineer regresyonun daha iyi sonuç verdiği; SVM, karar ağacı, ve yapay sinir ağları sınıflandırma algoritmalarında ise kullanılan sensitivity, precision, accuracy ve F-Score değerlendirme metriklerine bakıldığında en yüksek değerler yapay sinir ağlarını göstermektedir. Azroul vd. (2022) ayrıca sensitivity değerinin de su kalitesi tahmin çalışmalarında önemli olduğunu belirtmiştir.

Zaman serisi verilerindeki uzun vadeli eğilimleri belirlemek ve gelecekteki gelişmeleri tahmin etmek için temel bir yöntem olarak kullanılan trend analizi, çeşitli disiplinlerde, özellikle ekonomi, finans, sağlık, sosyal bilimler ve çevre bilimlerinde yaygın olarak uygulanır. Trend analizi, su kalitesi alanında parametrelerin zaman serisi verileri ile trendine bakılarak mevsimsellik açısından değerlendirilmesinde de kullanılmaktadır. Bu yöntem, istatistiksel modeller, makine öğrenimi algoritmaları ve veri görselleştirme teknikleri gibi farklı yaklaşımları kapsar (Montgomery vd., 2015). Akademik literatürde trend analizi, hem teorik hem de pratik boyutlarıyla ele alınmış ve farklı yöntemlerin etkinliği üzerine çok sayıda çalışma yapılmıştır. Su kalitesi parametrelerindeki (örneğin, çözünmüş oksijen, nitrat, fosfat, pH, bulanıklık) uzun vadeli eğilimlerin belirlenmesi ve gelecekteki değişimlerin öngörülmesi, iklim değişikliği, arazi kullanımı ve antropojenik etkilerin değerlendirilmesinde temel bir araçtır (Kundzewicz & Robson, 2004).

Box vd. (2015), finansal piyasalardaki fiyat hareketlerini analiz etmek için otoregresif entegre hareketli ortalama (ARIMA) modellerini kullanmışlardır. Çalışmada, ABD hisse senedi piyasalarındaki günlük kapanış fiyatlarının 10 yıllık bir dönemdeki verileri incelenmiştir. ARIMA modeli, verilerdeki mevsimsellik ve gürültüyü ayırıştırarak uzun vadeli trendleri belirlemede başarılı olmuştur. Sonuçlar, ARIMA modelinin kısa vadeli tahminlerde %85 doğruluk sağladığını ve portföy yönetimi için değerli bir araç olduğunu göstermiştir.

Deng vd. (2021), Çin'deki Pearl Nehri Deltasında 2008-2018 yılları arasında su kalitesi parametrelerindeki trendleri incelemiştir. Çalışmada, çözünmüş oksijen, kimyasal oksijen ihtiyacı ve amonyak azotu parametreleri analiz edilmiştir. Mann-Kendall testi ve Sen (2018)'in eğim tahmincisi kullanılmıştır. Sonuçlar, çözünmüş oksijen seviyelerinde artış, kimyasal oksijen ihtiyacı ve amonyak azotu seviyelerinde ise azalma trendi olduğunu göstermiştir. Bu trendler, nehir rehabilitasyon projeleri ve atık su arıtma tesislerinin etkinliği ile ilişkilendirilmiştir. Mann-Kendall testi, trendlerin %95 güven düzeyinde istatistiksel olarak anlamlı olduğunu doğrulamıştır (Deng vd., 2021).

Ho vd. (2019), Kanada'daki Goller Bölgesinde 1990-2015 yılları arasında su kalitesi parametrelerini tahmin etmek için yapay sinir ağı ve ARIMA modellerini karşılaştırmıştır. Çalışmada, pH, çözünmüş oksijen ve toplam fosfor verileri analiz edilmiştir. Yapay sinir ağı modeli, doğrusal olmayan ilişkileri modellemede ARIMA'ya göre daha yüksek doğruluk sağlamıştır. Sonuçlar, yapay sinir ağı modelinin toplam fosfor tahminlerinde %90 doğruluk oranına ulaştığını, ARIMA modelinin ise %82 doğruluk sağladığını göstermiştir. pH seviyelerinde istikrarlı bir trend gözlenirken, toplam fosfor seviyelerinde hafif bir azalma trendi tespit edilmiştir (Ho vd., 2019).

Parmar ve Bhardwaj (2015), Hindistan'daki Yamuna Nehri'nde 2000-2012 yılları arasında biyokimyasal oksijen ihtiyacı ve toplam koliform sayısındaki trendleri incelemiştir. Mann-Kendall testi ve doğrusal regresyon analizi kullanılmıştır. Sonuçlar, biyokimyasal oksijen ihtiyacı seviyelerinde anlamlı bir artış trendi (%3,5/yıl) ve toplam koliform sayısında ise istatistiksel olarak anlamlı bir artış olduğunu göstermiştir. Bu trendler, kentsel atık su deşarjları ve yetersiz arıtma altyapısı ile ilişkilendirilmiştir. Çalışma, su kalitesi yönetiminde acil önlemlerin gerekliliğini vurgulamıştır (Parmar & Bhardwaj, 2015).

Lutz vd. (2020), İspanya'daki Ebro Nehri'nde 1980-2018 yılları arasında nitrat konsantrasyonlarındaki trendleri analiz etmiştir. Çalışmada, Mann-Kendall testi, Sen (1968)'in eğim tahmincisi ve Pettitt testi kullanılmıştır. Sonuçlar, nitrat konsantrasyonlarında 1990'larda başlayan bir azalma trendini ortaya koymuş ve bu trend, tarımsal gübre kullanımının düzenlenmesiyle ilişkilendirilmiştir. Pettitt testi, 1992 yılında nitrat seviyelerinde bir kırılma noktası tespit etmiştir. Trendlerin istatistiksel anlamlılığı %99 güven düzeyinde doğrulanmıştır (Lutz vd., 2020).

Zhu vd. (2019), Çin'deki Taihu Gölü'nde 2005-2017 yılları arasında su kalitesi parametrelerini tahmin etmek için SVM ve rastgele orman modellerini kullanmıştır. Çalışmada, toplam azot, toplam fosfor ve klorofil-a konsantrasyonları analiz edilmiştir. Rastgele orman modeli, SVM'ye kıyasla daha yüksek tahmin doğruluğu (%92) sağlamıştır. Sonuçlar, toplam azot ve toplam fosfor seviyelerinde hafif bir azalma trendi olduğunu, ancak klorofil-a konsantrasyonlarında artış trendi gözlemlendiğini göstermiştir. Bu durum, ötrofikasyon riskinin devam ettiğini işaret etmiştir (Zhu vd., 2019).

Sprague vd. (2017), ABD'deki Mississippi Nehri'nde 1997-2015 yılları arasında bulanıklık ve askıda katı madde seviyelerindeki trendleri incelemiştir. Çalışmada, Mann-Kendall testi ve Yenilikçi trend analizi yöntemleri kullanılmıştır. Sonuçlar, askıda katı madde seviyelerinde %15'lik bir azalma trendi olduğunu ve bu trendin, erozyon kontrol önlemleriyle ilişkili olduğunu göstermiştir. Yenilikçi trend analizi yöntemi, düşük ve orta seviye bulanıklık

değerlerinde daha ayrıntılı trend desenleri ortaya koymuştur. Çalışma, su kalitesi iyileştirme politikalarının etkinliğini vurgulamıştır (Sprague vd., 2017).

Kim vd. (2022), Güney Kore'deki Han Nehri'nde 2010-2020 yılları arasında su kalitesi parametrelerini tahmin etmek için uzun kısa vadeli bellek (LSTM) sinir ağı modelini kullanmıştır. Çalışmada, çözünmüş oksijen, biyokimyasal oksijen ihtiyacı ve toplam fosfor parametreleri analiz edilmiştir. LSTM modeli, zaman serisi verilerindeki karmaşık desenleri modellemede %94 doğruluk sağlamıştır. Sonuçlar, biyokimyasal oksijen ihtiyacı seviyelerinde azalma, oksijen ihtiyacı seviyelerinde ise artış trendi olduğunu göstermiştir. Bu trendler, nehirdeki atık su arıtma kapasitesinin artmasıyla ilişkilendirilmiştir (Kim vd., 2022).

Vyshnevskiy vd. (2021), Ukrayna'daki Dnieper Nehri'nde 1990-2019 yılları arasında su kalitesi parametrelerindeki trendleri ve iklim değişikliğinin etkilerini incelemiştir. Çalışmada, Mann-Kendall testi, Pettitt testi ve çok değişkenli regresyon analizi kullanılmıştır. Analiz, oksijen ihtiyacı, nitrat ve fosfat parametrelerini kapsamıştır. Sonuçlar, nitrat seviyelerinde %10'luk bir azalma, fosfat seviyelerinde ise artış trendi olduğunu göstermiştir. Pettitt testi, 2005 yılında su kalitesi parametrelerinde bir kırılma noktası tespit etmiştir. Bu trendler, iklim değişikliği ve tarımsal faaliyetlerin kombine etkisiyle ilişkilendirilmiştir (Vyshnevskiy vd., 2021).

Najah vd. (2013), Malezya'daki Johor Nehri'nde 2003-2010 yılları arasında su kalitesi parametrelerindeki trendleri tahmin etmek için yapay sinir ağları kullanmıştır. Çalışmada, oksijen ihtiyacı, kimyasal oksijen ihtiyacı ve askıda katı madde parametreleri analiz edilmiştir. Yapay sinir ağları modeli, zaman serisi verilerindeki trendleri belirlemede %89 doğruluk sağlamıştır. Sonuçlar, oksijen ihtiyacı seviyelerinde hafif bir artış, kimyasal oksijen ihtiyacı ve askıda katı madde seviyelerinde ise azalma trendi olduğunu göstermiştir. Bu trendler, nehirdeki endüstriyel atık kontrol önlemleriyle ilişkilendirilmiştir. Çalışma, yapay sinir ağlarının trend tahmini için etkili bir araç olduğunu ortaya koymuştur (Najah vd., 2013).

Wang vd. (2020), Çin'deki Yangtze Nehri'nde 2008-2018 yılları arasında su kalitesi parametrelerindeki trendleri analiz etmek için k-gradient boosting decision tree modelini kullanmıştır. Çalışmada, nitrat, fosfat ve klorofil-a parametreleri incelenmiştir. k-gradient boosting decision tree modeli, trend analizi için zaman serisi verilerindeki uzun vadeli eğilimleri belirlemede %91 doğruluk sağlamıştır. Sonuçlar, nitrat ve fosfat seviyelerinde azalma trendi olduğunu, ancak klorofil-a seviyelerinde artış trendi gözlemlendiğini göstermiştir. Bu trendler, tarımsal faaliyetlerin düzenlenmesi ve atık su arıtma politikalarıyla ilişkilendirilmiştir (Wang vd., 2020).

1.1. SU KALİTESİ

Su canlıların yaşam kaynağıdır. Su olmazsa evrende hayat da olmaz. İnsanların yaşamlarını idame ettirebilmeleri için suya ihtiyaçları vardır. Bu suların sağlıklı sular olması, insan sağlığına zararlı olan maddeleri ihtiva etmemesi gerekmektedir. Çağımız ne kadar ilerlemiş olsa da ne kadar teknolojik çağda yaşıyor olsak da gelişmemiş ülkelerde yaşayan milyarlarca insan hala sağlıklı suya ulaşamamaktadır. UNESCO tarafından yayımlanan 2024 BM Dünya Su Kalkınma Raporuna göre güvenilir bir şekilde içme suyuna erişemeyen 2,2 milyar insanın var olduğu söylenmektedir (United Nations, 2024). Dünya Sağlık Örgütü'ne (WHO) göre, güvenli olmayan içme suyu kaynakları, küresel halk sağlığını tehdit eden en önemli çevresel faktörlerden biri olmaya devam etmektedir. Özellikle düşük ve orta gelirli ülkelerde yaşayan insanlar, mikrobiyolojik ve kimyasal açıdan kirli su kaynaklarına maruz kalmakta ve bu durum ciddi sağlık sorunlarına yol açmaktadır. World Health Organization (2022)'nin yayımladığı verilere göre, her yıl yaklaşık 1 milyon kişi, ishal gibi önlenebilir hastalıklar nedeniyle, doğrudan sağlıksız içme suyuna bağlı olarak yaşamını yitirmektedir. Geri kalmış ülkelerde yaşayan insanların bir kısmı içilebilir suya hala ulaşamazken, gelişmiş veya gelişmekte olan ülkelerde içilebilir sulara doğal kaynaklarından, arıtma veya sanitasyon yoluyla ulaşılabilesine rağmen su kirliliği ile mücadele edilmektedir.

Dünya nüfusundaki hızlı artış, iklim değişikliği, hatalı su yönetimi vb. nedenlerle içme ve kullanma suyuna olan ihtiyaç artarken; endüstrinin yaygınlaşması ile endüstriyel atıkların çoğalması, tarımsal atıklar, atmosfer kirliliği, doğal afetler vb. nedenler, içme ve kullanma suyu olarak insanların kullanımına sunulan suların kirliliğini artırmaktadır. Endüstriyel atık sular; ağır metaller, toksik organik bileşikler, boya maddeleri ve yüksek sıcaklıktaki sular gibi çeşitli kirleticiler içerir. Bu kirlilik, yalnızca ekosistemlerin dengesini bozmakla kalmaz, aynı zamanda tarımsal üretimi, içme suyu kalitesini ve su canlılarının yaşamını da doğrudan etkiler (Arif vd., 2021).

Memorial Sağlık Grubu'na (2025) göre su kirliliği; cilt sorunları, saç dökülmesi, şiddetli ishal, böbrek hasarı, bağışıklık sisteminin düşmesi, kolera, tifo, giardia, lejyoner hastalığı, salmonelloz, dizanteri, hepatit A gibi hastalıklara neden olabilmektedirler.

Görüldüğü üzere su kirliliğinin doğaya ve yaşayan tüm canlılara ciddi olumsuz etkileri vardır. Birçok nedenle kirlenen suların içme ve kullanma suyu olarak tüketiciye sunulması hastalıkları beraberinde getirecektir. Bu yüzden sanayide gerekli filtrasyonların yapılması, endüstriyel ve tarım atıklarının su ekosistemine karışmaması için çalışmalar yapılması ve gerekli tedbirlerin alınması gerekmektedir. Aksi halde bu suların tüketime sunulması halinde, insanlar ölümle sonuçlanabilecek hastalıklarla karşı karşıya kalacaktır.

İnsanların yaşamlarını devam ettirebilmeleri için suyun varlığı yetmemekte aynı zamanda bu suların sağlıklı, temiz, içilebilir, tatsız ve kokusuz olması, içerisinde insan sağlığını olumsuz yönde etkileyebilecek maddelerin bulunmaması gerekmektedir. Nerede yaşadığına bakılmaksızın dünyada yaşayan her insanın eşit bir şekilde sağlıklı su kullanma hakkı vardır. Her ülke vatandaşına sağlıklı su temin etmek zorundadır. Bunun için küresel ölçekte çalışmalar yapılmaktadır. Kirli sular, arıtma sistemleri ile sağlığa zararlı maddelerden arındırılarak kullanıma sunulmuştur. Ayrıca sağlıklı suya erişim için Dünya Sağlık Örgütü tarafından su kalite standartları kılavuzu yayınlanmış ve gelişmiş birçok ülke de kendi standartlarını belirleyerek insan sağlığı için önemli bir yeri olan suyun tüketici kullanımına sunulması sağlanmıştır.

Suyun sağlığı, suyun fiziksel, kimyasal, biyolojik ve radyolojik özelliklerine bağlıdır. B1’de istenilen aralıklarla da denetleme izlemesi yapılmaktadır.

Su kalite kontrolü yapılırken suyun yönetmelikte belirtilen parametre değerleri arasında yer alıp almadığına bakılır. Yapılan ölçümlerde yer alan her bir parametrenin insan sağlığı açısından büyük bir önemi vardır.

1.1.1. Suda Bulunan Mikrobiyolojik Parametreler

1.1.1.1. *Escherichia coli*

Suda *Escherichia coli* () bakterisinin bulunması, suda dışkı varlığının bir göstergesi olup enfeksiyona ve ishale neden olan bir bakteri türüdür. İnsani Tüketim Amaçlı Sular Hakkında Yönetmelik’e göre içme suyunda bu bakterinin bulunmaması gerekmektedir. kontaminasyonu, genellikle dışkı kaynaklı su veya gıda yoluyla bulaşır ve hemorajik kolit, ishal ve hemolitik üremik sendrom (HÜS) gibi hastalıklara neden olabilir (Nataro & Kaper, 1998).

1.1.1.2. Koliform

Koliform bakterinin suda bulunması, suda dışkı bulunduğunun göstergesidir. Bu bakterinin de mezkûr yönetmeliğe göre içme suyunda izin verilen değeri 0’dır. İçme suyunda bulunması ve kullanımı halinde mide bulantısı, kusma ve bağışıklığın düşmesi gibi insan sağlığına etkileri bulunmaktadır.

1.1.1.3. Enterokok

Suyun içerisinde bulunan enterokok bakterisi, suda insan ve hayvan dışkısının varlığına işaret eder. İnsani Tüketim Amaçlı Sular Hakkında Yönetmelik’e göre içme suyunda bulunmaması gereken bir bakteridir.

1.1.2. Suda Bulunan Kimyasal Parametreler

1.1.2.1. Antimon

Yarı metal olan antimonun içme suyunda belirli miktar üzerinde bulunması, mide bulantısı, karın ağrısı, kusma, ishal gibi sağlık sorunlarına neden olmaktadır. Yönetmeliğe göre sınır değeri 5 mg/L'dir.

1.1.2.2. Arsenik

Orman yangınları, tarım ilaçları veya endüstriyel atıklar gibi nedenlerle sulara karışan ve içme suyunda bulunan arseniğin bazı cilt rahatsızlıklarına ve hatta akciğer kanserine neden olan arseniğin içme suyundaki sınır değeri 0,01 mg/L'dir.

1.1.2.3. Benzen

Yakıtlarda bulunan, yanıcı ve aynı zamanda çözücü bir maddedir. İçme sularına karışması halinde insan sağlığına büyük zararı bulunmaktadır. Kanserojen bir madde olan benzen kemik iliği yetmezliğine de sebep olmaktadır. Sınır değeri 1,0 mg/L'dir.

1.1.2.4. Bor

İçme ve kullanma sularını kirleten maddelerin başında bor gelmektedir. Endüstriyel ve tarımsal faaliyetler sonucunda içme sularına karışan bor, mide bulantısı, karın ağrısı, kusma ve ishalle birlikte ciltte deri döküntülerine de neden olmaktadır. Yönetmeliğe göre sınır değeri 1 mg/L'dir.

1.1.2.5. Bromat

İçme ve kullanma sularında normalde bulunmayan, arıtma ve ozonlama sonrasında suda bulunan bromürün bromata dönüşmesi sonucunda suda var olmaktadır. Karaciğer ve böbrek yetmezliğine neden olan ve aynı zamanda kanserojen bir madde olan bromatın içme ve kullanma sularındaki sınır değeri 10 mg/L'dir.

1.1.2.6. Kadmiyum

Madencilik, endüstriyel atıklar sonucunda içme sularına karışan kadmiyum, kanserojen bir madde olup ishal, kusma, böbrek, karaciğer ve kemik hastalıkları gibi rahatsızlıklara neden olmaktadır. İçme ve kullanma sularındaki sınır değeri 5 mg/L'dir.

1.1.2.7. Krom

İçme ve kullanma sularında kromun bulunması cilt tahrişlerine, yanıklara ve zatürreye sebep olmaktadır. İçme ve kullanma sularında sınır değeri 50 mg/L'dir.

1.1.2.8. Bakır

Tarımda kullanılan pestisitler toprakta bakırın bulunmasının nedenlerindendir. Suda fazla miktarda bulunması suyun tadında değişikliğe sebep olmaktadır. İnsan sağlığı açısından bakırın insan vücudunda yeterince olmaması vücuttaki demirin işleyişinin bozulmasına neden olur. Vücuda yeterince bakırın olmamasının sağlığa zararı olmasıyla birlikte fazlalığı da zarar vermektedir. İçme ve kullanma sularındaki sınır değeri 2 mg/L'dir.

1.1.2.9. Kurşun

Toprakta bulunan kurşunun sulara sızması, eski tesisatlarda kullanılan borulardaki kurşunun suya geçmesi ile içme ve kullanma sularında bulunabilmektedir. İnsan sağlığı açısından çocukların fiziksel ve beyin gelişimini etkiler, hiperaktiviteye neden olur. Yönetmeliğe göre sınır değeri 10 mg/L'dir.

1.1.2.10. Cıva

Madencilik, endüstriyel atıklar, termik santraller, atmosferden kaynaklı nedenlerle sulara karışırlar. Maruz kalınması halinde işitme ve görme kaybına, akciğer, kas ve iskelet sorunlarına neden olmaktadır. İçme ve kullanma sularındaki sınır değeri 1 mg/L'dir.

1.1.2.11. Nikel

Endüstriyel atıklarla, topraktan sızıntılarla sulara karışan nikel baş ağrısı, mide bağırsak sorunları, astım, akciğer gibi sağlık sorunlarına neden olmaktadır. Yönetmeliğe göre sınır değeri 20 mg/L'dir.

1.1.2.12. Nitrat

Tarımsal atıklardan kaynaklanabilmektedir. Tarım alanlarının gübrenmesiyle sulama sonrası veya yağmur sonrası yer altına sızan suların su kaynaklarına karışması sonucu olabilmektedir. Yönetmeliğe göre suda nitrat oranı 50 mg/L'dir. Çalık vd. (2004) bu oranı 45 mg olarak belirtmişler ve bu miktarın fazla olması durumunda bebeklerde muthemoglobinemiye sebep olduğunu belirtmişlerdir.

1.1.2.13. Siyanür

Madencilik, endüstriyel atıklar veya atmosfere salınan siyanürün yağmur yoluyla yeryüzüne ve böylece su kaynaklarına sızmaları sonucunda içme ve kullanma sularına karışmaktadır. Maruz kalınması sonucunda solunum, tiroit ve sinir sistemine zarar veren siyanürün yönetmeliğe göre maksimum değeri 50 mg/L'dir.

1.1.2.14. Selenyum

Fosil yakıtlarda, volkanik materyallerde, topraklarda, bitkilerde bulunur. İnsan vücudunda fazla olması halinde bulantı, ishal, kusma gibi sağlık sorunlarına neden olmaktadır. Sınır değeri 10 mg/L'dir.

