



**BÜYÜK MENDERES HAVZASI SU KALİTE PARAMETRELERİNİN
İNCELENMESİ VE YAPAY ZEKÂ YÖNTEMLERİ İLE MODELLENMESİ**

Farid HASSANBAKİ GARABAGHI

ÇEVRE BİLİMLERİ ANA BİLİM DALI

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

TEMMUZ 2023

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Farid HASSANBAKI GARABAGHI

05/07/2023

BÜYÜK MENDERES HAVZASI SU KALİTE PARAMETRELERİNİN İNCELENMESİ VE YAPAY ZEKÂ YÖNTEMLERİ İLE MODELLENMESİ

(Doktora Tezi)

Farid HASSANBAKİ GARABAGHI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Temmuz 2023

ÖZET

Sürdürülebilir içme suyu sağlamak ve ekosistemi korumak için tatlı su kaynaklarının kalitesi sürekli izlenmelidir. Bu çalışma, Büyük Menderes Havzasını göz önünde bulundurarak su kalitesi sınıflandırması ve tahmininin otomasyonu doğrultusunda iki hedefi izlemektedir. Bu araştırmanın ilk amacı, su kalitesi sınıflandırması için doğru bir model ortaya koymaktır. Water Quality Index (WQI), su kalitesini sınıflandırıcı için kullanılmaktadır. Ancak, WQI'nin hesaplaması karmaşık ve zaman alıcıdır. Su kalitesi sınıflandırması için gelişmiş yöntemlerden birisi Machine Learning (ML) dir. Bu çalışmada, su kalitesi sınıflandırması için dört ML modelinin performansı incelenmiştir. Bir ML modelini hesaplama açısından uygun maliyetli hale getirmek için, en ilgili parametreler ele alınarak boyutsallık azaltma teknikleri kullanılır. Bu çalışmada, boyutsallığı azaltılmış veri setlerinin sınıflandırıcıların performansı üzerindeki etkisi değerlendirilmiştir. Bu çalışmada, farklı parametrelerden oluşan alt kümeler oluşumu için üç öznitelik seçici kullanılmıştır. Araştırmalar sonucunda, XGBoost modeli, tüm veri seti ile olağanüstü performans (Doğruluk: 96,9696%) göstermiştir. Sonuçlar, boyutsallık azaltmanın modellerin performansı üzerindeki olumsuz etkisini göstermektedir. Maliyet etkinliğine dayalı fizibilite çalışması, XGBoost modelinin, Backeard Feature Elimination Wrapper Method (BFE-WRAPM) tarafından önerilen veri seti ile en iyi sınıflandırıcı olarak kabul edilebileceğini (Doğruluk: %95.606) göstermektedir. Su yaşamı, Çözünmüş Oksijen (ÇO) seviyelerinden oldukça etkilenmektedir. Bunun için, gelecekteki ÇO seviyeleri için bir tahmin modeli oluşumu gereksinimdir. Bu tez çalışmasının ikinci aşaması, ÇO Tahmini için bir modelin oluşturulmasıdır. Bu doğrultuda, boyutsallık azaltma çalışması ile birlikte Random Forest (RF) ve Multilayer Perceptron (MLP) modellerinin ÇO tahmini performansı değerlendirilmiştir. RF modeli MLP'ye göre her iki veri kümesi ile (tüm veri seti ve boyutsallığı azaltılmış veri kümesi) üstün performans sergilemiştir. RF modeli için en iyi performans boyutsallığı düşürülmüş veri kümesi ile eğitildiğinde tespit edilmiştir. Her iki modelin Estimation Error Deviation'i (EED), boyutsallığı düşürülmüş veri seti ile önemli ölçüde azaldığı tespit edilmiştir. Sonuçlar, boyutsallık azaltmanın her iki modelin doğruluğu üzerindeki olumlu etkisini göstermektedir. Sonuç olarak, RF, ÇO tahmininde etkin bir model olarak kabul edilebilir.

Bilim Kodu : 90318
Anahtar Kelimeler : Su kalitesi, Büyük Menderes Havzası, Makine Öğrenimi, Sınıflandırma, Tahmin, Su Kalitesi Endeksi, Regresyon
Sayfa Adedi : 171
Danışman : Prof. Dr. Semra BENZER

ANALIZING THE WATER QUALITY PARAMETERS IN THE BÜYÜK MENDERES BASIN AND MODELING BY ARTIFICIAL INTELLIGENCE METHODS

(Ph.D Thesis)

Farid HASSANBAKI GARABAGHI

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2023

ABSTRACT

To provide sustainable drinking water and protect the ecosystem, the quality of freshwater resources must be monitored constantly. This research thesis aims to follow two objectives regarding the automation of water quality classification and prediction considering the Büyük Menderes Basin. The first objective of this research was to introduce an accurate model for water quality classification. Water Quality Index (WQI) has been used to classify water quality for years. However, calculating WQI is complicated and time-consuming. One of the advanced methods to automate water quality classification is machine learning (ML). In this study, the performance of four ML models were investigated for water quality classification. Besides, to make a ML model computationally cost-effective, dimensionality reduction techniques are used to detect the most relevant parameters. For this, the impact of the dimensionality-reduced datasets on the performance of the classifiers was evaluated. Three feature selection techniques were used to generate the subset of features with different combinations of parameters. XGBoost showed outstanding performance (accuracy: 96.9696%) with all the parameters. The results showed the adverse impact of dimensionality reduction on the performance of the classifiers. However, feasibility study based on cost-effectiveness showed that the XGBoost classifier can be considered the best classifier (accuracy: 95.606%) with the dataset suggested by the Backward Feature Elimination Wrapper Method (BFE-WRAPM). As aquatic life is highly influenced by Dissolved Oxygen (DO) levels so a prediction model for future levels of DO is a demand. The second phase of this thesis is to create a model for DO Prediction. To this end, the performance of the RF and Multilayer Perceptron (MLP) algorithms for DO prediction were evaluated along with the dimensionality reduction study. The results showed that the RF excelled MLP in performance with both datasets (the dataset of all and reduced variables). The best performance achieved by the RF with the dataset of reduced variables. The Estimation Error Deviation (EED) of both models declined significantly with the dataset of reduced variables. The results demonstrated the positive impact of the dimensionality reduction on the accuracy of both models. However, RF can be considered a robust model in predicting DO.

Science Code : 90318

Key Words : Water quality, Büyük Menderes basin, machine learning, classification, prediction, water quality index, regression

Page Number : 171

Supervisor : Prof. Dr. Semra BENZER

TEŞEKKÜR

Bu tez çalışmasının gerçekleşmesinde katkıda bulunan herkese teşekkürlerimi iletmek istiyorum. Eşim Homay Danaei Mehr'e, doktora tezi sürecindeki destekleri ve motivasyonu için minnettarım.

Danışman Hocam Prof. Dr. Semra BENZER'e, her zaman yol gösterici olması ve tez çalışmasının tamamlanmasında sağladığı büyük katkı için teşekkür ederim.

Çevre Bilimleri Bölüm Başkanı Prof. Dr. Çağlan GÜNAL'a, en zor anlarımda bana destek verdiği için minnettarım.

Doç. Dr. Recep BENZER'e, tez çalışmasında sağladığı bilgisayar bilimleri danışmanlığı için teşekkür ederim.

Tez İzleme Komitesi üyeleri Prof. Dr. Ali GÜL ve Prof. Dr. İlkay AÇIKGÖZ'e, yorumları ve değerlendirmeleriyle tezin kalitesini yükseltmeleri için teşekkür ederim.

Devlet Su İşleri Genel Müdürlüğü Etüt, Plan ve Tahsis Dairesi Başkanlığı'na, Büyük Menderes Havzası'nın su kalitesi verilerinin sağlanması ve diğer destekleri için teşekkür ederim.

Ayrıca, kendi aileme (anne ve babam) ve ikinci aileme (kayın pederim ve kayın validem) sonsuz maddi ve manevi destekleri için teşekkür etmek istiyorum.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ.....	xiii
SİMGELER VE KISALTMALAR.....	xvii
1. GİRİŞ.....	1
1.1. Su Kalitesi Parametreleri ve Standartları	2
1.1.1. Su kalitesi genel parametreler	2
1.1.2. Su kalitesi oksijenlendirme parametreleri.....	4
1.1.3. Su kalitesi besin madde parametreleri.....	5
1.1.4. Su kalitesi belirtici iz elementler (Metaller) ve inorganik kirleticiler	7
1.1.5. Diğer parametreler	10
1.2. Sınıflandırma ve Tahmin.....	14
1.2.1. Su kalitesi sınıflandırması	14
1.2.2. Su kalitesi parametreleri tahmini modellemesi	16
1.3. İstatistiksel, Makine Öğrenme (Machine Learning (ML)) Sınıflandırma ve Tahmin Modelleri.....	17
1.3.1. İstatistiksel sınıflandırma yaklaşımı ve modelleri.....	17
1.3.2. Makine öğrenme (Machine Learning (ML)) sınıflandırma yaklaşımı ve modelleri	21
1.4. ML Yaklaşımıyla Tahmin Modelleri	27
1.5. Öz Nitelik Seçimi (Feature Selection) ve Önemi	27
1.6. Projenin Kapsamı	28

	Sayfa
1.7. Çalışmanın Önemi ve Hedefi	32
2. LİTERATÜR İNCELEMESİ	37
2.1. İstatistiksel Yöntemlere Dayalı Sınıflandırma Modellemesi ve Tahmin Çalışmaları	37
2.2. Makine Öğrenme Yöntemlere Dayalı Sınıflandırma ve Tahmin Modellemesi Çalışmaları	40
3. MATERYAL VE METOD	47
3.1. Veri ve Veri Seti Oluşturulması	47
3.1.1. Kayıp veri inşası.....	52
3.1.2. WQI modelin kurulması ve su kalitesi sınıflarının belirlenmesi	54
3.2. Sınıflandırma ve Tahmin Çalışmalarında Kullanılan Öz Nitelik Seçimi Yöntemleri.....	57
3.2.1. Öznitelik seçimi FLTM yöntemi.....	58
3.2.2. Öznitelik seçimi sarmalayıcı yöntemi (Wrapper Method (WRAPM))	59
3.2.3. Gömülü öznitelik seçimi yöntemi (Embedded method (EMBM))	61
3.3. Sınıflandırma ve Tahmin Çalışmaları	62
3.3.1. Toplu öğrenme (Ensemble Learning (EL)).....	62
3.3.2. ANN ve MLP regresyon (Tahmin) modeli	71
3.4. Sınıflandırma Çalışmaları İçin Değerlendirme	72
3.5. Tahmin Çalışmaları İçin Değerlendirme Metrikleri	73
3.6. Eğitim, Doğrulama ve Deneme	74
3.6.1. HCV doğrulama tekniği	75
3.6.2. K-FCV tekniği.....	76
3.7. Ön Eğitim Çalışması	76
3.8. Sınıflandırma Çalışması Prosedürü.....	77
3.9. ÇO Tahmin Çalışması Prosedürü.....	80

	Sayfa
4. BULGULAR	83
4.1. Büyük Menderes Havzası'nın Genel Su Kalitesi Durumu	83
4.2. Sınıflandırma Çalışması Sonuçları.....	98
4.2.1. WQI sınıflandırması sonuçları	98
4.2.3. Öznitelik seçimi sonuçları.....	108
4.2.4. Sınıflandırma çalışması sonuçları	110
4.3. Çözünmüş Oksijen (ÇO) Konsantrasyonu Tahmin Modellemesi Çalışması Sonuçları	122
4.3.1. Öz nitelik seçimi sonuçları.....	122
4.3.2. Çözünmüş oksijen miktarı tahmin çalışması sonuçları.....	125
5. TARTIŞMA	135
5.1. Büyük Menderes Havzası Su Kalitesi Durumu Tartışması.....	135
5.2. Sınıflandırma Sonuçlarının Tartışması	137
5.3. Tahmin Çalışması Sonuçlarının Tartışılması	145
6. SONUÇ VE ÖNERİLER	151
KAYNAKLAR	155
ÖZGEÇMİŞ	171

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 1.1. Su Kalitesi Parametreleri Sınır Değerleri	13
Çizelge 3.1. Entegre Veri Kümesinin Şematik Örneği	48
Çizelge 3.2. Seçilen İstasyonlar.	49
Çizelge 3.3. Su Kalitesi Endeksi Denkleminde Kullanılan Parametrelerin Standart Değerleri.....	55
Çizelge 3.4. Standart Değerlere Göre Sn, K, Ve Wi Hesaplaması.	56
Çizelge 3.5. Ağırlığa Göre Su Kalitesi Sınıfları Aritmetik Su Kalitesi Endeksi Yöntemi.....	57
Çizelge 3.6. Şematik Birçok Sınıflı Cm.....	73
Çizelge 4.1. Wqı'leri Belirtilmiş Verilerin Şematik Veri Seti Çizelgesi (Msk: Mükemmel Su Kalitesi; İsk: İyi Su Kalitesi; Ksk: Kötü Su Kalitesi; : Çok Kötü Su Kalitesi; Tud: Tüketim İçin Uygun Degil)	99
Çizelge 4.2. Büyük Menderes Havzası Veri Seti İçin Uygulanan Öznitelik Seçimi Yöntemlerinin Sonuçları	109
Çizelge 4.3. Tüm Veri Seti Ve 5-Fcv Yöntemiyle Eğitildiği Doğrulama Sonuçları	110
Çizelge 4.4. Tüm Veri Seti Ve 10-Fcv Yöntemiyle Eğitildiği Doğrulama Sonuçları	111
Çizelge 4.5. Modellerin Tüm Veri Seti Ve %70-30 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları	111
Çizelge 4.6. Tüm Versi Seti Ve %75-25 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	112
Çizelge 4.7. Tüm Veri Seti Ve %80-20 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçlar	112
Çizelge 4.8. Bfe-Wrapm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve 5-Fcv Yöntemiyle Eğitildiği Doğrulama Sonuçları	113
Çizelge 4.9. Bfe-Wrapm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve 10-Fcv Yöntemiyle Eğitildiği Doğrulama Sonuçları	114
Çizelge 4.10. Bfe-Wrapm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %70-30 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	114

Çizelge	Sayfa
Çizelge 4.11. Bfe-Wrapm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %75-25 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	115
Çizelge 4.12. Bfe-Wrapm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %80-20 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	116
Çizelge 4.13. M1-Fltm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve 5-Fcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	116
Çizelge 4.14. M1-Fltm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve 10-Fcv Yöntemiyle Eğitilen Doğrulama Sonuçları	117
Çizelge 4.15. M1-Fltm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %70-30 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	117
Çizelge 4.16. M1-Fltm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %75-25 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	118
Çizelge 4.17. M1-Fltm Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %80-20 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	119
Çizelge 4.18. Rf-Embm Öznitelik Seçimi Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve 5-Fcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	119
Çizelge 4.19. Rf-Embm Öznitelik Seçimi Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve 10-Fcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	120
Çizelge 4.20. Rf-Embm Öznitelik Seçimi Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %70-30 Eğitim-Doğrulama Oranlı Hcv Eğitildiği Doğrulama Sonuçları	121
Çizelge 4.21. Rf-Embm Öznitelik Seçimi Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %75-25 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	121
Çizelge 4.22. Rf-Embm Öznitelik Seçimi Yöntemi'nin Önerdiği Parametrelerden Oluşan Altküme Ve %80-20 Eğitim-Doğrulama Oranlı Hcv Yöntemiyle Eğitildiği Doğrulama Sonuçları.....	122
Çizelge 4.23. Tahmin Modellerinin Performans Değerlendirme Sonuçları	125

Çizelge	Sayfa
Çizelge 5.1. Bfe-Wrapm Öznitelik Seçimi Yönteminin Önerdiği Alt Küme Ve 10-Fcv Uygulanmasıyla Xgboost Modelinin Cm Tablosu.....	141
Çizelge 5.2. M1-Fltm Öznitelik Seçimi Yöntemi'nin Önerdiği Alt Küme Ve 10-Fcv Uygulanmasıyla Xgboost Modelinin Cm Tablosu.....	142
Çizelge 5.3. Rf-Embm Öznitelik Seçimi Yöntemi'nin Önerdiği Alt Küme Ve 10-Fcv Uygulanmasıyla Xgboost Modelinin Cm Tablosu.....	143
Çizelge 5.4. Tüm Veriseti Ve 10-Fcv Uygulanmasıyla Xgboost Modelinin Cm Tablosu.....	144

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 1.1. Üç Boyutlu Uzayda Ve Çoklu Sınıf Ayırımında Ld Çalışma Şeması.....	18
Şekil 1.2. Bayes Kuralı Çalışma Şeması.....	19
Şekil 1.3. Knn'nin Çalışma Şeması	20
Şekil 1.4. Genetik Algoritmalar Çalışma Akış Şeması.....	22
Şekil 1.5. R1 Çalışma Akış Şeması	23
Şekil 1.6. Rote Learning Çalışma Akış Şeması.	24
Şekil 1.7. Tümevarım Ve Tümdengelim Öğrenme Yöntemleri Çalışma Akışı Şeması.	24
Şekil 1.8. Büyük Menderes Havzası Çevresel Altyapı Mevcut Durum Haritası.....	30
Şekil 1.9. Büyük Menderes Havzası Akarsuları Uzunlukları.	31
Şekil 3.1. Büyük Menderes Havzası Ve Bu Çalışma İçin Seçilen İstasyonlar.	51
Şekil 3.2. Lr Makine Öğrenme Algoritması İle Eksik Değer Atamasının Şematik Süreci.....	53
Şekil 3.3. Wrapm Öz Nitelik Seçimi Çalışma Akış Şeması.	60
Şekil 3.4. Rf Modelinin Çalışma Şeması	65
Şekil 3.5. Xgboost Modelinin Çalışma Şeması.	67
Şekil 3.6. Yapay Sinir Ağları Çalışma Akış Şeması.....	72
Şekil 3.7. K-Fcv Tekniğinin Çalışma Şeması.....	76
Ekil 3.8. Büyük Menderes Havzasi Su Kalitesi Siniflandirma Çalışmasi Akış Şeması	79
Şekil 4.1. Büyük Menderes Havzası'nın Çözünmüş Oksijen Durumu.....	84
Şekil 4.2. Büyük Menderes Havzası'nın Aylar Bazında Çözünmüş Oksijen Durumu ..	84
Şekil 4.3. Büyük Menderes Havzası'nın Aylar Bazında Sıcaklık Durumu	85
Şekil 4.4. Büyük Menderes Havzası'nın Koi Durumu	86
Şekil 4.5. Büyük Menderes Havzası'nın Aylar Bazında Koi Durumu	86
Şekil 4.6. Büyük Menderes Havzası'nın Boi-5 Durumu	87

Şekil	Sayfa
Şekil 4.7. Büyük Menderes Havzası'nın Aylar Bazında Boi-5 Durumu	88
Şekil 4.8. Büyük Menderes Havzası'nın Amonyum Durumu	89
Şekil 4.9. Büyük Menderes Havzası'nın Aylar Bazında Amonyum Durumu	89
Şekil 4.10. Büyük Menderes Havzası'nın Nitrat Durumu	90
Şekil 4.11. Büyük Menderes Havzası'nın Aylar Bazında Nitrat Durumu.....	90
Şekil 4.12. Büyük Menderes Havzası'nın Nitrit Durumu.....	91
Şekil 4.13. Büyük Menderes Havzası'nın Aylar Bazında Nitrit Durumu	91
Şekil 4.14. Büyük Menderes Havzası'nın Fosfat Durumu	92
Şekil 4.15. Büyük Menderes Havzası'nın Aylar Bazında Fosfat Durumu	92
Şekil 4.16. Büyük Menderes Havzası'nın Sülfat Durumu.....	93
Şekil 4.17. Büyük Menderes Havzası'nın Aylar Bazında Sülfat Durumu	93
Şekil 4.18. Büyük Menderes Havzası'nın Toplam Çözülmüş Katı Durumu.....	94
Şekil 4.19. Büyük Menderes Havzası'nın Aylar Bazında Tçk Durumu.....	94
Şekil 4.20. Büyük Menderes Havzası'nın Sodyum Durumu	95
Şekil 4.21. Büyük Menderes Havzası'nın Aylar Bazında Sodyum Durumu.....	95
Şekil 4.22. Büyük Menderes Havzası'nın Manganez Durumu.....	96
Şekil 4.23. Büyük Menderes Havzası'nın Aylar Bazında Manganez Durumu	96
Şekil 4.24. Büyük Menderes Havzası'nın Klorür Durumu.....	97
Şekil 4.25. Büyük Menderes Havzası'nın Aylar Bazında Klorür Durumu	97
Şekil 4.26. Büyük Menderes Havzası Su Kalitesi Genel Durumu	100
Şekil 4.27. Yıllar Bazlı Büyük Menderes Havzası'nda "Tüketim İçin Uygun Olmayan" Su Kalitesi Vaka Sayısı	101
Şekil 4.28. 2004 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	101
Şekil 4.29. 2005 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	102

Şekil	Sayfa
Şekil 4.30. 2006 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	103
Şekil 4.31. 2007 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	103
Şekil 4.32. 2008 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	104
Şekil 4.33. 2009 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	105
Şekil 4.34. 2010 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	105
Şekil 4.35. 2011 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	106
Şekil 4.36. 2012 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	106
Şekil 4.37. 2013 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	107
Şekil 4.38. 2014 Yılı İçin Büyük Menderes Havzası'nda Aylar Bazında "Tüketim İçin Uygun Olmayan" Su Kalitesi Vakaları.....	107
Şekil 4.39. Bfe-Wrapm Öznitelik Seçici Tarafından Seçilen Değişkenler Arasındaki Korelasyon Çizim Matrisi.....	124
Şekil 4.40. Tüm Veri Seti Ve %70-30 Eğitim-Deneme Oranlı Hcv Doğrulama Metodu Uygulanmasıyla Mlp Modelinin Gerçek Ve Tahmin Edilen Değerleri.....	126
Şekil 4.41. Boyutsallığı Düşürülmüş Veri Seti Ve %70-30 Eğitim-Deneme Oranlı Hcv Doğrulama Metodu Uygulanmasıyla Mlp Modelinin Gerçek Ve Tahmin Edilen Değerleri	127
Şekil 4.42. Tüm Veri Seti Ve 10-Fcv Doğrulama Metodu Uygulanmasıyla Rf Modelinin Gerçek Ve Tahmin Edilen Değerleri.....	127
Şekil 4.43. Boyutsallığı Düşürülmüş Veri Seti Ve 10-Fcv Doğrulama Metodu Uygulanmasıyla Rf Modelinin Gerçek Ve Tahmin Edilen Değerleri.....	128
Şekil 4.44. Tüm Veriseti İle Rf Tahmin Modelinin Gerçek Ço Değerine Karşı Tahmin Edilen Ço Konsantrasyonu.....	129
Şekil 4.45. Tüm Değişkenler İle Eğitilen Rf Modelinin Test Aşamasındaki Eed'sı.	129

Şekil	Sayfa
Şekil 4.46. Boyutsallığı Düşürülmüş Alt Küme İle Rf Tahmin Modelinin Gerçek Ço Değerine Karşı Tahmin Edilen Ço Konsantrasyonu.....	130
Şekil 4.47. Boyutsallığı Düşürülmüş Alt Küme İle Eğitilen Rf Modelinin Test Aşamasındaki Eed'sı.....	131
Şekil 4.48. Tüm Veriseti İle Mlp Tahmin Modelinin Gerçek Ço Değerine Karşı Tahmin Edilen Ço Konsantrasyonu.	132
Şekil 4.49. Boyutsallığı Düşürülmüş Alt Küme İle Mlp Tahmin Modelinin Gerçek Ço Değerine Karşı Tahmin Edilen Ço Konsantrasyonu.	132
Şekil 4.50. Tüm Veriseti İle Eğitilen Mlp Modelinin Test Aşamasındaki Eed'sı.	133
Şekil 4.51. Boyutsallığı Düşürülmüş Alt Küme İle Eğitilen Mlp Modelinin Test Aşamasındaki Eed'sı.....	134
Şekil 5.1. Farklı Doğrulama Yöntemleri Ve Bfe-Wrapm Öznitelik Seçiminin Önerdiği Altküme İle Modellerin Doğruluk Kıyaslaması.....	139
Şekil 5.2. Farklı Doğrulama Yöntemleri Ve M1-Fltm Öznitelik Seçiminin Önerdiği Altküme İle Modellerin Doğruluk Kıyaslaması.....	139
Şekil 5.3. Farklı Doğrulama Yöntemleri Ve Rf-Embm Öznitelik Seçiminin Önerdiği Altküme İle Modellerin Doğruluk Kıyaslaması.....	140
Şekil 5.4. Farklı Doğrulama Yöntemleri Ve Tüm Veriseti İle Modellerin Doğruluk Kıyaslaması	140
Şekil 5.5. Şekil 5.5. Tahmin Edilen Örneklerin Kabul Edilebilir Eed Aralığı	147

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklamalar
°C	Santigrat derece
Ca	Kalsiyum
Cl ⁻	Klorür
cm	Santimetre
Co	Kobalt
Fe	Demir
H	Hidrojen
ha	Hektar
K	Potasyum
km	Kilometre
km ²	Kilometre kare
km ³	Kilometre küp
L	Litre
Log	Logaritma
m ²	Metre kare
m ³	Metre küp
Mg	Magnezyum
Mg	Miligram
N	Nitrojen
Na ⁺	Sodyum
NH ₃	Amonyak
NH ₄ ⁺	Amonyum
NO ₂	Nitrit
NO ₃	Nitrat
nS	Nano Simens
O	Oksijen
P	Forfor

Simgeler**Açıklamalar** Φ

Fi

 PO_4

Fosfat

 Pt

Platinyum

 S

Sülfür

 Σ

Sigma

 Si

Silikon

 SO_4

Sülfat

 T

Sıcaklık

 α

Alfa

 β

Beta

 γ

Gama

 δ

Delta

 ε

Epsilon

 θ

Teta

 λ

Lambda

 ρ

Ro

 σ

Sigma

 Ω

Omega

 ν

Nü

 μg

Mikrogram

 μS

Mikro Simens

 ∂

Delta

Kısaltmalar**Açıklamalar****ABD**

Amerika Birleşik Devletleri

AdaBoost

Adaptive Boosting

AI

Artificial Intelligence

AL

Analog Learning

ANNs

Artificial Neural Networks

ARIMA

Autoregressive Integrated Moving Average

Kısaltmalar**Açıklamalar**

BC	British Columbia
BCMELP	British Columbia Minst. of Envi., Lands & Parks
BFE-WRAPM	Backward Feature Elimination Wrapper Method
BI	Bagging of Independents
BM	Bagging Method
BNET	Bayesian Networks
BOİ	Biyolojik Oksijen İhtiyacı
BPNN	Back Propagation Neural Network
CA	Clustering Analysis
CC	Conceptual Clustering
CCME	Canadian Council of Ministers of the Environment
CLM	Close Loop Method
CM	Confusion Matrix
CVAL	Cross Validation
CV-CPS	CV parameter selection
ÇO	Çözünmüş Oksijen
DA	Discriminant Analysis
DL	Deep Learning
DR	Dynamik Regression
DSİ	Devlet Su İşleri Genel Müdürlüğü
DT	Decision Tree
EBL	Explanation-based learning
EED	Estimation Error Deviation
EI	Ensemble Inducer
EL	Ensemble Learning
EMBM	Embedded Method
ENSL	Ensemble Learning
EPA	US Environmental Protection Agency
ETR	Extra Tree Regression
FCM	Fuzzy C-Means
FCMCA	Fuzzy C-Means Cluster Analysis
FCMDC	Fuzzy C-Means Data Clustering

Kısaltmalar**Açıklamalar****FLTM**

Filter Methods

GA

Genetic Algorithms

GBoost

Gradient Boosting Algorithm

GFS

Greedy Forward Selection

GO

Greedy Optimization

GP

Grid Partitioning

HACA

Hierarchical Agglomerative Cluster Analysis

HCV

Holdout Cross Validation

IR-WQI

Iran Water Quality Index

İst.