1.1.2.15. Akrilamid

Kanserojen bir madde olan akrilamidin yönetmeliğe göre sınır değeri 0,1 mg/L'dir.

1.1.2.16. Benzo (a) Piren

Açılımı polisiklik aromatik hidrokarbondur. Üzerine yapılmış birçok çalışma olan ve kanserojen etkiye sahip olduğu belirlenen bu maddenin içme ve kullanma sularında izin verilen sınır değeri 0.010 mg/L'dir.

1.1.2.17. 1,2-Dikloretan

Yanıcı, renksiz fakat kokusu olan klorlu bir hidrokarbondur. Yönetmeliğe göre sınır değeri 3 mg/L'dir.

1.1.2.18. Nitrit

Suda bulunan nitrit suyun kirlilik ölçüsünü göstermektedir. Su içerisinde organik madde olduğunun habercisidir. Yönetmeliğe göre kabul edilebilir nitrit miktarı 0,50 mg/L'dir.

1.1.2.19. Florür

İçme ve kullanma sularında florür miktarının parametre değeri 1,5 mg/L'dir. İnsan sağlığı açısından florür fazlalığı dişlerde kahverengi çizgilere neden olmaktadır.

1.1.2.20. Klor ve klorlu bileşikler

İçme ve kullanma sularında dezenfeksiyon gerekmesi halinde klor ve klorlu bileşikler kullanılmaktadır. Yönetmeliğe göre serbest klor düzeyi 0,2-0,5 mg/L aralığında olmalıdır.

1.1.2.21. Pestisitler

Tarımda kullanılan pestisitler, topraktan sızarak sulara karışma sonucu içme sularında bulunurlar. Cilt döküntüleri, baş dönmesi, baş ağrısı, mide bulantısı, kusma, titreme, panik atak gibi sağlık sorunlarına neden olmaktadır. Yönetmeliğe göre toplam pestisit sınır değeri 0,50 mg/L'dir.

Yönetmelikte yer alan mikrobiyolojik ve kimyasal maddelerin her birinin insan sağlığı üzerinde önemi vardır. Bu maddelerden bazılarının - gibi- insan vücuduna eser miktarda

alınmasının dahi sađlık aısından olumsuz etkileri bulunurken bazılarına insan vucudunun dođal bir řekilde ihtiyaı vardır. Ynetmelikteki bu maddelerin insan vucudunda az miktarda bulunması insan sađlıđına zarar vermekte iken aynı maddelerin fazla olması da bařka zararlara neden olmaktadır. Kimyasal ve mikrobiyolojik maddelerin ynetmelikteki parametre deđeri ime ve kullanma suları iin st sınır deđerini belirtmektedir.

1.2. MAKİNE ĐRENMESİ

Makine đrenmesi, yapay zekânın en nemli alt dallarından biridir ve verilerden đrenen algoritmaların geliřtirilmesini ierir (Mitchell, 1997). Makine đrenmesi, denetimli đrenme (rneđin regresyon ve sınıflandırma), denetimsiz đrenme (rneđin kmeleme) ve pekiřtirmeli đrenme gibi kategorilere ayrılır (Bishop, 2006). Denetimli đrenme algoritmaları, etiketlenmiř verilerle alıřarak giriř-ıkıř iliřkilerini modeller; denetimsiz đrenme, etiketlenmemiř verilerdeki gizli desenleri keřfeder; pekiřtirmeli đrenme ise dl-ceza mekanizmasıyla đrenir (Sutton & Barto, 2018). Yapay sinir ađları, rastgele orman, destek vektr makineleri (SVM) ve derin đrenme modelleri (rneđin, uzun kısa vadeli bellek-LSTM), makine đrenmesinin temel algoritmaları arasında yer alır (Hastie vd., 2009). Derin đrenme, ok katmanlı sinir ađları kullanarak karmařık veri setlerinde yksek dođruluk sađlar ve zellikle zaman serisi analizlerinde etkilidir (LeCun vd., 2015).

1.2.1. Makine đrenmesi Terminolojisi

1.2.1.1. Veri kmesi

Makine đreniminde kullanılacak olan algoritmanın eđitilmesi iin znitelik ve etiketten oluřan verilerin toplamıdır. Veri kmesi genellikle sayı, metin, grsel verilerden oluřmaktadır.

1.2.1.2. znitelikler

Veri kmesinde yer alan stunlar ierisindeki bađımsız deđiřkenlerin niteliđini ifade eder. Kısaca veri kmesindeki her bir stunu ifade eder. znitelikler llebilir veya gzlemlenebilirler. Bu deđiřkenler makinenin sınıflandırma veya tahmin yapması iin kullanılır.

1.2.1.3. Etiketler

Veri setinde yer alan bađımlı deđiřkenlerdir. Gzetimli đrenme problemlerinde makinenin eđitilmesinde nemlidir.

1.2.1.4. Veri ön işleme

Veri kümesini, model seçimi ve eğitimi öncesinde makine öğrenmesinin daha etkili ve güvenli olması için öznitelik silme, yeni öznitelik oluşturma, öznitelik seçme, aykırı değerleri silme vb. işlemler ile model eğitime hazırlama sürecidir. Veri ön işleme modelin performansını artırmak, karmaşıklığı azaltmak, aşırı uydurmayı engelleme bakımından önemlidir. Modelin veriyi daha iyi öğrenmesini ve eğitilmesini sağlamaktadır.

1.2.1.5. Veri bölme

Veri kümesinin eğitim, test ve doğrulama olmak üzere üçe veya daha fazla parçaya bölünmesi işlemidir. Veri kümesinin genellikle büyük bir kısmı modelin eğitimi için kullanılır. Model performansının doğru değerlendirilmesi açısından önemlidir.

1.2.1.6. Eğitim

Veri kümesinde yer alan veriler belirli oranda bölünerek makine öğrenme algoritmasının eğitilmesi için kullanılır. Modelin eğitiminin gerçekleştiği aşamadır.

1.2.1.7. Test

Model eğitildikten sonra geriye kalan veriler ile makine öğreniminin performansını değerlendirmek için algoritmanın test edilmesidir.

1.2.1.8. Model

Makine öğrenme algoritmalarında kullanılan matematiksel işlemlerdir.

1.2.1.9. Aşırı uyum

Makinenin eğitim aşamasında, modelin verileri ezber yapmasıdır. Bu durum tahminlerde doğruluk oranını olumsuz yönde etkiler.

1.2.1.10. Uyum sağlayamama

Makinenin eğitim aşamasında modelin verileri öğrenememesi, modele uyum sağlayamamasıdır.

1.2.1.11. Hiperparametre

Modelin eğitiminde modelin performansını etkileyen parametredir. Modelde aşırı uyum olmaması için doğru bir seçim yapılması açısından önemlidir.

1.2.1.12. P-value

Bu değeri modelleme yapıldıktan sonra veri kümesinde yer alan özneliklerin model sonucuna etkisini belirler. Özneliğin modele olan etkisinin anlamlı olup olmadığını bu değere bakarak anlayabiliriz. P-value değerinin 0,05'ten düşük olması beklenir.

1.2.1.13. Regularizasyon

Makine öğrenmesinde kullanılan modelin aşırı uyumu engellemesi için kullanılan bir yöntemdir.

1.2.1.14. Öğrenme algoritmaları

Makine öğrenmesinde kullanılan veri kümesindeki öznelikler ve etiketler arasındaki ilişkiyi anlamlandırmak ve modellemek için kullanılan matematiksel ve istatistiksel tekniklerdir.

Makine öğrenmesinin gözetimli öğrenme, gözetimsiz öğrenme ve pekiştirmeli öğrenme olmak üzere üç ana yöntemi vardır.

1.2.2. Gözetimli Öğrenme

Gözetimli öğrenme, veri kümesinin bölünme sonrasında eğitim için kullanılan veriler içerisindeki girdiler ile çıktılar arasındaki ilişkiyi öğrenmeyi amaçlayan bir öğrenme yaklaşımıdır. Buradaki girdiler öznelikleri çıktılar ise etiketleri ifade eder. Gözetimli öğrenme algoritmaları sınıflandırma ve regresyon olmak üzere iki başlık altında incelenebilir.

Sınıflandırma algoritmaları:

- K-En Yakın Komşu
- Karar Ağaçları
- Rastgele Orman
- Lojistik Regresyon
- Destek Vektör Makinesi
- Naive Bayes

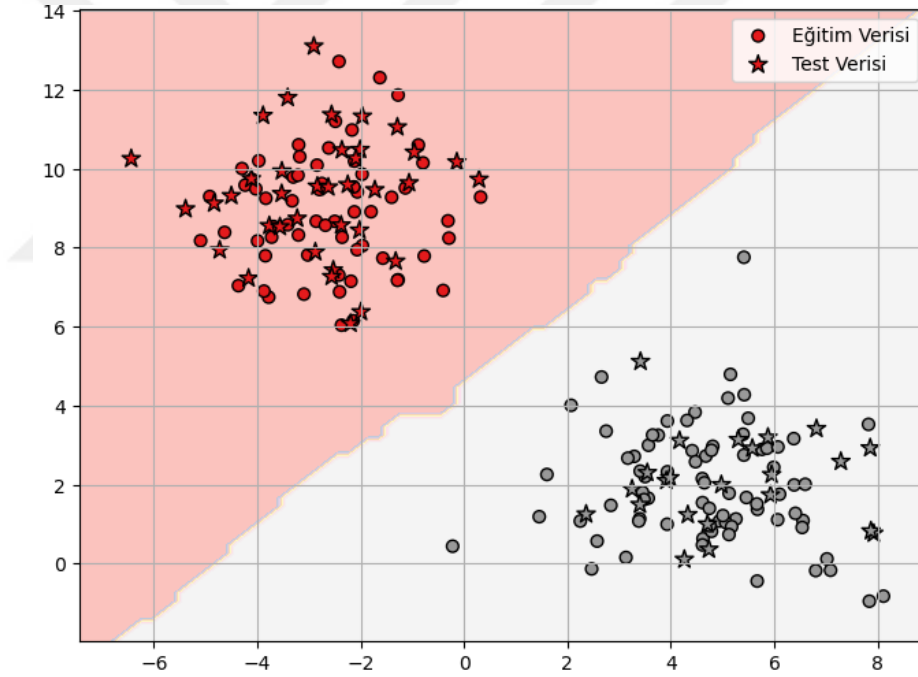
Regresyon algoritmaları:

- Lineer Regresyon
- Çoklu Lineer Regresyon
- Polinom Regresyon

1.2.2.1. K-en yakın komşu

K-en yakın komşu (knn), veri kümesinde bulunan veri noktasının hangi sınıfa ait olduğunu tahmin etmek için kullanılan bir algoritmadır. Hem regresyon hem de sınıflandırma problemlerinde kullanılır. Veri noktasının en yakınındaki komşu diğer veri noktalarının etiket değerlerine bakarak mesafe hesabı ile hangi sınıfa ait olduğunu tahmin eder. Hangi sınıfa ait olduğunu belirlemek için ise belirlenen k (komşu sayısı) değerine göre hesaplama yapılır. Veri noktasına en yakın k tane komşu veriye bakılarak en fazla hangi sınıftan komşuya yakınsa veri noktasının sınıfı ona göre belirlenir. Knn algoritmasında k değerinin seçimi modelin performansı ve verinin sınıflandırmasının doğru yapılması açısından önemlidir.

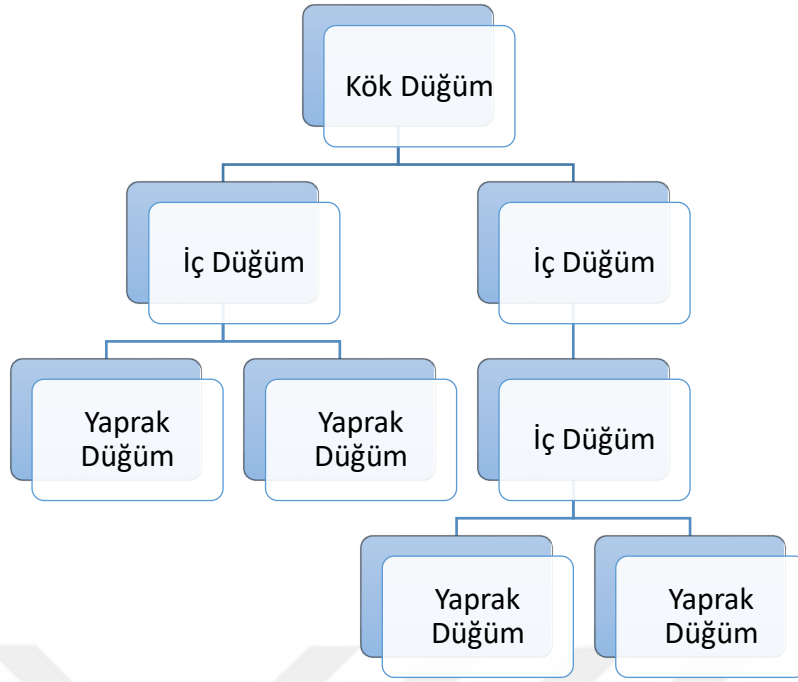
Knn algoritmasında mesafe hesabı, Öklid mesafesi, Manhattan mesafesi, Minkowski mesafesi ve Hamming mesafesi olmak üzere dört farklı şekilde hesaplanabilir. En çok kullanılan mesafe hesabı türü kuş uçuşu mesafeyi ifade eden öklid mesafe hesabıdır. Kodlamalarda genellikle default değeri olarak öklid mesafesi kullanılır.



Şekil 1: Knn Yöntemi

1.2.2.2. Karar ağaçları

Karar ağaçları, ağaç yapısına benzeyen veri kümesini bölerek iç içe sorgularla hedefe ulaşılmasını sağlayan algoritma türüdür. Karar ağaçları kök düğüm, dallar, iç düğümler ve yaprak düğümlerden oluşur. Yaprak düğüm tahminin yapıldığı, kararın verildiği son kısımdır.



Şekil 2: Karar Ağacı Yöntemi

Karar ağaçları ID3 (Iterative Dichotomiser 3), C4.5 ve CART (Classification and Regression Trees) olmak üzere üç farklı türü bulunmaktadır.

ID3

Kategorik verilerle çalışır. Orta ve küçük veri kümelerinde etkilidir. Hızı düşüktür. Düğümleri belirlemek için entropi ve bilgi kazancı algoritmalarını kullanır.

Entropi, düğümdeki veri noktası sınıflarının karmaşıklığını ölçen bir metriktir. Kısaca entropi karmaşıklık demektir. Entropi değeri aşağıdaki formül ile hesaplanmaktadır.

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (1)$$

$E(S)$: S özniteliği için karmaşıklık değerini

p_i : S sütunundaki i olayının gerçekleşme olasılığını ifade etmektedir.

Bilgi kazancı (Information gain), öznitelikler arasında hangi özniteliğin kök düğüm olarak seçilmesinin doğru olacağını, seçilen öznitelik verilerinin sınıflandırma homojenliğini değerlendirir. Bilgi kazancı değeri ne kadar yüksekse düğüm olarak seçilecek olan özniteliğin o kadar iyi bir bölünme kriteri olduğu anlamına gelir. Bilgi kazancı, aşağıdaki formül ile hesaplanır. Bu hesaplamalar her öznitelik için yaprak düğümler oluşuncaya kadar devam eder.

$$Gain(S, A) = E(S) - \sum_{v=values(A)} \left| \frac{S_v}{S} \right| E(S_v) \quad (2)$$

Gain (S,A): A özniteliğinin bilgi kazancını,

E(S): A özniteliği için karmaşıklık değerini

$\left| \frac{Sv}{S} \right|$: A özniteliği içerisinde yer alan v sınıfının oranını ifade etmektedir.

Her bir öznitelik için bilgi kazancı (gain) hesaplanır. Bilgi kazancı değeri yüksek çıkan öznitelik kök düğüm olarak belirlenir. Öznitelik içerisinde yer alan sınıflar iç düğümü oluşturur. Bu iç düğümler içerisinde yaprak düğüme ulaşan sınıf için işlem tamamlanmış olur. Ancak yaprak düğümün oluşmadığı iç düğümlerde ilgili sınıf için yeniden gain hesabı yapılır. Bu durum her dal için yaprak düğüm oluşuncaya kadar devam eder.

C4.5

ID3'ün bir üst versiyonudur diyebiliriz. Sürekli ve kategorik verilerde kullanılır. Orta ve büyük veri kümeleri olduğu durumlarda tercih edilir. Bilgi kazancı ve kazanç oranı algoritmalarını kullanır.

CART

Classification and Regression Trees'in baş harflerinden isimlendirilmiştir. C4.5 gibi sürekli ve kategorik verilerde kullanılır. Büyük veri kümelerinde tercih edilir. Gini impurity algoritmasını kullanır.

Gini impurity (Gini belirsizliği), veri noktalarının karmaşıklığını ifade eder. 0 ile 0.5 aralığında bir değer alır. Gini impurity değerinin 0 olması veri noktalarının aynı sınıfa ait olduğunu, yani karmaşıklığın olmadığını, 0.5 ise karmaşıklığın yüksek olduğunu ifade eder. Değer ne kadar düşükse karmaşıklık o kadar az demektir. Bu da iyi bir bölünme kriteri olarak kabul edilir. Aşağıdaki formül ile hesaplanır.

$$Gini\ Impurity = 1 - \sum p_i^2 \quad (3)$$

Gini impurity değeri her öznitelik içerisinde yer alan her sample sınıfı için ayrı ayrı hesaplanır. Her sınıfa ait gini impurity değerinin ağırlıklı ortalaması hesaplanarak özniteliğin gini impurity değeri hesaplanmış olur. Gini impurity değeri en düşük olan öznitelik kök düğüm olarak belirlenir. Sonraki adımlarda iç düğümler için aynı hesaplamalar tekrar yapılır. Yaprak düğümler tamamlanuncaya kadar işlem devam eder.

Karar ağaçlarında derin bir ağaç oluşumu aşırı uyuma (overfitting) neden olabilir. Bu yüzden derinliğin çok olması istenmez. Orta derinlikte bir karar ağacı modelin jenerikliği açısından daha iyi sonuç vermektedir.

1.2.2.3. Rastgele orman

Birden çok karar ağacının bir araya gelmesiyle oluşan ormana rastgele orman denir. Essamble learning olarak da adlandırılır. Bu öğrenme yönteminde farklı mimariye sahip modeller bir araya gelir. Aynı veya farklı makine öğrenme yöntemlerinin bir araya gelmesiyle oluşan bir makine öğrenme yöntemidir. Rastgele ormanlar aşırı uyuma (overfitting) karşı dirençlidir ve daha büyük veri kümelerinde, daha karmaşık problemlerde kullanılır. Hem sınıflandırma hem de regresyon problemlerinde tercih edilebilir.

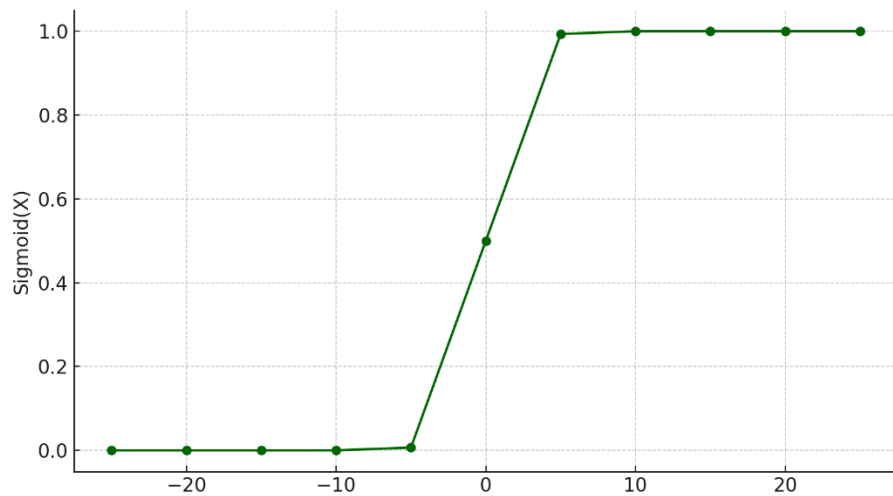
Sınıflandırma problemlerinde, karar ağaçları içerisinde tüm ağaçların tahmini değerlendirilir ve en iyi olan karar ağacı rastgele ormanın problem çözümünde kullanılır.

Regresyon problemlerinde ise karar ağacı için kullanılan her işlem, her durum rastgele ormanda da kullanılabilir.

İçerisinde çok fazla karar ağacı olduğu için eğitim süreci yavaştır. Performansın hızını veri kümesinin büyüklüğü de etkilemektedir.

1.2.2.4. Lojistik regresyon

Lojistik regresyon makine öğrenmesi alanında sınıflandırma problemlerini çözmek için kullanılır. Bağımlı değişkenin tahmini değerini olasılıksal olarak hesaplayarak sınıf belirlemeyi amaçlar. Lineer regresyon $(-\infty, +\infty)$ arasında değer alır. Lojistik regresyonda ise olasılık söz konusudur ve olasılık değerleri $(0, 1)$ aralığında yer alır. Bu nedenle temelde regresyon ile bir bağlantısı olmayan lojistik regresyon, lineer regresyona sigmoid fonksiyonunun eklenmesiyle birlikte oluşmuştur. Böylece değerler $(0,1)$ aralığına indirgenmiş olmaktadır.



Şekil 3: Lojistik Regresyonda Kullanılan Sigmoid Fonksiyonu

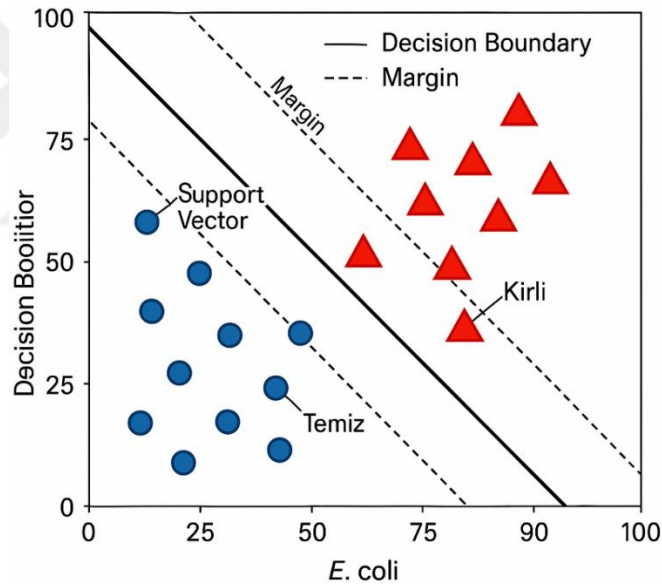
Lojistik regresyon bağımlı değişkenin kategorik olduğu durumlarda tercih edilir. Diğer makine öğrenme algoritmalarında sonuç belirlenirken lojistik regresyonda sonuç olayın gerçekleşme olasılığıdır. Lojistik regresyon kullanılabilmesi için ikili sınıflandırma (binary classification) olması yani etiketin iki farklı kategoriden oluşması gerekmektedir.