İstasyon

K-FCV

K-Fold Cross Validation

KM

K-Means

KNN

K-Nearest Neighbore

KOİ

Kimyasal Oksijen İhtiyacı

LASR

Lasso Regression

LD

Linear Discriminant

LDA

Linear Discriminant Analysis

LF

Loss Function

LGD

Logistics Discriminant

LLOSS

Log-Probability Loss Function

LogitBoost

Logistic Boosting

LS

Linear Separability

LSTM

Long-Short Term Memory

RBF

Radial Based Foction

RBFNN

Radial Based Foction Neural Networks

GRNN

General Regression Neural Network

MA

Metaheuristic Algorithms

MAE

Mean Absolute Error

MARC

Markov Chain

MI-FLTM

Mutual Information Filter Method

ML

Machine Learning

MRSM

Modified Response Surface Method

Kısaltmalar**Açıklamalar**

MLP	Multilayer Perceptron
MLPNN	Multilayer Perceptron Neural Network
MLR	Multi-Linear Regression
MSA	Multivariate Statistical Analysis
MWQI	Malaysian Water Quality Index
NB	Naive Bayes
NFIS	Neuro-Fuzzy Inference System
NSF	National Sanitation Foundation
ELM	Extreme Learning Machine
OP-ELM	Optimally Pruned Extreme Learning Machine
OSELM	Online Sequential Extreme Learning Machine
PCA	Principal Component Analysis
PCC	Pearson Correlation Coefficient
PCI	Pearson Correlation Index
PDF	Probability Density Function
PNN	Probabilistic Neural Networks
PR	Polynomial Regression
QD	Quadratic Discriminant
QRF	Quantitative Regression Forest
RBF	Random Bits Forest
REGT	Regression Tree
REPT	Reduced Error Pruning Tree
RF	LGR
RF	Random Forest
RFC	Random Filtered Classification
RF-EMBM	Random Forest Embedded Method
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RREG	Ridge Regression
SC	Subtractive Clustering
SC-ANFIS	Subtract. Clust. Neuro-Fuzzy Inference System
SDA	Spatial Discriminant Analysis

Kısaltmalar**Açıklamalar**

SL	Symbolic Learning
SRDD	Scottish Research and Development Department
SSL	Sub-Symbolic Learning
SUPL	Supervised Learning
SUPL	Supervised Learning
SVM	Support Vector Machine
AIG	Algorithm of the Innovative Gunner
TÇK	Toplam Çözünmüş Katı Madde
TSE	Türkiye Standartları Enstitüsü
UI	Unsupervised Induction
USL	Unsupervised Learning
WAWQI	Weighted Arithmetic Water Quality Index
WHO	World Health Organization
WQI	Water Quality Index
WRAPM	Wrapper Method
XGBoost	Extreme Gradient Boosting

1. GİRİŞ

Yeryüzünde suyun yaşamın en önemli gereksinimlerinden biri olduğu inkâr edilemez bir gerçektir. Dünyada "Mavi gezegen" olarak adlandırılan suların yaklaşık %90'ı yüzeysel sular olmasına rağmen, sadece küçük bir bölümü içme suyu olarak insanların kullanımına uygun durumdadır (Arabgol ve diğerleri, 2016; Ostad-Ali-Askari ve diğerleri, 2017).

Su kaynaklı hastalık salgınları açısından su kalitesi, insan sağlığı açısından önemlidir. Ayrıca, su kalitesi hastalık oranları üzerinde büyük bir etkiye sahiptir (Fewtrell ve Bartram, 2001). Dünya Sağlık Örgütü (World Health Organization (WHO)), tüm hastalıkların yaklaşık %80'inin su kaynaklı olduğunu bildirilmektedir (Dohare ve diğerleri, 2014). Bunun yanı sıra, yüzey suları kirlendiğinde temizlenmesi oldukça zor ve çoğu zaman iyileştirme maliyeti fayda oranı açısından çok yüksek olmaktadır (Arabgol ve diğerleri, 2016).

Günümüzde yüzey sularının ana kirlilik kaynakları, dolaylı veya dolaysız nüfus artışıyla ilişkili olan tarımsal ve endüstriyel bazı insan kaynaklı faaliyetlerdir. Bu gezegende yaşamın sürdürülebilmesi için su kaynaklarının sürdürülebilir bir şekilde yönetilmesi gerekmektedir (Dohare ve diğerleri, 2014; Motevalli ve diğerleri, 2019).

Yüzey suyunun kalitesinin sürekli olarak izlenmesi, insan sağlığı açısından büyük öneme sahiptir. Azot, Fosfor gibi besin maddeleri veya bunların türevleri su sistemlerine salındığında, istenmeyen sonuçlar olan ötrofikasyon meydana gelebilir. Besin zenginleşmesi alg patlamalarına neden olur ve su sistemlerinde Çözünmüş Oksijen (ÇO) seviyesi azalır. Sonuç olarak, sucul yaşamın yok olması gibi istenmeyen sonuçlar ortaya çıkabilir (Rozemeijer ve Broers, 2007; Varol ve Şen, 2012).

Anaerobik bakteri aktiviteleriyle Sülfat'ın indirgenmesi sonucu ortaya çıkan hidrojen Sülfür, yüzey su rezervuarlarında ötrofikasyonun bir göstergesi olarak değerlendirilebilir. İçme suyunda yüksek Sülfat konsantrasyonları sağlık sorunlarına yol açabilir (WHO, 2006; EPA, 2022). Klorür, sucul ortamlarda toksik etkileri nedeniyle önemli bir su kalitesi parametresi olarak bilinir. Klorür, çevrede serbest bir anyon olarak bulunabilir; bu nedenle Sodyum Klorür, Potasyum Klorür, Magnezyum Klorür gibi tuz oluşumlarının ana bileşeni olarak kabul edilir (Elphick ve diğerleri, 2011). Dolayısıyla, Klorür'ün su sistemlerinde çözünmüş

katırlarla ilişkili olduđu sonucuna varılabilir. Demir ve Manganez, yer kabuğunda bulunan iki temel elementtir. Erozyon yoluyla yüzey sulara salınmaları kaçınılmazdır. Bu iki elementin varlığı ciddi bir sağık sorunu olmasa da, yüksek konsantrasyonları suya istenmeyen tat ve renk verirken, anaerobik Demir-Manganez bakteri aktiviteleri ve tortu oluşumu nedeniyle pompaların tıkanmasına yol açabilir (Wong, 1984; WHO, 2006; Dalai ve diğlerleri, 2015). Su kalitesi parametrelerinin normal sınırların dışında olması çeşitli problemlere yol açabilir. Bu nedenle, bir havzadaki potansiyel kirlilik kaynaklarına bağılı olarak su kalitesi parametrelerinin tutarlı bir şekilde izlenmesi ve değlerlendirilmesi önemlidir. Bu, sucul yaşam döngüsünün sağılıklı bir şekilde devam etmesi ve insanlara güvenli ve sürdürülebilir içme suyu sağılanması için devletlerin sorumluluğundadır.

1.1. Su Kalitesi Parametreleri ve Standartları

Etkili bir havza yönetimi için elde edilmesi gereken en önemli bilgilerden biri, havzadaki su kaynaklarının kalitesidir. Su kalitesi, su içinde bulunan bileşenlerin miktarına bağılı olarak belirlenir. Bazı bileşenler su içinde bulunduğunda tehdit oluşturmazken, bazı su kalitesi parametreleri havzanın suyunu kullanılamaz hale getirebilir. Bu bileşenlerin sınırları, kullanım amacına bağılı olarak standartlarla belirlenir. Bu tez çalışmasında, Büyük Menderes Havzası ve Büyük Menderes Nehri içme suyu kalitesi, WHO (Dünya Sağık Örgütü) ve Türkiye Standartları Enstitüsü (TSE) TS-266 standartlarına göre incelenmiştir. WHO (2006) içme suyu kalitesi kılavuzunda birçok su kalitesi parametresine ait standart değler bulunmaktadır. Ancak Büyük Menderes Havzası'ndan elde edilen bazı parametreler WHO (2006) kılavuzunda yer almamaktadır. Bu nedenle, WHO (2006) İçme Suyu Kalitesi Kılavuzu'ndan elde edilemeyen su kalitesi parametreleri, TS-266 standardına göre değlerlendirilmiştir.

1.1.1. Su kalitesi genel parametreler

Su kalitesi değlerlendirmesinde incelenen parametreler ve genel şartlar:

pH

pH, suyun hidrojen iyonu konsantrasyonunu ifade eder. Daha spesifik olarak, pH, sulu bir çözeltideki hidrojen iyonu (H^+) konsantrasyonunun (mol/L) negatif logaritmasıdır (Oğuz ve diğerleri, 2019; Suter ve diğerleri, 2022) (Eşitlik 1.1)

$$pH = -\log_{10}(H^+) \quad (1.1)$$

pH, sudaki çoğu kimyasal ve biyolojik süreci etkiler. Bu nedenle, pH sucul habitatlarda tür dağılımlarını sınırlayan en önemli çevresel faktörlerden biridir. EPA, tatlı suda pH değerini 6,5 ile 9 aralığında önermektedir (Suter ve diğerleri, 2022). TS-266'ya göre pH değeri için bu aralık insanî tüketim amaçlı sular için 6,5 ile 9,5 arasındadır (TSE, 2005).

Sıcaklık

Sıcaklık, su gibi bir maddedeki termal enerjinin yoğunluğunu ifade eder. Akarsu ve nehirlerdeki sıcaklık, birçok atmosferik ve hidrolojik süreçten etkilenecek ısı hareketini etkiler. Sıcaklık, akarsuların temel sınıflandırmasında kullanılır (Sappington ve Norton, 2022).

Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde içme suyu olarak kullanılabilen yüksek kaliteli suların sıcaklığının $25^{\circ}C$ altında olması önerilmektedir (Orman ve Su İşleri Bakanlığı, 2015).

İletkenlik

İletkenlik, suyun elektrik akımını iletebilme yeteneğinin bir ölçüsüdür. Suyun iletkenliği, Klorür, Nitrat, Sülfat ve Fosfat gibi negatif yüklü anyonlar veya Sodyum, Magnezyum, Kalsiyum, Demir ve Alüminyum gibi pozitif yüklü katyonların varlığından etkilenir. İletkenlik, mikrosiemens ($\mu S/cm$) veya mikromhos ($\mu mho/cm$) cinsinden santimetre başına ölçülür. Damıtılmış suyun iletkenliği genellikle 0,5 ile 3 $\mu S/cm$ aralığındadır. İç tatlı sularla ilgili araştırmalar, balıkçılığı destekleyen akarsuların 150 ile 500 $\mu S/cm$ arasında bir iletkenliğe sahip olduğunu göstermektedir. Bu aralığın dışındaki iletkenlik, belirli balık türleri veya makro omurgasızlar için uygun olmayabilir. Endüstriyel sularda iletkenlik,

10000 $\mu\text{S}/\text{cm}$ 'ye kadar yüksek olabilir (EPA, 2012; Pal ve diğerleri, 2015). TS-266'ya göre iletkenlik insanî tüketim amaçlı sular için maksimum 2500 nS/cm olmalıdır (TSE, 2005).

Renk

İdeal olarak, içme suyu görünür bir renge sahip olmamalıdır (Davies-Colley ve diğerleri, 2003; Ritchie ve diğerleri, 2003; WHO, 2006; Gardner ve diğerleri, 2021). WHO'ya göre, renk seviyesinin genellikle tüketiciler tarafından kabul edilen düzeyin 15 TCU'nun altında olması gerektiği belirtilmektedir (WHO, 2006). TS-266'ya göre suyun rengi insanî tüketim amaçlı sular için maksimum 20 mg/L Pt-Co skalasında olmalıdır (TSE, 2005).

1.1.2. Su kalitesi oksijenlendirme parametreleri

Su kalitesi değerlendirmesinde incelenen oksijen seviyesi ile ilişkili parametreler:

Çözünmüş oksijen (ÇO)

Çözünmüş Oksijen (ÇO), suyun içerisinde bulunan oksijen gazının konsantrasyonunu ifade eder. Oksijen, türbülans etkisiyle atmosferden doğrudan absorpsiyon yoluyla suya geçer. Ayrıca, sucul bitkilerin fotosentezi sırasında saldığı oksijen de su tarafından emilir. Yeterli miktarda ÇO, aerobik su yaşamının büyümesi ve çoğalması için önemlidir. Çözünmüş Oksijen miktarı suyun sıcaklığı, besin maddeleri, sedimentler ve Amonyak gibi faktörlerden etkilenir. Yüksek sıcaklık, oksijenin su içindeki çözünürlüğünü azaltır, yani ılık su, soğuk suya göre daha az miktarda ÇO tutar. Yüksek besin maddeleri ise aşırı bitki büyümesine ve sonucunda solunum ve ayrışma nedeniyle ÇO seviyesinin düşmesine yol açabilir. Gömülü çökeller, ÇO'nun interstisyel alanlara nüfuz etmesini engelleyebilir. Amonyak oksitlenmesi (Nitrifikasyon), oksijen tüketir ve düşük oksijen seviyeleri, Nitrifikasyon'u engelleyerek Amonyak seviyelerini artırır (Sánchez ve diğerleri, 2007; Marcy ve diğerleri, 2021). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde içme suyu olarak kullanılabilecek yüksek kaliteli suların ÇO seviyesi en az 6 mg/L ve en uygun ÇO seviyesi ise 8 mg/L 'den fazla olması belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Kimyasal oksijen ihtiyacı (KOİ)

Kimyasal Oksijen İhtiyacı (KOİ), suda bulunan organik maddeleri oksitlemek için gereken ÇO miktarını ifade eder. Yüksek KOİ seviyesi, suyun içindeki ÇO'yu tüketebilir ve olumsuz çevresel etkilere yol açabilir (Lee ve diğerleri, 2016; Geerdink ve diğerleri, 2017). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde içme suyu olarak kullanılabilecek yüksek kaliteli suların KOİ'si " $50 \geq KOİ < 25$ " olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Biyolojik oksijen ihtiyacı (BOİ-5)

Biyolojik Oksijen İhtiyacı (BOİ), akarsularda organik maddelerin mikroorganizmalar tarafından ayrıştırılırken tüketilen oksijen miktarını ifade eder. BOİ aynı zamanda inorganik maddelerin kimyasal oksidasyonunu da ölçer. Bu, suyun içinden oksijenin kimyasal reaksiyonla çıkarılmasını sağlar. Genellikle 20 °C'de 5 gün süresince mikroorganizmalar tarafından tüketilen oksijen miktarını ölçmek için bir test kullanılır. BOİ doğrudan nehirlerde ve akarsularda ÇO miktarını etkiler. BOİ değeri ne kadar yüksekse, akıştaki oksijen tükenmesi o kadar hızlı olur. Bu da sucul yaşam formları için daha az oksijenin bulunduğu anlamına gelir. Yüksek BOİ sonuçları, düşük ÇO seviyeleriyle aynı sonuçları doğurur. Suda yaşayan organizmalar strese girer, boğulur ve ölür. Ardından, anaerobik mikroorganizmaların faaliyeti artar ve sonunda akarsu ölü bir bataklığa dönüşür (EPA, 2012; Lee ve diğerleri, 2016). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde içme suyu olarak kullanılabilecek yüksek kaliteli suların BOİ'si $8 \geq BOİ < 4$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

1.1.3. Su kalitesi besin madde parametreleri

Su kalitesi değerlendirmesinde suda bulunan besin madde miktarı ile ilgili parametreler :

Besinler, bitki büyümesi için gereken elementlerdir. Azot (N), Fosfor (P), Potasyum (K), Kalsiyum (Ca), Magnezyum (Mg), Sülfür (S) ve Silikon (Si) gibi elementleri içerirler. Azot, Fosfor ve Potasyum, birincil besinler olarak kabul edilir. Azot ve Fosfor, çoğu su ortamında ana sınırlayıcı besinlerdir. Besin maddelerinin sucul topluluklar üzerinde dolaylı olumsuz

etkileri, birincil üretim ve bitki/alg biyokütlesinin büyümesi ve birikimi üzerindeki etkileri yoluyla ortaya çıkar (Kurtz ve diğerleri, 2021).

Amonyum azotu (NH_4^+)

Amonyak terimi, iyonize olmayan (NH_3) ve iyonize (NH_4^+) türleri içerir. Çevredeki Amonyak metabolik, tarımsal ve endüstriyel işlemlerden ve Kloramin ile dezenfeksiyondan kaynaklanır. Amonyum Azotu'nun seviyesinin yüksek bulunduğu bölgeler genelde çiftlik hayvanlarının yoğun yetiştirildiği bölgelerdir. Sudaki Amonyak olası bakteri, kanalizasyon ve hayvan atıkları kirliliğinin bir göstergesidir. İçme suyundaki Amonyak'ın sağlık üzerinde olumsuz etkisi yoktur ve bu nedenle sağlık temelli bir kılavuz değeri önerilmemektedir (Johnston, 1991; WHO, 2006).

TS-266'ya göre sudaki Amonyum Azotu miktarı insanî tüketim amaçlı sular için 0,50 mg/L olmalıdır (TSE, 2005). Bu miktar Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde içme suyu olarak kullanılabilecek yüksek kaliteli sularda 0,20 mg/L'den az ve en fazla 1,00 mg/L olması gerektiği belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Nitrat ve nitrit azotu (NO_3^- & NO_2^-)

Nitrat ve Nitrit Azot döngüsünün bir parçası olan doğal olarak oluşan iyonlardır. Nitrat esas olarak inorganik gübrelerde kullanılır. Yeraltı ve yüzey sularındaki Nitrat konsantrasyonu normalde düşüktür, ancak Amonyak ve benzeri kaynakların oksidasyonunun bir sonucu olarak tarım alanlarından sızma, akış, insan veya hayvan atıklarından kaynaklanan kontaminasyonun bir sonucu olarak yüksek seviyelere ulaşabilir. Anaerobik koşullar Nitrit'in oluşumuna ve kalıcılığına neden olabilir (Kroupova ve diğerleri, 2005; WHO, 2006; Tan ve diğerleri, 2022). WHO'ya göre Nitrat'ın kılavuz değeri 50 mg/L'yi geçmemelidir. Bu değer Nitrit için 3 mg/L'dir (WHO, 2006). Bu miktar Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde içme suyu olarak kullanılabilecek yüksek kaliteli sularda Nitrat miktarı $10 \geq \text{Nitrat} < 5$ olarak belirtilmiştir. Nitrit için sınır değeri ise Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'ne göre $0,06 \geq \text{Nitrit} < 0,01$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Fosfor (P)

Fosfor, Çözünmüş Ortofosfat (PO_4^{3-}), çeşitli Organik Fosfor bileşikleri ve Partikülat Fosfor olarak bulunur. Su sistemlerindeki toplam Fosfor'un yaklaşık %25'inin biyolojik olarak kullanılabilir olduğu kabul edilir. Fosfor için en yaygın kullanılan ölçüm toplam Fosfor'dur (Parry, 1998; Kurtz ve diğerleri, 2021). Yapılan çalışmada Devlet Su İşleri Genel Müdürlüğü (DSİ) tarafından sağlanan veriler içerisinde ölçülmüş olan Fosfor türü Ortofosfat olduğu için bu bileşiğin sınır değeri bu tez çalışmasında kullanılmıştır.

Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'ne 2016 yılında ilave edilen Ek. 5'e göre Ortofosfat Fosforu ($PO_4\text{-P/L}$) konsantrasyonu içme suyu olarak kullanılabilir yüksek kaliteli sularda $0,16 \geq$ Fosfat $< 0,05$ olarak belirtilmiştir. Bu miktar TS-266'ya göre en fazla 0,4 mg/L olmalıdır.

1.1.4. Su kalitesi belirtici iz elementler (Metaller) ve inorganik kirleticiler

İnsan faaliyetleri doğal olarak metal bakımından zengin olmayan alanlarda metalleri yeniden dağıtır ve konsantre eder. Metaller havaya, suya ve toprağa salındıklarında su kütlelerine ulaşabilirler. Bazı metaller besin olarak gerekliyken, tüm metaller bir düzeyde toksik olabilir. Metaller suda yaşayan organizmaların hayatta kalmasını, üremesini ve davranışını etkileyen toksik konsantrasyonlarda bulunduklarında bozulmalar ortaya çıkar (Shallenberger, 1993; Shaw-Allen ve Suter, 2021).

WHO İçme Amaçlı su Kalitesi Klavuzu, TS-266, ve Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde bir havza yönetimi için dikkat edilmesi gereken ve havza su kalitesinin belirlenmesinde bir çok iz elementler listesi bulunmasına rağmen, bu çalışmada DSİ tarafından sağlanan Büyük Menderes Havzası su kalitesi verileri içinde bir çok iz elementler ve inorganik bileşiklere ilişkin veri yoktur. Bu tez çalışmasında veri setinde var olan kurum tarafından ölçülen iz elementler ve inorganik maddelerin kirletici olarak çevreye ve insan sağlığı üzerinde nasıl etkisi olabileceği incelenmiştir.

Klorür (Cl)

İçme suyundaki Klorür, doğal kaynaklardan, kanalizasyon ve endüstriyel atık sularından, buz çözücü tuz içeren kentsel akıştan ve tuzlu su girişinden kaynaklanır. İnsanların Klorür'e maruz kalmasının ana kaynağı, yiyeceklere eklenen tuzdur ve bu kaynaktan alınan miktar genellikle içme suyundan alınan miktarın çok üzerindedir. WHO'ya göre, içme suyundaki Klorür için insan sağlığını etkileyen bir kılavuz değeri önerilmemektedir. Bununla birlikte, yaklaşık 250 mg/L'yi aşan Klorür konsantrasyonu suya belirgin bir tat verebilir. Ayrıca, içme suyunda serbest Klorür miktarının 2-3 ppm'yi aşmaması gerektiği belirtilmektedir (WHO, 2006; Qin ve diğerleri, 2018; Pérez ve diğerleri, 2022; Mekki ve diğerleri, 2023). TS-266'ya göre sudaki Klorür miktarı insanî tüketim amaçlı sular için WHO'nun önerdiği miktar olmak üzere 250 mg/L olmalıdır. Bunun yanı sıra Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde Klorür değil Serbest Klor için bir klavuz değeri belirlenmiştir. Bu değer içme suyu olarak kullanılabilecek yüksek kaliteli sularda 10 µg/L olmalıdır. Fakat DSİ tarafından sağlanan verilerde Serbest Klor değil Klorür ölçülmüştür. Bu nedenle çalışmamızda Klorür standart değeri göz önünde bulundurularak Büyük Menderes Havzası su kalitesi değerlendirilmesinde kullanılmıştır.

Sülfat (SO₄)

Sülfatlar birçok mineralde doğal olarak bulunur ve ticari olarak özellikle kimya endüstrisinde kullanılır. Sülfatlar endüstriyel atıklar ve atmosferik çökeltme yoluyla suya karışır. Sülfatın en yüksek seviyeleri genellikle yeraltı sularında görülür. Mevcut veriler, içme suyunda insan sağlığı üzerinde olumsuz etkileri muhtemel olan bir Sülfat seviyesi tanımlamamaktadır (WHO, 2006; EPA, 2022; Barral ve diğerleri, 2022). WHO'ya göre Sülfat konsantrasyonları 500 mg/L'yi aşan içme suyu kaynakları hakkında sağlık yetkililerinin bilgilendirilmesi tavsiye edilmiştir (WHO, 2006). TS-266'ya göre sudaki Sülfat miktarı insanî tüketim amaçlı sular için 250 mg/L olmalıdır (TSE, 2005).

Demir (Fe)

Demir, yerkabuğunda en bol bulunan metallere biridir. Doğal tatlı sularda, demir seviyeleri genellikle 0,5 ile 50 mg/L arasında değişir. Demir, insan beslenmesinde önemli bir role sahiptir (WHO, 2006; Rahman ve Gagnon, 2014; Li ve diğerleri, 2016; Li ve

diğerleri, 2019). WHO'nun 1993 Kılavuzunda içme suyundaki Demir için İnsan Sağlığı üzerinde etki yaratan bir değer önerilmemektedir. WHO'nun 1993 kılavuzunda vücutta aşırı Demir depolanmaya karşı bir önlem olarak yaklaşık 2 mg/L'lik bir değerin aşmaması gereğinden bahsedilmektedir (WHO, 2006).

Manganez (Mn)

Manganez, yerkabuğunda en bol bulunan metallerden biridir ve genellikle Demir ile birlikte bulunur. Esas olarak Manganez, Demir ve çelik alaşımlarının imalatında, Potasyum permanganat olarak temizleme, ağartma ve dezenfeksiyon için oksidan ve çeşitli ürünlerde bileşen olarak kullanılır. (Mazloomi ve diğerleri, 2009). WHO İçme Amaçlı su Kalitesi Klavuzunda Manganez'in sınır miktarı 0,4 mg/L olarak belirtilmiştir (WHO, 2006). Türkiye Standartları TS-266'ya göre sudaki Manganez miktarı insanî tüketim amaçlı sular için 50 µg/L (TSE, 2005) olmalıdır. Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde bu miktar $0,5 \geq \text{Manganez} < 0,1$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Sodyum (Na⁺)

Sodyum tuzları, hemen hemen tüm yiyeceklerde (günlük maruziyetin ana kaynağı) ve içme suyunda bulunur (Bauder ve diğerleri, 2008). Sodyum tuzlarının insan sağlığı üzerindeki etkilerine yönelik herhangi bir kılavuz değeri önerilmemektedir. Ancak, suda bulunan Sodyum konsantrasyonu 200 mg/L'yi aştığında kabul edilemez bir tat oluşabilir (WHO, 2006; Vasanthavigar ve diğerleri, 2010). Bu miktar, TS-266'ya göre insanî tüketim amaçlı sular için aynı olmak üzere 200 mg/L dir (TSE, 2005).