Lojistik regresyonda ağırlık (w) ve bias (b) parametreleri bulunur. Ağırlık özneliliğin yani bağımsız değişkenin, bağımlı değişkeni etkileme büyüklüğünü ifade eder. Bias ise alt düzey olarak belirlenen bir değerdir. Lojistik regresyon aşağıdaki formül ile hesaplanır.

$$Z = x_1w_1 + x_2w_2 + b \quad (4)$$

1.2.2.5. Destek vektör makineleri

Destek vektör makineleri (SVM), sınıflandırma ve regresyon problemleri için kullanılan bir makine öğrenme algoritmasıdır. Doğrusal olmayan veri kümelerinde etkilidir. Kernel yöntemi kullanımı ile verileri daha yüksek boyutlu uzaylara dönüştürerek hiperdüzlemi bulur ve veriyi doğrusal olarak ayırabilir bir yapı haline getirir.



Şekil 4: Destek Vektör Makinesi Yöntemi

Doğrusal olmayan yapılara ihtiyaç duyulması halinde kernel yöntemi ile doğrusal yapılardan polinom yapılara dönüştürerek karmaşık problemler çözülebilir.

Büyük veri kümelerinde aşırı uydurmaya eğilimlidir.

1.2.2.6. Naive Bayes

Naive Bayes makine öğreniminde sınıflandırma problemlerinde kullanılan ve tahminler yapan olasılıksal bir yöntemdir. Bu yöntem Naive Bayes teoremine dayanır. Sınıf etiketlerinin birbirinden bağımsız olduğu varsayılır.



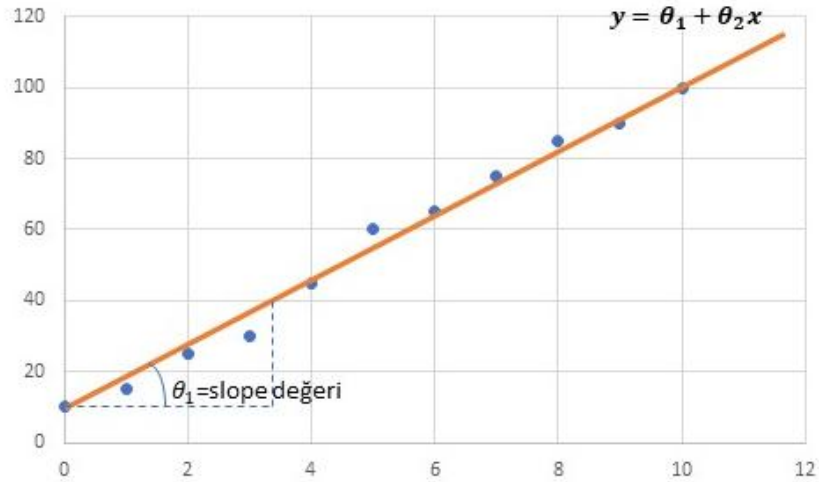
Şekil 5: Naive Bayes Yöntemi

Bu algorithmanda bir verinin tüm olasılıkları hesap edilir ve olasılık değeri yüksek olan durum için diğer verilerin gerçekleşme olasılığı hesaplanır. Hesaplama aşağıdaki formüle göre yapılır.

$$P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)} \quad (5)$$

1.2.2.7. Lineer regresyon

Lineer regresyon, iki veya daha fazla veri arasındaki ilişkiyi istatistiksel yöntemlerle modelleyen bir makine öğrenme yöntemidir. Bu yöntemde bağımsız değişkenin bağımlı değişkene etkisi tahmin edilmeye çalışılır.



Şekil 6: Lineer Regresyon Yöntemi

Lineer regresyon hesabında aşağıdaki formül kullanılır.

$$y = \theta_1 + \theta_2 x + \varepsilon \quad (6)$$

Burada;

y : bağımsız değişken,

θ_1 : bias değeri

θ_2 : eğim değeri

x : bağımsız değişken

ε : modelin tahmin edemediği farkı ifade eder.

Eğim değeri aşağıdaki formül ile bulunur.

$$\theta_2 = \frac{\overline{xy} - \bar{x} * \bar{y}}{\overline{x^2} - \bar{x}^2} \quad (7)$$

Bias değeri ise aşağıdaki formül ile bulunur.

$$\theta_1 = \bar{y} - \theta_2 * \bar{x} \quad (8)$$

Üç adet lineer regresyon türü vardır:

- Basit lineer regresyon
- Çoklu lineer regresyon
- Polinom regresyon

Basit lineer regresyonda bir tane bağımsız değişken vardır. Yukarıda formül ile ifade edilen lineer regresyon modeli basit lineer regresyondur.

Çoklu lineer regresyon, birden fazla bağımsız değişkenden oluşur. Aşağıdaki formül ile hesaplanır:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (9)$$

Burada;

y : bağımlı değişken,

$x_1, x_2, \dots, x_n$: bağımsız değişkenler,

$\beta_1, \beta_2, \dots, \beta_n$: regresyon katsayılarını ifade eder.

Polinom regresyon, doğrusal olmayan veri noktaları arasındaki ilişkiyi modellemek için kullanılır. Aşağıdaki formül ile hesaplanır:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_n x^n \quad (10)$$

Burada;

y : bağımlı değişkeni,

x : bağımsız değişkeni,

$\beta_0, \beta_1, \beta_2, \dots, \beta_n$: regresyon katsayılarını,

n : polinomun derecesini ifade eder.

Polinomun derecesi modelin karmaşıklığını belirlediğinden polinom regresyonda,

polinom derecesi hiperparametresinin seçimi önemlidir. Polinom derecesi gereğinden küçük bir değer seçilirse underfitting, büyük seçilirse overfitting durumu gerçekleşir.

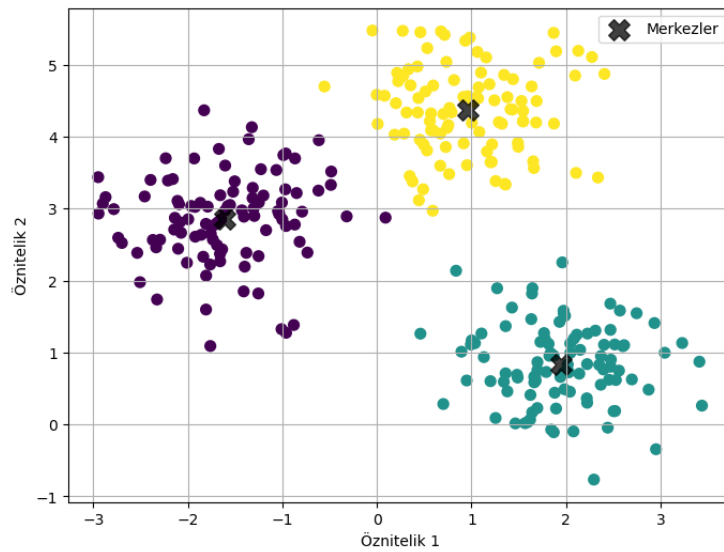
1.2.3. Gözetimsiz Öğrenme

Gözetimsiz öğrenmede çıktı değeri yani etiket bulunmaz. Burada amaç veri kümesindeki örüntüleri ve yapıyı anlamak, ilişkileri belirlemektir. Burada özniteliklere ait veri noktalarının kendi iç özelliklerine göre kümeleme işlemi yapılır. Birbirine yakın olduğu düşünülen veri noktaları aynı kümeye atanır. Temelde veri noktalarını gruplandırma işlemidir. Üç farklı yöntemi vardır. Bunlar;

- K-Ortalamalar Kümeleme (K-Means Clustering)
- Hiyerarşik Kümeleme
- Gürültülü Uygulamalar için Yoğunluk Tabanlı Uzaysal Kümeleme (DBSCAN)

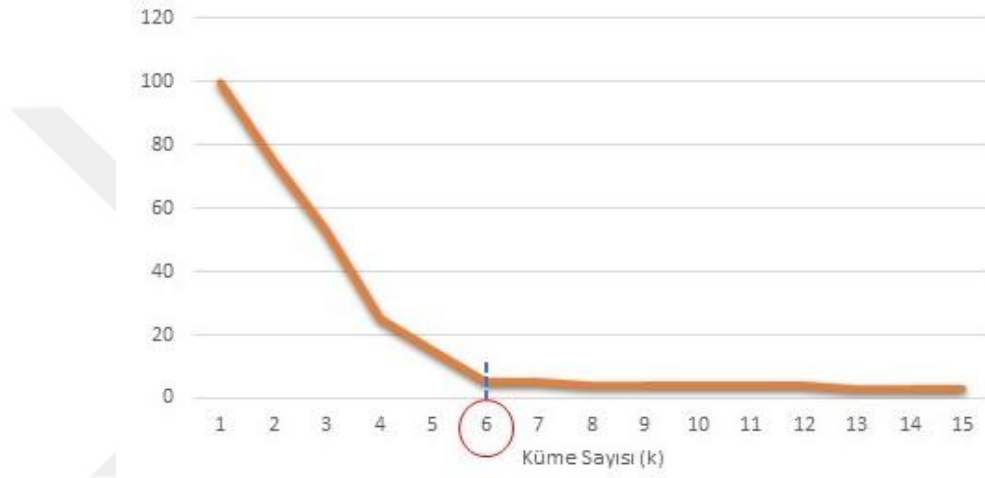
1.2.3.1. K-ortalamalar kümeleme

K-ortalamalar kümeleme, veri setinde yer alan benzer nitelikteki veri noktalarını belli bir sayıda (k) kümelere atamak için kullanılan bir yöntemdir. Bu yöntemde bir küme merkezi belirlenir ve her veri noktası için belirlenen merkeze uzaklık öklid mesafe hesabı ile hesaplanarak veri noktaları belirlenen k sayısı kadar kümelere gruplandırılır. Daha sonra her küme için küme merkezi küme içerisinde bulunan veri noktalarının ortalamasına göre yeniden hesaplanarak veri noktalarının kümelere yeniden atanması veya güncellenmesi yapılır. Bu durum veri noktalarını kümelere atamada herhangi bir değişiklik (güncelleme) olmayana kadar devam eder.



Şekil 7: K-Ortalamalar Kümeleme Yöntemi

K-ortalamlar kümeleme yönteminde, sınıflandırma yapılırken küme sayısı elbow metodu ile hesaplanır. Elbow metodundaki k değeri küme sayısını temsil eder. Bu metotta farklı k değerleri denenerek optimum k değeri belirlenir. Bu değer belli bir k değerinden sonra değişmemeye başlar. Değişimin durduğu ilk k değeri küme sayısı değeri olarak belirlenir. Şekil 8’de görüldüğü gibi, k değeri 6’ten sonra paralel değerleri almaya başlamıştır. Dirsek şeklini alan grafikte dirseğin olduğu kısma denk gelen küme sayısı değeri k değeri olarak belirlenir. k değeri daha yüksek bir değer de seçilebilir fakat bu durum her veri noktası için hesaplama yapmayı gerektireceğinden modelin performansını hız yönünden olumsuz yönde etkileyecektir. Bu bakımdan k değeri model performansını optimum düzeye çıkaracak şekilde seçilmelidir.



Şekil 8: Elbow Methodu (Dirsek Yöntemi)

1.2.3.2. Hiyerarşik kümeleme

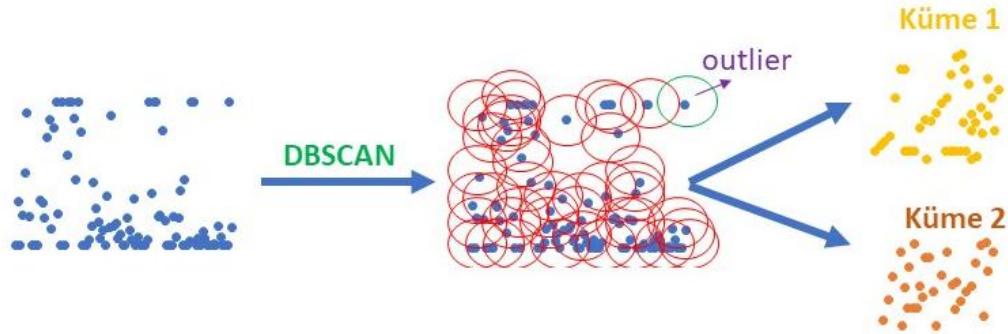
Hiyerarşik kümeleme, veri setindeki veri noktalarını bir hiyerarşik yapı içerisinde gruplandırır. Veri noktalarının benzerlikleri veya birbirlerine olan uzaklıkları hesaplanarak kümelerin birleştirilmesi veya bölünmesiyle hiyerarşik yapı oluşur. Birleştirici hiyerarşik kümeleme ve bölücü hiyerarşik kümeleme olmak üzere iki tür hiyerarşik kümeleme vardır. Birleştirici hiyerarşik kümelemede veri noktalarının her biri bir küme olarak başlar ve benzerlik veya yakınlıklarına göre kümelerin birleşmesiyle hiyerarşik yapı oluşur.

Bölücü hiyerarşik kümelemede ise veri kümesindeki bütün veri noktaları başlangıçta tek bir küme olarak başlar daha sonra benzerlikleri veya mesafeleri doğrultusunda kümelere ayrılırlar.

1.2.3.3. Gürültülü uygulamaların yoğunluğa dayalı mekânsal kümelenmesi (DBSCAN)

DBSCAN (Density-Based Spatial Clustering of Applications with Noise), veri noktalarının yoğunluğuna ve birbirlerine olan mesafelerine göre küme oluşturur. DBSCAN’da bir epsilon (ϵ) ve minpoint hiperparametreleri vardır. ϵ yarıçap değerini ifade eder. Her veri

noktasının etrafına ϵ yarıçapında bir çember çizilir. Bu çember sınırları içerisinde minimum içermesi gereken veri noktası minpoint olarak adlandırılır. Belirlenen minpoint kadar veri noktası içermesi halinde küme oluşur. Bu işlem her veri noktası için hesaplanır.



Şekil 9: DBSCAN Yöntemi

1.2.4. Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme, bir yapay zekâ modelinin veri kümesinden değil de bir ortamdan deneme-yanılma yöntemi ile öğrenme yöntemidir. Pekiştirmeli öğrenmede ajan, bir ortamda bir hamle gerçekleştirir, o hamle sonucunda bir durum değişikliği olmuş olur ve bu durum değişikliği sonrasında ajan ortamdan olumlu veya olumsuz bir geri bildirim alır. Deneme-yanılma yöntemiyle öğrenen ajan, bu eylemlerin sonuçlarını gözlemleyerek en uygun hamlenin ne olacağını öğrenir. Pekiştirmeli öğrenmede Q-learning öğrenme tekniği kullanılır.

1.2.5. Karmaşıklık Matrisi

Karmaşıklık matrisi sınıflandırma problemlerinde sonuçları değerlendirmek için kullanılır. Makine öğrenmesinde geliştirilen sınıflandırma modellerinin başarısı, yalnızca genel doğruluk oranıyla değil, aynı zamanda farklı hata türlerinin incelenmesiyle de değerlendirilmelidir. Bu bağlamda, karmaşıklık matrisi (confusion matrix), modelin hangi sınıfları doğru veya yanlış sınıflandırdığını gösteren önemli bir araçtır (Powers, 2011).

İkili sınıflandırma problemlerinde, karmaşıklık matrisi genellikle dört temel bileşenden oluşur: doğru pozitif (TP), yanlış pozitif (FP), doğru negatif (TN) ve yanlış negatif (FN). Bu bileşenler aracılığıyla modelin duyarlılık (sensitivity (recall)), kesinlik (precision), F1-skoru ve doğruluk (accuracy) gibi metrikleri hesaplanabilir (Sammur & Webb, 2010). Karmaşıklık matrisi tablosu Tablo 1’de gösterilmektedir.

Tablo 1: Karmaşıklık Matrisi

		Gerçek Değerler	
		C1	C2
Tahmin Edilen Değerler	C1	TP (True Positive)	FN (False Negative)
	C2	FP (False Positive)	TN (True Negative)

TP: Gerçek değeri pozitif olan ve doğru (pozitif) tahmin edilen değer sayısını

FN: Gerçek değeri pozitif olan ve yanlış (negatif) tahmin edilen değer sayısını

FP: Gerçek değeri negatif olan ve yanlış (pozitif) tahmin edilen değer sayısını

TN: Gerçek değeri negatif olan ve doğru (negatif) tahmin edilen değer sayısını ifade eder.

Burada FP ve FN yanlış yapılmış olan tahminlerdir. Karmaşıklık matrisinde yer alan bileşenlere bakılarak doğruluk, duyarlılık, kesinlik, özgüllük değerleri hesaplanır ve model değerlendirmesi yapılır. Hangi modelin uygun olduğuna bu değerlendirme sonucunda karar verilir.

Accuracy, modelin yaptığı doğru tahminlerin tüm tahminlere oranını ifade eden temel bir performans ölçütüdür. Modelin ne kadar doğru tahminde bulunduğunu genel olarak gösterir (Sathyanarayanan & Tantri, 2024).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Sensitivity, aynı zamanda True Positive Rate (TPR) veya Recall olarak da bilinir. Bu metrik, modelin gerçek pozitifleri (yani gerçekten olumlu olanları) doğru tanıma oranını gösterir. Modelin gerçekten pozitif olanları pozitif olarak tanıma başarısıdır (Sathyanarayanan & Tantri, 2024).

$$Sensitivity = \frac{TP}{TP + FN} \quad (12)$$

Specificity, bir sınıflandırma modelinin negatif sınıfı doğru tahmin etme yeteneğini ölçen bir performans metriğidir. Diğer bir ifadeyle, modelin gerçekten negatif olanları doğru olarak tahmin etme başarısıdır (Sathyanarayanan & Tantri, 2024).

$$Specificity = \frac{TN}{TN + FP} \quad (13)$$

Precision, modelin pozitif tahminlerinin ne kadarının gerçekten doğru olduğunu gösterir

(Sathyanarayanan & Tantri, 2024).

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

1.3. TREND VE MEVSİMSELLİK ANALİZİ

Zaman serisi analizi, zamanla sıralı verilerin desenlerini anlamak ve gelecekteki eğilimleri öngörmek için kullanılan güçlü bir istatistiksel yaklaşımdır (Hyndman & Athanasopoulos, 2018). Bu analiz, özellikle çevresel verilerde trendlerin, mevsimsel desenlerin ve rastgele dalgalanmaların belirlenmesi için kritik öneme sahiptir. Trend, zaman serisinin genel yönünü (artış, azalma veya sabitlik) ifade eder ve genellikle doğrusal veya doğrusal olmayan modellerdir (Brockwell & Davis, 2016).

Zaman serisi analizlerinde genel olarak kullanılan yöntemlerden biri lineer regresyondur. Lineer regresyon hesabı $y = \theta_1 + \theta_2 x$ formülü kullanılarak yapılır. Burada θ_2 değeri eğimi gösterir. Eğim değeri olarak adlandırılır. Eğim değeri zaman değerindeki artışın bağımlı değişken üzerindeki etkisini belirler. Eğim değeri 0'dan büyükse trend yukarı doğru, 0'dan küçükse trend aşağı doğrudur (Kutner vd., 2005).

$$\theta_2 = \frac{n \sum (x_i y_i) - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (15)$$

$$\theta_1 = \frac{\sum y_i - m \sum x_i}{n} \quad (16)$$

Bu formüller, verinin zaman içindeki eğilimini en iyi şekilde temsil eden doğruyu bulur. Modelin uygunluğunu değerlendirmek için R^2 (açıklanan varyans oranı) ve p-value metrikleri kullanılır. R^2 , modelin veriyi açıklama gücünü gösterir ve R^2 değeri 1'e ne kadar yakınsa, modelin bağımlı değişken üzerindeki açıklama gücü o kadar fazladır. Bu katsayı, modelin ne ölçüde başarılı bir şekilde değişkenliği yakaladığını gösteren önemli bir metriktir (Field, 2013). P-value, eğimin sıfırdan farklı olup olmadığını test eder ($p < 0.05$, istatistiksel anlamlılık için eşik değer olarak kabul edilir) (Montgomery vd., 2012). P-value değeri 0,05'ten küçük ise trendin anlamlı olduğunu gösterir. Yani zamana bağlı olarak bağımlı değişkenin artma ya da azalma eğiliminde olduğunu ifade eder (Montgomery vd., 2012).

1.3.1. Mevsimsellik Analizi ve Zaman Serisi Ayırıştırma

Mevsimsellik, zaman serisinde düzenli ve periyodik dalgalanmaları ifade eder (Hyndman & Athanasopoulos, 2018). Çevresel verilerde mevsimsellik, sıcaklık, yağış veya tarımsal faaliyetler gibi faktörlerden kaynaklanabilir. Mevsimselliği analiz etmek için zaman serisi ayırıştırma yöntemi yaygın olarak kullanılır. Bu yöntem, zaman serisini üç temel bileşene

ayırır:

Trend: Uzun vadeli eğilim.

Mevsimsellik: Periyodik desenler.

Kalıntılar (Residüeller): Rastgele veya açıklanamayan dalgalanmalar.

Zaman serisi ayrıştırma, toplamsal (additive) veya çarpımsal (multiplicative) modellerle gerçekleştirilebilir. Toplamsal model, $y_t = T_t + S_t + R_t$ formunda ifade edilir; burada T_t , S_t ve R_t sırasıyla trend, mevsimsellik ve kalıntılardır (Brockwell & Davis, 2016).

Zaman serisi analizlerinde mevsimsel etkilerin gücünü değerlendirmek amacıyla genellikle mevsimsel bileşenin standart sapmasının artık bileşenin (remainder) standart sapmasına oranı kullanılmaktadır (Cleveland vd., 1990; Hyndman & Athanasopoulos, 2018).