Toplam çözünmüş katı madde (TÇK)

TÇK, inorganik tuzları (esas olarak Kalsiyum, Magnezyum, Potasyum, Sodyum, bikarbonatlar, Klorür'ler ve Sülfat'lar) ve suda çözünen az miktarda organik madde içerir. İçme suyundaki TÇK doğal kaynaklardan, kanalizasyondan, kentsel akış ve endüstriyel atık sulardan gelmektedir (Bauder ve diğerleri, 2008; Mazloomi ve diğerleri, 2009). WHO'ya göre TÇK'nın içme suyu ile alınması durumunda olası sağlık etkilerine ilişkin güvenilir veriler mevcut değildir ve insan sağlığı üzerinde etki yaratan bir kılavuz değeri

önerilmemektedir. Ancak, WHO'nun 1984'te yayınlanan İçme Suyu Kalitesi Kılavuzu'nun ilk baskısında, suyun tadı ile ilgili hususlara dayalı olarak TÇK için belirlenen değer 1000 mg/L olarak önerilmektedir (WHO, 2006).

1.1.5. Diğer parametreler

DSİ tarafından sağlanan Büyük Menderes Havzası su kalitesi verileri arasında bulunmayan yada değerlendirilmeyecek kadar oldukça az sayıda ölçülmüş fakat su kalitesi değerlendirmesi açısından dikkate alınması gereken parametreler:

Alüminyum (Al)

Alüminyum, en bol bulunan metalik elementlerden biridir ve yerkabuğunun yaklaşık %8'ini oluşturur (WHO, 2006; DeForest ve diğerleri, 2018). İçme suyuna katkısı göz önüne alındığında, toplam Alüminyum alımının %5'ten daha az olduğu belirtilmektedir. Alüminyum için WHO'nun 1984 yılında yayınlanan İçme Suyu Kalitesi Kılavuzu'nda 0,2 mg/L'lik bir kılavuz değer belirlenmiştir (WHO, 2006). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde yüksek kaliteli sularda Alüminyum konsantrasyonunun 0,3 mg/L'den az olması önerilmektedir (Orman ve Su İşleri Bakanlığı, 2015). Bu miktar TS-266'ya göre insanî tüketim amaçlı sular için 200 µg/L dir (TSE, 2005).

Arsenik (As)

Arsenik, genellikle Arsenik Sülfür'ü, metal Arsenat'ları ve Arsenit'ler şeklinde yerkabuğunda yaygın olarak dağılmıştır. Arsenik'e maruz kalan kişilerin yanı sıra, ağız yoluyla yiyecek ve içecek alımı Arsenik'e en önemli maruziyet yoludur. WHO tarafından yayınlanan İçme Suyu Kalitesi Kılavuzu'nda Arsenik için 0,01 mg/L'lik bir sınır kılavuz değeri belirlenmiştir (Nickson ve diğerleri, 2005; WHO, 2006). Bu miktar TS-266'ya göre insanî tüketim amaçlı sular için 10 µg/L belirlenmiştir (TSE, 2005). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde yüksek kaliteli sularda Arsenik konsantrasyonunun $0,05 \geq \text{Arsenik} < 0,02$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Nikel (Ni)

Nikel, başta paslanmaz çelik ve Nikel alaşımlarının üretiminde kullanılmaktadır. WHO tarafından yayınlanan İçme Suyu Kalitesi Kılavuzu'nda Nikel için 0,07 mg/L'lik bir sınır kılavuz değeri belirlenmiştir (WHO, 2006; WHO, 2021). Bu değer TS-266'ya göre insanî tüketim amaçlı sular için 20 µg/L olarak belirlenmiştir (TSE, 2005). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde yüksek kaliteli sularda Nikel konsantrasyonu $0,05 \geq \text{Nikel} < 0,02$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Kadmiyum (Cd^{2+})

Kadmiyum, çelik endüstrisinde ve plastik üretiminde kullanılmaktadır. Ayrıca, pil bileşiklerinde yaygın olarak kullanılmaktadır. Kadmiyum, atık sularla çevreye salınır ve gübrelerden ve yerel hava kirliliğinden kaynaklanan yaygın kirlilikle ilişkilidir (WHO, 2006; De Vries ve McLaughlin, 2013). WHO tarafından yayınlanan İçme Suyu Kalitesi Kılavuzu'nda Kadmiyum için 0,003 mg/L'lik bir sınır kılavuz değeri belirlenmiştir (WHO, 2006). Bu değer TS-266'ya göre insanî tüketim amaçlı sular için 5 µg/L olarak belirlenmiştir (TSE, 2005). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde yüksek kaliteli sularda Kadmiyum konsantrasyonu $0,005 \geq \text{Kadmiyum} < 0,002$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Kurşun (Pb)

Kurşun, özellikle Kurşun-asit piller, lehim ve alaşımların üretiminde ve petrol endüstrisinde kullanılmaktadır. İçme suyu, Kurşun alımının en büyük kaynağıdır (Jarvis ve diğerleri, 2018; Roy ve Edwards, 2019). WHO tarafından yayınlanan İçme Suyu Kalitesi Kılavuzu'nda Kurşun için 0,01 mg/L'lik bir sınır kılavuz değeri belirlenmiştir (WHO, 2006). Bu değer TS-266'ya göre insanî tüketim amaçlı sular için aynı olmak üzere 10 µg/L olarak belirlenmiştir (TSE, 2005). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde yüksek kaliteli sularda Kurşun konsantrasyonu $0,02 \geq \text{Kurşun} < 0,01$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Cıva (Hg^{2+})

Cıva, özellikle Klor'un elektrolitik üretiminde, elektrikli cihazlarda, diş amalgamlarında ve çeşitli cıva bileşiklerinin hammaddesi olarak kullanılmaktadır. Kirlenmemiş içme suyunda, hemen hemen tüm cıvanın Hg^{2+} formunda olduğu düşünülse de, inorganik cıvanın tatlı suda ve deniz suyunda metilasyonunun gerçekleştiği gösterilmiştir (Walters ve diğerleri, 2011; Rodrigues ve diğerleri, 2013). WHO tarafından yayınlanan İçme Suyu Kalitesi Kılavuzu'nda, inorganik cıva için 0,006 mg/L'lik bir sınır kılavuz değeri belirlenmiştir (WHO, 2006). Bu değer TS-266'ya göre insanî tüketim amaçlı sular için 1 μ g/L olarak belirlenmiştir (TSE, 2005). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde yüksek kaliteli sularda Kurşun konsantrasyonu $0,0005 \geq \text{cıva} < 0,0001$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Siyanür

Siyanürler, özellikle bazı gelişmekte olan ülkelerde, belirli gıdalarda bulunabilir ve genellikle endüstriyel kontaminasyonun bir sonucu olarak ara sıra içme suyunda tespit edilirler (Nyamunda, 2017). WHO tarafından yayınlanan İçme Suyu Kalitesi Kılavuzu'nda, siyanür için 0.07 mg/L'lik bir sınır kılavuz değeri belirlenmiştir (WHO, 2006). Bu değer TS-266'ya göre insanî tüketim amaçlı sular için 50 μ g/L olarak belirlenmiştir (TSE, 2005). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde yüksek kaliteli sularda siyanür konsantrasyonu $0,05 \geq \text{siyanür} < 0,01$ olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Toplam koliform

Toplam Koliform bakteriler, 24 saat içinde 35–37 °C'de laktoz fermentasyonu ve asit veya aldehit üretimi ile nispeten yüksek konsantrasyonlarda safra tuzlarının bulunduğu büyüeyebilen, çok çeşitli aerobik ve fakültatif olarak anaerobik, gram negatif, spor oluşturmeyen basilleri içerir (WHO, 2006). Koliform bakterilerin *Escherichia*, *Citrobacter*, *Klebsiella* ve *Enterobacter* cinslerine ait olduğu kabul edilir (Divya ve Solomon, 2016). Toplam Koliform grubu hem dışkı hem de çevresel türleri içermektedir (WHO, 2006). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi

Yönetimi Yönetmeliği'nde yüksek kaliteli sularda Toplam Koliform konsantrasyonu 20,000 $\geq$ Toplam Koliform < 100 olarak belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015).

Belirtilen su kalitesi parametrelerinin WHO tarafından yayınlanan İçme Suyu Kalitesi Kılavuzu, Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği, ve TS-266'ya göre insanî tüketim amaçlı sular için belirtilen kılavuz değerleri Çizelge 1'de verilmiştir. Bu parametrelerin bazıları DSİ tarafından sağlanan veriler arasında ya bulunmadığı yada bir havza su kalitesi değerlendirilmesinde kullanılamayacak kadar düzensiz ölçüldüğü için bu tez çalışmasında kullanılan veriler arasında bulundurulmamıştır. Ancak, bu parametreler bir akarsuyun kalitesi değerlendirilmesinde göz önünde bulundurulması gerektiğinden ismi geçen parametreler açıklanmıştır.

Çizelge 1.1. Su kalitesi parametreleri sınır değerleri

Sayı	Parametre	Birim	WHO (içme suyu kılavuzu)	TS-266 (insanî tüketim amaçlı sular kılavuzu)	Orman ve Su İşleri Bakanlığı yönetmeliği (Yüzeysel su kalitesi yönetimi yönetmeliği)
1	pH	---	---	6,5-9,5	6 -9
2	Sıcaklık	°C	---	---	≤ 25
3	İletkenlik	nS/cm	---	2500	$1000 \geq$ yada < 400 $\mu\text{S/cm}$
4	Renk	mg/L	---	20	---
5	ÇO	mg/L	---	---	$6 \leq$ yada > 8
6	KOİ	mg/L	---	---	$50 \geq$ yada < 25
7	BOİ	mg/L	---	---	$8 \geq$ yada < 4
8	Amonyum	mg/L	---	0,50	$1,00 \geq$ yada $< 0,2$
9	Nitrat	mg/L	50	---	$10 \geq$ yada < 5
10	Nitrit	mg/L	3	---	$0,06 \geq$ yada $< 0,01$
11	Fosfor	mg/L	---	0,4	$0,16 \geq$ yada $< 0,05$
12	Klorür	mg/L	250	250	---
13	Sülfat	mg/L	500	250	---
14	Demir	mg/L	2	---	$1000 \geq$ yada < 300
15	Mangan	mg/L	0,4	0,05	$0,5 \geq$ yada $< 0,1$
16	Sodyum	mg/L	200	200	---
17	TÇK	mg/L	1000	---	---
18	Alüminyum	mg/L	0,2	0,2	$\leq 0,3$

Çizelge 1.2. (devam) Su kalitesi parametreleri sınır değerleri

19	Arsenic	mg/L	0,01	0,01	$0,05 \geq \text{yada} < 0,02$
20	Nikel	mg/L	0,07	0,02	$0,05 \geq \text{yada} < 0,02$
21	Kadmiyum	mg/L	0,003	0,005	$0,005 \geq \text{yada} < 0,002$
22	Kurşun	mg/L	0,01	0,01	$0,02 \geq \text{yada} < 0,01$
23	Cıva	mg/L	0,006	0,001	$0,0005 \geq \text{yada} < 0,0001$
24	Siyanür	mg/L	0,07	0,05	$0,05 \geq \text{yada} < 0,01$
25	Toplam koliform	membran	---	---	$20,000 \geq \text{yada} < 100$

1.2. Sınıflandırma ve Tahmin

1.2.1. Su kalitesi sınıflandırması

Son yıllarda, gelişmekte olan ülkeler su temini ve sanitasyonu iyileştirmeye çalışırken su kalitesini korumada önemli sorunlarla karşılaşmışlardır. Gelişmiş ülkeler besin zenginleştirme ve su kaynaklarının ötrofikasyonu ve artan nüfusa su ve atık su hizmetlerinin sağlanması gibi sorunlar karşısında su kalitelerinin durumunu korumak veya iyileştirmek için mücadele etmektedirler (Uddin ve diğerleri, 2021).

Bir havzanın su kalitesini korumak için en önemli gereksinim, havzadaki suyun kalitesi hakkında yeterli bilgiye sahip olmaktır. Etkili bir su havzası yönetimi için, havzadaki su kalitesinin sürekli olarak izlenmesi gerekmekte ve su kalitesi standart değerlerin altına düştüğünde veya sınır değerlere yaklaştığında, kaliteyi iyileştirmek için ciddi kararlar alınmalıdır.

Su kalitesi endeksi (Water Quality Index (WQI))

Su kalitesinin yönetimi, değerlendirilmesi ve sentezlenmesi büyük su kalitesi veri kümelerinin toplanması ve analizi açısından zor olabilir (Poonam ve diğerleri, 2013). Birçok ülkenin su kalitesi değerlendirmek için yönetmeliği ve farklı parametreler için belirlenen kılavuz değerleri olmasına rağmen, bir su kütlesinin genel değerlendirilmesi zordur. Örneğin, Orman ve Su İşleri Bakanlığı Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde bulunan her bir parametre değerine göre belli bir su kalitesi sınıfında yer almaktadır. Bir havzadaki suların genel kalitesine tek bir parametre ile karar vermek doğru bir

değerlendirme olmayabilir. Bir havzanın su kalitesini belirlemek için tespit edilen parametrenin ölçüm değerleri veri setinde bulunabilir. Bu parametrelerin hangi su kalitesi sınıfına ait olduğuna Orman ve Su İşleri Bakanlığı Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'ne göre incelendiğinde üç parametrenin üçüncü sınıf su kalitesi, beş parametrenin birinci sınıf su kalitesi ve iki parametrenin dördüncü sınıf su kalitesine ait olduğu değerlendirildiğinde, bu havzanın genel su kalitesi hangi sınıfa girdiği konusunda karışıklıklar oluşmaktadır.

Geleneksel ve sadece yönetmeliklere dayanarak bir havza yada her hangi bir su kütlesinin genel su kalitesini belirlemenin zor olduğu görülmektedir (Uddin ve diğerleri, 2021).

Su kalitesi sınıflandırması için, tüm su kalitesi parametrelerini bir arada değerlendirmek amacıyla birçok araç geliştirilmiştir. WQI (Su Kalite İndeksi), su kütlesinin kalitesini tek bir değerle ifade etmeyi sağlayan bir modeldir. WQI modelleri, zaman ve mekan bazında değişen geniş su kalitesi veri setlerinin analizine olanak sağlayan bir toplama işlevine dayanmaktadır (Alexakis, 2022). WQI modelleme yöntemleri, su yönetimi ve tedarik kurumları için cazip hale gelmiştir, çünkü kullanımları nispeten kolaydır ve karmaşık su kalitesi veri setlerini anlaşılması kolay tek bir su kalitesi ölçüm değerine dönüştürürler (Van ve diğerleri, 2022). WQI sınıflandırma modellemesi su kütlesi kullanım amacı ve bölgeye göre değişiklikler gösterebilmektedir. Dünya'da kullanılan toplam 35 adet WQI sınıflandırma modeli mevcuttur (Poonam ve diğerleri, 2013; Uddin ve diğerleri, 2021).

Horton, 1960 yılında, çoğu su kütlesinde önemli kabul edilen 10 su kalitesi parametresine dayanan ilk WQI modelini geliştirmiştir (Horton, 1965). Brown, Ulusal Sağlık Vakfı (National Sanitation Foundation (NSF)) desteğiyle, 142 su kalitesi uzmanından oluşan bir panelin parametre seçimi ve ağırlıklandırma hakkında bilgi verdiği Horton'un WQI modeli NSF-WQI'nin daha titiz bir versiyonunu geliştirdi (Rown ve diğerleri, 1972). Diğer birçok WQI modeli NSF-WQI'ye dayanarak geliştirilmiştir. 1973'te İskoç Araştırma ve Geliştirme Departmanı (Scottish Research and Development Department (SRDD)), aynı zamanda bir şekilde Brown'ın modeline dayanan SRDD-WQI'yi geliştirdi ve onu nehir suyu kalitesinin değerlendirilmesi için kullanmıştır (Scottish Development Department, 1975). Bascaron Endeksi (Bascaron Index (1979)), Ev Endeksi (Home Index (1986)) ve Dalmatia Endeksi (Dalmatia Index), SRDD-WQI'nin sonraki türevleridir. Bir diğer önemli gelişme ise British Columbia Çevre, Arazi ve Parklar Bakanlığı (British Columbia Ministry of Environment,

Land and Parks (BCMELP)) tarafından 90'lı yılların ortalarında geliştirilen ve Kanada'nın British Columbia eyaletindeki birçok su kütlesinin kalite durumunu değerlendirmek için kullanılan British Columbia WQI (BCWQI)'dir (Zandbergen ve Hall, 1998). Kanada Çevre Bakanlar Konseyi'nin (Canadian Council of Ministers of the Environment (CCME)) Su Kalitesi Yönergeleri Görev Grubu, BCWQI modelinin gözden geçirilmesi ve revizyonunun ardından 2001 yılında CCME WQI'yi geliştirdi (Hurley ve diğerleri, 2012). Son zamanlarda Liou Index, Malezya Endeksi (Malaysia Index) (İbrahim ve Kutty, 2013) ve Almeida Endeksi (Almeida Index) gibi modeller de geliştirilmiştir. WQI modelleri dünyanın birçok yerinde kullanılmıştır. WQI modelleri tüm ana su kütlelerine uygulanmış olsa da, uygulamaların %82'si nehir suyu kalitesini değerlendirmek için olmuştur (Poonam ve diğerleri, 2013; Uddin ve diğerleri, 2021).

1.2.2. Su kalitesi parametreleri tahmini modellemesi

Bir su havzası su kalitesinin belirlenmesi, havzadaki suların hem insani amaçlı tüketim hemde havza ekosistemi korunması açısından çok önemlidir (Ilić ve diğerleri, 2022). Fakat WQI yöntemi ile su kalitesi belirlenmesi birçok su kalitesi parametresi ölçümüne bağlıdır. Bu parametrelerin bazılarının ölçümü zordur. Bazılarının ölçümü pahalı ve bazılarının ise miktarı ölçülebilecek sınırların altında olabilir. Belirtilen nedenlerden dolayı su kalitesi parametreleri tahmin modellemesinin önemi ortaya çıkmaktadır (Asadollah ve diğerleri, 2021). Tahmin modellerinin oluşturulmasındaki amaç gelecekte ele alınan parametrenin değeri için model ortaya koymaktır (Chen ve diğerleri, 2020). Örneğin, ÇO bir su kütlesinde en önemli parametrelerden birisi olarak su ekosistemi hakkında çok önemli bilgiler vermektedir. Sudaki ÇO miktarı belli bir konsantrasyonun altına düşerse sucul canlıların ölümüne ve yok oluşuna yol açmakla kalmaz, anaerobik mikroorganizmaların çoğalması ve baskın olmasına yol açar ve ardından ötrofikasyon ortaya çıkarak göller ötrofik göle ve nehirler bataklıklara dönüşür. Yani ÇO tahmini modeli bir su kütlesinin sağlığı açısından çok önemli olduğu açıkça bellidir. Ayrıca tahmin modelleri kullanılarak daha az zaman harcayarak, daha az maliyetle bir havza suyu hakkında gelecek için oldukça efektif ve başarılı su havzası yönetimi sağlanabilir (Heddam ve Kisi, 2017; Ouma ve diğerleri, 2020). Bu nedenle tahmin modelleri sınıflandırma modellerinin yanında havza yönetimi için kullanılmalıdır.

1.3. İstatistiksel, Makine Öğrenme (Machine Learning (ML)) Sınıflandırma ve Tahmin Modelleri

Sınıflandırma öğeleri, önceden belirlenmiş kategorilere dönüştürme süreci olarak tanımlanır ve bu süreçte bir eğitim veri kümesi kullanılır. ML sınıflandırma işlemi denetimli bir öğrenme algoritması olarak da kabul edilir. Bir sınıflandırma prosedürü, yeni durumlarda bu tür yargıları tekrar tekrar yapmak için genel resmi ortaya koyan yöntemdir. Gerçek sınıfların bulunduğu bir veri kümesinden bir sınıflandırma prosedürünün oluşturulması, çeşitli şekillerde örüntü tanıma, ayırım veya Denetimli Öğrenme (Supervised Learning (SUPL)) olarak da adlandırılmıştır. SUPL, bilimde, sanayide ve ticarete karmaşık ve genellikle çok kapsamlı verileri sınıflandırma veya karar verme aşamalarında kullanılmaktadır (Learned-Miller, 2014). Sınıflandırma görevi için çok çeşitli yaklaşımlar bulunmaktadır. Bunları istatistiksel yaklaşım, ML ve Yapay Sinir Ağları (Artificial Neural Networks (ANNs)) olarak ifade edebiliriz.

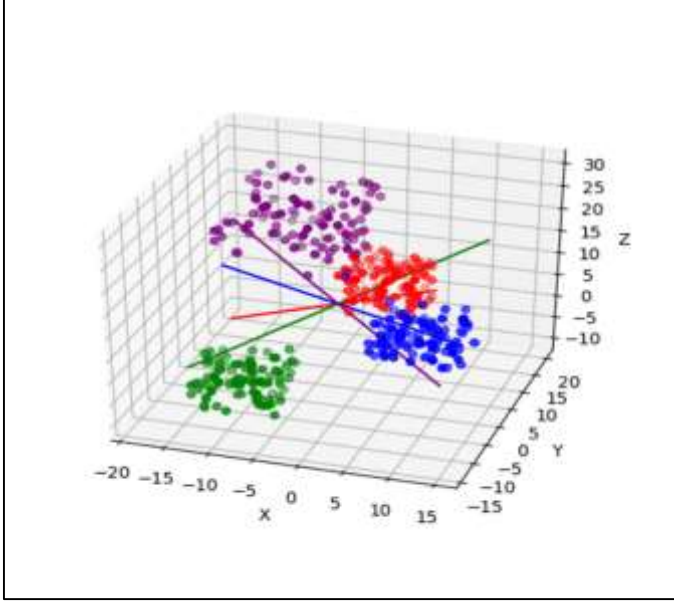
1.3.1. İstatistiksel sınıflandırma yaklaşımı ve modelleri

İstatistik çalışmalarında, sınıflandırma üzerine iki aşama bulunmaktadır. İlki “klasik” aşama, Fisher’in Lineer Ayrımcı’sı (Linear Discriminant (LD)) ve türevleri üzerinde yoğunlaşmıştır. İkincisi, “modern” aşama, daha esnek sınıflandırma modelinden yararlanan, her bir sınıf içindeki özelliklerin ortak dağılımının bir tahmini çalışan yaklaşımdır. İstatistiksel yaklaşımlar genellikle, basit bir sınıflandırmadan ziyade her bir sınıfta bulunma olasılığını sağlayan açık bir temel olasılık modeline göre karakterize edilmektedir. Genellikle, geleneksel teknikler istatistikçiler tarafından kullanılır. Kullanılan tekniklerde değişkenlerin seçimi, dönüşümü gibi yapılandırmalarda bazı değişkenlere müdahaleler yapıldığı varsayılır (Michie ve diğerleri, 1994). Olasılık varsayımı gerektirmeyen Fisher’in LD’sini olabirliğe dayalı yöntemler üç yöntem ile ele alınmaktadır. Bunlar, LD, İkinci dereceden Ayrımcı (Quadratic Discriminant (QD)) ve Lojistik Ayrımcı’dır (Logistics Discriminant (LGD)) dır.

Doğrusal ayırma analizi (Linear Discriminant Analysis (LDA))

Yüksek boyutlu verileri düşük boyutlu bir uzaya yansıtan boyutsallık azaltma ve sınıflandırma için iyi bilinen bir yöntemdir. Buradan verilerin maksimum sınıf ayrılabilirliği

elde edilmektedir. LDA'daki türetilmiş özellikler, katsayıların dönüşüm matrisinden olan orijinal özelliklerin doğrusal kombinasyonlarıdır. Klasik LDA'da optimal projeksiyon, sınıf içi mesafeyi en aza indirmekte ve sınıflar arası mesafeyi de aynı anda en üst düzeye çıkarmaktadır. Böylece maksimum sınıf ayrımı sağlanmaktadır (Şekil 1.1) (Ye, 2007).



Şekil 1.1. Üç boyutlu uzayda ve çoklu sınıf ayrımında LD çalışma şeması.

Maksimum olasılık (Maximum Likelihood)'a göre LD

İstatistiksel algoritmaların kullanılmasının gerekçesi, olasılık dağılımlarının dikkate alınması olup LD prosedüründe buna dayanmaktadır. A_i sınıfı örnekleri için öznitelik vektörlerinin bağımsız olduğu ve Olasılık Yoğunluk Fonksiyonu (Probability Density Function (PDF)) f_i ile belirli bir olasılık dağılımını takip ettiği varsayılmaktadır. Daha sonra, PDF, $f_i(x)$ 'in en büyük olduğu sınıfa x öznitelik vektörüne sahip yeni bir nokta olarak atanır. Buna Maximum Likelihood yöntemi denir (Day ve diğerleri, 1967).

İkinci dereceden ayırıcı (Quadratic Discriminant (QD))

İkinci Dereceden Ayırıcı (Quadratic Discriminant (QD)), LD'ye benzer, ancak iki ayırım bölgesi arasındaki sınır ikinci dereceden bir yüzeydir. Eşit kovaryans matrislerin varsayımına bırakıldığında, normal dağılımlarla Maximum Likelihood argümanında ikinci dereceden bir yüzey (Quadratic Space), örneğin, elipsoid, hiperboloid vb. elde edilir. Bu tür

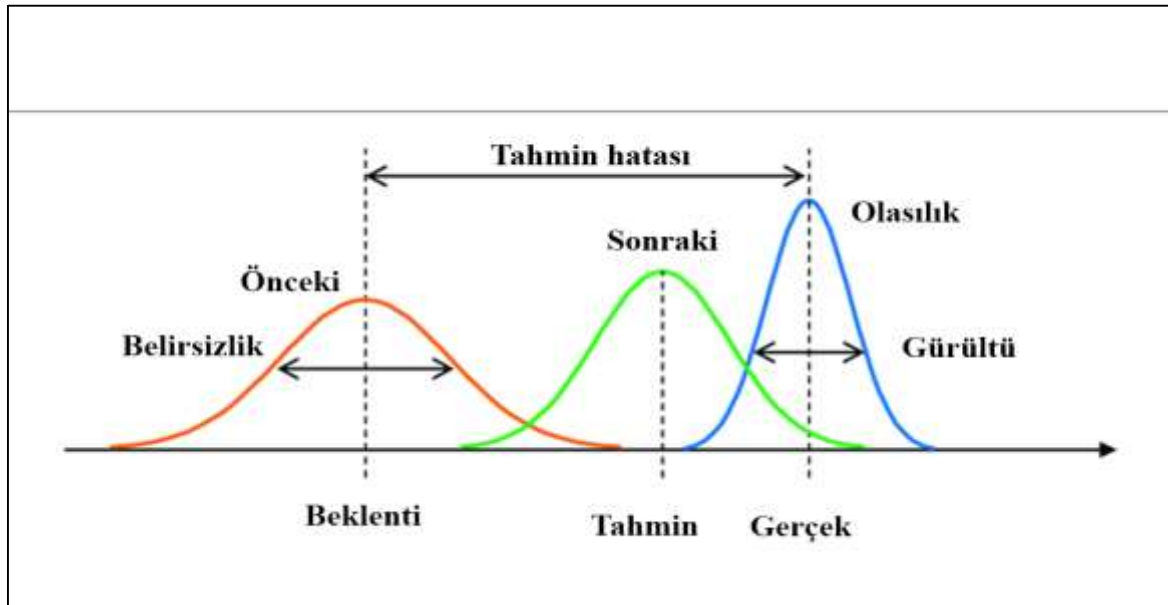
ayırım, bir sınıf için öznitelik değerleri kümesinin bir dereceye kadar diğerini çevrelediği sınıflandırmalarla ilgilenmektedir (Michie ve diğerleri, 1994; Srivastava ve diğerleri, 2007).