Bu oran:

$$\text{Mevsimsellik gücü} = \frac{\text{std}(S_t)}{\text{std}(R_t)} \quad (17)$$

Formülü ile hesaplanır. 0.6'nın üzerindeki değerler güçlü mevsimsellik, 1.0'ın üzerindeki ise çok güçlü mevsimsellik olarak yorumlanmaktadır. Bu yaklaşım, hem teorik temeli STL (Seasonal-Trend decomposition using Loess) yöntemine dayanan çalışmalarda hem de feasts gibi modern zaman serisi paketlerinde kullanılmaktadır (Hyndman & Athanasopoulos, 2018).

Otokorelasyon fonksiyonu (ACF) grafikleri, mevsimsel desenlerin periyodikliğini görselleştirmek için kullanılmıştır. ACF, bir zaman serisinin kendi gecikmeli kopyalarıyla korelasyonunu ölçer ve düzenli pikler, mevsimsel periyotları gösterebilir (Box vd., 2015).

İKİNCİ BÖLÜM

2. MATERYAL VE YÖNTEM

Su kalitesinin yapay zekâ teknolojileri kullanılarak tahmin edilebilmesi için öncelikle su analizi verilerine ihtiyaç vardır. Makine öğrenimi uygulamalarından önce veriler ön işleminden geçer. Bunun için veri bilimi alanından faydalanılır. Veri analizi, verilerin birbirleri arasındaki ilişkiyi belirleyerek verileri anlamlandırma bilimi olarak tanımlanabilir. Veri analizi bu çalışmaların önemli bir parçasıdır.

Makine öğrenme yöntemleri ile tahmin çalışmalarında verilerin büyüklüğü tahminin genel doğruluk başarı oranı açısından önemlidir. Verinin çokluğu ile tahminin genel doğruluk başarı oranı arasında doğru orantı vardır.

Bu çalışmada Bayburt ili şebeke sularına ait yıllık yapılan mikrobiyolojik ve kimyasal analiz sonuçları verileri ile Bayburt ili umumî çeşme sularının aylık yapılan mikrobiyolojik analiz sonuçları verileri kullanılarak makine öğrenme yöntemleri uygulanarak su kalitesi tahmini ve parametreler arasındaki ilişkilerin analizleri yer almaktadır.

Makine öğrenme yöntemleri çalışmaları için python dağıtım sistemi Anaconda kullanılmıştır. Kodlamalar Anaconda içerisinde yer alan Spyder (Python 3.12) version 5.0 geliştirici arayüz kullanılarak python programlama dili ile yapılmıştır.

2.1. BAYBURT ŞEBEKE SULARI ANALİZ VERİLERİ İLE MAKİNE ÖĞRENMESİ YÖNTEMLERİ

Makine öğrenme yöntemleri ile su kalitesi tahmin çalışması için Bayburt ilinde içme ve kullanıma sunulan suların Bayburt Halk Sağlığı Laboratuvarı ile Erzurum Halk Sağlığı Laboratuvarında yapılan su analizi sonuçları veri kümesi olarak kullanılmıştır. Veri kümesinde “İnsani Tüketim Amaçlı Sular Yönetmeliği’nde Ek-1 (a)’da yer alan , enterokok, koliform bakteri, 22 °C’de koloni sayımı mikrobiyolojik parametreleri ile Ek-1 (b)’de yer alan akrilamid, antimon, arsenik, benzen, benzo (a) piren, bor, bromat, kadmiyum, krom, bakır, siyanür, 1,2-dikloretan, epikloridin, florür, kurşun, cıva, nikel, nitrat, nitrit, pestisitlerddt, toplam pestisitler, polisiklik aromatik hidrokarbonlar, selenyum, tetrakloretan ve trikloretan, trihalometanlar-toplam ve vinil klorür kimyasal parametreleri öznitelik olarak ve sonuç sütunu ise etiket olarak yer almaktadır.

Çalışmada makine öğrenme yöntemlerini uygulamaya geçmeden önce veriler analiz edilerek makine öğrenimine hazır hale getirilmiştir. Çalışmada kullanılan veri kümesi 1180 adet sample’dan oluşmaktadır. Veri kümesinin etiket verileri “uygun” ve “uygun değil” olmak üzere

iki adet kategorik veriden oluşmaktadır. Bu verilerin 685 adeti “uygun” ve 495 adeti “uygun değil” şeklindedir. Özniteliklerin bir kısmında veri bulunmamaktadır. Veri bulunmayan bu alanlar, python’da ortalama stratejisi ile impute edilerek doldurulmuştur. Veri kümesindeki kategorik olan etiket verileri kodlama ile numerik verilere dönüştürülmüştür. Yapılan veri ön işleme sonucunda düzenlenmiş veriler ile numerik verileri içeren yeni bir veri kümesi oluşturularak makine öğrenme uygulamaları için hazır hale getirilmiştir.

Veri kümesi 29 bağımsız değişken ve bir adet bağımlı değişkenden oluşmaktadır. Veri kümesindeki antimon, arsenik, benzen, bor, bromat, kadmiyum, krom, bakır, siyanür, kurşun, cıva, nikel, nitrat, selenyum, akrilamid, benzo (a), dikloretan, epikloridin, florür, nitrit, toplam pestisitler, tetra, toplam trihalometanlar, vinil klorür, toplam ddt, toplam pah, enterokok, koliform, *E. coli* öznitelikleri bağımsız değişken; sonuç etiketi ise bağımlı değişken olarak belirlenmiştir. Verilerin %80’i eğitim, %20’si ise test verisi olarak kullanılmıştır.

2.1.1. Lojistik Regresyon Yöntemi

Lojistik regresyon yöntemi uygulanarak yöntemin doğruluk değeri hesaplanmıştır. Yöntemin doğruluk değeri hesaplanırken karmaşıklık matrisinden faydalanmıştır. Karmaşıklık matrisi sonucu Tablo 2’de gösterilmiştir.

Tablo 2: Lojistik Regresyon Karmaşıklık Matrisi

		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	83 (True Positive)	20 (False Negative)
	C2	7 (False Positive)	126 (True Negative)

2.1.2. Knn Yöntemi

Knn yöntemi uygulanarak farklı k değerleri ile hesaplamalar yapılmıştır. Hesaplamalarda k değeri 3’ten büyük seçildiğinde aynı değerleri verdiği için en küçük değer olan k=3 seçilerek sınıflandırma yapılmıştır. Minkowski, euclidian mesafe hesapları aynı sonucu vermiştir. En yakın 3 komşu veri kümesine mesafe hesabı sonucunda karmaşıklık matrisi Tablo 3’teki şekilde hesaplanmaktadır.

Tablo 3: Knn Karmaşıklık Matrisi -Minkowski/Euclidean Mesafe Hesabı

Minkowski, Euclidean		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	86 (True Positive)	17 (False Negative)
	C2	2 (False Positive)	131 (True Negative)

Manhattan mesafe hesabı ile k=3 değeri seçildiğinde Tablo 4'deki değeri vermektedir.

Tablo 4: Knn Karmaşıklık Matrisi- Manhattan Mesafe Hesabı

Manhattan		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	86 (True Positive)	18 (False Negative)
	C2	1 (False Positive)	131 (True Negative)

2.1.3. SVM Yöntemi

Destek vektör makine yöntemi her bir kernel tipi için uygulanarak karmaşıklık matrisi hesabı ile doğruluk oranları hesaplanmıştır. Yüksekten düşüğe doğru sırası ile linear, RBF, sigmoid ve polinomial kernel tipi doğruluk oranları hesaplanmıştır. Lineer kernel tipi uygulanarak oluşturulan karmaşıklık matrisi verileri Tablo 5'te yer almaktadır.

Tablo 5: SVM Karmaşıklık Matrisi-Linear Metrik

Linear		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	91 (True Positive)	12 (False Negative)
	C2	7 (False Positive)	126 (True Negative)

Polinomial kernel tipi uygulanarak oluşturulan karmaşıklık matrisi verileri Tablo 6'da yer almaktadır.

Tablo 6: SVM Karmaşıklık Matrisi-Polinomial Metrik

Polinomial		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	53 (True Positive)	50 (False Negative)
	C2	1 (False Positive)	132 (True Negative)

RBF kernel tipi uygulanarak oluşturulan karmaşıklık matrisi verileri Tablo 7’de yer almaktadır.

Tablo 7: SVM Karmaşıklık Matrisi-RBF Metrik

RBF		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	84 (True Positive)	19 (False Negative)
	C2	3 (False Positive)	130 (True Negative)

Sigmoid kernel tipi uygulanarak oluşturulan karmaşıklık matrisi verileri Tablo 8’de yer almaktadır.

Tablo 8: SVM Karmaşıklık Matrisi-Sigmoid Metrik

Sigmoid		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	84 (True Positive)	19 (False Negative)
	C2	13 (False Positive)	120 (True Negative)

2.1.4. Naive Bayes Yöntemi

Gaussian naive bayes yöntemi kullanılarak yapılan makine öğreniminde elde edilen karmaşıklık matrisi Tablo 9’da gösterilmektedir.

Tablo 9: Naive Bayes Karmaşıklık Matrisi

		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	83 (True Positive)	20 (False Negative)
	C2	9 (False Positive)	124 (True Negative)

2.1.5. Karar Ağacı Yöntemi

Karar ağacı yöntemi uygulanarak entropi ve gini metriği ile yapılan makine öğrenme yöntemine ait karmaşıklık matris hesabı verileri Tablo 10'da gösterilmektedir. Entropi ile yapılan makine öğrenmesi modelinin doğruluk oranı daha yüksek çıkmıştır.

Tablo 10: Karar Ağacı Karmaşıklık Matrisi-Entropi

Entropi		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	102 (True Positive)	1 (False Negative)
	C2	1 (False Positive)	132 (True Negative)

Tablo 11:Karar Ağacı Karmaşıklık Matrisi-Gini

Gini		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	101 (True Positive)	2 (False Negative)
	C2	4 (False Positive)	129 (True Negative)

2.1.6. Rastgele Orman Yöntemi

Rastgele orman yönteminde default ayarları olan $n_estimators=10$ olarak seçilmiştir. Bu metrik modelin kaç tane karar ağacından oluşacağını belirler. Entropi ve gini metriği ayrı ayrı uygulanmıştır. Gini metrisi ile uygulanan modelde doğruluk değeri daha yüksek çıkmıştır. Rastgele orman yöntemine ait karmaşıklık matrisi sonuçları Tablo 12 ve Tablo 13'te yer almaktadır.

Tablo 12: Rastgele Orman Karmaşıklık Matrisi-Gini

Gini		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	99 (True Positive)	4 (False Negative)
	C2	2 (False Positive)	131 (True Negative)

Tablo 13: Rastgele Orman Karmaşıklık Matrisi-Entropi

Entropi		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	100 (True Positive)	3 (False Negative)
	C2	4 (False Positive)	129 (True Negative)

2.2. BAYBURT İLİ UMUMÎ ÇEŞME SULARI ANALİZ VERİLERİ İLE MAKİNE ÖĞRENMESİ YÖNTEMLERİ

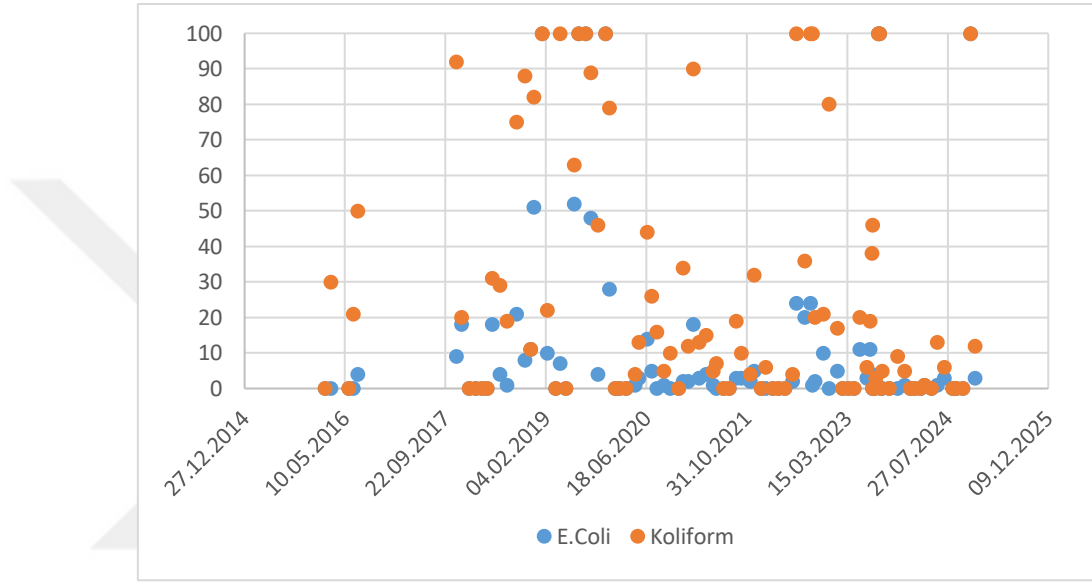
Bayburt ilinde hanelerde kullanıma sunulan sular su kirliliği, suyun tadı ve renginin iyi olmaması vb. nedenlerle içilememektedir. Bu nedenle insanlar hazır su, arıtma cihazı veya umumî çeşmeler ile içme sularına ulaşabilmektedirler. Bayburt'ta 2014-2016 yılları arasında alt yapı çalışmaları yapılmış ve hanelere ulaşan sularda iyileştirmeye gidilmiştir. Ancak birçok hane halkı içme sularında yapılan iyileştirmelere rağmen hala umumî çeşmelerden temin ettikleri suları içme suyu olarak kullanmaktadır. Bayburt İl Sağlık Müdürlüğü, içme suyu olarak kullanılan umumî çeşme sularının Bayburt İl Sağlık Müdürlüğü Halk Sağlığı Laboratuvarında yapılan su mikrobiyolojisi analiz sonuçlarını her ay yayınlamaktadır. Böylece suyun temiz veya kirli olma durumuna göre halkın içme suyu olarak kullandıkları umumî çeşme sularının durumu ile ilgili halk bilgilendirilmiş olmaktadır. Çalışmanın bu kısmında içme suyu olarak kullanılan umumî çeşmelerin su kalitesi analiz sonuçlarından faydalanılarak makine öğrenme yöntemleriyle su kalitesi tahmini çalışmaları yapılacaktır. Ayrıca korelasyona hesabı ile parametrelerin birbirleriyle olan ilişkileri incelenecektir.

Çeşmelerin 2014 yılından başlamak üzere günümüze kadar su mikrobiyolojisi analizleri her ay yapılmaktadır. Mikrobiyolojik parametrelerden *E. coli* ve koliform bakterilerine bakılmaktadır.

Çalışma için Bayburt ilinde en çok kullanılan umumî çeşmelerden Gordon Çeşmesi,

Kışla Çeşmesi, Tarım Çeşmesi ve Narkazan Çeşmesine ait su mikrobiyolojisi analiz sonuçlarına ilişkin veriler kullanılacaktır. Su mikrobiyolojisi analizleri yapılırken 100 ml suda *E. coli* ve koliform bakteri koloni sayısına bakılmaktadır. İnsani Tüketim Amaçlı Kullanılan Sular Yönetmeliği'ne göre izin verilen maksimum değer 0'dır. Suda *E. coli* veya koliform bakteri koloni sayısının bulunması durumunda miktarı ne olursa olsun analiz sonucu kirli olarak çıkmaktadır.

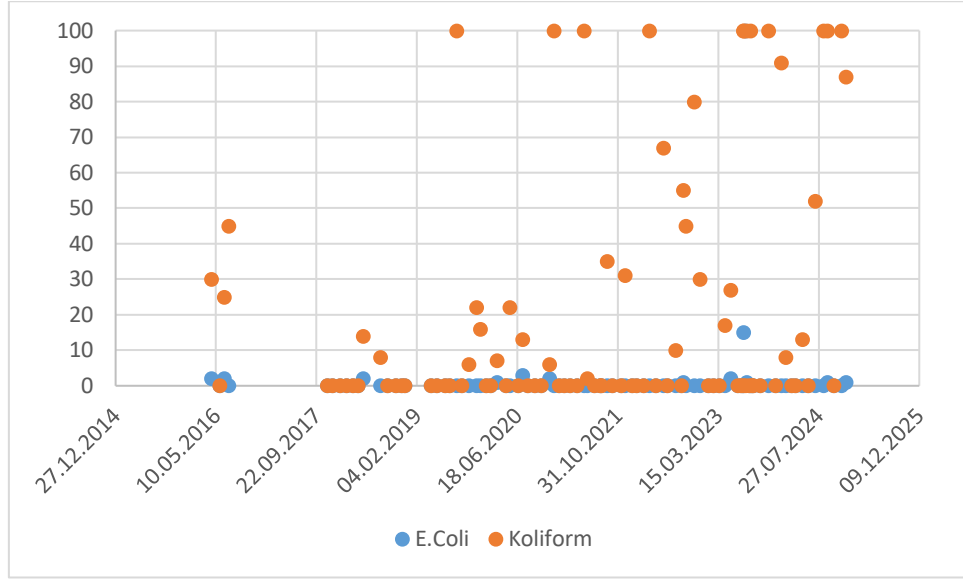
Gordon çeşmesi su mikrobiyolojisi analiz sonuçlarına ilişkin grafik Şekil 10'da gösterilmektedir.



Şekil 10: Gordon Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları

Gordon çeşmesi su mikrobiyolojisi analizi 2016-2024 yılları arasında yapılan analizleri içermektedir. Belli zamanlarda sular kirli çıkmıştır. Grafik incelendiğinde en fazla koliform bakteriden kaynaklı su kirliliğinin olduğu anlaşılmaktadır. Yapılan analizlerin yaklaşık %57'sinde *E. coli* bakterisine, %67,34'ünde ise koliform bakteriye rastlanmıştır. Su analizlerinin %67,34'ünde sular kirli çıkmıştır.

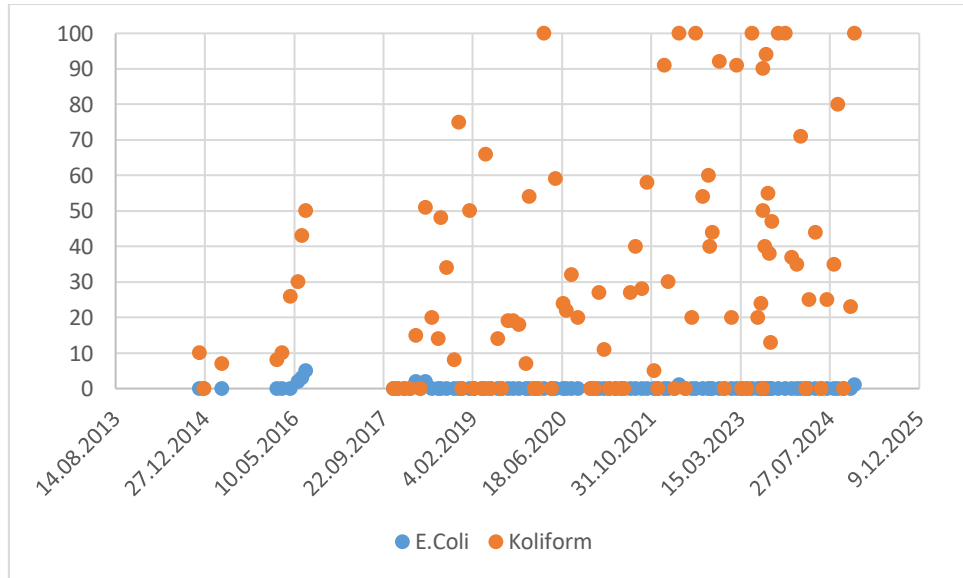
Kışla çeşmesi su mikrobiyolojisi analiz sonuçlarına ilişkin grafik Şekil 11'de yer almaktadır.



Şekil 11: Kışla Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları

Kışla çeşmesi su mikrobiyolojisi analizi 2016-2024 yılları arasında yapılan analizleri içermektedir. Belli zamanlarda sular kirli çıkmıştır. Grafik incelendiğinde en fazla koliform bakteriden kaynaklı su kirliliğinin olduğu anlaşılmaktadır. *E. coli* bakterisine ise Gordon çeşmesine nispeten daha az rastlandığı görülmektedir. Yapılan analizlerin yaklaşık %13,83'ünde *E. coli* bakterisine, %42,55'inde ise koliform bakteriye rastlanmıştır. Su analizlerinin %43,6'sında sular kirli çıkmıştır.

Tarım çeşmesi su mikrobiyolojisi analiz sonuçlarına ilişkin grafik Şekil 12'de gösterilmektedir.

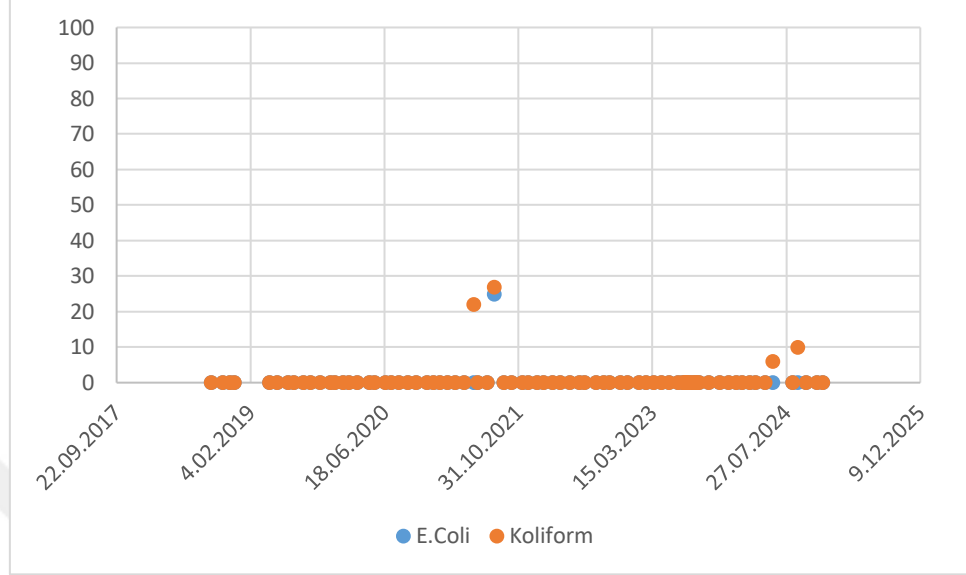


Şekil 12: Tarım Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları

Tarım çeşmesi su mikrobiyolojisi analizi 2014-2024 yılları arasında yapılan analizleri içermektedir. Belli zamanlarda sular kirli çıkmıştır. Grafik incelendiğinde en fazla koliform bakteriden kaynaklı su kirliliğinin olduğu anlaşılmaktadır. Yapılan analizlerin yaklaşık

%6,86'sında *E. coli* bakterisine, %70,59'unda ise koliform bakteriye rastlanmıştır. Su analizlerinin %70,59'unda sular kirli çıkmıştır.