Lojistik ayırmacı (Logistic Discriminant (LGD))

Lojistik Ayırmacı (Logistic Discriminant (LGD)), sınıfları ayırmak için bir hiperdüzlem (hyperplane) seçerek çalışır. Ancak iyi bir ayırım için kriterin değiştirilmesi gerekmektedir. Fisher'in LD'si, ikinci dereceden (Quadratic) bir maliyet fonksiyonunu optimize ederken, LGD'de ise maksimize edilen bir koşullu olasılık durumu bulunur. Bununla birlikte, pratikte genellikle ikisi arasında çok az fark vardır ve LD, LGD'ye göre iyi başlangıç değerleri sağlar (Press ve Wilson, 1978; Kurita ve diğerleri, 2009).

Bayes kuralı

Olabilirlik oranlarına (Likelihood Ratios) dayalı yöntemler, eşit olmayan yanlış sınıflandırma maliyetlerini ve/veya eşit olmayan önceki olasılıkları içerecek şekilde uyarlanır. Bayes kuralı, parametrik ve parametrik olmayan durumlarda PDF tahmin edicileri oluşturmak için bir yöntem sağlar (Şekil 1.2) (Barron, 1987; Johnson ve diğerleri, 2022).



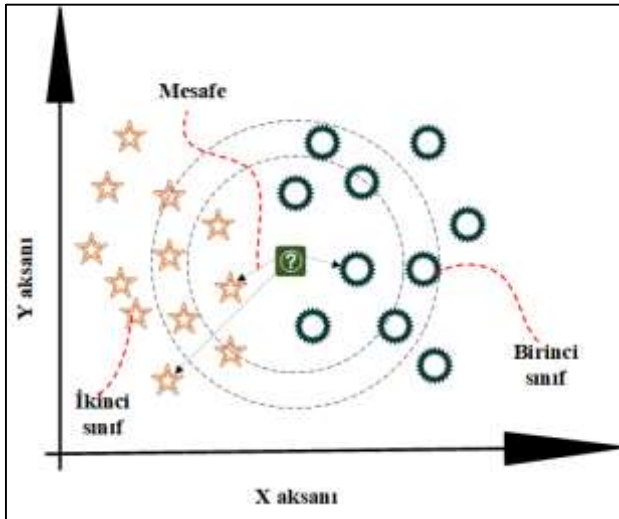
Şekil 1.2. Bayes Kuralı çalışma şeması

Naif bayes (Naive Bayes (NB))

Bayes sınıflandırıcıları, özellik vektörü tarafından tanımlanan belirli bir örneğe en olası sınıfı atamaktadırlar. Bu yöntemde çalışan presedür, her bir özneliliğin her sınıfa ait olma olasılığının, öznelilikler arasında korelasyon yapılmadan bağımsız olarak ele alınmasını sağlar. NB'nin, metin sınıflandırma, tıbbi teşhis ve sistem performans yönetimi gibi birçok pratik uygulamada etkili olduğu görülmektedir (Rish, 2001; Benzer ve diğerleri, 2022).

K-En yakın komşu (K-Nearest Neighbore (KNN))

K-En Yakın Komşu (K-Nearest Neighbor (KNN)), en temel ve basit sınıflandırma yöntemlerinden biridir. KNN, verilerin dağılımı hakkında çok az veya hiç önceden bilgi sahibi olunmadığı durumlarda kullanılan bir sınıflandırma yöntemidir. KNN, parametrik tahminlere güvenmek için yeterli bilginin olmadığı veya belirlemenin zor olduğu durumlarda Ayrım Analizi (Discriminant Analysis (DA)) yapma ihtiyacından ortaya çıkmıştır. KNN hata oranı, Bayes hata oranının iki katıyla sınırlanmıştır (Şekil 1.3) (Peterson, 2009).



Şekil 1.3. KNN'nin çalışma şeması

1.3.2. Makine öğrenme (Machine Learning (ML)) sınıflandırma yaklaşımı ve modelleri

Genel anlamda öğrenme, kişinin bir görevi yapabilme yeteneğini geliştirme süreci olarak tanımlanmaktadır. Öğrenme, çıkarım yapma ve ezberlemenin birleşik süreçleridir. Çıkarım yeni bilginin üretilmesiyle sonuçlanmaktadır (Dutton ve Conroy, 1997).

ML, bilgisayarı akıllı hale getirmenin temel yoludur. ML, insan öğrenme etkinliklerini simüle etmek için bilgisayarların nasıl kullanılacağını, yeni bilgiler ve yeni beceriler elde etmek, mevcut bilgileri tanımlamak, performansı ve başarıyı sürekli iyileştirmek için bilgisayarların kendini geliştirme yöntemlerini inceleyen bir konudur. İnsan öğrenmesi ile karşılaştırıldığında ML daha hızlı öğrenir ve öğrenmenin sonuçlarını daha kolay yayar. Bu nedenle, insanın ML alanındaki ilerlemesi, bilgisayarların kapasitesini artıracak ve böylece insan toplumu üzerinde bir etkisi olacaktır (Wang ve diğerleri, 2009).

Öğrenme, bir dizi farklı soyutlama düzeyinde modellenenebilir. İki tür öğrenme vardır. Bunlardan birisi bilgi edinimi anlamına gelen *bilgi düzeyinde öğrenme* ve hızlandırılmış öğrenme anlamına gelen *sembol düzeyinde öğrenme* (mevcut bilgi daha verimli kullanılır) dir. Genel olarak ML modelleri sembolik (symbolic) veya alt-sembolik (sub-symbolic) yada başka bir deyimle sayısal olarak öğrenme modellerine ayrıştırmaktadır (Dutton ve Conroy, 1997).

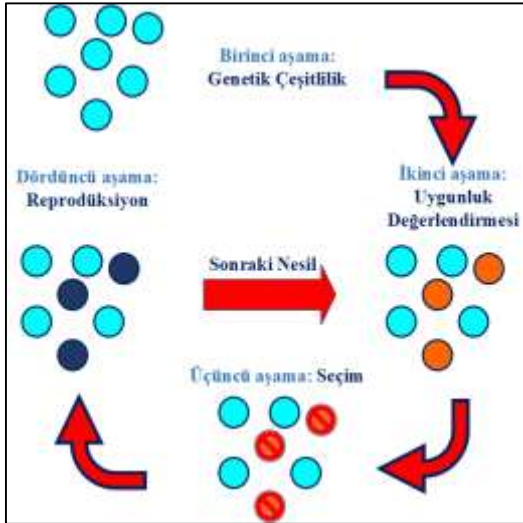
Alt-sembolik öğrenme (Sub-Symbolic Learning (SSL))

Yapay sinir ağları (Artificial Neural Networks (ANNs))

İnsan beyninin öğrenme modelini taklit eden ANN'ler, mantıklı kararlar vermek için karmaşık bilgileri işlemek için üretilmiştir. ANN, yapay nöronlar adı verilen bağlı birimler veya düğümler koleksiyonuna dayanır. ANN, biyolojik bir beyindeki nöronları gevşek bir şekilde modeller. Her bağlantı diğer nöronlara bir sinyal iletebilir. Yapay bir nöron sinyalleri alır, ardından bunları işler ve kendisine bağlı nöronlara sinyal gönderebilir. Bir bağlantıdaki "sinyal" gerçek bir sayıdır ve her bir nöronun çıktısı, girdilerinin toplamının doğrusal olmayan bir fonksiyonu tarafından hesaplanır (Hopfield, 1988).

Genetik algoritmalar (Genetic Algorithms (GA))

Son yıllarda, ekonomi, mühendislik, politika, yönetim ve mühendislik gibi farklı alanlardan kaynaklanan karmaşık problemlerin çözümünde Metasezgisel Algoritmalar (Metaheuristic Algorithms (MA)) kullanılmaktadır. Yoğunlaştırma ve çeşitlendirme, MA'nın temel unsurları olarak sayılmaktadır. Ele alınan problemlerin etkili bir şekilde çözmek için bu unsurlar arasındaki uygun denge kurulmalıdır. MA'lar tek çözüm ve popülasyon tabanlı olmak üzere iki kategoriye ayrılmaktadır. İyi bilinen popülasyon tabanlı MA'lardan bazıları GA olarak tanımlanmaktadır. MA'lar arasında GA, biyolojik evrim sürecinden esinlenen bir algoritmadır. GA, Darwin'in doğada en uygun olanın hayatta kalması teorisi üzerine dayanmaktadır (Şekil 1.4) (Lambora ve diğerleri, 2019; Katoch ve diğerleri, 2021).



Şekil 1.4. Genetik algoritmalar çalışma akış şeması.

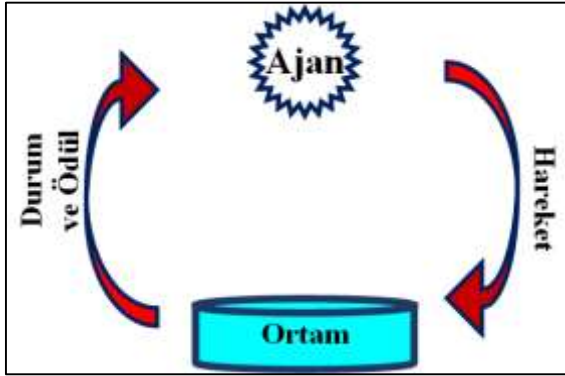
İstatistiksel yaklaşımlara dayalı ML

İstatistiksel yaklaşımlar sınıflandırma çalışmalarının temel yöntemlerinden biridir. ML yöntemleri ile sınıflandırma çalışmalarında istatistiksel ML yaklaşımları kullanılmakta ve önemli rol oynamaktadırlar. İstatistiksel teknikler, genel olarak bir popülasyon hakkında sonuçlar çıkarmanın mümkün olduğu bir örneği tanımlamak için olasılıksal bir modelin oluşturulmasını gerektirir. ML topluluğunun ilgisini çeken, çoğunlukla verilere dayalı hipotezlerin oluşturulduğu endüktif çıkarımla ilgilidir. Burada çıkarım, belirli bir gözlem örneğindeki olasılıklara doğrudan dayanmaktadır. Değişkenler arasındaki ilişkileri

araştırmak için çok sayıda teknik bulunmaktadır. Örneğin regresyon, iki değişken arasındaki ilişkinin doğasını belirlemeye çalışır (Örnek: Bir veri dağılım diyagramına bir eğri sığdırmak). İstatistiksel yöntemler, kullanılan veriler hakkında varsayımlarda bulunmaları bakımından genellikle parametrikdir (Dutton ve Conroy, 1997).

Güçlendirmeli öğrenme (Reinforcement Learning (RL))

Güçlendirmeli Öğrenme (Reinforcement Learning (RL)), doğal ve yapay sistemlerin, eylemlerin onları bir durumdan diğerine yönlendirdiği ve ödüllere yada cezalara yol açabileceği davranışların sonuçlarını tahmin ve optimize etmeyi nasıl öğrenebileceğini incelemektedir. Bu uygulama etoloji, ekonomi, psikoloji ve kontrol teorisi gibi disiplinlerarası konu olarak ortaya çıkar. Matematiksel psikoloji araştırma alanından doğan RL, çözümlerin nitel ve nicel hesaplama düzeyinde modellerin ortaya çıkmasını sağlar. RL çok şeffaf bir şekilde gömülü gibi görülmesi, çeşitli paradigmalar ve sistemlerde toplanan davranışsal ve sinirsel veri zenginliği hakkında hipotezler ve bunların geriye dönük ve tahmine dayalı yorumlarını yapmak için onu çok daha hızlı bir şekilde kullanmayı mümkün kılmaktadır (Şekil 1.5) (Dayan ve Niv, 2008; Sutton ve Barto, 2018).



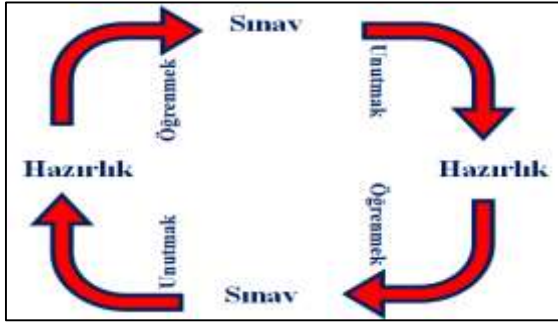
Şekil 1.5. RL çalışma akış şeması

Sembolik öğrenme (Symbolic Learning (SL))

Teorik anlamda gerekli bilgi miktarı ve gerçekleştirilen çıkarım düzeyi farklılık gösteren bir dizi sembolik öğrenme türü bulunmaktadır. SL’de çıkarım düzeyi ne kadar yüksekse, türetilen bilgi potansiyel olarak daha karmaşık ve/veya hacimli hale gelir. SL’nin temel modelleri ezber (Rote), anlatılarak öğrenme, tümdengelim (deduction) ve tümevarım (Induction) dır (Chen, 1995).

Ezberci öğrenme (Rote Learning)

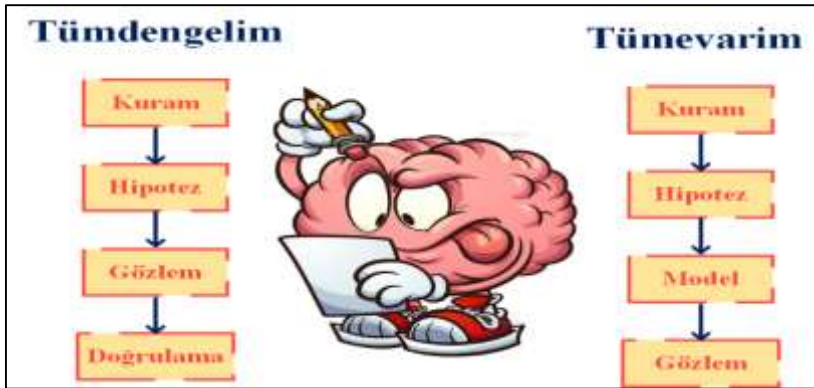
Rote Learning bir hafıza olarak bilgiyi depolar ve gerektiğinde bilgiye kolayca ulaşır. Ezberci (Rote) sistemde bilgiye erişim daha istikrarlı ve yönlendiricidir. Rote Learning prosedürü aşırı işlem gerektirmez. Öğrenme sisteminin uygulama kısmı soyut bir fonksiyondur (Şekil 1.6) (Xue ve Zhu, 2009; Wang ve diğerleri, 2009).



Şekil 1.6. Rote Learning çalışma akış şeması.

Tümdengelim (Deduction)

Tümdengelim, temel bir mantıksal “gerçeği koruma” işlemidir. $P \rightarrow Q$ gibi bir “kural” ve bir P gerçeği verildiğinde, Q’nun da doğru olduğu sonucuna varılabilir. Bu tür kurallar bir teorinin temelini oluşturur. Herhangi bir teori uygulama alanına uygulanabilir ve ideal olarak eksiksiz ve doğru olmalıdır. Tümdengelimli teoriler, zorunlu olarak belirli bir alana göre uyarlanır ve teorinin faydası, mevcut alan giderek amaçlanan alandan uzaklaştıkça azalır. Bu öğrenme türü aynı zamanda Analitik Öğrenme (Analytical Learning) olarak da bilinir (Şekil 1.7) (Xue ve Zhu, 2009).



Şekil 1.7. Tümevarım ve Tümdengelim öğrenme yöntemleri çalışma akışı şeması.

Tümdengelim (Açıklamaya dayalı öğrenme (Explanation-based learning (EBL)))

Yoruma dayalı öğrenmeye Açıklamaya Dayalı Öğrenme (Explanation-based learning (EBL)) denir. EBL, alandaki bilgi ve kavram bilgisinin öğrendiği mevcut örnekleri analiz eder ve cevaplar. Örneğin cevaplama sürecini devreden bir neden-sonuç açıklama ağacını eğitir. Yeni bilgi edinme sürecinde, doğayı, simge olgusunu ve iç ilişkileri açıklayarak yeni bilgiler öğrenir. Sistem öğrenirken, önce eğitim örneklerin kavramının nedenini bulmak, ardından ispatı işlemsel kurallarla genişletmek, son olarak hedef kavramın evrensellik tanımını elde etmektir (DeJong ve Mooney, 1986).

Tümdengelim (Analog Öğrenme (Analog Learning (AL)))

Analoji, nesneler arasındaki benzerliği açık ve net bir şekilde tanımlayabilir. AL, benzer şeyleri karşılaştırarak öğrenmeyi gerçekleştirir. Örneğin, bir profesör, öğrencilere anlaşılması zor olan yeni bir kavramı öğretmek istediğinde, öğrencilerin ustalaştığı bazı örnekler arasında ve yeni kavrama benzer şekilde bir mecaz kullanır. Benzetme yoluyla derin bir anlayış içerir.

AL'nin anahtarı, benzer tanımlama ve benzer dönüşüm yöntemidir. Benzer tanıma dayalı nesne, AL'nin amacı ile değişmektedir. Öğrenmenin amacı yeni problemleri çözmek, yeni sorunların durumu ile eski problemler arasındaki ilişkiye dayanmalıdır (Wang ve diğerleri, 2009; Xue ve Zhu, 2009).

Tümevarım (Induction)

Tümevarım genellikle "Arama Öğrenme (learning as search)" olarak tanımlanır. Öğrenme, özel bilgi temsili olması yanında bir bilgi alanı içinde araştırma olarak tanımlanır. Bu hipotez tüm karmaşık davranışların arama yöntemi olarak nitelendirilebilir. Bu da aramanın buluşsal ve hedefe yönelik olduğunu, yani akılda bir performans görevi olduğunu varsaymaktadır (Mooney, 2000).

Tümevarım denetimsiz öğrenme (Unsupervised Learning (USL))

Denetimsiz tümevarım (Unsupervised Induction (UI)), kavram oluşturma olarak da bilinmektedir. Bir algoritmanın sınıflandırılmamış örneklerle sağlandığını veya elde edildiğini ve hangi örneklerin hangi sınıflara ait olduğunu ve sınıfların ne olduğuna karar vermesi gerektiğini ifade eder. Bu alan yine kümeleme ve keşif olarak alt bölümlere ayrılabilir (Ghahramani, 2004).

Kavramsal kümeleme (Conceptual Clustering (CC))

Kümeler, sınıf içi benzerliği yüksek ve sınıflar arası benzerliği düşük olan örnek gruplarıdır. Sorun, örnekleri kümelemek ve sonra onları karakterize etmektir. Bu da iki aşamalı bir aramaya yol açar. Burada ilk olarak arama bir küme uzayı aracılığıyla ve sonra kavramlar alanı ile yapılır (Pérez-Suárez ve diğerleri, 2019). Kümeleme algoritmaları, örnekler ve/veya kümeler arasındaki farkı veya benzerliği yargılamak için bir mesafe işlevi kullanır (Dutton ve Conroy, 1997).

Keşfederek öğrenme (Learning by discovery)

Bu tür öğrenme, gözlem yoluyla öğrenme (Learning from Observation) olarak da bilinir. Bir algoritma çevresini gözlemler ve kendi başına “kavramlar” oluşturur. Bu sınıftaki algoritmalar ya basitçe çevreyi gözlemleyebilir (Pasif Gözlemlleme) ya da çevrede değişikliklere neden olabilir ve etkileri gözlemleyebilir (Aktif Gözlemlleme) (Svinicki, 1998).

Tümevarım (Denetimli Öğrenme (Supervised Learning (SUPL))

SUPL ön bilgilerden gelen verileri kategorize ettiği için denetimli bir öğrenme tekniğidir. Her test örneğinin sınıfına, özelliklerin birleştirilmesi ve eğitim verilerinden her sınıf için ortak olan kalıpların bulunmasıyla karar verilir. Sınıflandırma iki aşamada yapılır. İlk olarak, eğitim veri setine bir sınıflandırma algoritması uygulanır ve daha sonra model performansını ve doğruluğunu ölçmek için çıkarılan model etiketli bir test veri seti ile doğrulanır. Doğruluğu belirlemek için çeşitli yöntemler mevcuttur. Bunları, yeniden ikame (Substitution,), eğitmek ve test (Train and Test), rastgele örnekleme (Random Sampling),

tekrarlanan rastgele örnekleme (Repeated Random Sampling), K-Kat Çapraz Doğrulama (K-Fold Cross Validation (KFCV)), ve 0.632 önyükleme (Bootstrap) olarak sayabiliriz (Singh ve diğerleri, 2016).

1.4. ML Yaklaşımıyla Tahmin Modelleri

Bir veri hakkında ön bilgi yoksa ve ML algoritmasından bu veriyi bir sınıfa sokulması veya veri değeri tahmin edilmesi istenirse bu tarz öğrenme USL sınıfına girer ve algoritma gözlem yoluyla çevresini gözlemler ve kendi başına “kavramlar” oluşturur. ML yaklaşımıyla tahmin modelleri bu kavram ile çalışır.

1.5. Öz Nitelik Seçimi (Feature Selection) ve Önemi

Veri kümelerindeki önemli bilgileri keşfetmek için akıllı algoritmaların geliştirilmesini gerekir. Veri modelleri, veri madenciliği görevlerine bağlı olarak, genellikle sınıflandırma, regresyon ve kümeleme alanlarında oluşturulur. Veri kümelerinin ön işleminin neden ele alındığını açıklayacak olursak:

- Daha verimli analiz elde etmek için veri kümesinin boyutunun küçültülmesi,
- Veri setinin seçilen analiz yöntemine en uygun şekilde uyarlanmasıdır.

Veri kümesi boyutunu küçültme, özellik kümesini küçültme veya örnek kümesini küçültme olmak üzere iki yoldan biriyle gerçekleştirilebilir. Veri kümesindeki aşırı örnek sayısı karşı karşıya gelmesi durumunda ele alınan modelde kullanılan çoklu veri ezberci bir model oluşumuna yol açar. Bu da doğrulama veri kümelerinde kötü sonuçlara yol açabilir. Birçok özelliğe sahip veri kümelerinden modeller oluşturmak, hesaplama açısından zordur. Bunların hepsi, araştırmacıları, özellik kümesini azaltmak için birçok yöntem önermeye yönlendirir. Öznitelik seçimi, indirgeme, öznitelik çıkarma (dönüşüm) ve öznitelik seçimi süreçleri aracılığıyla gerçekleştirilir.

Yeni özellik seti veriseti varyans kapsamı gibi özellikler dikkate alınarak kolayca azaltılabilir. Öznitelik seçimi orijinal öznitelik kümesinden dönüştürme yapılmadan küçük bir öznitelik alt kümesini alma ve analiz hedefine göre doğrulama sürecidir. Seçim süreci,

amaca, eldeki kaynaklara ve istenen optimizasyon düzeyine bağı olarak çeşitli şekillerde gerçekleştirilebilir.

Özellik kümesi azaltma, amaca göre ilişkili özellik düzeyi ve fazlalık terimlerine dayanır. Daha spesifik olarak, bir özellik genellikle dört şekilde kategorize edilir:

- Son derece alakalı,
- Zayıf ilgili ancak gereksiz değil
- Alakasız
- Gereksiz

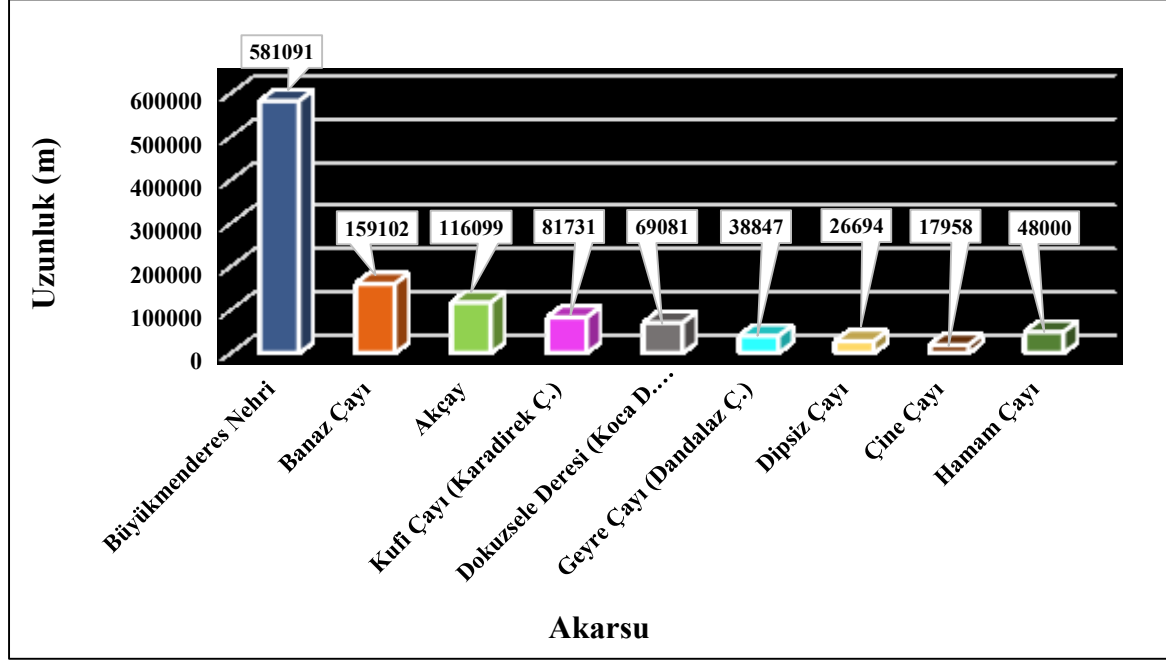
Optimal bir özellik alt kümesi için güçlü şekilde alakalı özellik her zaman gereklidir. Burada orijinal koşullu hedef dağılımı etkilenmeden kaldırılamaz. İlgisiz veya zayıf ilgili özellik, optimal bir alt küme için her zaman gerekli olmayabilir. Bu belirli koşullara bağı olabilir. Alakasız özelliklerin eklenmesi gerekli değildir. Gereksiz özellikler, zayıf ilgili ancak hedef dağılımı bozulmayacak şekilde bir dizi başka özellik ile tamamen değiştirilebilen özelliklerdir. Bu nedenle, çok değişkenli durumda fazlalık her zaman denetlenir, oysa bireysel özellikler için alaka düzeyi belirlenir. Özellik seçiminin amacı, alaka düzeyini en üst düzeye çıkarmak ve boyutsallığı en aza indirmektir. Genellikle sadece ilgili özelliklerden oluşan bir özellik alt kümesi bulma hedeflenmektedir (Chandrashekar ve Sahin, 2014; Jović ve diğerleri, 2015; Khaire ve Dhanalakshmi, 2019). Öznitelik seçimi yöntemleri genelde üç ana bölüme ayrılmaktadır. Bunlar, Filtre, sarmalayıcı, ve gömülü öznitelik seçimi yöntemleri olarak adlandırılmıştır. Oysa her bir ana yöntemin uygulama metoduna göre mevcut bir çok alt yöntem bulunmaktadır.

1.6. Çalışmanın Kapsamı

Tez çalışmasında su kalitesi değerlendirilmek ve su kalitesi sınıfı belirlenmek üzere Türkiye'nin 25 su havzası arasından seçilen 24,976 km²'lik yüz ölçümü ile en büyük su kütlesine sahip ve yıllık ortalama 3,03 km³'lük yağış özelliği ile Büyük Menderes Nehri seçilmiştir. Bu havza Türkiye'nin güneybatısında yer alan etrafı Küçük Menderes, Gediz, Sakarya ve Burdur Nehir Havzası ve Ege Denizi ile çevrilidir. Ege bölgesinin en uzun nehri olmak üzere 584 km'lik uzunluğu ile Büyük Menderes Nehri kaynağı havzanın doğusunda

yer alan yüksek dağlardır (Koç, 2010; Çevre ve Şehircilik Bakanlığı, 2016; Akyıldız ve Duran, 2021). Büyük Menderes Nehri Işıklı Gölü'nü dolduran sularla beslenir. Uşak'tan katılan Banaz Çayı ve Muğla'dan Çine Çayı sularını bünyesine katarak 2.600.967 ha'lık bir havzaya adını vererek Ege Denizi'ne dökülür (Şekil 1.8). Büyük Menderes Nehri'nin kollarından birisi olan Kufi Çayı Kumalar eteğinden başlar ve farklı sıcak su kaynaklarıyla büyür. Honaz Dağı, Kaklık ve Kocabağı yöreleri suları Çürüksu Çayı ile birleşerek, Gökpınar Çayı'nı da alarak Sakarya yakınından Büyük Menderes Nehri ile birleşir. Başta Geyre Çayı ismi alan Dandalaz Çayı Karacasu'nun güney-doğusundan Babadağ eteklerinden gelen Işıklar deresi ve Akyar deresi sularının toplanma sonucu oluşur ve Kuyucak yakınından Büyük Menderes ile birleşir. Tavas Ovası'na bakan yamaçlardan gelen Yenidere ile birleşen Akçay'ı Karıncadağ ve Mardan arası vadileri geçerek Yenipazar yakınından Büyük Menderes'e boşalır. Pınarbaşı'ndan çıkan suların Bencik ve Kamış derelerinde toplanmasıyla oluşan Çine Çayı Sağnan, Gökçay ve Madran Dere'lerini içine alır ve Çiftlikburnu yakınından Büyük Menderes'e boşalır. Aydın ili sınırları içinde Büyük Menderes'e katılan pek çok dere de vardır. Bunlar, sağ yandan katılan Feslek, Kestel, Malgaç, Köşk, Musluca, İmamköy, Tabakhane, Karagöz, Yalkı, İkizdere, Alangüllü, Kocadere, Dalama, Çakırbeyli, Çakmar, Koçarlı, Sarıçay Dereleri'dir. Bozdağ ve Sandıras dağlarından Yukarı Akçay Eskere Ovası'ndan geçerek diğer kaynakları da içine alır ve Tavas Kızılcabölük yakınlarından çıkan Yenidere Çayı'nı da içine alarak Aydın Bozdoğan ilçesindeki Kemer Barajı'nı besler. Murat Dağı'ndan doğan ve Yukarı Büyük Menderes Havzası'nı oluşturan Banaz Çayı Uşak ilinin en önemli akarsuyudur. Diğer önemli bir kol olan Dokuzsele Deresi de Uşak il merkezinden geçerek, Ulubey ilçesinden sonra Banaz Çayı ile birleşir ve Denizli ili sınırları içerisinde bulunan Adıgüzel Baraj Gölü'ne karışarak Büyük Menderes Nehri'ne karışmaktadır. Adıgüzel Barajı'nı besleyen diğer bir yan kol da Hamam Çayı'dır. Büyük Menderes havzasında yer alan Doğu Menteşe dağlarından doğan Dipsiz Çayı Yatağan Ovası'nı sular (TÜBİTAK, 2013).

Büyük Menderes Havzası'nda isimleri geçen akarsuların uzunlukları Şekil 1.9'da göstermektedir.



Şekil 1.9. Büyük Menderes Havzası akarsuları uzunlukları.

Türkiye'nin en kalabalık yedinci nehir havzası olmak üzere Büyük Menderes Nehri Havzası içinde Aydın, Denizli, Uşak, Afyonkarahisar, Muğla, İzmir, Isparta, Burdur, Kütahya, Manisa olmak üzere on şehir vardır. Türkiye topraklarının %3.2'si Büyük Menderes Havzası'na aittir (Çevre ve Şehircilik Bakanlığı, 2016; Boyacıoğlu ve Boyacıoğlu, 2017).

Bu havzada en başta gelen su tüketicileri tarım (yaklaşık 1,600.000 m³/gün) mesken (yaklaşık 350,000 m³/gün) ve endüstriyel tesislerdir (yaklaşık 150,000 m³/gün) (Boyacıoğlu, 2006; Çevre ve Şehircilik Bakanlığı, 2016).

Büyük Menderes Havzası'nın en büyük arazi kullanıcısı, toplam havzanın yaklaşık %44'ünü kapsayan tarımdır. Ardından havzanın yaklaşık %33 arazisini kapsayan yarı-doğal alanlar gelmekte , arazinin yaklaşık %20'si ise ormanlar ve nihayet sadece %1'ini ise yüzeysel sular kapsamaktadır (Çevre ve Şehircilik Bakanlığı, 2016; Boyacıoğlu ve Boyacıoğlu, 2017).

TÜBİTAK tarafından 2010 yılında Büyük Menderes Havzası hakkında yayınlanmış olan raporda havza su potansiyelinin toplamda $2.933 \cdot 10^9 \text{ m}^3/\text{yıl}$ olmak üzere, $2.00 \cdot 10^9 \text{ m}^3/\text{yıl}$ 'ı yüzeysel ve yaklaşık $933 \cdot 10^6 \text{ m}^3/\text{yıl}$ 'ı yeraltı suyuna aittir. Mevcut yüzey suyu potansiyeli ise 2010 yılı itibarıyla toplam su potansiyeli $1.70 \cdot 10^9 \text{ m}^3/\text{yıl}$ olmak üzere $1.00 \cdot 10^9 \text{ m}^3/\text{yıl}$ 'ı yüzeysel su, yeraltı suyu potansiyeli ise yaklaşık $700 \cdot 10^6 \text{ m}^3/\text{yıl}$ olmuştur (TÜBİTAK, 2013; Çevre ve Şehircilik Bakanlığı, 2016).

Yüzeysel sular nehirler, göller, kıyı ve geçiş suları olmak üzere dört kısımda incelenmektedir. Büyük Menderes Havzası'nda toplam 66 su kütlesinden ise 39'u nehir, 21'i göl, 4'ü yeraltı suyu, 1'i geçiş suyu ve 1'i kıyı suyu kütlesidir. Büyük Menderes Havzası Türkiye'nin 25 su havzası arasında besin madde açısından hassas su havzası olduğu ve su kalitesinin iyileştirmesi doğrultusunda çalışmaların devam ettiği belirlenmiştir (Çevre ve Şehircilik Bakanlığı, 2016; Tarım ve Orman Bakanlığı, 2018).

Büyük Menderes Havzası'nda belirlenmiş olan eli altı nehirin %60'ından fazlası (35 tanesi) insan aktiviteleri sonucu değişmiştir. Bu havzada tespit edilmiş olan 14 gölün Bafa Gölü hariç hepsi baraj gölleridir. Bu havza başlıca tarımsal, endüstriyel ve evsel kirlilik ile tehdit edilmektedir. Ağır metal kirliliği, pestisit kirliliği, kentsel atık suların ve diğer besin maddelerinin uygunsuz deşarjı, kontrolsüz jeotermal atık su deşarjı alandaki başlıca kirleticilerdir. Havzadaki başlıca kirlilik kaynakları, evsel ve tarımsal aktiviteler sonucu meydana gelir ve endüstriyel kaynaklar nedeniyle nispeten sınırlı oranda kirlilik vardır (TÜBİTAK, 2013; Çevre ve Şehircilik Bakanlığı, 2016).

1.7. Çalışmanın Önemi ve Hedefi

İnsanlar çeşitli nedenlerle göllere, nehirlere ve kıyılara ilgi duyarlar. Temiz su, içme, sulama, sanayi, ulaşım, eğlence, balıkçılık, avcılık, biyolojik çeşitliliğin desteklenmesi ve saf estetik zevk için çok önemli bir kaynaktır. İnsanlık tarihi boyunca su, kirleticileri yıkamak ve seyreltmek için kullanılmıştır. Kirlitici girdiler son yıllarda artmış ve birçok nehir, göl ve kıyı okyanusunun su kalitesini düşürmüştür. Bu hayati su kaynaklarının bozulması, doğal sistemlerin, bileşenlerinin türlerinin ve sağladıkları kolaylıkların kaybı olarak ölçülebilir. Su kıtlığı giderek yaygınlaşmaktadır. Gelecekte bu durumun daha şiddetli hale gelmesi muhtemeldir. Su kıtlığı ve kötü su kalitesi birbiriyle bağlantılıdır, çünkü kirlilik kaliteli su tedarikini azaltır ve kullanım için su arıtma maliyetlerini artırır. Kirliliği önlemek, su

kaynaklarını artırmanın en uygun maliyetli yollarından birisidir (Carpenter ve diğerleri, 1998).

Türkiye'nin su kaynakları çok olmasına rağmen, 1,500-1,700 m³/kişi-yıl su kapasitesi ile su stresi yaşayan bir ülkedir. Yıl içinde beş yâda altı ay ile sınırlı olan yağışlı dönemler, ülke toprakları çoğunun yarı kurak iklim bölgesi içine atmaktadır. Bu nedenle, su yönetimi ve iklim değişikliği çalışmaları Türkiye için ayrıca önem arz etmektedir (Çevre ve Şehircilik Bakanlığı, 2016).

Sürdürülebilir su yönetimi için bir temel olarak hidrolojik döngünün gözlenmesi ve değerlendirilmesi esastır. Bir temel taşı olarak bu işlem su kaynaklarının sürdürülebilir kullanımı açısından da önemlidir. Etkili bir su havzası yönetimi ve yüzey suları, yeraltı suları, drenaj suları, barajlar, göller gibi su kaynakları açısından ise su kaynakları miktarı ve kalitesi, arazi kullanımı, su kaynakları kullanımı ve mansap-memba ilişkileri gözlenmesi ve değerlendirilmesi esastır. Burada başta gelen hedefin havzadaki su kaynaklarının çok işlevli kullanımı yanı sıra havzadaki suların ekolojik ve fiziksel süreçlerin korunması ve havzadaki bütün sektörlerin ve toplulukların ihtiyaçlarının dikkate alınmasıdır. Bu nedenle her parametre dikkatli bir şekilde ölçülüp değerlendirilmelidir (Koç, 2010; Çevre ve Şehircilik Bakanlığı, 2016).

Büyük Menderes Havzası Türkiye'nin yedinci en kalabalık havzası olmak üzere başta tarım ve ardından endüstri en çok su tüketici sanayilerdir (Boyacıoğlu, 2006; Çevre ve Şehircilik Bakanlığı, 2016). Bu iki sanayinin en çok su tüketiciler olduğunu düşünüldüğünde havza su kirliliğinin bu iki kaynaktan geldiği sonucuna varılabilir.

Nehirlere, göllere ve okyanuslara kimyasal girdiler, noktasal veya noktasal olmayan kaynaklar olarak sınıflandırılır (Loague ve Corwin, 2006). Belediye kaynaklı kanalizasyon gibi noktasal kaynaklardan kirletici deşarjlar arıtma tesisleri, zaman içinde çok az değişkenlik göstererek sürekli olma eğilimindedir. Deşarj ve kimyasal konsantrasyonları periyodik olarak tek bir yerde ölçülerek izlenebilirler. Sonuç olarak, noktasal kaynakların ölçülmesi ve düzenlenmesi nispeten basittir ve genellikle kaynakta işleme tabi tutularak kontrol edilebilir. Noktasal olmayan girdiler de sürekli olabilir, ancak daha sıklıkla aralıktır ve mevsimlik tarımsal faaliyetlerle veya yoğun yağış veya büyük inşaat gibi düzensiz olaylarla bağlantılıdır. Noktasal olmayan girdiler genellikle geniş arazi

alanlarından elde edilir ve karadan, yeraltından veya atmosfer yoluyla alıcı sulara taşınır. Sonuç olarak, noktasal olmayan kaynakların ölçülmesi zordur. Noktasal olmayan kirliliğin kontrolü, arazi yönetimi uygulamalarına ve kirleticilerin atmosfere salınmasının kontrolüne odaklanır ve milyonlarca insanın günlük faaliyetlerini etkileyebilir. Ek olarak, çoğu ötrofik suyun restorasyonu, Fosfor ve Azotun noktasal olmayan girdilerinin azaltılmasını gerektirir. Aşırı Fosfor ve Azot girdilerinin neden olduğu ötrofikasyon, göllerde, nehirlerde, haliçlerde ve kıyı okyanuslarında yaygın ve büyüyen bir sorundur. Ötrofikasyon, gelişmiş ülkelerin haliçlerinde ve kıyı denizlerinde de yaygındır ve hızla genişlemektedir. Azot girdisi, çoğu ılıman nehir ağzı ve kıyı ekosisteminde birincil üretimi sınırlayan ve ötrofikasyona neden olur. Ötrofikasyonda Azot ana faktör olmasına rağmen, Fosfor aynı zamanda ötrofikasyona katkıda bulunan önemli bir elementtir (Carpenter ve diğerleri, 1998).

Büyük Menderes Havzası'nda tarım sanayinin birincil kirleticisi kaynağını düşünerek havzadaki suların ötrofikasyona maruz kalma potansiyeli açıkça ortaya çıkmaktadır. Azot ve Fosfor gibi ötrofikasyona neden olan elementlerin noktasal olmayan kaynaklardan geldiği ve bırakılma noktasının belli olmadığı için arıtımı ve hatta kontrolü zordur (Harper, 1992; Loague ve Corwin, 2006; Smith ve Schindler, 2009). Buda, havza su kalitesinin besin madde açısından sürekli gözlemmesi ve kontrol dışı olduğu takdirde gerekli önlemler alınmasına önem arz etmektedir.

Ötrofikasyonun ayrıca sucul ekosistemler üzerinde birçok olumsuz etkisi vardır. Ötrofikasyonun en belirgin sonucu, suyun balıkçılık, eğlence, sanayi, tarım ve içme suyu için kullanımını engellemesidir. Aşırı çoğalan bitkiler oksijen kıtlığına neden olur ve ardından balıkların ölümüne yol açar. Ötrofikasyon, tatlı ve deniz sularındaki su bitki yatakları ve tropik kıyıların mercan resifleri dahil olmak üzere habitatların kaybına neden olur. Ötrofikasyon, sudaki biyolojik çeşitliliğin kaybında bir faktördür (Carpenter ve diğerleri, 1998; Smith ve Schindler, 2009).

Ötrofikasyon sonucu ortaya çıkan oksijen azalması ve ardından gelen olumsuz sonuçlar literatürde ortaya konulmuştur (Harper, 1992). Bu nedenle ÇO bir havza suyu sağlığını gösteren bir faktör olarak sürekli ölçülmeli ve gözetlenmeli birinci su kalitesi parametresidir. Ötrofikasyon ayrıca bir tek oksijene bağlı olan sucul yaşamı sınırlamaz bir su kaynağını içme suyu açısından da olumsuz etkiler (Smith ve Schindler, 2009).

Bu çalışmasının hedefi Büyük Menderes Havzası su yönetimine yardımcı olacak şekilde, havzanın su kalitesinin değerlendirilmesi ve insani su tüketimi açısından su kirliliğinin önlenmesi için efektif ve verimli bir model önerilmesidir. WQI su kalitesi parametrelerini göz önünde bulundurarak genel su kalitesini modellenmektedir.

WQI su kalitesi parametrelerini göz önünde bulundurarak genel su kalitesini modellenmektedir. Fakat WQI yöntemi hesaplama açısından çok zaman alıcı ve zahmetlidir. Yani bir havzadaki su kalitesi verileri elde edildikten sonra önce bir veri kümesi oluşturmalıdır. Daha sonra her bir parametrenin standart değerine göre ağırlığı hesaplanmalıdır. Ağırlıklar toplandıktan sonra toplam ağırlık değerinin aralığına göre su kalitesi sınıfı belirlenmelidir. Buda oldukça zaman, hesaplama açısından maliyetli bir işlemdir.

Son yıllarda ML yöntemleri bir çok alanda maliyetli işlemleri basitleştirmek adına kurumlar ve karar mercilerine yardımcı olmakta ve uygulanmaktadır (Michie ve diğerleri, 1994). Bir ML modeli girdi veriler ve sonuçlarıyla eğitilir (SUPL yöntemi). Genel ve basitleştirilmiş anlamda parametreler ve sonuçlar arasında bir model oluşturur (Singh ve diğerleri, 2016). Daha sonra test veri kümesiyle denetlenir. Denetleme veri kümesi kendisi ikiye ayrılır. Bir bölümüyle girdi verileri ve parametreleriyle örneğin bir sınıflandırma işlemini yapmak için sınıf tahmini yapar. Bu aşamanın sonunda ML modelinin sınıflandırdıkları gerçek sınıflar ile denetlenir. Daha sonra denetleme veri kümesinin ikinci parçasıyla ML modelinin doğruluk testi yapılır ve modelin performansı hesaplanır. Burada önemli olan konu ML modeli bir paket programı haline geldiğinde sürekli öğrenmekte olmasıdır. Yani böyle bir paket programı oluşturulduktan sonra bir kurumun veri toplama departmanı her hangi bir veri kümesi oluşturması gerekmeden sadece veri değerlerini girdiğinde ML modeli verileri bir arada analiz yapar, öğrenir, denetler ve sonuç olarak her hangi bir karmaşık hesaplama yapmadan öğrendikleri ve denetlediklerine dayalı bir sınıfa atar. Bu çalışmanın en önemli hedefi Büyük Menderes Havzası su kalitesi sınıflandırması için yüksek performanslı ML modeli önermektir. ML modeli kapsamında farklı alanlarda çok iyi performans gösteren Toplu Öğrenme metotlarına dayalı (Ensemble Learning (EL)) dört ML sınıflandırma modelinin performansıda ele alınmıştır. Bu çalışmanın sonuçlarına göre elde edilen en verimli model paket programı haline getirilerek Büyük Menderes Havzası için bir otomatik ve güvenilir model oluşturulabilir. Bu çalışmanın bulguları ve sonuçarı doğrultusunda önerilen model başka havzaların su kalitesi verileri üzerinde de denetlenebilir ve

arastirmalardaki sonuçların performansının yüksek olması durumunda Türkiye'nin tüm havzalarının su kalitesi modellemesi içinde bir sürekli çalışan, öğrenen ve denetlenmekte olan referans model haline gelebilir. Sonuç olarak bu çalışma Türkiye Cumhuriyeti su yönetimi, ekonomisi ve insan sağlığına çok önemli faydalar sağlayabilir.

Ayrıca, bir havzadaki suların yönetimini basitleştiren ve önleyici ani kararlar alınmasına yardımcı olan havza su kirliliğinin en önemli göstergelerinden olan ÇO değerleri için yüksek verimli bir tahmin modeli oluşturulmasıdır. Çözünmüş Oksijen bir su havzası yönetimi için denetlenen ilk su kalitesi parametrelerinden birisidir. Düşük ÇO değerleri bir su sisteminde kirliliğin göstergesi olabilir. Bu nedenle ÇO değerleri sürekli izlenmeli ve denetlenmelidir. ML kavramında kullanılan bu parametre için bir tahmin modeli oluşturulursa ilerleyen aylarda yada yıllarda muhtemel ÇO değeri tahmin edilerek tespit edilir. Tahmin edilen ÇO değeri düşükse karar mercileri tarafından bir havzanın su kalitesini iyileştirmek için önleyici kararlar alınabilir. Bu çalışmada Büyük Menderes Havzası için yüksek performanslı ML modeli önermenin yanında ilgili havzada ÇO tahmini doğrultusunda ML yöntemlerine dayalı tahmin modeli önerilmesi planlanmaktadır.

Bu çalışmanın bulgularına dayanarak, ilgili bakanlıklar tarafından paket program olarak hazırlanacak olan modeller, havzanın su yönetimi kararlarını basitleştirerek, Türkiye Cumhuriyeti ekonomisi ve halk sağlığı açısından önemli faydalar sağlayabilir. Bu modeller, zaman, hesaplama ve maliyet açısından avantajlar sunarak, etkili bir şekilde kullanılabilir.

2. LİTERATÜR İNCELEMESİ

2.1. İstatistiksel Yöntemlere Dayalı Sınıflandırma ve Tahmin Modellemesi Çalışmaları

Boyacıoğlu ve diğerleri (2005), faktör analizi yöntemi ile Türkiye’de Büyük Menderes Havzası su kalitesi parametreleri arasında ilişkiyi bularak bu havzanın önemli olan kirleticilerini belirlemeye çalışmışlardır. Çalışmanın sonucunda kentsel ve tarımsal kirleticiler Büyük Menderes Havzası’nın en önemli kirleticileri olarak belirtilmektedir.

Boyacıoğlu (2006), faktör analizi yöntemi ile Türkiye’de Büyük Menderes Havzası su kalitesini etkileyen faktörleri bulmaya çalışmıştır. Bu çalışmanın sonucunda düşük akış döneminde tarımsal faaliyetlerin Büyük Menderes Havzası’nın su kalitesi negative yönde etkileyen faktör olduğu ortaya konulmuştur. Ancak araştırmada yüksek akışlı dönemlerde kentsel atıklardan gelen kirlilik bu havzanın su kalitesini etkilemekte olduğu belirtilmektedir.

Koç (2010), MODQUAL istatistiksel modelleme yöntemi ile Türkiye’de Büyük Menderes Havzası su kalitesi durumunu ortaya koymaya çalışmıştır. Bu çalışma sonucunda havza genelinde organik kirlenme söz konusu olduğu ortaya konulmuştur. Havzanın genelinde su kalitesinin BOİ, KOİ, Amonyak, Nitrit ve Nitrat açısından III. ve IV. su kalitesi sınıfında yer aldığı söylenmiştir.

Boyacıoğlu ve Boyacıoğlu (2017), çok değişkenli istatistiksel yöntem ve Keşif Faktörü analizi yaparak Türkiye’de Büyük Menderes Havzası su kalitesini etkileyen değişkenler arasındaki ilişkiyi bulmaya ve daha sonra bu havzanın su kalitesini etkileyen önemli faktörleri bulmaya çalışmışlar. Bu çalışmanın sonucunda Büyük Menderes Havzası’nın organik kirlenmeye maruz kaldığı ortaya konulmuştur.

Avila ve diğerleri (2018), Yeni Zelanda’da Wallacetown’daki Oreti Nehri’nin, ML yöntemleri ile su kalitesi sınıfı tahmin modellemesini çalışmışlardır. Bu çalışmada, su kalitesini sadece *E.coli* açısından değerlendirilmiştir. Sürekli ve kategorik sınıflara ayrılmış olan modeller arasında NB, Çoklu Lineer Regresyon (Multi-Linear Regression (MLR)), Dinamik Regresyon (Dinamik Regression (DR)), Regresyon Ağacı (Regression Tree

(REGT)), Markov Zinciri (Markov Chain (MARC)), Sınıflandırma Ağacı (Classification Tree), Rasgele Orman (Random Forest (RF)), Lojistik Regresyon (Logistic Regression (LGR)), Ayrımcı Analizi (Discriminant Analysis (DA)) ve Bayes Ağları (Bayesian Networks (BNET)) gibi modeller göze çarpmaktadır. Sonucunda LDA ve BN yöntemleri en düşük hata ile en yüksek performansı sergilediği bildirilmektedir.

Ayla Bilgin (2018), DA yöntemi kullanarak Türkiye’de Çoruh Nehri su kalitesinde etken en önemli parametreleri ayırmaya çalışmıştır. Kanada Çevre Bakanlar Konseyi Su Kalitesi Endeksi (Canadian Council of Ministers of The Environment Water Quality Index (CCME-WQI)) ile su kalitesi sınıflandırılmıştır. Ayrıca bu çalışmada kümeleme yöntemi kullanarak tekrardan su kalitesi sınıflarını gruplara ayrılmıştır.

Kusnandar ve diğerleri (2019) tarafından , QDA kullanarak Kirlilik Endeksi yöntemi ile üç sınıfa ayrılmış su kalitesi endekslerini sınıflandırmaya çalışmışlardır. Bu çalışmada, çalışma kapsamı ve alanı 61 nehir ve kanal içeren Endonezya’da Pontianak Şehri olduğu belirtilmektedir.

Kükrer ve Mutlu (2019) tarafından , önce Ağırlıklı Aritmetik Su Kalitesi Endeksleme yöntemi (Weighted Arithmetic Water Quality Index (WAWQI)) ile Türkiye’nin Kızılırmak Havzası’nda yer alan Saraydüzü Baraj Gölü su kalitesini sınıflara ayrılmıştır. Sonrasında Çok değişkenli istatistiksel analizler (Multivariate Statistical Analysis (MSA)) yöntemini, su kalitesi sınıflandırması analizinde kullanmışlardır.