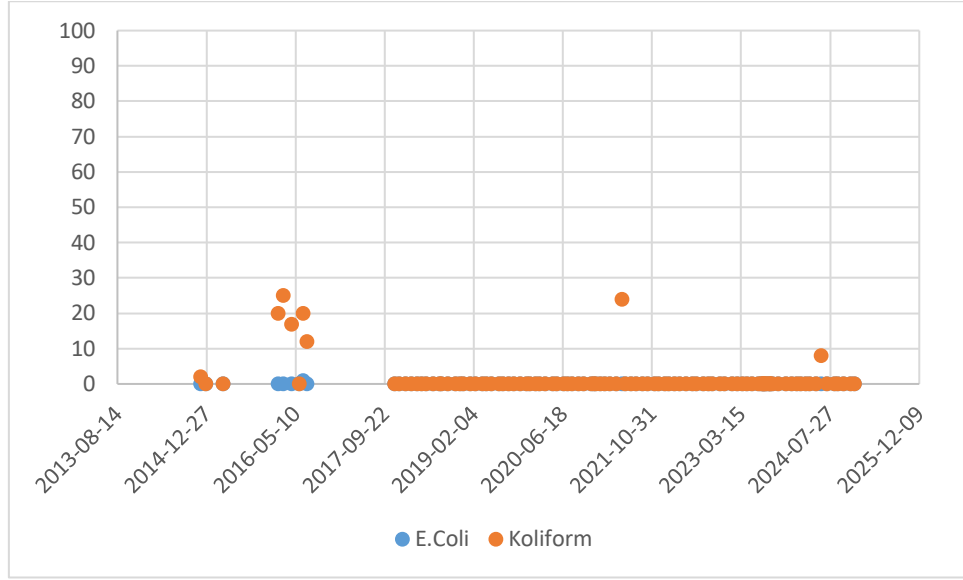
Aşağı Narkazan çeşmesi su mikrobiyolojisi analiz sonuçlarına ilişkin grafik Şekil 13'te gösterilmektedir.



Şekil 13: Aşağı Narkazan Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları

Aşağı Narkazan çeşmesi su mikrobiyolojisi analizi 2018-2024 yılları arasında yapılan analizleri içermektedir. Belli zamanlarda sular kirli çıkmıştır. Grafik incelendiğinde en fazla koliform bakteriden kaynaklı su kirliliğinin olduğu anlaşılmaktadır. Yapılan analizlerin yaklaşık %1,26'sında *E. coli* bakterisine, %5,06'sında ise koliform bakteriye rastlanmıştır. Su analizlerinin %5,06'sında sular kirli çıkmıştır.

Yukarı Narkazan çeşmesi su mikrobiyolojisi analiz sonuçlarına ilişkin grafik Şekil 14'te gösterilmektedir.



Şekil 14: Yukarı Narkazan Çeşmesi Su Mikrobiyolojisi Analiz Sonuçları

Yukarı Narkazan çeşmesi su mikrobiyolojisi analizi 2014-2024 yılları arasında yapılan analizleri içermektedir. Belirli zamanlarda sular kirli çıkmıştır. Özellikle 2014-2016 yılları arasında kirlilik vardır. Grafik incelendiğinde en fazla koliform bakteriden kaynaklı su kirliliğinin olduğu anlaşılmaktadır. 2017 yılından itibaren kirliliğin büyük ölçüde azaldığı sonraki yıllarda sadece birkaç defa analiz sonuçlarında suyun kirli çıktığı görülmüştür. Yapılan analizlerin yaklaşık %0,97'sinde *E. coli* bakterisine, %7,77'sinde ise koliform bakteriye rastlanmıştır. Su analizlerinin %7,77'sinde sular kirli çıkmıştır.

Çeşmeler içerisinde *E. coli* bakterisine en az Yukarı Narkazan çeşmesinde rastlanıldığı görülmektedir. Koliform bakteriye ise en az Aşağı Narkazan çeşmesinde rastlanmıştır. En az kirlilik oranının ise Aşağı Narkazan çeşmesine ait olduğu görülmektedir.

Bayburt'ta yer alan umumî çeşmelerden Aşağı Narkazan, Yukarı Narkazan, Tarım, Gordon ve Kışla çeşmelerinin aylık yapılan su analiz sonuçları makine öğrenmesinde veri kümesi olarak kullanılmıştır. Veri kümesi; *E. coli*, koliform ve sonuç sütunlarından oluşmaktadır. Makine öğrenmesi uygulamalarında *E. coli* ve koliform öznitelikleri ile sonuç etiket verilerinden faydalanılmıştır. Veri kümesi 476 adet sample'dan oluşmaktadır. Veri kümesinin etiket verileri “kirli” ve “temiz” olmak üzere iki adet kategorik veriden oluşmaktadır. Bu verilerin 285 adeti “temiz” ve 191 adeti “kirli” şeklindedir. Sonuç etiketi kategorik veri olduğu için numerik veriye dönüştürülmüştür. Bu aşamadan sonra veri kümesi makine öğrenim uygulamaları için hazır hale gelmiştir. Verilerin %80'i eğitim %20'si ise test verisi olarak kullanılmıştır.

2.2.1. Lojistik Regresyon Yöntemi

Lojistik regresyon verilere uygulandıktan sonra tahminin doğruluğu karmaşıklık matrisi yöntemi ile hesaplanmıştır. Karmaşıklık matrisi sonuçlarına ilişkin veriler Tablo 14'te yer almaktadır.

Tablo 14: Umumî Çeşmeler Lojistik Regresyon Karmaşıklık Matrisi

		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	65 (True Positive)	0 (False Negative)
	C2	3 (False Positive)	28 (True Negative)

2.2.2. Knn Yöntemi

Knn yöntemi uygulanarak farklı k değerleri ile hesaplamalar yapılmıştır. Hesaplamalarda k değeri 3'ten büyük seçildiğinde aynı değerleri verdiği için en küçük değer olan k=3 seçilerek sınıflandırma yapılmıştır. En yakın 3 komşu değere bakarak yapılan sınıflandırmada tahmin doğruluğu karmaşıklık matrisi yöntemi ile hesaplanmıştır. Karmaşıklık matrisi Tablo 15'te gösterilmektedir. Minkowski, euclidean, manhattan mesafe hesapları ayrı ayrı kullanılmış hepsinde de aynı sonucu (Tablo 15) vermiştir. Hamming mesafe hesabı ile yapılan hesaplamada doğruluk oranı diğer hesaplamalara göre daha düşük çıkmıştır.

Tablo 15: Umumî Çeşmeler Knn Yöntemi Karmaşıklık Matrisi

		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	65 (True Positive)	0 (False Negative)
	C2	1 (False Positive)	30 (True Negative)

2.2.3. SVM Yöntemi

Destek vektör makine yöntemi, kernel tipi lineer, RBF, sigmoid ve polynomial seçilerek ayrı ayrı uygulanmıştır. Yapılan sınıflandırmada doğruluk oranı karmaşıklık matrisi ile hesaplanarak sonuçları Tablo 16, Tablo 17'de gösterilmiştir. Linear, RBF, sigmoid aynı doğruluk oranlarını vermiş ve polynomial kernel tipine göre doğruluk oranı daha yüksek çıkmıştır.

Tablo 16: Umumî Çeşmeler SVM Yöntemi Karmaşıklık Matrisi (Linear, RBF, Sigmoid)

Linear, RBF, sigmoid		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	65 (True Positive)	0 (False Negative)
	C2	1 (False Positive)	30 (True Negative)

Tablo 17: Umumî Çeşmeler SVM Yöntemi Karmaşıklık Matrisi (Polinomial)

Polynomial		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	65 (True Positive)	0 (False Negative)
	C2	19 (False Positive)	12 (True Negative)

2.2.4. Naive Bayes Yöntemi

Naive bayes yöntemi ile makine öğrenmesinde gauss algoritması kullanılmıştır. Elde edilen karmaşıklık matrisi Tablo 18’de gösterilmektedir.

Tablo 18: Umumî Çeşmeler Naive Bayes Karmaşıklık Matrisi

		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	65 (True Positive)	0 (False Negative)
	C2	0 (False Positive)	31 (True Negative)

2.2.5. Karar Ağacı Yöntemi

Karar ağacı yönteminde entropi ve gini metriği ile ayrı ayrı uygulanan makine öğrenme yönteminde karmaşıklık matrisi Tablo 19’da gösterildiği gibi hesaplanmış ve her iki metrikle uygulanan modelde sonuçlar aynı çıkmıştır.

Tablo 19: Umumî Çeşmeler Karar ağacı Yöntemi Karmaşıklık Matrisi

Entropi, Gini		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	65 (True Positive)	0 (False Negative)
	C2	0 (False Positive)	31 (True Negative)

2.2.6. Rastgele Orman Yöntemi

Rastgele orman yönteminde default ayarları olan $n_estimators=10$ olarak seçilmiştir. Entropi ve gini metriği ayrı ayrı uygulanmıştır. Her ikisinde de sonuçlar aynı çıkmıştır. Rastgele orman yöntemine ait karmaşıklık matrisi sonuçları Tablo 20’de yer almaktadır.

Tablo 20: Umumî Çeşmeler Rastgele Orman Yöntemi Karmaşıklık Matrisi

Entropi, Gini		Gerçek Değerler	
		C1	C2
Tahmin Değerleri	C1	65 (True Positive)	0 (False Negative)
	C2	0 (False Positive)	31 (True Negative)

2.3. BAYBURT İLİ UMUMÎ ÇEŞMELERİ TREND VE MEVSİMSELLİK ANALİZİ

Mikrobiyolojik analiz sonuçlarının zaman serisi perspektifinden değerlendirilmesi, çevresel kontaminasyonun dinamiklerini anlamak açısından önemlidir.

Analizde, *E. coli* ve koliform bakterilerin zaman serisi verileri kullanılmıştır. Trend analizi için lineer regresyon yöntemi uygulanmış, mevsimsellik ve kalıntı analizleri ise zaman serisi ayrıştırma teknikleriyle gerçekleştirilmiştir. İstatistiksel anlamlılık $p < 0.05$ düzeyiyle değerlendirilmiş, açıklayıcılık ise R^2 değerleriyle ölçülmüştür. Grafikler; orijinal veriler, trend, mevsimsellik ve kalıntı bileşenlerini görselleştirmektedir.

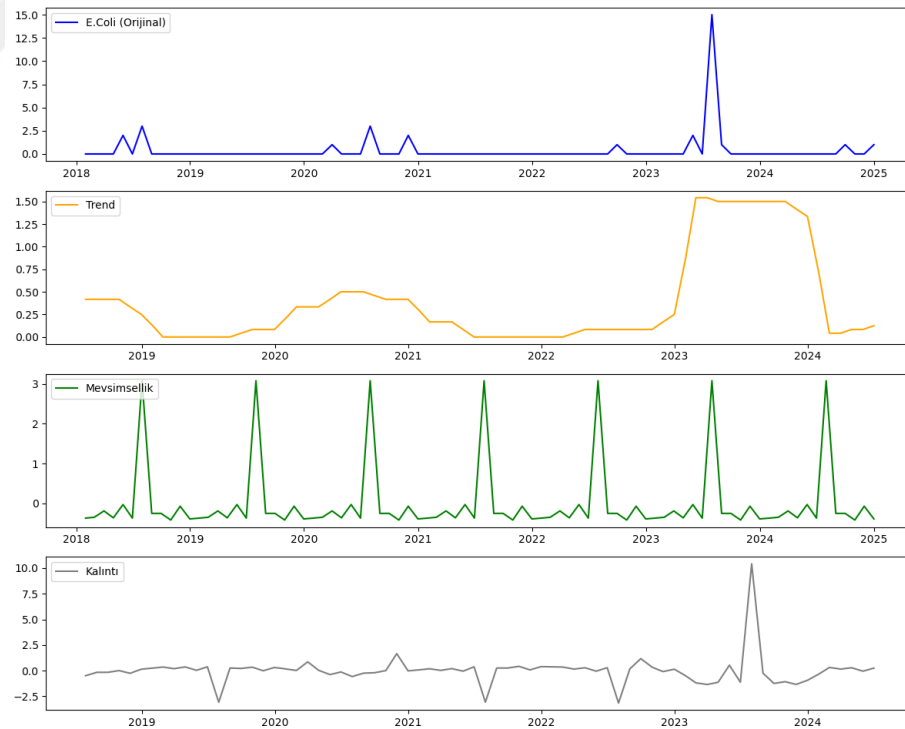
Trend analizleri yapılırken seriler, statsmodels kütüphanesinin `seasonal_decompose` fonksiyonu ile ayrıştırılmış, her bir bileşen (gerçek veriler, trend, mevsimsellik ve kalıntı) grafiksel olarak incelenmiştir. Veriler aylık düzeyde olduğu için frekans değeri 12 olarak belirlenmiştir. Uygulanan trend ve mevsimsellik analizlerine ait sonuçlar Tablo 21’de gösterilmektedir.

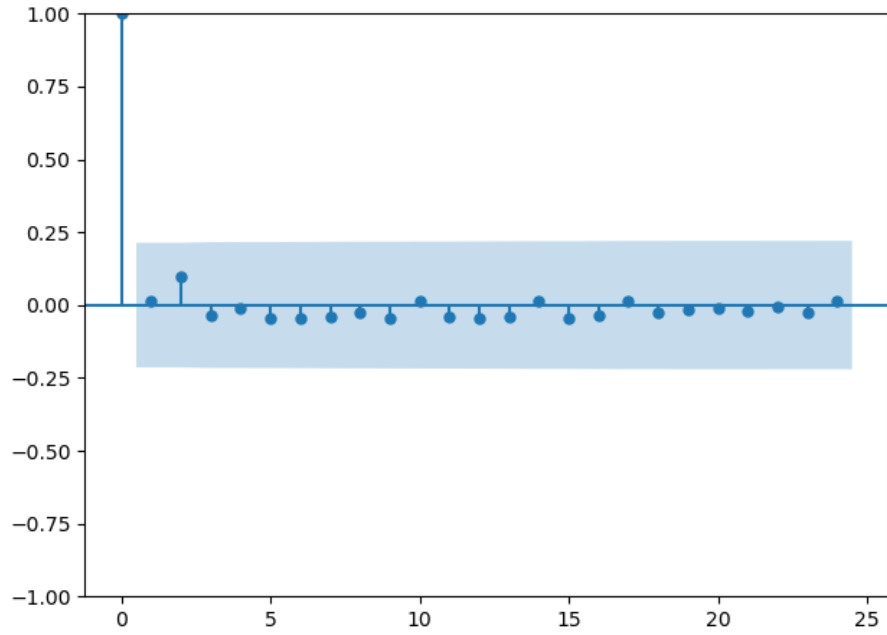
Tablo 21: Trend ve Mevsimsellik Analiz Sonuçları

<i>E. coli</i> Trend Analizi/Mevsimsellik Sonuçları							
Çeşme Adı	Eğim	Kesme Noktası	R ²	P-Value	Anlamlılık	Mevsimsellik Gücü	Mevsimsellik
Aşağı Narkazan	-0.0001	0.4755	0.0007	0.8304	Anlamlı Değil	0.5012	Yok
Gordon	-0.0067	20,2879	0.0395	0.0698	Anlamlı Değil	0.4885	Yok
Kışla	0.0002	0.1414	0.0067	0.4608	Anlamlı Değil	0.6326	Yok
Tarım	-0.0001	0.1580	0.0225	0.1733	Anlamlı Değil	0.4699	Yok
Yukarı Narkazan	0.0000	0.0000	0.0000	1.000	Anlamlı Değil	-	-

Koliform Trend Analizi/Mevsimsellik Sonuçları							
Çeşme Adı	Eğim	Kesme Noktası	R ²	P-Value	Anlamlılık	Mevsimsellik Gücü	Mevsimsellik
Aşağı Narkazan	0.0002	0.6571	0.0012	0.7768	Anlamlı Değil	0.4912	Yok
Gordon	-0.0123	41,1088	0.0723	0.0134	Anlamlı	0.5744	Yok
Kışla	0.0186	-1,8146	0.1539	0.0002	Anlamlı	0.4091	Yok
Tarım	0.0130	14,7144	0.0897	0.0056	Anlamlı	0.3863	Yok
Yukarı Narkazan	0.0002	0.1731	0.0020	0.6881	Anlamlı Değil	0.4260	Yok

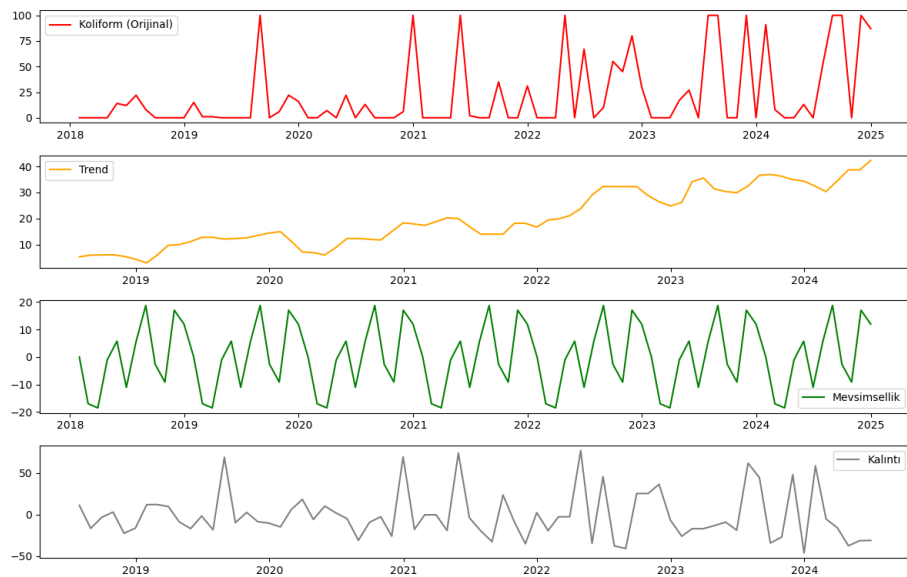
Kışla çeşmesi *E. coli* trend analizi grafiği Şekil 15’de otokorelasyon grafiği ise Şekil 16’da gösterilmektedir.

**Şekil 15:** Kışla Çeşmesi *E. coli* Trend Analizi

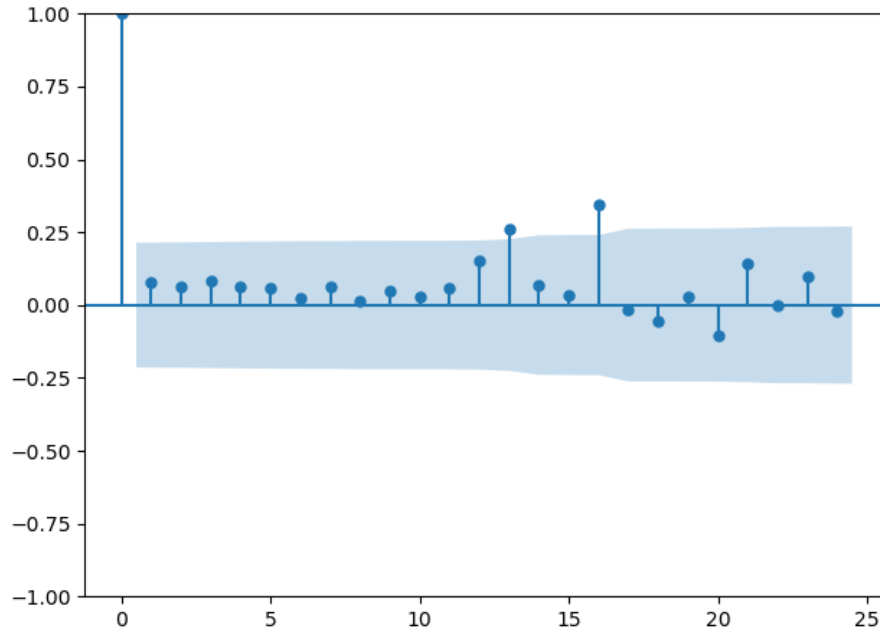


Şekil 16: Kışla Çeşmesi *E. coli* Otokorelasyon Grafiği

Tablo 21’de belirtildiği üzere, Kışla çeşmesindeki *E. coli* konsantrasyonlarına ilişkin trend analizi istatistiksel olarak anlamlı bir sonuç vermemiştir (eğim = 0.0002, $R^2 = 0.0067$, $p = 0.4608$). Şekil 15’deki grafik, 2018-2025 yılları arasında *E. coli* konsantrasyonlarında belirgin bir artış veya azalma olmadığını göstermektedir. Orijinal veri grafiği (Şekil 15, üst panel), 2019, 2021 ve 2024 yıllarında ani artışlar (pikler) sergilemesine rağmen, trend çizgisi (Şekil 15, ikinci panel) neredeyse düz bir seyir izlemektedir. Mevsimsellik grafiği (Şekil 15, üçüncü panel), mevsimsel bir örüntü olmadığını doğrulamakta olup, bu bulgu Tablo 21’deki mevsimsellik gücü (0.6326) ve mevsimsellik yok sonucuyla uyumludur. Otokorelasyon grafiği (Şekil 16), verilerin otokorelasyonunun düşük olduğunu ve belirgin bir periyodik yapı bulunmadığını göstermektedir.