Hamed (2019), Mısır’da Nil Nehri su kalitesi sınıflarını Bulanık C-Ortalamlar Kümeleme Analizi (Fuzzy C-Means Cluster Analysis (FCMCA)) yöntemi ile yedi su kalitesi sınıfına ayırmıştır. Ayrıca arştırmacı veriseti gürültüsünü azaltmak için PCA yöntemi ile boyutunu düşürmeye çalışmıştır. Sınıflandırma için bu çalışmada K-Ortalamlar (K-Means (KM)) algoritması kullanılmıştır.

Samsudin ve diğerleri (2019), Uzaysal Diskriminant Analizi (Spatial Discriminant Analysis (SDA)) yöntemini Batı Malezya Yarımadası bölgesi (Doğu ve Batı Kıyısı’ndan oluşan) su kalitesi veriseti parametreleri sayısını azaltarak veriseti gürültüsü (Noise) ve boyutunu azaltmak için kullanmışlardır. Çalışmanın devamında deniz WQI yöntemi ile boyutu azaltılmış verilerin su kalitesi sınıfları belirlenmiştir. MLR istatistiksel sınıflandırma

yöntemini ANN makine öğrenme sınıflandırma yöntemi ile karşılaştırmışlardır. Çalışmanın sonucunda ANN su kalitesi sınıflandırmasında MLR'ye göre daha efektif yöntem olduğunu belirtmişlerdir.

Ustaoglu ve diğerleri (2020), PCA, Pearson korelasyon Endeksi (Pearson Correlation Index (PCI)) ve Kümeleme Analizi (Clusterin Analysis (CA)) yöntemini Türkiye'nin Doğu Karadeniz Havzası'nda yer alan Turnasuyu Çayı'nın en önemli su kalitesi parametrelerini bulmak için parametreler arasındaki ilişkiyi bularak kullanmışlardır. Ayrıca bu çalışmada WQI su kalitesi sınıflandırması için kullanılmıştır.

Yılmaz ve diğerleri (2020), Türkiye'de Büyük Menderes Havzası su kalitesini sekiz istasyonda WQI yöntemi ile değerlendirme ve sınıflandırmaya çalışmışlar. Değerlendirmeler sonucu Büyük Menderes Havzası'nın üst akış bölgelerinde su kalitesinin "İyi" su kalitesi sınıfında olduğu ortaya konulmuştur. Ancak bu havzanın orta akış bölgelerinde su kalitesi "Kötü" ve "Çok Kötü" sınıfına girdiği bildirilmiştir. Bunun nedeni Büyük Menderes Havzası'nın orta akış bölgelerinde kentsel, tarımsal ve endüstriyel faaliyetlerin yaygın olduğu bildirilmiştir.

Varol (2020), WQI yöntemi ile Türkiye'nin Malatya İli'nde bulunan Sürgü Çayı'nın su kalitesi sınıflandırmasını yapmıştır. Daha sonra CA yöntemi ile su kalitesi endekslerini göz önünde bulundurarak örnekleme alanlarını kirlilik açısından üç genel bölgeye ayırmıştır. Zamansal Ve Mekânsal Diskriminant Analizi yöntemini (Spatio-Temporal Discriminant Analysis (STDA)) ayrıca parametrelerin kirlilik üzerinde etkilerini değerlendirmek için kullanmıştır.

Zhu ve diğerleri (2021), Çin'in Taowang antiklinal bölgesinde bir su kaynağının su kalitesi parametreleri arasındaki ilişkiyi bulmak için Fisher'in Diskriminant modelini kullanarak Diskriminant Fonksiyon (Discriminant Function) modelini oluşturmuşlardır.

Yukarıda değerlendirilen çalışmalar 2005-2022 yılları arasında istatistiksel yöntemlerle su kalitesi değerlendirmesi, sınıflandırması ve tahmini çalışmaların örnekleridir. Bu çalışmaların hiç birisi geleceğe yönelik havza yönetiminde karar makamlarına yardımcı olacak bir şekilde her hangi bir model ortaya koymamışlardır. Hazır mevcut yöntemlerle sadece belli bir bölgenin yüzeysel sularının kalitesini değerlendirme açısından literatüre

katkı sunmuşlardır. Ancak son yıllarda bilim dünyasında Yapay Zekâ (Artificial Intelligence (AI)) konseptine dayalı ML modelleri gibi daha modern ve geleceğe yönelik modeller de araştırmacılar tarafından sunulmuştur. Bu yöntemlerin verimliliği ve güvenilir olmaları hali hazırda tartışılmakta olup bu yöntemlerin istatistiksel yöntemlere göre hızlı, doğruluğu yüksek ve maliyet verimli oldukları elde edilen sonuçlara bakılarak görülmektedir. Ancak her bir problemi çözmek için bu algoritmaların verimlilikleri araştırılmalı ve çözüm odaklı bir model sunulmalıdır.

2.2. Makine Öğrenme Yöntemlere Dayalı Sınıflandırma ve Tahmin Modellemesi Çalışmaları

Faruk (2010), Otoregresif Entegre Hareketli Ortalama (Autoregressive Integrated Moving Average (ARIMA)) ve ANN modellerinden hibrit bir model geliştirerek Türkiye’de Büyük Menderes Havzası su kalitesi verileri ile Bor, ÇO, ve sıcaklık tahmini doğrultusunda bir çalışma yapmıştır. Tahmin çalışması sonucunda korelasyon katsayısı değerlerini Bor için 0,902, ÇO için 0,893 ve sıcaklık için 0,909 bulmuş ve geliştirilmiş olan modelin tahmin çalışmalarında verimli bir model olduğunu ortaya koymuştur.

Camejo ve diğerleri (2013), DT, ANN ve KNN makine öğrenme algoritmalarının performansını Canada’nın Ontario ili ve Orta Potekiz’in su kalitesi sınıflandırmasında karşılaştırmışlardır. Bu çalışmada WAWQI yöntemi içme suyu amaçlı su kalitesi sınıflandırması için kullanılmıştır. Bu çalışmada literatür araştırmalı öz nitelik seçimi yapılmıştır. Sonuç olarak KNN algoritması %98,02 doğruluk oranıyla su kalitesi sınıflandırmasında diğer algoritmalara göre daha üstün performans göstermiştir.

Modaresi ve Araghinejad (2014), SVM, KNN ve olasılıksal sinir ağları (Probabilistic Neural Networks (PNN)) ML yöntemlerinin performansını İran’ın Tahran İli’nin ana akiferinin WQI sınıflandırmasında değerlendirmişlerdir. Değerlendirmeler sonucunda SVM doğrulama aşamasında en düşük hatayla en etkili algoritma olarak önerilmiştir.

Mohammadpour ve diğerleri (2015), Malezya’da bir sulak alanın WQI tahmini için SVM ve ANN yöntemlerini kullanmışlardır. Bu çalışmada FLTM öz nitelik seçimi yöntemi kullanılan algoritmaların hassasiyetini belirtmek için kullanılmıştır. Çalışma sonucunda SVM yönteminin ANN’ye üstün geldiği belirtilmiştir.

Danades ve diğerkleri (2016), KNN ve SVM modellerinin performansını su kalitesi sınıflandırmasında deęerlendirmişler. Sumitomo ve Nemerow (1970) tarafından ABD, Texas Üniversitesi'nde geliştirilmiş olan Kirlilik Endeksi (Pollution Index) yöntemi su kalitesi endekslemesi için kullanılmıştır. Cross Validation yöntemi doğrulama aşamasında modellerin hatasını belirlemek için kullanılmıştır. Çalışma sonucunda SVM modeli KNN'ye göre daha üstün performans sergilemiştir.

Hameed ve diğerkleri (2017), ANN modellerini Malezya'da iki nehrin su kalitesini sınıflandırmak için kullanmışlardır. Bu çalışmada Malezya su kalitesi endeksi (Malaysian Water Quality Index (MWQI)) su kalitesi endekslemesi için kullanılmıştır. Sonuçta Radyal Tabanlı Fonksiyon (Radial Based Function (RBF)) içeren ANN ile geri yayılım (Backpropagation) yöntemi ile çalışan ANN'ye göre daha üstün performans sergilediğı bildirilmektedir.

Heddam ve Kişı (2017), ABD'nin farklı havzaları akarsularından toplanan veriler ile ÇO tahmini doğrultusunda çalışma yapmışlardır. Bu çalışmada, farklı aktivasyon fonksiyonları uygulamasıyla farklı Aşırı Öğrenme Makinesi (Extreme Learning Machine (ELM)) algoritmaları ÇO deęerleri tahmini doğrultusunda kullanılmıştır. Çalışmanın sonucunda genel olarak Optimally Pruned Extreme Learning Machine (OP-ELM) modeli diğerk kullanılan ELM modellerine göre daha üstün performans sergilediğı belirtilmektedir.

Wang ve diğerkleri (2017), uzun ve kısa süreli hafızalı sinir ağı (Long-Short Term Memory (LSTM)) derin öğrenme yöntemini Çin'in Taihu Gölü'nde ÇO ve Fosfor tahmini için kullanmışlardır. Bu çalışmada LSTM Derin Öğrenme (Deep Learning (DL)) yöntemini geleneksel iki ANN yöntemi (Geri Yayılım Sinir Ağı (Back Propagation Neural Network (BPNN)) ve Çevrimiçi Sıralı Aşırı Öğrenme Makinesi (Online Sequential Extreme Learning Machine (OSELM))) ile karşılaştırılmıştır. Çalışma sonucunda LSTM Neural Network yöntemi ÇO ve Fosfor tahmini için diğerk iki yönteme göre daha üstün performans sergilediğı görülmektedir.

Dezfooli ve diğerkleri (2018), PNN, KNN ve SVM modellerinin performansını İran'da Arvandrood Nehri WQI sınıflandırmasında kıyaslamışlardır. Çalışmanın sonucunda PNN modeli diğerk iki modele göre daha üstün performans sergilediğı belirtilmektedir.

Haghiabi ve diğerleri (2018), ANN, SVM ve grup veri işleme yöntemi performanslarını İran'da Tireh Nehri su kalitesi parametreleri tahmininde değerlendirmişlerdir. Bu çalışmada tek bir hedef parametre tahmini yerine tüm su kalitesi parametreleri sırayla tahmin edilmiştir. Bu çalışmada 9 parametreden sırayla birisi hedef parametresi olarak seçilerek algoritma diğer 8 parametre ile eğitilerek çıkarılan bir parametrenin tahmini doğrultusunda uygulanmıştır. Çalışmanın sonucunda SVM diğer yöntemlere göre daha üstün performans sergilediği bildirilmektedir.

Keshtegar ve Heddam (2018), ABD'nin farklı havzalarına ait su kalitesi parametreleri verileri ile ÇO günlük değerleri tahmini doğrultusunda ML yöntemleri ile tahmin modelleri oluşturmaya çalışmışlar. Bu çalışmada, Modified Response Surface Method (MRSM) ve Multilayer Perceptron Neural Network (MLPNN) tahmin modelleri olarak kullanılmış ve ÇO değerleri tahmini doğrultusunda performansları kıyaslanmıştır. Çalışmanın sonucunda MRSM modeli MLPNN modeline göre daha üstün performans sergilediği belirlenmiştir.

Tiwari ve diğerleri (2018), Çıkarmalı Kümeleme Tabanlı Uyarlanabilir Nöro-Bulanık Çıkarım Sistemi (Subtractive Clustering Based Neuro-Fuzzy Inference System (SC-ANFIS)) ve Bulanık C-Ortalamlar (Fuzzy C-Means (FCM)) uyarlanabilir nöro-bulanık çıkarım sistemi (FCM-ANFIS) ile iki ANN hibrit modelini Hindistan'da Satluj Nehri su kalitesi tahmin modeli oluşturmak için kullanmışlardır. Bu çalışmada ayrıca WAWQI yöntemi su kalitesi endekslerini belirlemek için kullanmıştır. Çalışma sonucunda SC-ANFIS modeli FCM-ANFIS'ye göre daha üstün performans sergilediği tespit edilmiştir.

Yaseen ve diğerleri (2018), Uyarlanabilir Nöro-Bulanık Çıkarım Sistemini (Neuro-Fuzzy Inference System (NFIS)) üç hibrit modeli kombinasyonu (Bulanık C-Ortalamlar Veri Kümeleme (Fuzzy C-Means Data Clustering (FCMDC)), Izgara Bölme (Grid Partitioning (GP)) ve Eksiltici Kümeleme (Subtractive Clustering (SC))) ile Malezya'da Selangor Nehri'nin WQI tahmini için kullanmışlardır. Çalışmada korelasyon çıkarma yöntemi ile (FLTM) en etkin parametreleri belirtmek için bir öz nitelik çalışması da yapılmıştır. Çalışmanın sonucunda optimum koşullarda SC uyarlanabilir NFIS algoritması diğer iki yönteme göre üstün performans sergilediği bildirilmektedir.

Ahmed ve diğerleri (2019), Pakistan'da Rawal Gölü'nün WQI sınıflandırması için MLR, Polinom Regrasyon (Polynomial Regression (PR)), RF, Gradyan Artırma Algoritması

(Gradient Boosting Algorithm (Gboost)), SVM, Sırt Regresyonu (Ridge Regression (RREG)), Kement Regresyonu (Lasso Regression (LASR)), Elastik Net Regresyon (Elastic Net Regression (ENR)), Çok Katmanlı Algılayıcılar (Multilayer Perceptron (MLP)) sinir ağıları ML yöntemleri kullanmış ve performanslarını değerlendirmişlerdir. Çalışma sonucunda MLP sinir ağıları diğer modellere göre daha üstün performans sergilediği görülmektedir.

Csábrági ve diğerleri (2019), Macaristan'da Tisza Nehri'nin ÇO değerlerinin ML yöntemleri ile tahmini doğrultusunda çalışma yapmışlar. Bu çalışmada MLR, Radial Basis Function Neural Network (RBFNN) ve General Regression Neural Network (GRNN) tahmin modelleri olarak kullanılmış ve performansları ÇO değerleri tahmini doğrultusunda kıyaslanmıştır. Çalışmanın sonucunda GRNN modeli diğer modellere daha üstün performans sergilediği belirtilmektedir.

Ho ve diğerleri (2019), DT yöntemi ile Malezya'da Klang Nehri su kalitesi sınıflandırması yapmışlardır. Bu çalışmada manuel bir öznitelik seçimi çalışması da yapılmıştır. Bu çalışma sonucunda 5 parametre öğrenme aşamasında dâhil edilince en yüksek sınıflandırma performansı sonucu elde edilmiştir.

Isiyaka ve diğerleri (2019), Malezya'da Kinta Nehri WQI tahmini için ANN yöntemini uygulamışlardır. Bu çalışmada en kritik su kalitesi izleme alanları belirtmek için Hiyerarşik Aglomeratif Küme Analizleri (Hierarchical Agglomerative Cluster Analysis (HACA)) yöntemi kullanılmıştır. Bu çalışmada ayrıca bir öznitelik seçimi çalışması da yapılmıştır. Çalışma sonucunda varimax döndürmeden önce temel bileşen puanlarının kullanılması ile sekizinci düğümde en iyi tahmin performansı tespit edilmiştir.

Kadam ve diğerleri (2019), ANN modelini Hindistan'da Shivganga Nehri WQI tahmini için kullanmışlardır. Bu çalışmada MLR yöntemi parametreler arası ilişkiyi ve parametrelerin önem derecesini bulmak için kullanılmıştır. Bu çalışmanın sonucunda ANN, WQI tahmini doğrultusunda kullanılabilir bir yöntem olduğu belirtilmiştir.

Sulaiman ve diğerleri (2019), Malezya'da Malacca Boğazı Kıyısı su kalitesi sınıfı tahmini doğrultusunda ANN yöntemini kullanmışlardır. Bu çalışmada algoritmayı eğitmek ve

yöneltmek amacıyla su kalitesi sınıflarını belirlemek için MWQI yönetimi kullanılmıştır. Çalışma sonucunda ANN modeli performansı %80,00 olarak belirtilmiştir.

Aldhyani ve diğerleri (2020), çalışmalarında Hindistan'ın bir çok nehrinden toplanan veriler ile WQI tahmini ve sınıflandırması yapmışlardır. Bu çalışmada Otoregresif Sinir Ağı (Autoregressive Neural Network (ARNN)) ve LSTM su kalitesi tahmini için kullanılmıştır. Ayrıca üç geleneksel ML modeli (SVM, KNN, ve NB) sınıflandırıcı olarak kullanılmıştır. Çalışma sonucunda ARNN modeli su kalitesi tahmini doğrultusunda LSTM modeline göre daha üstün performans sergilemiştir. Sınıflandırma modelleri arasında ise SVM diğer iki geleneksel sınıflandırma modeline göre daha üstün performans sergilediği sonucuna varılmıştır.

Bui ve diğerleri (2020), 4 bağımsız (MLP, M5P, RT ve Azaltılmış Hata Budama Ağacı (Reduced Error Pruning Tree (REPT)), 12 hibrit veri madenciliği algoritması (Bağımsızların Torbalama (Bagging Of Independents (BI), CV Parametre Seçimi (CV Parameter Selection (CV-CPS)) ve Rastgele Filtrelenmiş Sınıflandırma (Random Filtered Classification (RFC))), yöntemlerini İran'da Talar Nehri'nin WQI sınıflandırması için kullanmışlardır. Bu çalışmada İran Su kalitesi endeksi (Iran Water Quality Index (IR-WQI)) su kalitesi endeksleri belirlemede kullanılmıştır. Bu çalışmada bir öz nitelik seçimi çalışması da yapılmıştır. Çalışma sonucunda BART hibrit modeli diğer yöntemlere göre daha üstün performans sergilemiştir.

Chen ve diğerleri (2020), Çin'de 10 milli nehir ve gölden toplanan veriler ile su kalitesi sınıfı tahmini doğrultusunda 10 adet ML modeli performansını değerlendirmişlerdir. Bu çalışmada su kalitesi sınıflarını belirtmek için Çin Halk Cumhuriyeti yüzeysel sular kılavuzu kullanılmıştır. Bu çalışmada ayrıca hesaplama ve maliyet uygunu yapmak amaçlı bir öz nitelik seçimi çalışması da yapılmıştır.

Gaya ve diğerleri (2020), MLR, uyarlanabilir NFIS, ve ANN modelleri performansını Hindistan'da Yamuna Nehri WQI tahmini için kullanmışlardır. Bu çalışmada rasgele çıkarma yöntemi ile de bir öz nitelik seçimi yapılmıştır. Çalışma sonucunda ANN diğer yöntemlere göre daha üstün performans sergilemiştir.

Zhu ve Heddam (2020), Çin’de Yangtze Nehri ÇO değerlerinin ML yöntemleri ile tahmini doğrultusunda çalışma yapmışlar. Bu çalışmada, ELM ve MLPNN tahmin modelleri olarak kullanılmıştır. Çalışmanın sonucunda MLPNN modeli ÇO tahmini doğrultusunda ELM modeline göre daha üstün performans sergilemiştir.

Ahmed ve Lin (2021), ABD’nin farklı eyaletlerinden üç nehirin su kalitesi verileri ile ÇO değerlerinin ML yöntemleri ile tahmini doğrultusunda çalışma yapmışlar. Bu çalışmada QRF ve MLPNN modelleri ÇO değerleri tahmini doğrultusunda kullanılmıştır ve performansları kıyaslanmıştır. Çalışma sonucunda QRF modeli ÇO tahmini doğrultusunda MLPNN modeline göre daha üstün performans sergilemiştir.

Asadollah ve diğerleri (2021), Hong Kong’da Lam Tsuen Nehri su kalitesi parametrelerinin değerleri tahmini doğrultusunda Ekstra Ağaç Regresyon (Extra Tree Regression (ETR)), SVR ve DT regresyon modellerini kullanmış ve performanslarını değerlendirmişlerdir.

Dehghani ve diğerleri (2022), ABD’nin Cumberland Nehri ÇO değerlerinin ML yöntemleri ile tahmini doğrultusunda çalışma yapmışlar. SVM modeli hem bağımsız hemde hibrid formlarda ÇO değerleri tahmini doğrultusunda kullanılmıştır. Çalışmanın sonucunda SVM’nin Algorithm of the Innovative Gunner (AIG) ile hibridi diğer tahmin modellerine göre daha üstün performans sergilemiştir.

Malek ve diğerleri (2022), Malezya’da Kelantan Nehri WQI sınıflandırması doğrultusunda DT, ANN, KNN, NB, SVM, RF, GBoost modelleri performansını değerlendirmişlerdir. Çalışma sonucunda GBoost modeli diğer modellere göre daha üstün performans sergilemiştir.

Ilić ve diğerleri (2022), NB makine öğrenme yönteminin performansını Sırbistan’da Vojvodina İli’nin akarsuları WQI sınıflandırmasında değerlendirmişlerdir. Çalışma sonucunda bu model test veri seti için ayrılan 68 veriden 64 tanesini doğru tahmin ederek, bu tür çalışmalarda kullanılabilir olarak önerilmektedir.

3. MATERİYAL VE METOD

3.1. Veri ve Veri Seti Oluşturulması

Bu çalışmanın amacı doğrultusunda, Türkiye'de Devlet Su İşleri Genel Müdürlüğü (DSİ) tarafından sağlanan su kalitesi verileri kullanılmıştır. Havzanın su kalitesini belirlemek amacıyla Türkiye Devlet Su İşleri Genel Müdürlüğü (DSİ) tarafından 1985-2014 yılları arasında havzada bulunan 142 istasyona ait aylık düzenli aralıklarla ölçülen veriler kurumdan temin edilmiştir. Bu araştırmada kurumdan izinle alınan veriler kullanılmıştır.

Büyük Menderes Havzası'nda bulunan bu istasyonlarda kurum tarafından Arsenik, Bor, Baryum, Kalsiyum, Kadmiyum, Kobalt, Krom, Bakır, Demir, Cıva, Potasyum, Sodyum, Alüminyum, Magnezyum, Manganez, Nikel, Kurşun, Çinko, Klorür gibi inorganik maddeler ölçülmüştür. Ancak Arsenik, Bor ve Baryum gibi bu inorganik maddelerin bazılarının ölçümü senede bir yâda iki kere yapılmıştır. Bu tarz ölçüm genel bir su kalitesi değerlendirilmesi açısından bir sınıflandırma modeli oluşturulması için ciddi hataların ortaya çıkmasına sebep olacağından bu araştırmada veri setinden çıkarılmıştır. Su kalitesi değerlendirmelerinde kullanılan ve önem arz eden oksijenlendirme parametreleri arasında Çözünmüş Oksijen (ÇO), Beş Günlük Biyokimyasal Oksijen İhtiyacı (BOİ-5) ve Kimyasal Oksijen İhtiyacı (KOİ) değerleri oluşturulan veri setine eklenmiştir. Aynı şekilde besin madde kirliliğini gösteren Amonyak Azotu (NH_4), Nitrit (NO_2), Nitrat (NO_3), Fosfat (PO_4) ve Sülfat (SO_4) verileri de veri setine eklenmiştir. Kullanılabilecek bir şekilde ölçülmüş olan diğer su kalitesi parametreleri olan Sıcaklık, pH, iletkenlik ve toplam katı madde değerleri de veri setine eklenmiştir (Çizelge 3.1).

Araştırmada 2004-2014 yılları arasında Şubat, Nisan, Haziran, Ağustos, Ekim, Aralık aylarında düzenli olarak ölçülen 10 istasyondan (Çizelge 3.2 ve Şekil 3.1) elde edilen parametrelere ait veriler kullanılmıştır.

Büyük Menderes Havzası'nda baskın sanayinin tarım sanayi olduğu göz önünde alınırsa DSİ tarafından sağlanan verilerde besin madde kirlenmesini gösterecek olan tüm parametreler elde edilmiştir. Ancak Arsenik, Nikel, Bor vs. gibi su kalitesi açısından bazı önemli inorganik maddelerin ölçülmediği yada düzenli olarak ölçülmediği tespit edilmiştir.

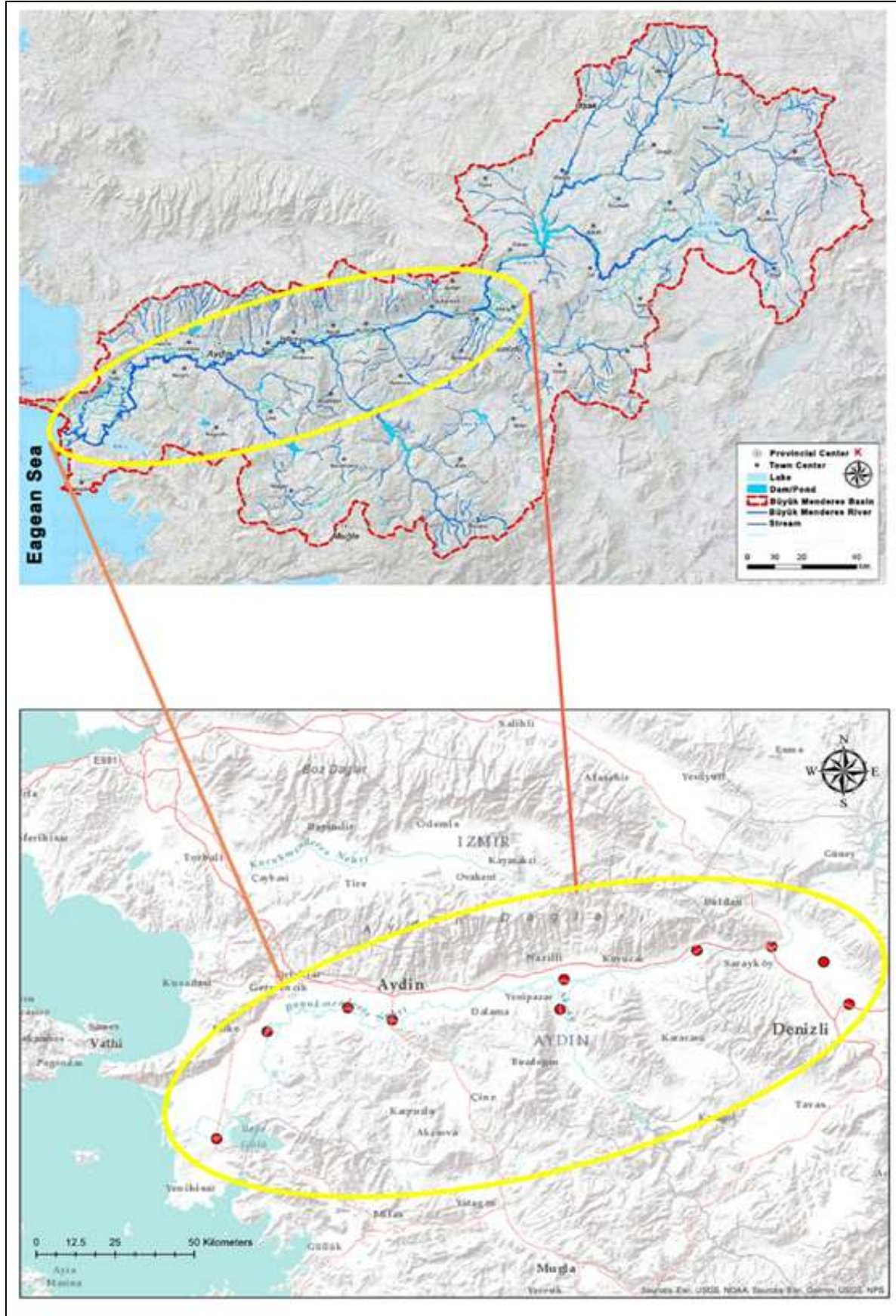
Çizelge 3.2. Seçilen istasyonlar.