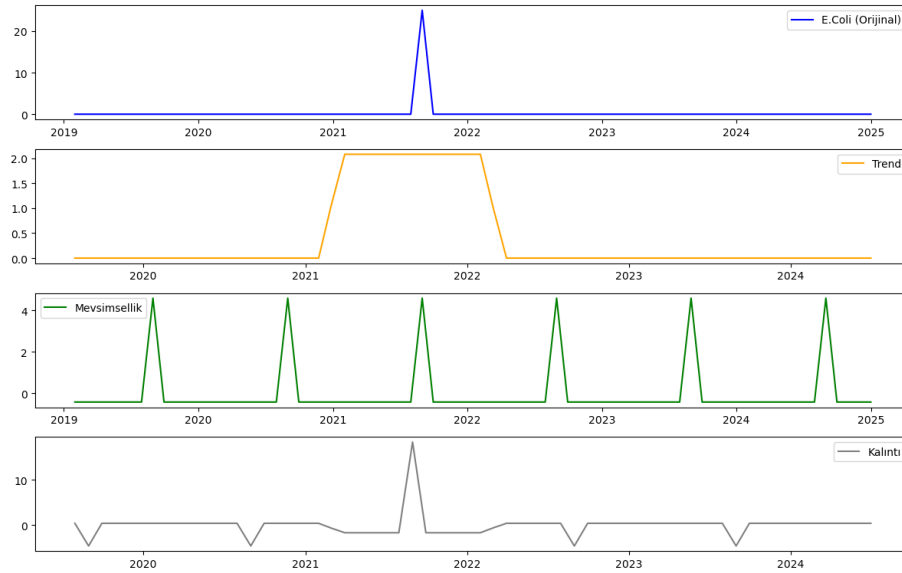


Şekil 17: Kışla Çeşmesi Koliform Trend Analizi

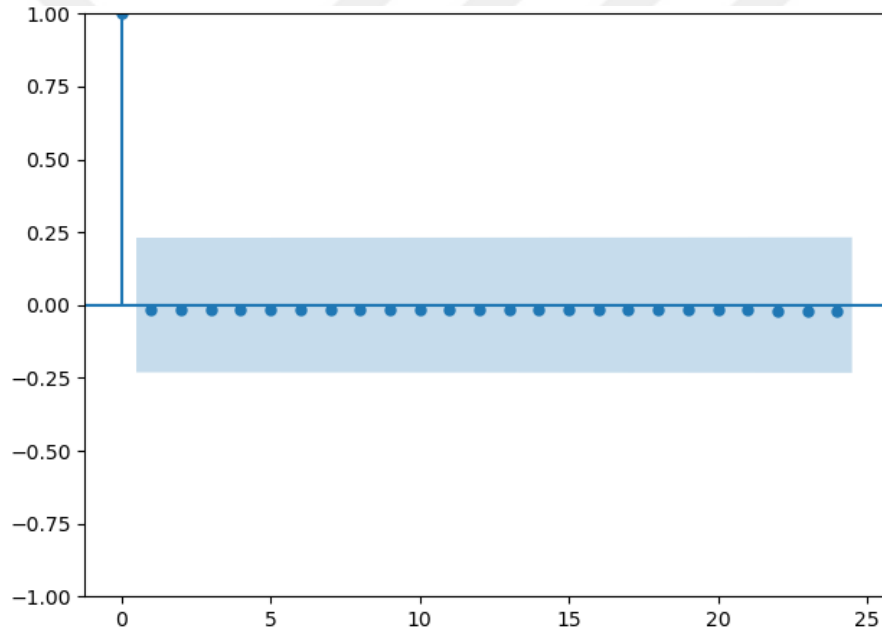


Şekil 18: Kışla Çeşmesi Koliform Otokorelasyon Grafiği

Kışla çeşmesi koliform trend analizi grafiği Şekil 17’de otokorelasyon grafiği ise Şekil 18’de gösterilmektedir. Koliform bakterilere ilişkin trend analizi, Kışla çeşmesinde istatistiksel olarak anlamlı bir artış trendi ortaya koymaktadır (eğim = 0.0186, $R^2 = 0.1539$, $p = 0.0002$). Şekil 17’deki grafik, 2018-2025 yılları arasında koliform konsantrasyonlarının genel bir artış eğilimi gösterdiğini doğrulamaktadır. Orijinal veri grafiği (Şekil 17, üst panel), 2019, 2021 ve 2023 yıllarında belirgin pikler sergilerken, trend çizgisi (Şekil 17, ikinci panel) pozitif bir eğimle zamanla artış göstermektedir. Mevsimsellik grafiği (Şekil 17, üçüncü panel), Tablo 21’deki mevsimsellik yok sonucuna paralel olarak herhangi bir mevsimsel örüntü olmadığını göstermektedir (mevsimsellik gücü = 0.4091). Otokorelasyon grafiği (Şekil 18), koliform verilerinin otokorelasyonunun düşük olduğunu ve periyodik bir yapının bulunmadığını teyit etmektedir.



Şekil 19: Aşağı Narkazan Çeşmesi *E. coli* Trend Analizi

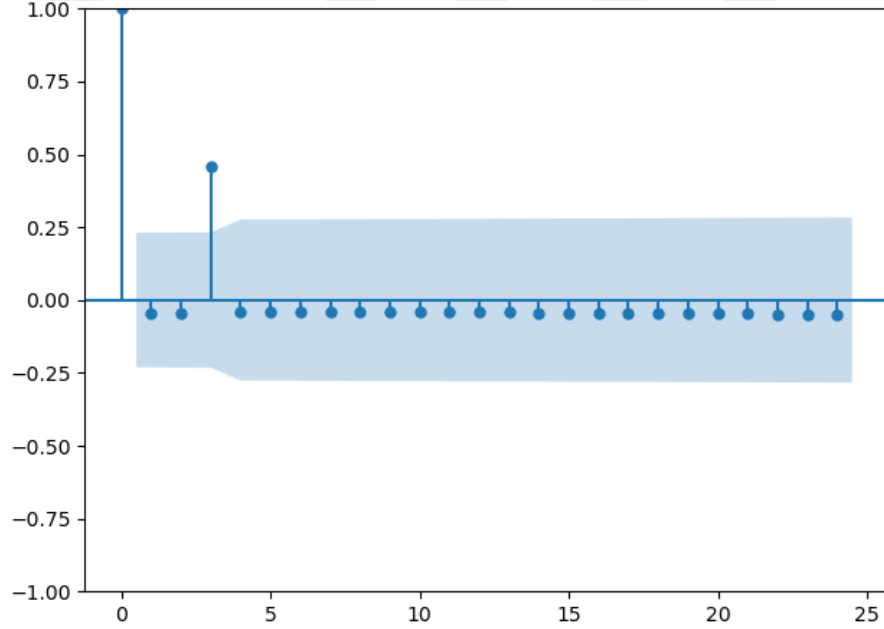


Şekil 20: Aşağı Narkazan Çeşmesi *E. coli* Otokorelasyon Grafiği

Aşağı Narkazan çeşmesine ait *E. coli* trend analizi Şekil 19’da gösterilmektedir. *E. coli* için yapılan analizde, belirgin bir trend bileşeni görülmemektedir. Aşağı Narkazan ’da *E. coli* kontaminasyonu, seyrek piklerle (2020 ve 2022) karakterizedir ve anlamlı bir trend veya mevsimsellik göstermemektedir. Tablo 21’deki sonuçlarla (eğim=-0.0001, $p= 0.8304$, mevsimsellik gücü= 0.5012) uyumlu olarak, kontaminasyon olay bazlı olup çevresel girişlerle ilişkilendirilebilir.



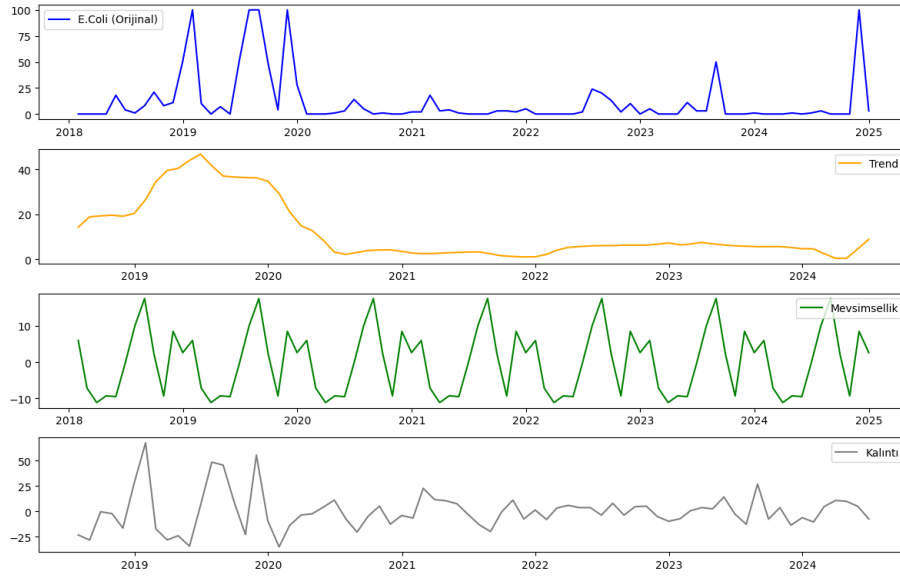
Şekil 21: Aşağı Narkazan Çeşmesi Koliform Trend Analizi



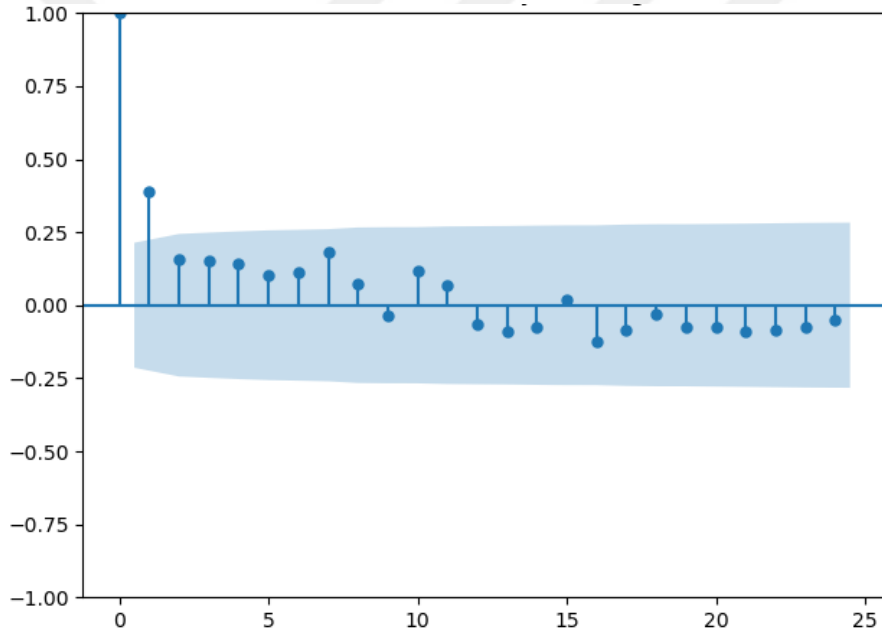
Şekil 22: Aşağı Narkazan Çeşmesi Koliform Otokorelasyon Grafiği

Aşağı Narkazan 'da Koliform kontaminasyonu, *E. coli* den daha sık piklerle (2021, 2024) karakterizedir, ancak anlamlı bir trend veya mevsimsellik göstermemektedir. Tablo 21'deki sonuçlarla (eğim= 0.0002, $p = 0.7768$, mevsimsellik gücü= 0.4912) uyumlu olarak, kontaminasyon düzensiz olup çevresel faktörlerle ilişkilendirilebilir.

Tablo 21'deki Aşağı Narkazan sonuçları, hem *E. coli* hem de koliform için anlamlı bir trend veya mevsimsellik olmadığını göstermektedir.

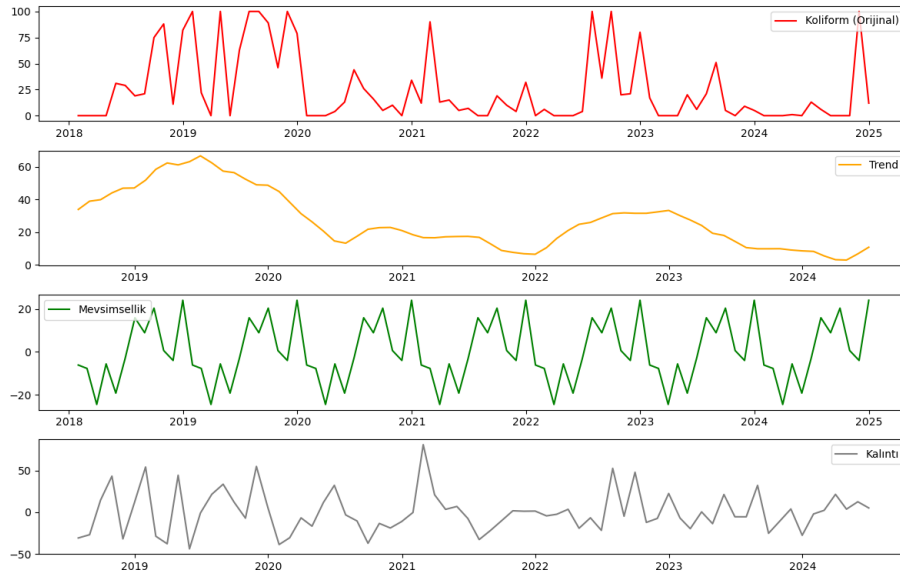


Şekil 23: Gordon Çeşmesi *E. coli* Trend Analizi

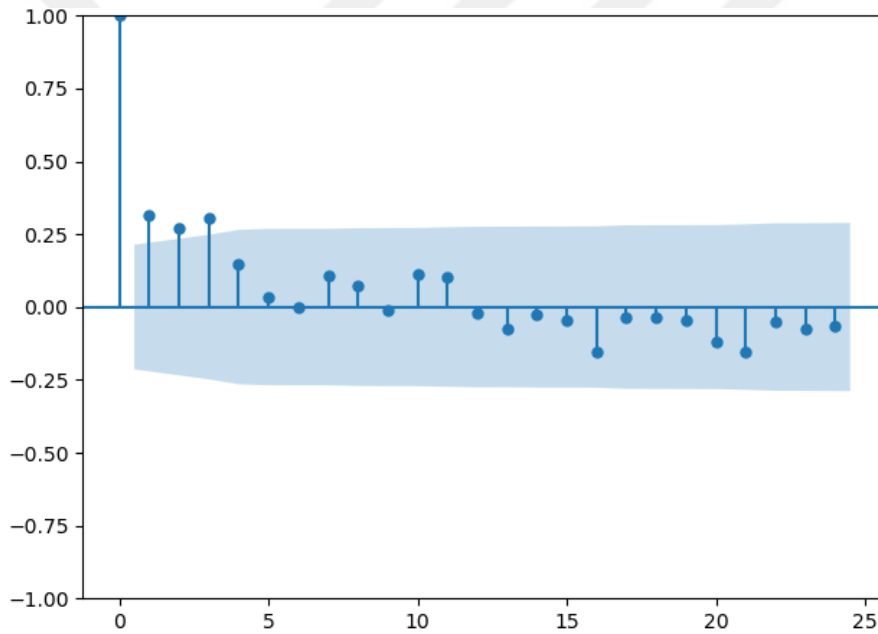


Şekil 24: Gordon Çeşmesi *E. coli* Otokorelasyon Grafiği

Tablo 21’de belirtildiği üzere, Gordon çeşmesindeki *E. coli* konsantrasyonlarına ilişkin trend analizi istatistiksel olarak anlamlı bir sonuç vermemiştir (eğim=-0.0067, $R^2 = 0.0395$, $p = 0.0698$). Şekil 23’teki grafik, 2018-2025 yılları arasında *E. coli* konsantrasyonlarında belirgin bir artış veya azalma olmadığını göstermektedir. Orijinal veri grafiği (Şekil 23, üst panel), 2019 ve 2024 yıllarında ani artışlar sergilemesine rağmen, trend grafiği (Şekil 23, ikinci panel) hafif bir düşüş eğilimi göstermektedir; ancak bu eğilim istatistiksel olarak anlamlı değildir ($p > 0.05$). Mevsimsellik grafiği (Şekil 23, üçüncü panel), Tablo 21’deki mevsimsellik yok sonucuyla uyumlu olarak herhangi bir mevsimsel örüntü olmadığını göstermektedir (mevsimsellik gücü = 0.4885). Otokorelasyon grafiği (Şekil 24), verilerin otokorelasyonunun düşük olduğunu ve belirgin bir periyodik yapı bulunmadığını teyit etmektedir.



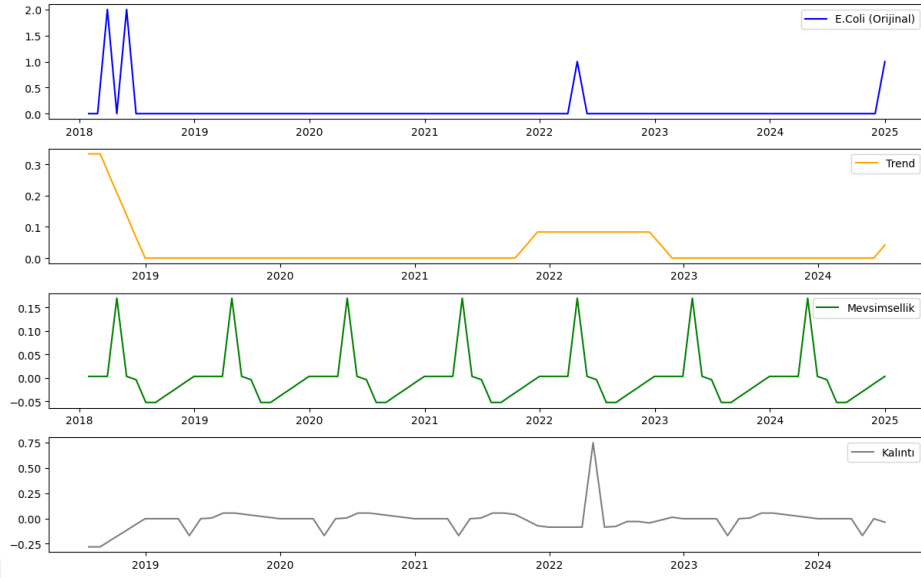
Şekil 25: Gordon Çeşmesi Koliform Trend Analizi



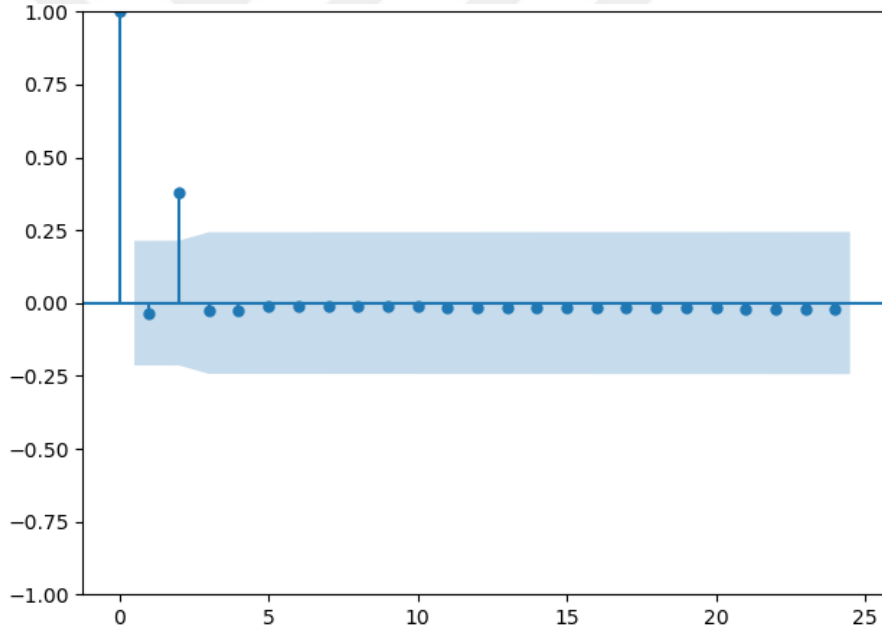
Şekil 26: Gordon Çeşmesi Koliform Otokorelasyon Grafiği

Koliform bakterilere ilişkin trend analizi, Gordon çeşmesinde istatistiksel olarak anlamlı bir azalma trendi ortaya koymaktadır (eğim=-0.0123, $R^2=0.0723$, $p=0.0134$). Şekil 25’deki grafik, 2018-2025 yılları arasında koliform konsantrasyonlarının genel bir düşüş eğilimi sergilediğini göstermektedir. Orijinal veri grafiği (Şekil 25, üst panel), 2019 ve 2021 yıllarında belirgin pikler sergilerken, trend grafiği (Şekil 25, ikinci panel) zamanla azalan bir eğilim göstermektedir. Mevsimsellik grafiği (Şekil 25, üçüncü panel), Tablo 21’deki mevsimsellik yok sonucuna paralel olarak herhangi bir mevsimsel örüntü olmadığını doğrulamaktadır (mevsimsellik gücü = 0.5744). Mevsimsellik olmaması, bakteriyel konsantrasyonların mevsimsel faktörlerden ziyade sürekli bir kirlilik kaynağı veya iyileştirme çabalarına bağlı olabileceğini düşündürmektedir.

Otokorelasyon grafiği (Şekil 26), koliform verilerinin otokorelasyonunun düşük olduğunu ve periyodik bir yapının bulunmadığını göstermektedir.