	İstasyon No	İstasyon Adı	Bölge	Havza	X	Y
1	07-21-00-001	Gökpınar Çayı-Akhan Regülatörü	21.Böl. Md. Aydın	Büyük Menderes	688175	4188007
2	07-21-00-002	Büyükenderes Nehri-Sarayköy Köprüsü	21.Böl. Md. Aydın	Büyük Menderes	668510	4202450
3	07-21-00-003	Büyükenderes Nehri-Nazıllı Köprüsü	21.Böl. Md. Aydın	Büyük Menderes	616797	4193041
4	07-21-00-004	Büyükenderes Nehri-Aydın Köprüsü	21.Böl. Md. Aydın	Büyük Menderes	573960	4182106
5	07-21-00-005	Büyükenderes Nehri-Koçanlı Köprüsü	21.Böl. Md. Aydın	Büyük Menderes	562807	4185150
6	07-21-00-006	Büyükenderes Nehri-Söke Regülatörü	21.Böl. Md. Aydın	Büyük Menderes	542660	4178840
7	07-21-00-018	Çürüksu-Yukarı Samlı	21.Böl. Md. Aydın	Büyük Menderes	681600	4198810
8	07-21-00-045	Büyükenderes Nehri-Akçay Regülatörü	21.Böl. Md. Aydın	Büyük Menderes	615876	4185311
9	07-21-00-055	Büyükenderes Nehri-Feslek Regülatörü	21.Böl. Md. Aydın	Büyük Menderes	649844	4201157
10	07-21-00-096	Büyükenderes Nehri-Söke Mılas Karayolu Köprüsü	21.Böl. Md. Aydın	Büyük Menderes	530116	4150969

Tüm su kalitesi verilerinde bulunması gereken ve su kalitesinde insan sağlığı açısından aşırı önem arz eden Fekal Koliform ve pestisit miktarı değerleri de bu havzada ya ölçülmemiş yada düzensiz ölçüldüğü belirlenmiştir.

Elde edilen kullanılabilir veriler dikkatlice incelenmiş ve ölçümlerin yoğun olduğu yıllar ve aylar belirlendikten sonra, 2004-2014 yılları arasındaki Şubat, Nisan, Haziran, Ağustos,

Ekim, Aralık aylarına ait veriler kullanılarak birleşik bir veri seti oluşturulmuştur. Veri entegrasyonu adı verilen veri ön işlemenin önemli bir aşaması olan farklı kaynaklardan gelen verilerin birleştirilmesi, herhangi bir çalışmanın amacına hizmet etmektedir (Kumar ve Manjula, 2012). Birleşik veri seti örneği, Çizelge 3.1'de gösterilmektedir. Ancak, yeni veri setinin oluşturulmasının ardından ciddi bir kayıp veri problemi tespit edilmiştir.



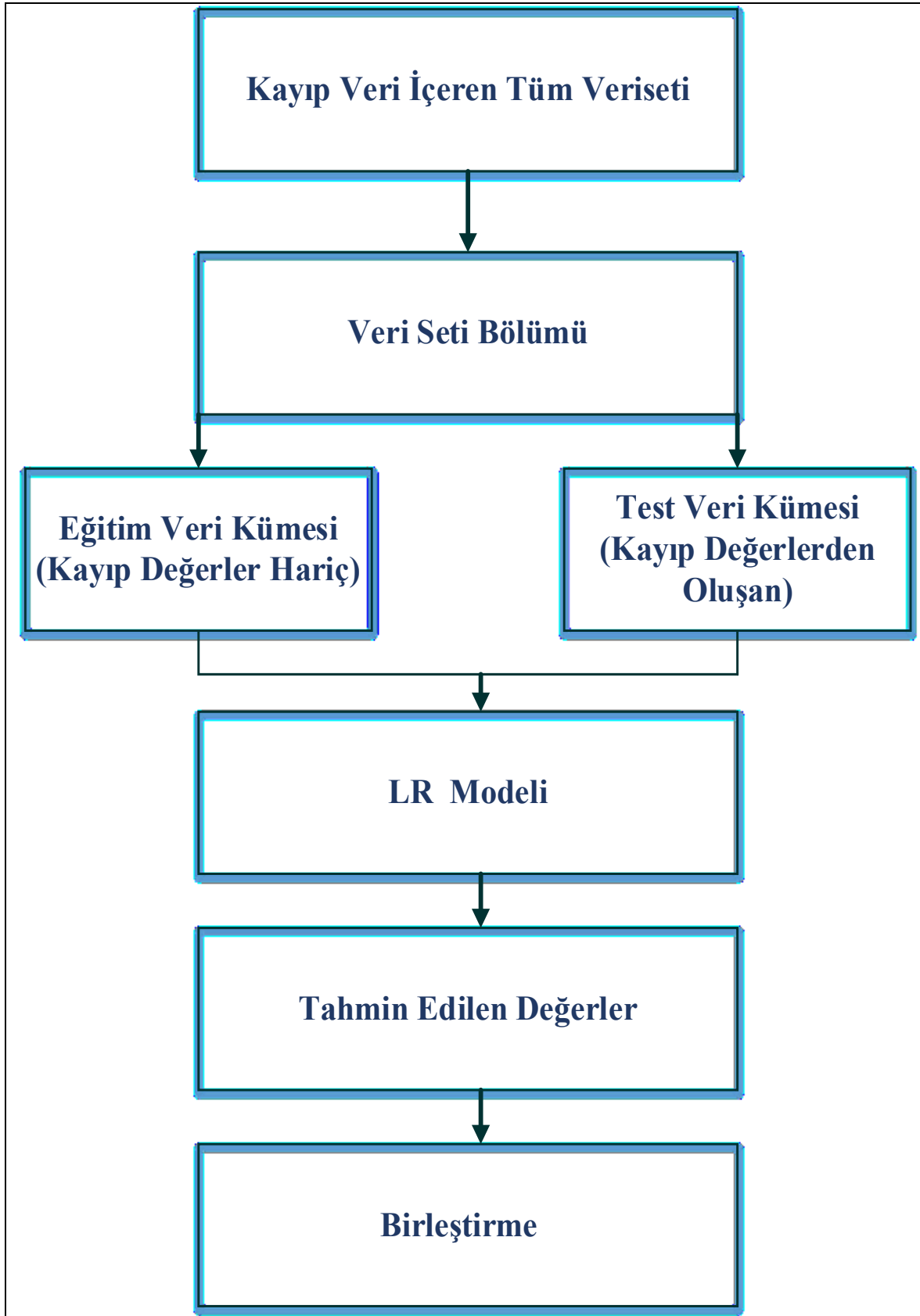
Şekil 3.1. Büyük Menderes Havzası ve bu çalışma için seçilen istasyonlar.

3.1.1. Kayıp veri inşası

Çevresel çalışmalarda, sürekli süreçlerin gözlemlenmesinin mümkün olmaması nedeniyle kayıp değerlerle karşılaşmak kaçınılmaz bir durumdur. Ayrıca, eksik değerlere sahip bozuk bir veri seti, veri analizi sürecindeki veri tutarsızlığı nedeniyle yanlış sonuçlara sebep olabilir. Birçok çalışmada veri temizliği için kullanılan sıfır atama, ortalama atama, medyan atama, regresyon atama gibi birçok kayıp değer atama yöntemi olmasına rağmen, makine öğrenmesi algoritmalarının kayıp değerleri tahmin etme ve veri kümelerini tamamlama yeteneği mantıksal değerlere sahip olması, son yıllarda bu teknikleri veri temizlemede kullanmak için araştırmacıların dikkatini çekmiştir. Bu çalışmada kapsamında, kayıp değer tahmini için kullanılan birçok makine öğrenmesi algoritmasından, LR algoritması seçilmiştir. Bu algoritma tüm veri setinin %10'undan fazlasının kayıp değerler olduğu eksik değerlerin tahmin edilmesindeki etkinliği ve uygulanmasının basitliği nedeniyle seçilmiştir. LR algoritması, bağımsız değişkenler (X) dikkate alınarak yapılan regresyona dayalı olarak hedef değeri (Y) tahmin ederek basitçe bir regresyon görevi gerçekleştirir (Eşitlik 3.1) (Kumar ve Manjula, 2012; Yozgatligil ve diğerleri, 2013; Sim ve diğerleri, 2015; Ighalo ve diğerleri, 2020).

$$Y_i = \beta_0 + \beta_i X_i + \varepsilon_i \quad (3.1)$$

Kayıp değer ataması için LR modeli geliştirmenin şematik çerçevesi Şekil 3.2'de gösterilmektedir. Burada veri seti açıkça tahmin süreci için test veri setini oluşturmak için tüm veri setinden eksik değerler çıkarılarak eğitim ve test veri setlerine bölünmüştür. Daha sonra LR modeli, değişkenler arasında regresyon yapmak ve eğitim verilerinden öğrenmesi sonucunda test veri setindeki değerleri tahmin etmek için eğitim veri seti tarafından eğitilmiştir. Sonunda, tahmin edilen değerler, kayıp değerler olmadan bütün bir veri seti oluşturmak için eğitim veri seti ile birleştirilmiştir.



Şekil 3.2. LR makine öğrenme algoritması ile eksik değer atamasının şematik süreci.

3.1.2. WQI modelin kurulması ve su kalitesi sınıflarının belirlenmesi

Veriler temizlendikten sonra, toplu bilgiyi azaltarak genel kullanıma yönelik suyun genel kalitesini tanımlamak için tek bir mantıksal dönüşüm hazırlanmıştır. Bu amaca hizmet edecek bir araç, su sistemlerinin genel durumunu tanımlamak için dünya çapında ulusal ve uluslararası kuruluşlar tarafından geliştirilen fonksiyonlar olan WQI kullanılmaktadır. Genel olarak, WQI işlevleri, bir su sistemindeki olası kontaminasyona ve ayrıca uzmanların görüşlerine bağlı olarak, ilk olarak ilgilenilen parametrelerin seçildiği dört aşamalı bir süreci takip etmektedir. Daha sonra tek değerli bir alt endeks oluşturmak için konsantrasyonlar dönüştürülür. Üçüncü aşama, her bir su kalitesi parametresi için birim ağırlığın hesaplanmasıdır. Son olarak, WQI'yı gösteren bir değer, önceden hesaplanmış alt endekslere ve birim ağırlıklara dayalı olarak hesaplanır (Tyagi ve diğerleri, 2013; Madalina ve Gabriela, 2014; Uddin ve diğerleri, 2021).

Bu çalışmada, Büyük Menderes Havzası'nın 2004-2014 yılları arasında 10 istasyonda iki aylık aralıklarla gözlemlenen 16 su kalitesi parametresine göre su kalitesi sınıf etiketlerini oluşturmak için WAWQI yöntemi seçilmiştir (Eşitlik 3.2).

$$WAWQI = \sum Q_i W_i / \sum W_i \quad (3.2)$$

Bu denklemde,

Q_i , i 'inci parametrenin kalite derecesi ve W_i ağırlık birimidir. Bu değerler sırayla denklem (Eşitlik 3.3) ve (Eşitlik 3.4) ile hesaplanırlar.

$$Q_i = 100 * [(V_i - V_0) / (S_i - V_0)] \quad (3.3)$$

Bu denklemde,

V_i , i 'inci parametresinin tahmini konsantrasyonu ve V_0 , i 'inci parametresinin ideal değeridir.

$$W_i = K / S_i \quad (3.4)$$

Bu denklemde,

S_i , i 'inci parametrenin tavsiye edilen standart değeridir ve K , Eşitlik 3.5 ile hesaplanmak üzere orantı sabitidir.

$$K = \frac{1}{\sum 1/S_i} \quad (3.5)$$

Bu çalışmada, WQI hesaplamasına dâhil edilen yukarıda bahsedilen parametreler için önerilen standart değerler WHO İçme Suyu Rehberi (WHO, 2006) ve Türk Standartlarından (TS-266) İnsan Tüketimi Amaçlı Su (TSE, 2005) alınmıştır (Çizelge 3.3).

Çizelge 3.3. Su kalitesi endeksi denkleminde kullanılan parametrelerin standart değerleri.

Parametre	Standart	Standart değer
pH	TS-266	8,5
Sıcaklık	TS-266	25
İletkenlik	TS-266	1000
ÇO	TS-266 (Ortalama)	7
KOİ	TS-266 (Ortalama)	50
BOİ-5	TS-266 (Ortalama)	8
NH ₄	TS-266	1
NO ₂	WHO, 2006	0,2
NO ₃	WHO, 2006	50
Klorür	WHO, 2006	250
Fe	WHO, 2006	2
Mn	WHO, 2006-TS-266	0,4
Na	WHO, 2006	200
SO ₄	WHO, 2006	500
TÇK	WHO, 2006	1000
PO ₄	TS-266	0,4

Bu çalışmada kullanılan parametrelerin standart değerleri tek bir kılavuzda bulunmamaktadır (Çizelge 3.4). Bazı parametrelere de sunulan referans değerler tek bir değer olarak verilmeyip referans aralıklar arasında belirtilmektedir. Bu problemi ortadan kaldırmak için bu çalışmada standart değerler elde etmek için iki kılavuz kullanıldı (WHO Kılavuzu ve Türkiye Standartları TS-266). Aralık olarak verilen standart değerlerin ortalaması alınmıştır (TS-266 (Ortalama)). Bazı standart değerler hem WHO kılavuzu hemde TS-266'da aynı olmak (WHO, 2006-TS-266) üzere alınmıştır (Çizelge 3.4).

Bu aşamadan sonra yukarıdaki eşitlikler kullanılarak K , S_i ve W_i değerleri belirlenmiştir (Çizelge 3.4). Çizelge 3.4'de görüldüğü gibi Q_i hesaplamasında kullanılan parametrelerin

ideal değerlerine ihtiyaç bulunmaktadır. pH, sıcaklık ve ÇO dışında yüksek kaliteli bir suda herhangi bir harici madde bulunması istenilmemektedir (Safsu mantığı). Ancak yüksek kaliteli bir su oksijen doygunu (ÇO miktarı en üst değer yani 14 mg/L olmalı) olmalı ve pH ve sıcaklık değerleri de standartlara göre sırayla 7 ve 14° C olmalıdır. Bu nedenle bu çalışmada bu üç parametre dışında Q_i hesaplamasında diğer parametrelerin ideal değeri sıfır olarak alınmıştır.

Çizelge 3.4. Standart değerlere göre S_n , K , ve W_i hesaplaması.

Parametre	Standart	Standart değeri (S_n)	$1/S_n$	$\Sigma 1/S_n$	$k=1/(\Sigma 1/S_n)$	$W_i=k/S_n$	İdeal değeri (V_0)
pH	TS-266	8,5	0,117647059	11,9785042	0,083482878	0,009821515	7
Sıcaklık	TS-266	25	0,04	11,9785042	0,083482878	0,003339315	16
İletkenlik	TS-266	1000	0,001	11,9785042	0,083482878	8,34829E-05	0
ÇO	TS-266-Ortalama	7	0,142857143	11,9785042	0,083482878	0,011926125	14
KOİ	TS-266-Ortalama	50	0,02	11,9785042	0,083482878	0,001669658	0
BOİ-5	TS-266-Ortalama	8	0,125	11,9785042	0,083482878	0,01043536	0
NH ₄	TS-266	1	1	11,9785042	0,083482878	0,083482878	0
NO ₂	WHO	0,2	5	11,9785042	0,083482878	0,417414388	0
NO ₃	WHO	50	0,02	11,9785042	0,083482878	0,001669658	0
Klorür	WHO	250	0,004	11,9785042	0,083482878	0,000333932	0
Fe	WHO	2	0,5	11,9785042	0,083482878	0,041741439	0
Mn	WHO-TS-266	0,4	2,5	11,9785042	0,083482878	0,208707194	0
Na	WHO	200	0,005	11,9785042	0,083482878	0,000417414	0
SO ₄	WHO-TS-266	500	0,002	11,9785042	0,083482878	0,000166966	0
TÇK	WHO	1000	0,001	11,9785042	0,083482878	8,34829E-05	0
PO ₄	TS	0,4	2,5	11,9785042	0,083482878	0,208707194	0

Ağırlıklı aritmetik su kalitesi endeksinde W_i Çizelge 3.5'e göre hesaplanmış ve toplamı 1 olduğu için bu eşitlikte sadece her bir parametre için W_n ve Q_n değeri ilgili eşitliklere göre hesaplanmış ve sonrasında veri setinde her bir sıra için (bir sırada yıl, ay, istasyon ve ilgili parametreler bulunmakta ve her bir sıra bir veriyi göstermektedir) her bir parametre için bu iki değerlerin çarpımı ve sıra içinde bu çarpımların aritmetik toplamı bir rakama dönüşmüştür. Bu rakam su kalitesi endeksi ağırlıklıların aritmetik toplamıdır. WAWQI fonksiyonunun sonucuna göre, su kalitesi Çizelge 3.5'e göre beş ana sınıfa ayrılmıştır.

Çizelge 3.5. Ağırlığa Göre Su Kalitesi Sınıfları Aritmetik Su Kalitesi Endeksi Yöntemi

WQI Değer Aralığı	Sınıf
0-25	Mükemmel Su Kalitesi (Excellent Quality)
26-50	İyi Su Kalitesi (Good Quality)
51-75	Kötü Su Kalitesi (Bad Quality)
76-100	Çok Kötü Su Kalitesi (Very Bad Quality)
100 ve üstü	Tüketim için Uygun Değil (Unfit for Consumption)

Bu aşamada ağırlıkların aritmetik toplamı veri seti üzerinde yazılan bir yazılım kodu ile her bir sırayı belli bir yıla, aya ve istasyona ait su kalitesini Çizelge 3.5'ye göre sınıflara ayrılmıştır.

3.2. Sınıflandırma ve Tahmin Çalışmalarında Kullanılan Öz Nitelik Seçimi Yöntemleri

Öznitelik seçim yöntemleri, yalnızca ilgisiz özelliklerin ortadan kaldırılmasına katkıda bulunan özelliklerin boyutlarını azaltmakla kalmaz, aynı zamanda ana sınıf/sınıflarla ilgili en etkili özellik alt kümesini de seçmektedir. Bu nedenle, veri kümesinin boyutunu küçültür, eğitim süresini kısaltır, sınıflandırmanın doğruluğunu artırır ve aşırı uyum sorunlarını azaltır. Öznitelik seçiminden kaynaklanan önemsiz küçük bir alt küme sınıflandırıcıyı yanlış yönlendirebileceğinden, en önemli öznitelikleri seçmek önemlidir. FLTM, WRAPM ve EMBM yöntemleri, özellik seçim yöntemlerinin üç kategorisi olarak sıralanır (Bolón-Canedo ve diğerleri, 2012; Mehr and Polat, 2022).

3.2.1. Öznitelik seçimi FLTM yöntemi

FLTM'ler, sıralayarak değişken seçimi için temel kriter olarak değişken sıralama tekniklerini kullanır. Sıralama yöntemleri basit olmaları nedeniyle kullanılmaktadır ve pratik uygulamalar için iyi başarı göstermektedirler. Bu işlemler sırasında değişkenleri puanlamak için uygun bir sıralama kriteri ve eşiğin altındaki değişkenleri çıkarmak için eşik değeri kullanılır. Sıralama yöntemleri, daha az ilgili değişkenleri filtrelemek için sınıflandırmadan önce uygulandıkları için FLTM'lerdir. Benzersiz bir özelliğin temel özelliği, verilerdeki farklı sınıflar hakkında faydalı bilgiler içermesidir. Bu özellik, farklı sınıfları ayırt etmede özelliğin kullanışlılığının bir ölçümünü sağlayan özellik alaka düzeyi olarak da tanımlanabilir. Burada bir özelliğin ilgililiği konusu gündeme getirilmelidir, yani bir özelliğin veri veya çıktı ile alaka düzeyini nasıl ölçülebilir. Bir tanım, bir özelliğin sınıf etiketlerinden koşullu olarak bağımsız olması durumunda alakasız olarak kabul edilebilmesidir. Eğer bir özellik alakalı olursa, giriş verilerinden bağımsız olabileceğini ancak sınıf etiketlerinden bağımsız olamayacağını, yani sınıf etiketleri üzerinde hiçbir etkisi olmayan özelliğin atılabileceğini belirtir. Özellikler arası korelasyon, benzersiz özelliklerin belirlenmesinde önemli bir rol oynar. Pratik uygulamalar için, temel dağılım bilinmemektedir ve sınıflandırıcı doğruluğu ile ölçülür. Bu nedenle, farklı özellik kümelerini kullanarak aynı sınıflandırıcı doğruluğunu elde etmek mümkün olabileceğinden, optimal bir özellik alt kümesi benzersiz olmayabilir (Law ve diğerleri, 2004; Chandrashekar ve Sahin, 2014). FLTM yönteminde, ML algoritmaları uygulamak yerine, özniteliklerin öznitelikler arasındaki uzaklık veya korelasyon gibi özellikleri ve özniteliklerin diğerleri arasındaki sıralaması dikkate alınarak alt küme seçilir. FLTM yöntemleri, öznitelik seçiminde istatistiksel yöntemlerden yararlanması nedeniyle hızlı ve ölçeklenebilir yöntemler olarak kabul edilmektedir. Ancak özelliklerin bağımlılığını ve sınıflandırıcılarla etkileşimini gözden kaçırmalar (Khaire ve Dhanalakshmi, 2019).

Karşılıklı bilgi filtre yöntemi (Mutual Information Filter Method (MI-FLTM))

$I(X; Y)$, bilgi teorisi açısından X ve Y arasındaki karşılıklı bağımlılığı hesaplayan X ve Y'nin iki rastgele değişkeni arasındaki karşılıklı bilgiyi gösterir. Karşılıklı bilgi (MI), doğrusallık veya süreklilik gibi değişkenler arasında herhangi bir bağımlılığı dikkate almadığından, korelasyon katsayısı gibi doğrusal yöntemlerden daha geneldir ve bu da onu daha avantajlı

bir yöntem haline getirmektedir. Eşitlik 3.6, X ve Y değişkenlerinin karşılıklı bilgisini gösterir.

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (3.6)$$

X ve Y'nin marjinal olasılık kütle fonksiyonu, $p(x) = \Pr\{X = x\}$, $x \in X$, ve $p(y) = \Pr\{Y = y\}$, $y \in Y$ olarak gösterilir. Sonuç olarak, $I(X; Y)$ 'nin sürekli alanı Eşitlik 3.7'de sunulmaktadır.

$$I(X; Y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (3.7)$$

$p(x, y)$, X ve Y'nin birleşik olasılık yoğunluk fonksiyonunu gösterir ve $p(x)$ ve $p(y)$, X ve Y'nin marjını göstermektedir. Ayrıca, karşılıklı bilginin yüksek değeri, X ve Y arasındaki yoğun bağımlılığı gösterir. MI matrisinin hesaplanması, simetrik bir $m \times m$ matrisi ile gösterilir (Eşitlik 3.8), burada μ_{ij} , veri kümesinin (D) iki özniteliği arasındaki karşılıklı bilgisidir (Sefidian ve Daneshpour, 2019).

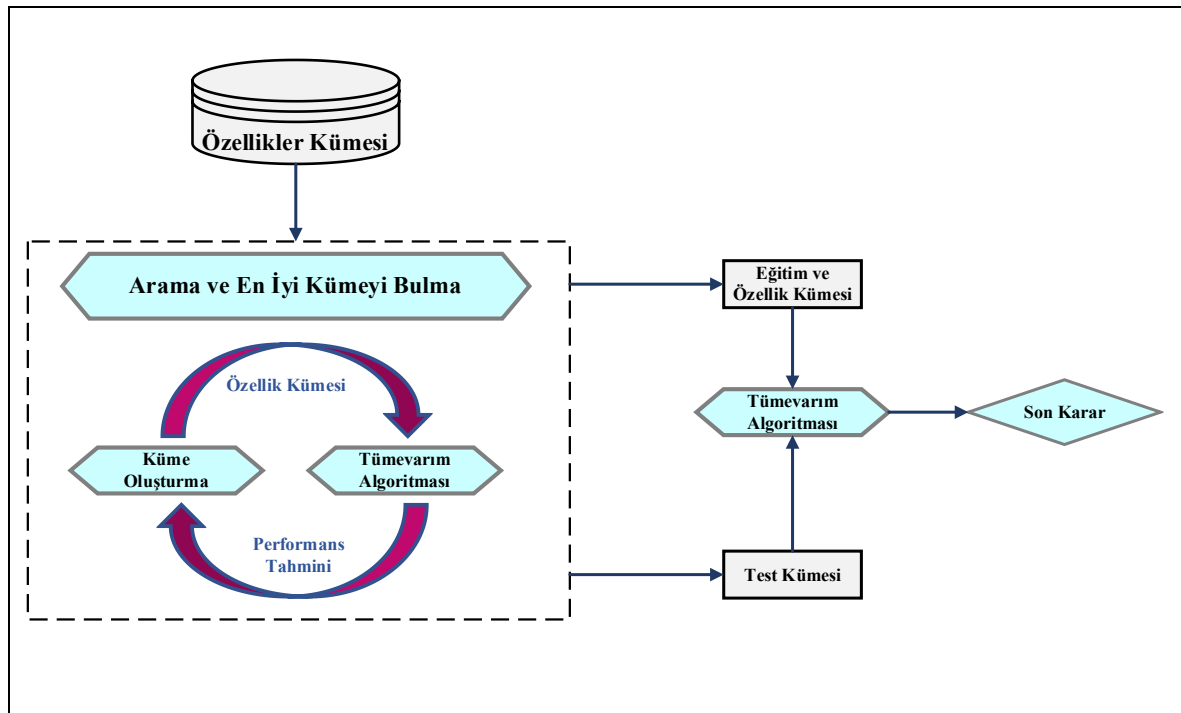
$$\mu(D) = \begin{bmatrix} 1 & \mu_{12} & \cdots & \mu_{1m} \\ \mu_{21} & 1 & \cdots & \mu_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{m1} & \mu_{m2} & \cdots & 1 \end{bmatrix} \quad (3.8)$$

3.2.2. Öznitelik seçimi sarmalayıcı yöntemi (Wrapper Method (WRAPM))

Yakın döngü (Close Loop Method (CLM)) öznitelik seçim yöntemi olarak adlandırılan Wrapper Method (WRAPM) yönteminde öznitelik seçici ML algoritması ile çalışır ve performansı sınıflandırma algoritmasının doğruluk oranı veya hata oranı ile değerlendirilir. WRAPM), öznitelik seçimi sürecinde denetimli öğrenme algoritmasını içermektedir. Bu yöntem, alt küme değerlendirme tekniğine göre özellikleri sıralamaktadır. Özellikler seçilirken özellikler arasındaki korelasyon ve bağımlılıklar dikkate alınır (Khaire ve Dhanalakshmi, 2019). Sonuç olarak, WRAPM yöntemi, özniteliklerin en önemli alt kümesini seçmede öğrenme algoritmasına dayandığından, sınıflandırıcının hata oranı göze çarpmamalıdır. WRAPM'ler, değişken alt kümesini değerlendirmek için bir kara kutu olarak tahmin ediciyi ve amaç fonksiyonu olarak tahmin edici performansını kullanır

(Chandrashekar ve Sahin, 2014). Tümevarım algoritmasının etrafını sarmak (Şekil 3.3) ve en iyi, en seçkin öz nitelikleri seçmek, sınıflandırma yada tahmin sürecinin doğruluk ve hata oranını kullanarak öz nitelikleri değerlendirmekte olan WRAPM öz nitelik seçimi yaklaşımı belirli bir uyarılmış sınıflandırıcının tahmin hatasını azaltır (Kohavi ve John, 1997; Zebari ve diğerleri, 2020). WRAPM yönteminin dezavantajı, geniş boyutlu uzaydan optimal kümenin aranmasından dolayı hesaplama maliyeti verimli değildir. Destek Vektör Makinesi (Support Vector Machine (SVM))- Özyinelemeli Özellik Eliminasyonu ve Greedy Forward Selection (GFS) stratejisi, WRAPM'in bazı örnekleridir (Khaire ve Dhanalakshmi, 2019).