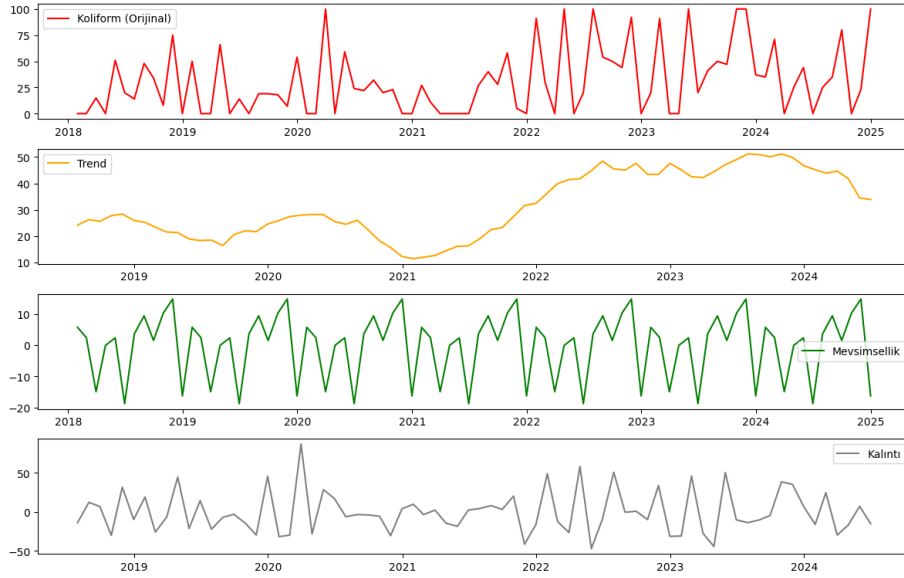
Şekil 27: Tarım Çeşmesi *E. coli* Trend Analizi



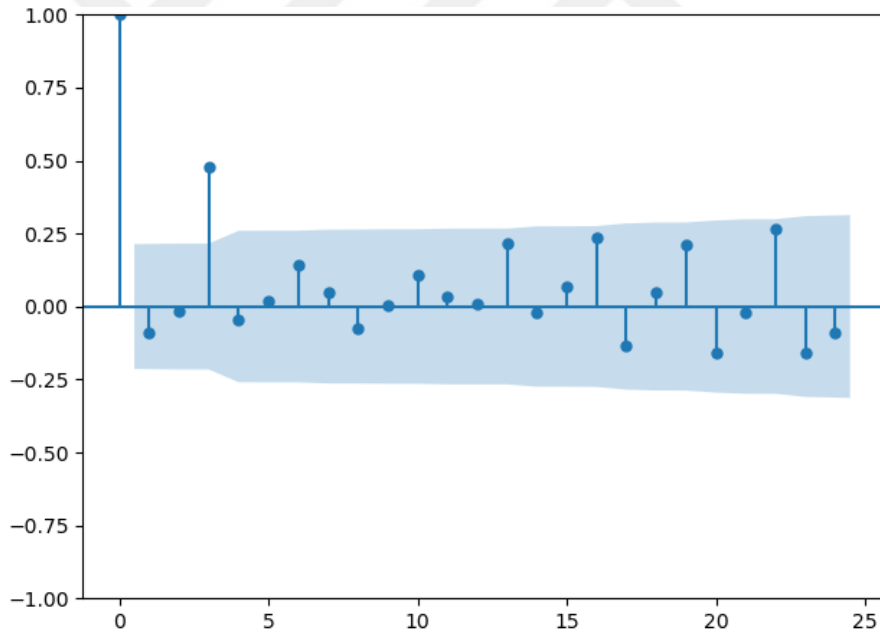
Şekil 28: Tarım Çeşmesi *E. coli* Otokorelasyon Grafiği

Tablo 21’de belirtildiği üzere, Tarım çeşmesindeki *E. coli* konsantrasyonlarına ilişkin trend analizi istatistiksel olarak anlamlı bir sonuç vermemiştir (eğim=-0.0001, $R^2=0.0225$, $p=0.1733$). Şekil 27’deki grafik, 2018-2025 yılları arasında *E. coli* konsantrasyonlarında belirgin bir artış veya azalma olmadığını göstermektedir. Orijinal veri grafiği (Şekil 27, üst panel), 2018, 2021 ve 2024 yıllarında ani artışlar sergilemesine rağmen, trend grafiği (Şekil 27, ikinci panel) neredeyse düz bir seyir izlemektedir. Bu da istatistiksel olarak anlamlı bir trend olmadığını doğrulamaktadır ($p > 0.05$). Mevsimsellik grafiği (Şekil 27, üçüncü panel), Tablo 21’deki mevsimsellik yok sonucuyla uyumlu olarak herhangi bir mevsimsel örüntü olmadığını

göstermektedir (mevsimsellik gücü = 0.4699). Şekil 28'deki otokorelasyon grafiği, verilerin otokorelasyonunun düşük olduğunu ve belirgin bir periyodik yapı bulunmadığını teyit etmektedir.



Şekil 29: Tarım Çeşmesi Koliform Trend Analizi

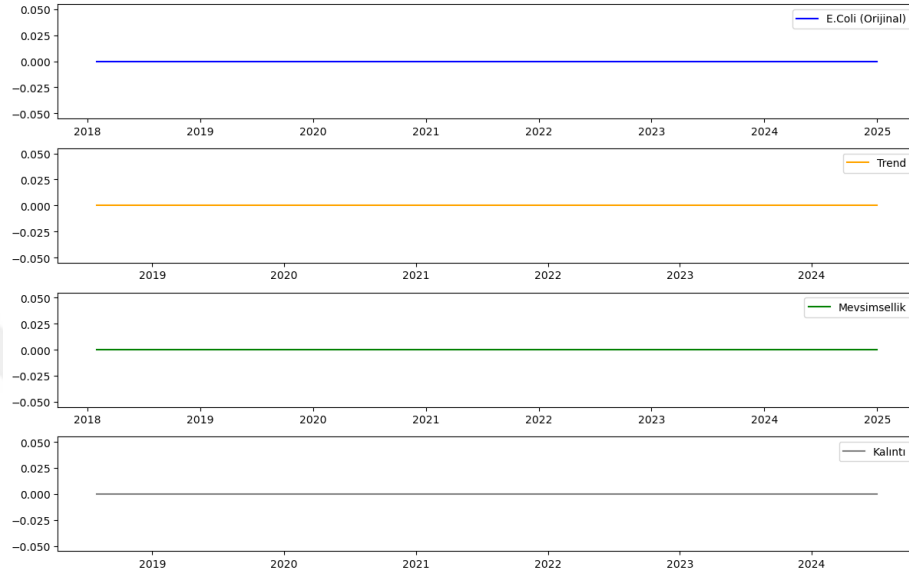


Şekil 30: Tarım Çeşmesi Koliform Otokorelasyon Grafiği

Koliform bakterilere ilişkin trend analizi, Tarım çeşmesinde istatistiksel olarak anlamlı bir artış trendi ortaya koymaktadır (eğim = 0.0130, $R^2 = 0.0897$, $p = 0.0056$). Şekil 29'deki grafik, 2018-2025 yılları arasında koliform konsantrasyonlarının genel bir artış eğilimi sergilediğini göstermektedir. Orijinal veri grafiği (Şekil 29, üst panel), 2019 ve 2021 yıllarında belirgin pikler sergilerken, trend grafiği (Şekil 29, ikinci panel) zamanla artan bir eğilim göstermektedir. Mevsimsellik grafiği (Şekil 29, üçüncü panel), Tablo 21'deki mevsimsellik yok sonucuna paralel olarak herhangi bir mevsimsel örüntü olmadığını doğrulamaktadır.

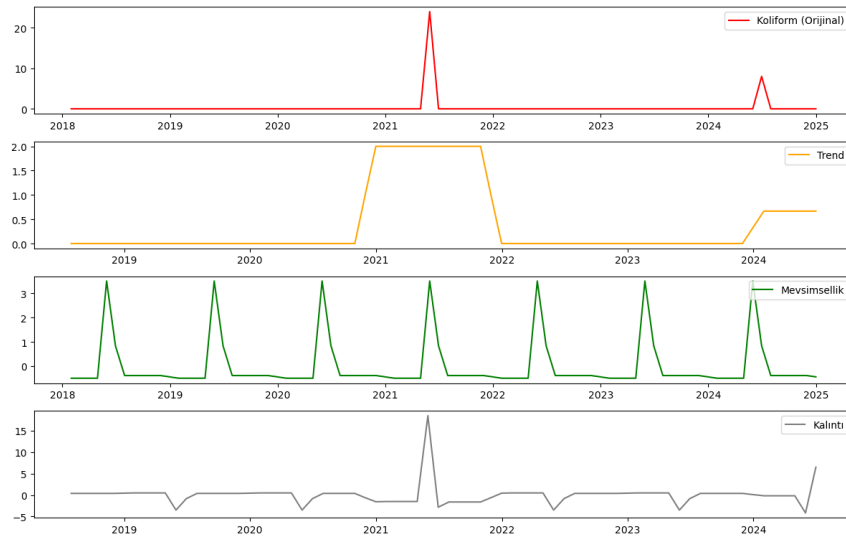
(mevsimsellik gücü = 0.3863). Mevsimsellik olmaması, bakteriyel konsantrasyonların mevsimsel faktörlerden ziyade sürekli bir kirlilik kaynağına bağlı olabileceğini düşündürmektedir. Bu bulgular, su kalitesinin düzenli olarak izlenmesi ve potansiyel kirlilik kaynaklarının araştırılması gerektiğini vurgulamaktadır.

Otokorelasyon grafiği (Şekil 30), koliform verilerinin otokorelasyonunun düşük olduğunu ve periyodik bir yapının bulunmadığını göstermektedir.

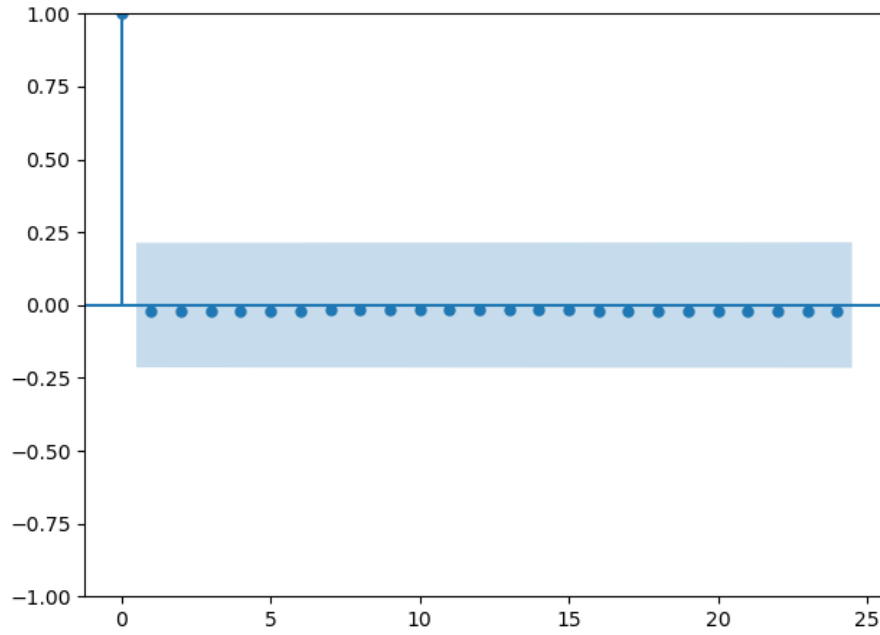


Şekil 31: Yukarı Narkazan Çeşmesi *E. coli* Trend Analizi

Tablo 21’de belirtildiği üzere, Yukarı Narkazan çeşmesindeki *E. coli* konsantrasyonlarına ilişkin trend analizi yapılamamıştır (eğim = 0.0000, $R^2 = 0.0000$, $p = 1000$). Bu durum, veriler sıfır olduğu için varyansının da sıfır olduğu, dolayısıyla istatistiksel bir analiz için uygun olmadığını göstermektedir.



Şekil 32: Yukarı Narkazan Çeşmesi Koliform Trend Analizi



Şekil 33: Yukarı Narkazan Çeşmesi Koliform Otokorelasyon Grafiği

Koliform bakterilere ilişkin trend analizi, Yukarı Narkazan çeşmesinde istatistiksel olarak anlamlı bir trend göstermemiştir (eğim = 0.0002, $R^2 = 0.0020$, $p = 0.6881$). Şekil 32’teki grafik, 2018-2025 yılları arasında koliform konsantrasyonlarında belirgin bir artış veya azalma olmadığını göstermektedir. Orijinal veri grafiği (Şekil 32, üst panel), 2021 ve 2023 yıllarında ani artışlar sergilemesine rağmen, trend grafiği (Şekil 32, ikinci panel) neredeyse düz bir seyir izlemektedir, bu da istatistiksel olarak anlamlı bir trend olmadığını doğrulamaktadır ($p > 0.05$). Mevsimsellik grafiği (Şekil 32, üçüncü panel), Tablo 21’deki mevsimsellik yok sonucuyla uyumlu olarak herhangi bir mevsimsel örüntü olmadığını göstermektedir (mevsimsellik gücü = 0.4260). Otokorelasyon grafiği (Şekil 33), koliform verilerinin otokorelasyonunun düşük olduğunu ve periyodik bir yapının bulunmadığını teyit etmektedir.

Bu bölümde, Bayburt ili umumî çeşmelerine ait aylık ölçülen veriler ile *E. coli* ve koliform bakteri üzerinde zamana bağlı yapılan trend ve mevsimsellik analizi gerçekleştirilmiştir. Her bir çeşme için basit regresyon uygulanarak eğim, kesme noktası, R^2 , p-değeri ve istatistiksel anlamlılık düzeyleri hesaplanmıştır. Regresyon analizi sonucunda eğimin pozitif yönde ya da negatif yönde olması bakterinin zamanla arttığı veya azaldığını ifade ederken, artan ya da azalan trendin anlamlı olup olmadığı değerlendirilmesi p-değerine bakılarak yapılmıştır. Tüm çeşmelerin p-değeri 0.05’in üzerinde bulunmuş, bu da trend eğimlerinin anlamlı olmadığını göstermiştir.

Bayburt ili umumî çeşmelerinin zaman serisine bağlı olarak yapılan analizlerinde anlamlı bir eğilim veya güçlü bir mevsimsellik tespit edilememiştir. Yapılan analizler sonucunda; koliform bakterilerin mevsimsel değişimlerden etkilenecek çevredeki varlığını artırabileceği, buna karşılık *E. coli*’nin ise ölçülen zaman aralığında belirgin bir mevsimsel yapı

sergilemediđi sylenabilir. Bu tr farklılıklar, bakteriyel kontaminasyon kaynaklarının farklı olabileceđine iřaret etmektedir.



SONUÇ VE ÖNERİ

Bayburt ili şebeke suları aylık yapılan su analizlerinde mikrobiyolojik ve kimyasal parametreler için su analizi yapılmaktadır. Kontrol analizi veya denetleme analizi çerçevesinde yıllık olarak veya gerekli görülmesi halinde kimyasal, mikrobiyolojik ve radyoaktif analizler yapılmaktadır.

Bayburt ili umumî çeşmeleri halk tarafından evde içme suyu olarak da kullanılmakta olduğu için suyun kalitesi büyük önem arz etmektedir. Bayburt İli Sağlık Müdürlüğü İnsani Tüketim Amaçlı Kullanılan Sular Yönetmeliği'nde belirtilen parametrelerden *E. coli* ve koliform bakteri koloni sayılarına aylık olarak bakmakta, suyun kirli veya temiz olduğu sonucuna bu iki bakterinin varlığına göre karar vermekte ve sonuçları yayınlayarak halk ile paylaşmaktadır.

Çalışma kapsamında Erzurum Halk Sağlığı Laboratuvarı ve Bayburt Halk Sağlığı Laboratuvarında yapılan 2020 yılı ve sonrasına ait Bayburt ili şebeke suları analiz verileri ile Gordon Çeşmesi, Kışla Çeşmesi, Tarım Çeşmesi, Aşağı Narkazan Çeşmesi ve Yukarı Narkazan Çeşmesinin 2014 yılından sonraki verilerine ulaşılmış olup iki ayrı veri kümesi ile çalışılmıştır.

Çalışmada öncelikle Bayburt ili umumî çeşme sularına ait veri analizi sonuçlarından faydalanılarak verilerin korelasyonuna bakılmış, *E. coli* ile koliform bakteri arasında 0.408193 oranında bir ilişki olduğu, Bayburt ili şebeke sularına ait veri analizi sonuçlarında yer alan kimyasal parametrelerin korelasyon hesabı yapıldığında ise birçok maddenin birbirleriyle ilişkili olduğu anlaşılmıştır. Parametrelere ait korelasyon sonuçları Ek-1'de yer alan tabloda gösterilmektedir.

Bayburt ili şebeke sularının analiz verileri kullanılarak uygulanan makine öğrenmesi yöntemlerine ait sonuçlar Ek-2/a'da, Bayburt ili umumî çeşme suları analiz verileri kullanılarak uygulanan makine öğrenmesi yöntemlerine ait sonuçlar ise Ek-2/b'de yer almaktadır.

Veri tahminleri için lojistik regresyon yöntemi, knn, SVM, naive bayes, karar ağacı ve rastgele orman yöntemi uygulanmıştır.

Bayburt ili şebeke suları analiz verileri ile yapılan makine öğrenme yöntemleri sonuçlarına (Ek-2/a) bakıldığında en iyi doğruluk (accuracy) oranının 0,99 ile karar ağacı olduğu görülmektedir.

Bayburt ili umumî çeşme suları analiz verileri ile yapılan makine öğrenme yöntemi sonucuna (Ek-2/b) bakıldığında en iyi doğruluk (accuracy) oranlarının 1 ile karar ağacı, naive bayes ve rastgele orman yöntemlerinin olduğu anlaşılmaktadır.

Suyun kalitesinin insan sađlıđı üzerinde b y k etkisi vardır. Kirli olan bir suyun temiz olarak tahmin edilmesi toplumda b y k ve toplu sađlık sıkıntılarına neden olacaktır.  zellikle enterokok, *E. coli*, koliform gibi bakterilerin i me ve kullanıma sunulan sularda sık g r lmesi su kalitesi tahmin  alıřmalarında en uygun modelin se imi a ısından  nemlidir. Tahminin dođruluk oranının y ksek olmasının yanında i ilebilir olmayan suyu, i ilebilir olarak en az tahmin eden modelin se ilmesi gerekmektedir. Burada  nemli olan, dođruluk oranının y ksek olmasından ziyade uygun olmayan suların hatalı tahminleridir. Uygun olmayan sularda en az hata tahmini yapan modele specificity deđerinden ya da false positive rate deđerinden bakılabilir. Bu iki deđer birbirinin tersidir. Model se iminde specificity deđerı en y ksek olan ya da false positive rate oranı en d ř k olan model se ilmelidir. Bayburt ili řebeke suları i in makine  đrenme y ntemlerine bakıldıđında en y ksek specificity deđerinin 0,99 ile karar ađacı, knn ve SVM y ntemleri olduđu g r lmektedir. Bayburt ili umum   eřme suları i in makine  đrenme y ntemlerine bakıldıđında ise karar ađacı, naive bayes ve rastgele orman y ntemlerinin 1 sonucu ile en y ksek specificity deđerlerine sahip olduđu g r lmektedir.

Sonuc olarak;

Bayburt ili řebeke suları i in makine  đrenme y ntemlerinde specificity deđerleri 0,99 ile aynı olan karar ađacı, knn ve SVM y ntemlerinden; karar ađacı y nteminin dođruluk oranının 0,99 olduđu, knn y nteminin dođruluk oranının 0,92 olduđu, SVM y nteminin dođruluk oranının ise 0,78 olduđu, bu bakımdan en dođru modelin karar ađacı y ntemi olduđu anlaşılmaktadır.

Bayburt ili umum   eřme suları i in makine  đrenme y ntemlerinde naive bayes, rastgele orman ve karar ađacı y ntemlerine ait sonu ların aynı olması dolayısıyla her    y ntemin de kullanılabilir olduđu verilerden anlaşılmıřtır.

Yapılan trend ve mevsimsellik analiz  alıřmalarında;

Tablo 21’de yer alan veriler, beř farklı  eřmede *E. coli* ve koliform bakterilerine iliřkin trend ve mevsimsellik analizlerini g stermektedir. *E. coli* konsantrasyonları a ısından, hi bir  eřmede istatistiksel olarak anlamlı bir trend g zlenmemiřtir ($p > 0.05$), eđim deđerleri olduk a d ř k ve R^2 deđerleri modelin a ıklayıcılıđının zayıf olduđunu g stermektedir. Koliform bakteriler i in ise Gordon (eđim = -0.0123, $p = 0.0134$), Kıřla (eđim = 0.0186, $p = 0.0002$) ve Tarım (eđim = 0.0130, $p = 0.0056$)  eřmelerinde istatistiksel olarak anlamlı trendler tespit edilmiř, Gordon’da azalma, Kıřla ve Tarım’da ise artıř g zlenmiřtir; Ařađı ve Yukarı Narkazanı’da ise anlamlı bir trend bulunmamıřtır ($p > 0.05$). T m  eřmelerde mevsimsellik analizi, bakteriyel konsantrasyonlarda mevsimsel bir  r nt  olmadıđını ortaya koymuř (mevsimsellik g c : 0.3863-0.6326; Mevsimsellik: Yok), bu da kontaminasyonun mevsimsel

faktörlerden ziyade çevresel veya antropojenik kaynaklı olabileceğini düşündürmektedir (Edberg vd., 2000).

Makine öğrenmesi yöntemleri ile yapılan tahminlerin doğru çalışması için daha fazla veri ile çalışılması gerekmektedir. Veriler ne kadar büyük olursa, eksik veri ne kadar az olursa doğruluk oranı da o kadar yüksek ve güvenilir olacaktır. Makine öğrenme yöntemleri kullanılarak su kirliliği tahmini yapmak için daha fazla parametreye ait verilerin daha sık yapılan analiz sonuçlarına ihtiyaç duyulmaktadır.

Çalışmada su analizi verileri incelenirken izleme noktalarına ait kimlik numara değerlerinin olmaması çalışmanın yapılmasını zorlaştırmıştır. Su kalitesi halk sağlığı açısından önemlidir ve bu bakımdan tekrarlayan çalışmaların olması ihtimalinin de yüksek olduğu tahmin edilmektedir. İlerde yapılabilecek çalışmalar için her izleme noktasının, her umumî çeşmenin kimlik numaralandırmasının yapılması su analizi sonuçlarının arşivlenmesi gerektiği düşünülmektedir. Ayrıca umumî çeşmelere ait kimyasal analizlerin yapılmadığı anlaşılmış olup, çeşme sularının evlerde içme ve kullanma suyu olarak kullanıldığı düşünüldüğünde kimyasal analizlerinin de yapılması gerektiği düşünülmektedir.

KAYNAKÇA

- Abuzir, S. Y., & Abuzir, Y. S. (2022). Machine Learning for Water Quality Classification. *Water Quality Research Journal*, 57(3). <https://doi.org/10.2166/wqrj.2022.004>
- Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., & García-Nieto, J. (2019). Efficient Water Quality Prediction Using Supervised Machine Learning. *Water*, 11(11), 2210. <https://doi.org/10.3390/w11112210>
- Arif, M., Ilyas, A., Amin, M., Aslam, S., Rasheed, R., & Rasool, A. (2021). Water Pollution: A Silent Threat to Environmental Sustainability. *Environmental Science and Pollution Research*, 28(3), 26324–26330. <https://doi.org/10.1007/s11356-021-12436-4>
- Azrou, M., Mabrouki, J., Fattah, G., Guezzaz, A., & Aziz, F. (2022). Machine Learning Algorithms for Efficient Water Quality Prediction. *Modeling Earth Systems and Environment*, 8(2), 2793–2801. <https://doi.org/10.1007/s40808-021-01266-6>
- Berners-Lee, T. (1999). Weaving the web: The original design and ultimate destiny of the World Wide Web. HarperCollins.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Zaman serisi analizi: Tahmin ve kontrol* (5. baskı). Wiley.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). Springer. <https://doi.org/10.1007/978-3-319-29854-2>
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
- Bulut, O. (2020). *İçme suyu şebekelerinde oluşan su kayıp ve kaçakları belirlemek için kullanılan scada sisteminin değerlendirilmesi: Erzincan ili örneği*. (Yüksek lisans tezi). Yükseköğretim Kurulu Ulusal Tez Merkezi'nden edinilmiştir. (Tez No. 641867).
- Castells, M. (2009). *The rise of the network society* (2nd ed.). Wiley-Blackwell.
- Ceruzzi, P. E. (2003). *A history of modern computing* (2nd ed.). MIT Press.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, 6(1), 3–73.
- Çalık, E., Menteş, Y., Karadağ, F. & Dayıoğlu, H. (2004). İçme Suyunun Sağlık Açısından Değerlendirilmesi. *Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, (6), 17–26.
- Deng, Y., Zhang, T., Wang, Y., & Liang, L. (2021). Long-Term Trends and Variations in Water Quality of the Pearl River Delta, China. *Environmental Science and Pollution Research*, 28(15), 18950–18962. <https://doi.org/10.1007/s11356-020-11929-6>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.
- Fuchs, C. (2017). *Social media: A critical introduction* (2nd ed.). Sage.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Ho, J. C., Michalak, A. M., & Pahlevan, N. (2019). Widespread Global Increase in Intense Lake

- Phytoplankton Blooms since the 1980s. *Nature*, 574(7780), 667–670. <https://doi.org/10.1038/s41586-019-1648-7>
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>
- Kaddoura, S. (2022). Evaluation of Machine Learning Algorithm on Drinking Water Quality for Better Sustainability. *Sustainability*, 14(18), 11478. <https://doi.org/10.3390/su141811478>
- Kim, S., Kim, Y., & Kang, N. (2022). Water Quality Prediction in the Han River Using Deep Learning Techniques. *Journal of Hydrology*, 610, 127893. <https://doi.org/10.1016/j.jhydrol.2022.127893>
- Kundzewicz, Z. W., & Robson, A. J. (2004). Change Detection in Hydrological Records—A Review of the Methodology. *Hydrological Sciences Journal*, 49(1), 7–19. <https://doi.org/10.1623/hysj.49.1.7.53993>
- Kuşçu, A., & Kılıç, İ. (2023). Su Kalite İndeksinin Makine Öğrenme Yöntemleri ile Tahmini. *Bilimsel ve Akademik Araştırmalar Dergisi*, 7(13), 1–18. <https://bibad.gen.tr/index.php/bibad/article/view/444>
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill/Irwin.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lutz, S. R., Trauth, N., & Musolff, A. (2020). Long-Term Trends and Seasonality of Nitrate Concentrations in A Regional Agricultural Catchment. *Science of the Total Environment*, 714, 136729. <https://doi.org/10.1016/j.scitotenv.2020.136729>
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine*, 27(4), 12–14.
- Memorial Sağlık Grubu. (2025, Ocak 21). Su kirliliği. *Memorial*. <https://www.memorial.com.tr/saglik-rehberi/su-kirliligi> adresinden edinilmiştir.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting* (2nd ed.). Wiley. <https://doi.org/10.1002/9781118745113>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.
- Najah, A., El-Shafie, A., Karim, O. A., & El-Shafie, A. H. (2013). Application of Artificial Neural Networks for Water Quality Prediction. *Neural Computing and Applications*, 22(Ek 1), 187–201. <https://doi.org/10.1007/s00521-012-0940-3>
- Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., & Al-Shamma'a, A. (2022). Water Quality Classification Using Machine Learning Algorithms. *Journal of Water Process Engineering*, 48, 102920. <https://doi.org/10.1016/j.jwpe.2022.102920>
- Nataro, J. P., & Kaper, J. B. (1998). Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews*, 11(1), 142–201. <https://doi.org/10.1128/CMR.11.1.142>
- O'Reilly, T. (2005). *What is Web 2.0? Design patterns and business models for the next generation of software*. O'Reilly Media. Retrieved from <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>

- Özçelik, R. (2020). Makine Öğrenmesi ile Su Kalitesi Tahmini: Türkiye’de Bir Uygulama. *Çevre Bilimleri ve Teknolojisi Dergisi*, 17(4), 123–130.
- Panigrahi, N., Patro, S. G. K., Kumar, R., Omar, M., Tran, T. N., Nguyen, L. G., Bui, T. T., & Nguyen, T. T. (2023). Groundwater Quality Analysis and Drinkability Prediction Using Artificial Intelligence. *Earth Science Informatics*, 16(2), 1701–1725. <https://doi.org/10.1007/s12145-023-00977-x>
- Parmar, K. S., & Bhardwaj, R. (2015). Statistical, Time Series, and Fractal Analysis of Water Quality of River Yamuna At Kalpi for Forecasting and Management. *Environmental Monitoring and Assessment*, 187(3), 124. <https://doi.org/10.1007/s10661-015-4338-5>
- Poudel, D., Shrestha, D., Bhattarai, S., & Ghimire, A. (2023). Comparison of Machine Learning Algorithms in Statistically Imputed Water Potability Dataset. *Journal of Innovations in Engineering Education*, 5(1), 38–46. <https://doi.org/10.3126/jiee.v5i1.42265>
- Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- Sammut, C., & Webb, G. I. (Eds.). (2010). *Encyclopedia of machine learning*. Springer.
- Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, 27(4s), 4023–4031. <https://africanjournalofbiomedicalresearch.com/index.php/AJBR>
- Sen, P. K. (1968). Estimates Of The Regression Coefficient Based on Kendall’s Tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resources Research*, 54(11), 8558–8593. <https://doi.org/10.1029/2018WR022643>
- Sprague, L. A., Hirsch, R. M., & Aulenbach, B. T. (2017). Trends in Nutrient and Sediment Concentrations and Loads in Major River Basins of the United States. *Journal of the American Water Resources Association*, 53(3), 665–684. <https://doi.org/10.1111/1752-1688.12520>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
- T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. (2021). *Ulusal yapay zekâ stratejisi 2021–2025*. <https://cbddo.gov.tr/ulusal-yapay-zeka-stratejisi/>
- Turan, T. (2023) Makine Öğrenmesi Algoritmaları ile Su Kalitesi ve İçilebilirlik Tahmini. *Uluborlu Mesleki Bilimler Dergisi*, 6(2), 65–80.
- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- United Nations. (2024). *The United Nations World Water Development Report 2024: Water for prosperity and peace*. UNESCO. <https://www.unwater.org>
- Vyshnevskiy, M., Kopytko, R., Kharlamova, N., & Grygoruk, M. (2021). Trends in Water Quality in the Dnieper River Basin (Ukraine) in Relation to Climate Change (1990–2019). *Environmental Monitoring and Assessment*, 193, Article 731. <https://doi.org/10.1007/s10661-021-09538-3>

- Wang, J., Zhang, X., & Yang, L. (2020). Water Quality Forecasting in the Yangtze River Using Gradient Boosting Decision Trees. *Environmental Science and Pollution Research*, 27(15), 18025–18037. <https://doi.org/10.1007/s11356-020-08234-7>
- Wilde, S. A., Valley, J. W., Peck, W. H., & Graham, C. M. (2001). Evidence from Detrital Zircons for the Existence of Continental Crust and Oceans on the Earth 4.4 Gyr Ago. *Nature*, 409(6817), 175–178. <https://doi.org/10.1038/35051550>
- World Health Organization. (2022). *Drinking-water*. <https://www.who.int/news-room/fact-sheets/detail/drinking-water>
- World Health Organization (2023, Eylül 13). *Drinking-water*. <https://www.who.int/news-room/fact-sheets/detail/drinking-water>
- Yılmaz, A. (2015). Küresel Isınmanın Dünya Su Rezervleri Üzerindeki Etkileri. *Kent Akademisi*, 8(22), 63-72.
- Yurtsever, M., & Emeç, M. (2023). Potable Water Quality Prediction Using Artificial Intelligence and Machine Learning Algorithms for Better Sustainability. *Ege Akademik Bakış*, 23(2), 265–278. <https://doi.org/10.21121/eab.1252167>
- Zhu, M., Wang, J., Yang, X., & Zhang, Y. (2019). Forecasting Water Quality Parameters in Lake Taihu Using Machine Learning Models. *Water Resources Research*, 55(11), 9502–9519. <https://doi.org/10.1029/2019WR025189>

EKLER

Ek 1-a: Bayburt İli Şebeke Sularının Kimyasal Parametrelerine Ait Korelasyon Tablosu

Korelasyon	Antimon	Arsenik	Benzen	Bor	Bromat	Kadmiyum	Krom	Bakır	Kurşun
Antimon	1,000	0,195	0,841	0,280	0,509	0,838	0,606	0,104	0,691
Arsenik	0,195	1,000	0,222	0,090	0,121	0,222	0,292	0,015	0,200
Benzen	0,841	0,222	1,000	0,344	0,602	0,997	0,747	0,128	0,823
Bor	0,280	0,090	0,344	1,000	0,155	0,343	0,243	0,093	0,288
Bromat	0,509	0,121	0,602	0,155	1,000	0,600	0,445	0,080	0,499
Kadmiyum	0,838	0,222	0,997	0,343	0,600	1,000	0,745	0,156	0,821
Krom	0,606	0,292	0,747	0,243	0,445	0,745	1,000	0,087	0,648
Bakır	0,104	0,015	0,128	0,093	0,080	0,156	0,087	1,000	0,169
Kurşun	0,691	0,200	0,823	0,288	0,499	0,821	0,648	0,169	1,000
Civa	0,742	0,193	0,919	0,341	0,558	0,917	0,706	0,114	0,754
Nikel	0,074	0,031	0,064	0,025	0,041	0,064	0,040	0,329	0,053
Nitrat	-0,039	0,063	-0,038	0,023	-0,028	-0,038	0,027	-0,013	-0,028
Selenyum	0,782	0,206	0,969	0,340	0,587	0,967	0,745	0,126	0,796
Akrilamid	0,754	0,215	0,941	0,331	0,554	0,939	0,717	0,119	0,777
Benzo (a)	0,786	0,220	0,980	0,341	0,590	0,978	0,749	0,156	0,815
Dikloretan	0,786	0,220	0,980	0,341	0,590	0,978	0,749	0,157	0,815
Epikloridin	0,760	0,214	0,948	0,333	0,558	0,946	0,723	0,119	0,779
Florür	0,150	0,360	0,109	0,060	0,094	0,122	0,133	0,138	0,089
Nitrit	0,787	0,211	0,975	0,344	0,591	0,973	0,750	0,155	0,811
Toplam Pestisitler	0,214	0,061	0,262	0,364	0,144	0,262	0,198	0,061	0,214
Tetra	0,798	0,215	0,943	0,310	0,555	0,945	0,713	0,172	0,773
Toplam Trihalometanlar	0,282	0,114	0,333	0,095	0,205	0,332	0,242	0,116	0,267
Vinil Klorür	0,789	0,203	0,927	0,307	0,545	0,925	0,704	0,114	0,765
Toplam DDT	0,362	0,065	0,453	0,182	0,322	0,452	0,372	0,080	0,417
Toplam PAH	0,738	0,185	0,915	0,335	0,557	0,913	0,700	0,132	0,750
Enterokok	0,044	-0,021	0,052	-0,031	0,038	0,052	0,041	-0,009	0,085
Koliform	0,034	-0,013	0,053	0,109	0,008	0,053	0,075	-0,018	0,077
<i>E. coli</i>	0,068	0,001	0,086	0,017	0,061	0,086	0,052	0,037	0,105

Ek 1-b: Bayburt İli Şebeke Sularının Kimyasal Parametrelerine Ait Korelasyon Tablosu

Korelasyon	Civa	Nikel	Nitrat	Selenyum	Akrilamid	Benzo (a)	Dikloretan	Epikloridin	Florür
Antimon	0,742	0,074	-0,039	0,782	0,754	0,786	0,786	0,760	0,150
Arsenik	0,193	0,031	0,063	0,206	0,215	0,220	0,220	0,214	0,360
Benzen	0,919	0,064	-0,038	0,969	0,941	0,980	0,980	0,948	0,109
Bor	0,341	0,025	0,023	0,340	0,331	0,341	0,341	0,333	0,060
Bromat	0,558	0,041	-0,028	0,587	0,554	0,590	0,590	0,558	0,094
Kadmium	0,917	0,064	-0,038	0,967	0,939	0,978	0,978	0,946	0,122
Krom	0,706	0,040	0,027	0,745	0,717	0,749	0,749	0,723	0,133
Bakır	0,114	0,329	-0,013	0,126	0,119	0,156	0,157	0,119	0,138
Kurşun	0,754	0,053	-0,028	0,796	0,777	0,815	0,815	0,779	0,089
Civa	1,000	0,051	-0,039	0,916	0,877	0,913	0,917	0,884	0,089
Nikel	0,051	1,000	-0,005	0,055	0,062	0,065	0,065	0,057	0,031
Nitrat	-0,039	-0,005	1,000	-0,034	0,027	-0,039	-0,039	0,028	0,377
Selenyum	0,916	0,055	-0,034	1,000	0,925	0,962	0,967	0,933	0,095
Akrilamid	0,877	0,062	0,027	0,925	1,000	0,951	0,946	0,988	0,117
Benzo (a)	0,913	0,065	-0,039	0,962	0,951	1,000	0,995	0,948	0,127
Dikloretan	0,917	0,065	-0,039	0,967	0,946	0,995	1,000	0,948	0,127
Epikloridin	0,884	0,057	0,028	0,933	0,988	0,948	0,948	1,000	0,104
Florür	0,089	0,031	0,377	0,095	0,117	0,127	0,127	0,104	1,000
Nitrit	0,915	0,060	-0,049	0,965	0,942	0,985	0,985	0,943	0,117
Toplam Pestisitler	0,242	0,017	0,030	0,261	0,257	0,265	0,265	0,256	0,029
Tetra	0,873	0,068	0,014	0,921	0,947	0,944	0,939	0,944	0,169
Toplam Trihalometanlar	0,305	0,045	0,012	0,326	0,322	0,328	0,328	0,325	0,050
Vinil Klorür	0,863	0,055	0,009	0,910	0,927	0,925	0,925	0,934	0,137
Toplam DDT	0,417	0,004	-0,128	0,450	0,430	0,452	0,452	0,433	-0,044
Toplam PAH	0,860	0,051	-0,033	0,907	0,896	0,915	0,915	0,902	0,070
Enterokok	0,045	-0,002	-0,044	0,049	0,042	0,052	0,052	0,040	-0,038
Koliform	0,044	-0,017	-0,106	0,041	0,056	0,053	0,052	0,056	-0,061
<i>E. coli</i>	0,078	0,005	-0,051	0,082	0,079	0,087	0,086	0,078	-0,017

Ek 1-c: Bayburt İli Şebeke Sularının Kimyasal Parametrelerine Ait Korelasyon Tablosu

Korelasyon	Nitrit	Toplam Pestisitler	Tetra	Toplam Trihalometanlar	Vinil Klorür	Toplam DDT	Toplam PAH	Enterokok	Koliform	<i>E. coli</i>
Antimon	0,787	0,214	0,798	0,282	0,789	0,362	0,738	0,044	0,034	0,068
Arsenik	0,211	0,061	0,215	0,114	0,203	0,065	0,185	-0,021	-0,013	0,001
Benzen	0,975	0,262	0,943	0,333	0,927	0,453	0,915	0,052	0,053	0,086
Bor	0,344	0,364	0,310	0,095	0,307	0,182	0,335	-0,031	0,109	0,017
Bromat	0,591	0,144	0,555	0,205	0,545	0,322	0,557	0,038	0,008	0,061
Kadmiyum	0,973	0,262	0,945	0,332	0,925	0,452	0,913	0,052	0,053	0,086
Krom	0,750	0,198	0,713	0,242	0,704	0,372	0,700	0,041	0,075	0,052
Bakır	0,155	0,061	0,172	0,116	0,114	0,080	0,132	-0,009	-0,018	0,037
Kurşun	0,811	0,214	0,773	0,267	0,765	0,417	0,750	0,085	0,077	0,105
Civa	0,915	0,242	0,873	0,305	0,863	0,417	0,860	0,045	0,044	0,078
Nikel	0,060	0,017	0,068	0,045	0,055	0,004	0,051	-0,002	-0,017	0,005
Nitrat	-0,049	0,030	0,014	0,012	0,009	-0,128	-0,033	-0,044	-0,106	-0,051
Selenyum	0,965	0,261	0,921	0,326	0,910	0,450	0,907	0,049	0,041	0,082
Akrilamid	0,942	0,257	0,947	0,322	0,927	0,430	0,896	0,042	0,056	0,079
Benzo (a)	0,985	0,265	0,944	0,328	0,925	0,452	0,915	0,052	0,053	0,087
Dikloretan	0,985	0,265	0,939	0,328	0,925	0,452	0,915	0,052	0,052	0,086
Epikloridin	0,943	0,256	0,944	0,325	0,934	0,433	0,902	0,040	0,056	0,078
Florür	0,117	0,029	0,169	0,050	0,137	-0,044	0,070	-0,038	-0,061	-0,017
Nitrit	1,000	0,262	0,929	0,325	0,926	0,453	0,917	0,046	0,049	0,087
Toplam Pestisitler	0,262	1,000	0,255	0,029	0,252	0,025	0,252	-0,010	-0,020	-0,006
Tetra	0,929	0,255	1,000	0,344	0,961	0,422	0,872	0,047	0,046	0,082
Toplam Trihalometanlar	0,325	0,029	0,344	1,000	0,337	0,188	0,298	-0,008	-0,060	-0,001
Vinil Klorür	0,926	0,252	0,961	0,337	1,000	0,418	0,866	0,039	0,034	0,072
Toplam DDT	0,453	0,025	0,422	0,188	0,418	1,000	0,421	0,036	0,104	0,101
Toplam PAH	0,917	0,252	0,872	0,298	0,866	0,421	1,000	0,038	0,043	0,070
Enterokok	0,046	-0,010	0,047	-0,008	0,039	0,036	0,038	1,000	0,370	0,467
Koliform	0,049	-0,020	0,046	-0,060	0,034	0,104	0,043	0,370	1,000	0,443
<i>E. coli</i>	0,087	-0,006	0,082	-0,001	0,072	0,101	0,070	0,467	0,443	1,000

Ek 2-a: Bayburt İli Şebeke Sularına Ait Makine Öğrenme Yöntemleri Sonuçları

Bayburt İli Şebeke Suları Mikrobiyolojik ve Kimyasal Parametre Analizi Makine Öğrenme Yöntemleri Sonuçları											
Kullanılan Yöntemler	T(positive)	F(negative)	F(positive)	T(negative)	Sensitivity	Specificity	Precision	Accuracy	Error Rate	False Positive Rate	Prevalance
Karar Ağacı Yöntemi (Entropi)	102	1	1	132	0,99	0,99	0,99	0,99	0,01	0,03	0,44
Karar Ağacı Yöntemi (Gini)	101	2	4	129	0,98	0,97	0,96	0,97	0,03	0,03	0,44
Knn Yöntemi (Manhattan)	86	18	1	131	0,83	0,99	0,99	0,92	0,08	0,01	0,44
Knn Yöntemi (Minkowski-Euclidean)	86	17	2	131	0,83	0,98	0,98	0,92	0,08	0,02	0,44
Lojistik Regresyon Yöntemi	83	20	7	126	0,81	0,95	0,92	0,89	0,11	0,05	0,44
Naive Bayes Yöntemi	83	20	9	124	0,81	0,93	0,90	0,88	0,12	0,07	0,44
Rastgele Orman Yöntemi (Entropi)	100	3	4	129	0,97	0,97	0,96	0,97	0,03	0,03	0,44
Rastgele Orman Yöntemi (Gini)	99	4	2	131	0,96	0,98	0,98	0,97	0,03	0,02	0,44
SVM Yöntemi (Linear)	91	12	7	126	0,88	0,95	0,93	0,92	0,08	0,05	0,44
SVM Yöntemi (Polinomial)	53	50	1	132	0,51	0,99	0,98	0,78	0,22	0,01	0,44
SVM Yöntemi (RBF)	84	19	3	130	0,82	0,98	0,97	0,91	0,09	0,02	0,44
SVM Yöntemi (Sigmoid)	84	19	13	120	0,82	0,90	0,87	0,86	0,14	0,10	0,44

Ek 2-b: Bayburt İli Çeşme Sularına Ait Makine Öğrenme Yöntemleri Sonuçları

Bayburt İli Çeşme Suları Analizi Makine Öğrenme Yöntemleri Sonuçları											
Kullanılan Yöntemler	T(positive)	F(negative)	F(positive)	T(negative)	Sensitivity	Specificity	Precision	Accuracy	Error Rate	False Positive Rate	Prevalance
Karar Ağacı Yöntemi (Entropi-Gini)	65	0	0	31	1,00	1,00	1,00	1,00	0,00	0,00	0,68
Knn Yöntemi (Minkowski-Euclidean-Manhattan)	65	0	1	30	1,00	0,97	0,98	0,99	0,01	0,03	0,68
Lojistik Regresyon Yöntemi	65	0	3	28	1,00	0,90	0,96	0,97	0,03	0,10	0,68
Naive Bayes Yöntemi	65	0	0	31	1,00	1,00	1,00	1,00	0,00	0,00	0,68
Rastgele Orman Yöntemi (Entropi, Gini)	65	0	0	31	1,00	1,00	1,00	1,00	0,00	0,00	0,68
SVM Yöntemi (Linear, RBF, Sigmoid)	65	0	1	30	1,00	0,97	0,98	0,99	0,01	0,03	0,68
SVM Yöntemi (Polynomial)	65	0	19	12	1,00	0,39	0,77	0,80	0,20	0,61	0,68

ÖZ GEÇMİŞ

İlkokulu Şanlıurfa’da, ortaokul ve liseyi Osmaniye’de okudu. 2010 yılında Adıyaman Besni Meslek Yüksekokulu Bilgisayar Programcılığı bölümünden mezun oldu. 2010-2014 yılları arasında Halk eğitim merkezinde kurs öğreticiliği yaptı. 2015 yılında Çağ Üniversitesi Matematik-Bilgisayar (İngilizce) bölümünden mezun oldu.