Ek olarak, çoğu durumda, WRAPM yöntemi, FLTM yönteminden daha yüksek bir doğruluk oranı elde ettiği gibi, avantajlı bir öz nitelik seçme yöntemi olarak da kabul edilir. WRAPM, hesaplaması karmaşıktır ve aşırı uyum sorunlarına eğilimlidir (Zebari ve diğerleri, 2020). Amaç, sınıflandırıcı yada tahmin edici performansını en üst düzeye çıkaran n değişken arasından r boyutunda bir alt küme bulmaktır ($r < n$). Nihai karar kuralında yalnızca $r < n$, ve girdi değişkenlerinin çalıştırılması istendiği verilen yöntem, r özelliklerinin en iyi alt kümesini bulmaya çalışır. Bu yöntem en büyük sınıf ayrımı marjına yol açan r özelliklerini seçmeye çalışmaktadır (Maldonado ve Weber, 2009).



Şekil 3.3. WRAPM Öz Nitelik Seçimi çalışma akış şeması.

Geriye dönük özellik eleme sarmalayıcı yöntem (Backward Feature Elimination-based Wrapper Method (BFE-WRAPM))

İleriye yönelik öznitelik seçim yöntemine (Forward Feature Elimination-based Wrapper Method (FFE-WRAPM)) göre daha verimli olan BFE-WRAPM öznitelik seçimi yönteminde, her yinelemede örneklerin önem puanlarına tüm verilerle karar verilir ve en az öneme sahip öznitelikler elenir (Pan ve diğerleri, 2009). Bu yöntem, “r” özellik kalana kadar her seferinde bir özelliği kaldırır. Sonuç olarak, modelin performansını artırır. Son olarak, kalan tüm özellikler birleştirilir ve özellikler kaldırılarak daha fazla iyileştirme tanımlayana kadar tür yeniden başlar. Kaldırılan özellik, ortadan kaldırılması $W^2(\alpha)$ varyasyonunu en aza indiren özelliktir (Maldonado ve Weber, 2009):

$$W^2(\alpha) = \sum_{i,s=1}^m \alpha_i \alpha_s y_i y_s K(X_i, X_s) \quad (3.9)$$

Bu, modelin tahmin yeteneğinin bir ölçüsüdür ve marjla ters orantılıdır. Özellik eleme, aşağıdaki prosedür uygulanarak yapılır:

a) Bir α çözümü verildiğinde, her p özelliği için hesaplanması

$$W_{-p}^2(\alpha) = \sum_{i,s=1}^m \alpha_i \alpha_s y_i y_s K(X_i^{-p}, X_s^{-p}) \quad (3.10)$$

Bu eşitlikte,

X_i^{-p} , p özelliği kaldırılmış i eğitim nesnesi anlamına gelir.

b) En küçük $|W^2(\alpha) - W_{-p}^2(\alpha)|$ değerine sahip özelliğin kaldırılması

3.2.3. Gömülü öznitelik seçimi yöntemi (Embedded method (EMBM))

EMBM özellik seçim algoritması, yerleşik yöntem olarak da bilinir. Bu yöntemde en etkili alt kümeyi seçmek için öznitelik değerlendiriciye öğrenme algoritmasının bir parçası olarak EMBM öznitelik seçme yöntemi yardımcı olur. ANNs ve farklı türdeki DT gömülü yöntemlerin en yaygın öğrenme algoritmalarıdır. EMBM öznitelik seçiciler, FLTM ve

WRAPM yöntemlerine kıyasla, daha az hesaplama süreci ve öğrenme algoritması kullanarak öznitelik seçimine bağımlı olmaları nedeniyle daha optimum sonuçlar elde edebilirler (Chen ve diğerleri, 2020). Farklı problemlerin çözümünde faydalı sonuçları ve genelleştirilmesi nedeniyle bu çalışmada EMBM öznitelik seçim yöntemi olarak RF uygulanmıştır.

İlk adımda, torbalama yöntemiyle (Bagging Method (BM)) eğitim kümesindeki öznitelikler arasından farklı alt kümeler üretilir ve seçilen her bir öznitelik için farklı DT'ler oluşturulur. DT büyümesi, her bir düğümün bölünmesiyle ilgilidir. Bu nedenle düğümler, özniteliklerin aday alt kümeleri arasından rastgele olarak seçilen seçili özniteliklere göre bölünür. Daha sonra, düğümlerin ayrılır, budama yapılmadan büyütülen farklı ağaçlar üretilir ve her ağaç bir sınıflandırıcı olarak kabul edilir. Son aşamada, veri kümesindeki tüm özellikler arasından özelliklerin alt kümelerini seçen rastgele bir orman (RF) oluşturan tüm ağaçlar bağlanır (Zhou ve diğerleri, 2016).

3.3. Sınıflandırma ve Tahmin Çalışmaları

3.3.1. Toplu öğrenme (Ensemble Learning (EL))

Ensemble Learning (EL), genellikle SUPL makine öğrenimi görevlerinde bir karar vermek için birden çok uyarıcıyı birleştiren yöntemler için bir şemsiye terimdir. Temel öğrenen olarak da adlandırılan indükleyici, girdi olarak bir dizi etiketli örnek alan ve bu örnekleri genelleştiren bir model (örneğin bir sınıflandırıcı veya regresör) üreten bir algoritmadır. Üretilen model kullanılarak yeni etiketlenmemiş örnekler için tahminler oluşturulabilir. Bir topluluk indükleyicisi (Ensemble Inducer (EI)), herhangi bir türde ML algoritması olabilir (örneğin, DT, ANN, LR modeli, vb.). EL'nin ana öncülü, birden fazla modeli birleştirerek, tek bir indükleyicinin hatalarının muhtemelen diğer indükleyiciler tarafından telafi edileceği ve sonuç olarak topluluğun genel tahmin performansının tek bir indükleyicinininkinden daha iyi olacağıdır (Dong ve diğerleri, 2020). EL yöntemlerinin genellikle tahmin performansını iyileştirmesinin dört nedeni vardır:

- a) *Fazla uydurmadan kaçınma*: Sadece az miktarda veri mevcut olduğunda, bir öğrenme algoritması, tüm eğitim verilerini mükemmel bir şekilde tahmin ederken, görülmeyen örnekler için zayıf tahminler yapan birçok farklı hipotez bulmaya eğilimlidir. Farklı

hipotezlerin ortalamasını almak, yanlış bir hipotez seçme riskini azaltır ve bu nedenle genel tahmin performansını iyileştirir.

- *Hesaplama avantajı*: Yerel olarak çalışan tek öğrenen aramalar yerel optimumda takılabilir.
- *Birleştirme*: Birkaç öğrenme birleştirilerek, EL yöntemleri yerel bir minimum elde etme riskini azaltır.
- *Temsil*: Optimal hipotez, herhangi tek modelin uzayının dışında olabilir. Farklı modelleri birleştirilerek, arama alanı genişletilebilir ve dolayısıyla veri alanına daha iyi uyum sağlanabilir (Sagi ve Rokach, 2018).

EL modeli kurulumunda, n örnek ve m özellikten oluşan bir veri kümesi verildiğinde ($D = \{X_i, Y_i\} (|D| = n, X_i \in R^m, Y_i \in R)$) bir topluluk öğrenme modeli φ , K sayıda indükleyicileri toplayan bir G toplama işlevi kullanır. $\{f_1, f_2, \dots, f_k\}$ aşağıdaki gibi tek bir çıktıyı tahmin etmeye yöneliktir (Sagi ve Rokach, 2018):

$$\hat{Y}_i = \varphi(X_i) = G(f_1, f_2, \dots, f_k) \quad (3.11)$$

Bu eşitlikte,

Regresyon problemleri için $\hat{Y}_i \in R$ ve sınıflandırma problemleri için $\hat{Y}_i \in Z$ kullanılır. Bu genel çerçeve göz önüne alındığında, EL modeli oluşturmak, katılımcı modelleri eğitmek için bir metodoloji seçimini ve indükleyicilerin çıktılarını birleştirmek için uygun bir süreç seçmeyi içerir (Sagi ve Rokach, 2018).

EL yöntemleri iki ana çerçeveye ayrılabilir: Bağımlı çerçeve ve bağımsız çerçeve. Bağımlı çerçevede, her bir indükleyicinin çıktısı, bir sonraki indükleyicinin yapısını etkiler. Bu çerçevede, önceki yinelemelerde üretilen bilgi, bir sonraki yinelemede öğrenmeye rehberlik eder. Bağımsız çerçevede, her bir indükleyici, diğer indükleyicilerden bağımsız olarak inşa edilir (Sagi ve Rokach, 2018; Sefidian ve Daneshpour, 2019).

AdaBoost sınıflandırma modeli (Adaptive Boosting (AdaBoost))

AdaBoost Sınıflandırıcısı, aynı zamanda uyarlamalı güçlendirme olarak da adlandırılır (Freund ve Schapire, 1997). Bu yöntem en yaygın olarak kullanılan EL yöntemlerinin bir başka kategorisidir. Her zayıf öğrenilen unsurun performansını artırmak için sıralı bağımsız tekil hipotezlerin (Ordered Independent Single Hypotheses) kombinasyonu kavramını kullanır. Sonuç olarak, tüm bireysel sınıflandırıcılardaki eğitim setinin dağılımı değiştirilerek eğitim hatasının azaltılmasına katkıda bulunulur ve eğitim sürecinin sonunda her yinelemede ağırlıklar düzenlenir. Doğru tahmin edilen örneklerin ağırlığı azaltılacak ve EL'nin nihai sonucu, tüm tek sınıflandırıcıların sonuçlarının birleşimine dayanacaktır.

X giriş veri alanını ve Y olası sınıf etiketleri kümesini gösterebilir. İki sınıf durumunu ele alıyoruz. Burada, $Y = \{-1, +1\}$ ve $X = \mathbb{R}^n$ olduğunu varsayalım. Amacımız, $x \in X$ öznitelik vektörüne verilen (doğru) y sınıf etiketini hesaplayan bir $F: X \rightarrow Y$ eşleme fonksiyonu oluşturmaktır. Ayrıca, eğitim için bir dizi etiketli verinin mevcut olduğu durum da dikkate alınır (Vezhnevets ve Vezhnevets, 2005):

$$(X_1, Y_1), \dots, (X_n, Y_n); x_i \in X; y_i \in Y \quad (3.12)$$

AdaBoost tabanlı tüm teknikler, üstel hata fonksiyonunu en aza indirmek için açgözlü optimizasyon (Greedy Optimization (GO)) yöntemleri olarak kabul edilir (Vezhnevets ve Vezhnevets, 2005).

$$\sum_{i=1}^N e^{-y_i F(x_i)} \quad (3.13)$$

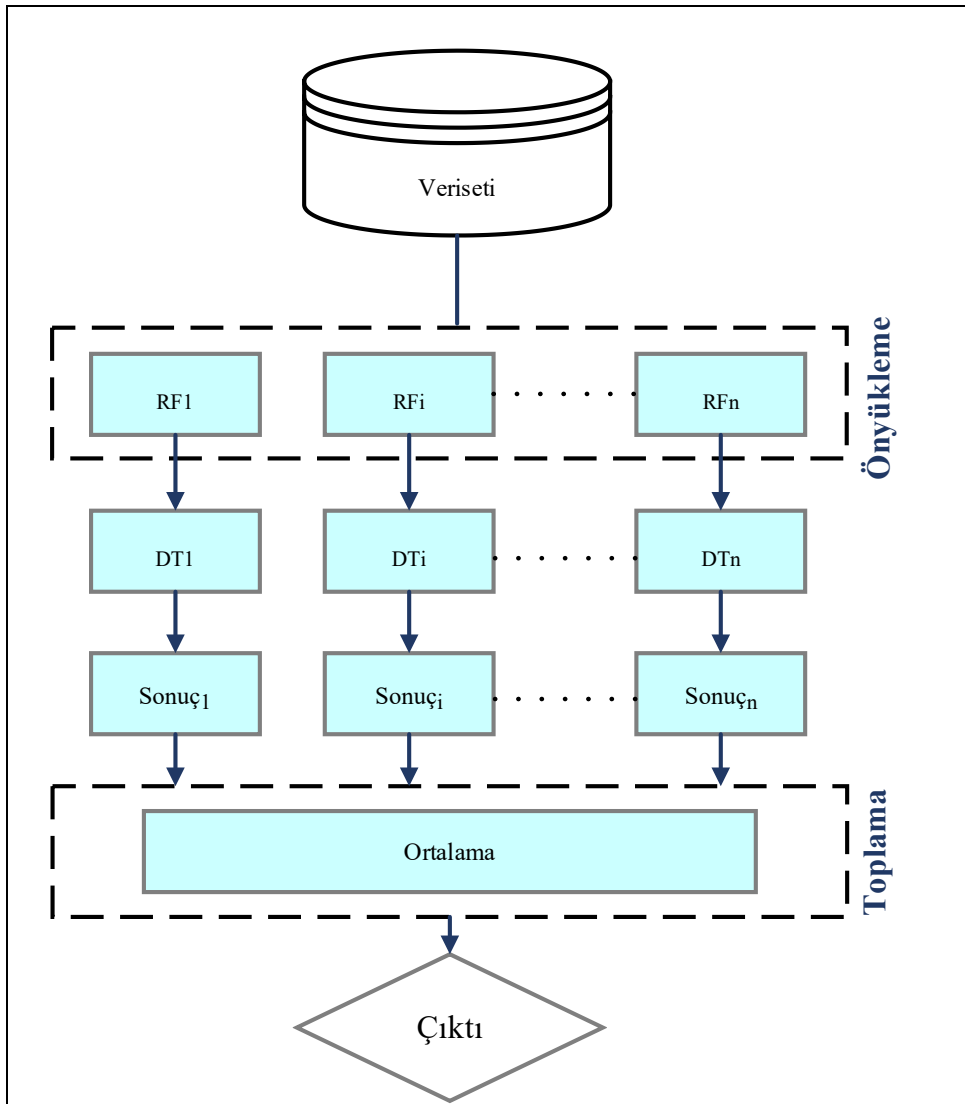
Bu denklemde, $F(x)$ yapılandırılmış sınıflandırıcı kararıdır. Boosting'in iyi yaklaşıklık yetenekleri teorik olarak kanıtlanmış olsa da, etkileyici genelleme yetenekleri hala yalnızca deneysel olarak onaylanmakta ve iyileştirmeler için gelecekteki araştırmalar için yeterli alan bırakmaktadır (Vezhnevets ve Vezhnevets, 2005).

Random forest (RF) sınıflandırma ve regresyon modeli

Torbalama tekniği (Bagging Method (BM)) ile çalışan bir EL modeli olan RF, genellikle DT modelleri olan birkaç temel modelden oluşmaktadır ($h(X; \Theta_k)$, $k=1, \dots, K$) (Şekil 3.4). Bu

model, özellikle veri kümesinin çok sayıda aykırı değerden ve gürültüden (Noise) oluştuğu durumlarda, sınıflandırma ve regresyon problemlerini çözmektedir. Bu modelde, her ağaç modeli, ormandaki tüm ağaçlar arasında eşit dağılımla, Θ_k veri kümesinin rastgele bir örneği tarafından beslenir. Ağaç modelleri tarafından elde edilen sonuçlar, ormanın çıktısı olmak için, p uzunluğundaki girdi vektörüne (x) dayalı olarak eşit ağırlıklarla toplanır ve ortalaması alınır (Breiman, 2001; Segal, 2004).

Sınıflandırma modellerinde sonuç bir sınıfa hitap ederken regresyon modellerinde sonuç sayısaldir (Y), eğitim verileri $n(p+1)$ - demetler $(x_1, y_1), \dots, (x_n, y_n)$ içeren rastgele vektörün (X, Y) dağılımından rastgele seçilir. RF tahmini ağırlıksız bir ortalamadır (Breiman, 2001; Segal, 2004).



Şekil 3.4. RF modelinin çalışma şeması

$$\bar{h}(x) = \frac{1}{K} \sum_{k=1}^K h(x; \theta_k) \quad (3.14)$$

Eğer $k \rightarrow \infty$:

$$E_{X,Y} \left(Y - \bar{h}(X) \right)^2 \rightarrow E_{X,Y} (Y - E_{\theta} h(X; \theta))^2 \quad (3.15)$$

Eşitlik 3.15'in sağ tarafı, RF'nin tahmin hatasını veya bazen PE_t^* olarak gösterilen sözde genellemeyi temsil eder. Burada yakınsama, RF'nin aşırı uyum problemlerini çözmedeki gücünü ortaya çıkarır (Breiman, 2001; Segal, 2004).

Her $h(X; \theta)$ ağacı için ortalama tahmin hatası aşağıdaki gibi tanımlanır:

$$PE_t^* = E_{\theta} E_{X,Y} (Y - h(X; \theta))^2 \quad (3.16)$$

Tüm Θ için ağaçların tarafsızlığı varsayılırsa

$$PE_f^* < \bar{\rho} PE_t^* \quad (3.17)$$

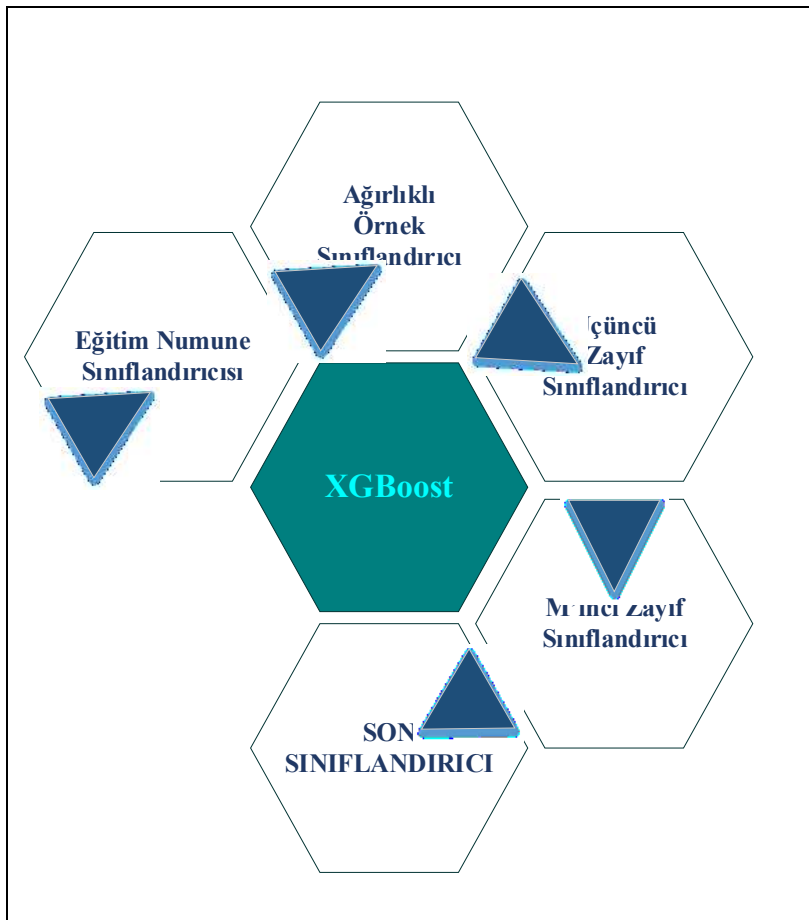
Burada $\bar{\rho}$, bağımsız $\theta, \hat{\theta}$ için $Y - h(X; \theta)$ ve $Y - h(X; \hat{\theta})$ artıkları arasındaki ağırlıklı korelasyonları temsil eder (Breiman, 2001; Segal, 2004).

RF yöntemi tarafından kullanılan BM, tahmin hatasını azaltır ve sonuç olarak ortalama alma, varyansı azaltır (Breiman, 2001; Segal, 2004).

Extreme gradient boosting (XGBoost) sınıflandırma modeli

XGBoost, EL tabanlı yaklaşımın bir parçası olan bir güçlendirme tekniğidir. Dağıtılmış ML topluluğuna aittir. Bu araç o kadar etkilidir ki, veri tabanındaki veri değerlerinin her birini kontrol eder. Bu yöntemin tanıtılmasından önce, bazı DT'lere aynı verilerin sağlandığı ve her bir DT'nin eğitilmesine ve her bir DT'nin ortalama doğruluğunun hesaplanmasına yol açan RF ismi verilen bir teknik kullanılırdı. XGBoost durumunda, sıralı EL tekniği olarak da adlandırılan sıralı bir DT oluşturulur. Bu yöntemde, veritabanındaki veri değerlerinin her birine, daha sonraki analizler için bir DT tarafından seçilme olasılığını tanımlayan bir ağırlık

değeri atanır. Başlangıç ağırlık değeri her veri değeri için aynıdır ve analize göre değişir. Veri değerlerinin ilk geçişinin sonuçları, önceki sonuçları koruyan ve bunun üzerine inşa edilen yeni bir sınıflandırma modelinin oluşturulmasına yardımcı olur ve bu süreç, son sınıflandırıcı oluşana kadar devam eder. Bu yöntem derinlik açısından çok karmaşıktır ve dolayısıyla bu ağaçları artırma yöntemi, eğitim verilerine uyması için daha iyi bir yeteneğe sahip olduğundan, varyansta düşük ve yüksek bir sapma içeren bir sonuç sağlamakla birlikte, RF yöntemi, yüksek ve düşük bir sapma içeren bir sonuç sağlar. XGBoost süreci Şekil 3.5'de gösterilmektedir.



Şekil 3.5. XGBoost modelinin çalışma şeması.

m sayıda özellik ve n sayıda örnek içeren bir veri kümesini ele alalım, burada, $D = \{(X_1, Y_1)\} (|D| = n, X_1 \in R^m, Y_1 \in R, D$ çıktıyı tahmin etmek için K toplama fonksiyonlarını kullanan bir ağacın topluluk modelidir (Eşitlik 3.18).

$$\hat{y}_1 = \varphi(x_1) = \sum_{k=1}^K f_k(x_1), f_k \in F \quad (3.18)$$

Eşitlik 3.18'de $F = \{f(x) = w_q(x)\} (q : R_m \rightarrow T, w \in R^T)$ ağaçların bir parçası olan uzay olarak tanımlanırken, q ağacı temsil eder. Yukarıdaki eşitlikte ağaçtaki yaprak sayısı T ile temsil edilmektedir. Burada her f_k bir ağaç yapısına (bağımsız), yani q 'ya ve yaprak ağırlıklarına yani w 'ye karşılık gelmektedir. Belirli bir örnek için, yaprakları sınıflandırmak için ağaçlarda (q ile verilen) karar kuralları kullanılır ve nihai tahmin, karşılık gelen yapraklardaki (w ile verilen) puan toplanarak hesaplanır. Modelde kullanılan fonksiyonlar kümesi hakkında bilgi sağlamak için Eşitlik 3.19'daki düzenlenileştirilmiş amaç minimize edilmiştir.

$$L(\varphi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (3.19)$$

Yukarıdaki eşitlik, $\Omega(f_k) = \gamma T + 12 \lambda ||w||$ ve fark, "diferansiyellenebilir" bir dışbükey kayıp fonksiyonu (Loss Function (LF)) olan l kullanılarak hedef y_i ile tahmin $\hat{y}_i$ arasındaki ölçülür. Ω , modelin karmaşıklığını azaltır. Son öğrenilen ağırlıkları yumuşattığı için ekstra düzenleme süresi nedeniyle aşırı uydurma önlenir. Bu nedenle basit ve kestirimci fonksiyonları kullanan model seçilecektir. Düzenleme parametresini sıfıra ayarladıktan sonra hedef, geleneksel gradyan ağacı güçlendirmesi olan varsayılan ayarlanır.

Modeli toplamalı bir şekilde eğitmek için, $\hat{y}_i^{(t)}$, t 'inci iterasyonun i 'inci örneği olduğunu var sayalım ve aşağıdaki amacı f_t ekleyerek minimize edilirse

$$L^{(t)} = \sum_{i=1}^n l(\hat{y}_i, \hat{y}_i^{t-1} + f_t(X_i)) + \Omega(f_t) \quad (3.20)$$

Eşitlik 3.20'de, f_t 'nin eklenmesi, hedefi hızlı bir şekilde optimize eden model doğruluğunu iyileştirir ve ikinci dereceden yaklaşıklık gereklidir.

$$L^{(t)} = \sum_{i=1}^n \left(l(\hat{y}_i, \hat{y}_i^{t-1}) + g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i) \right) + \Omega(f_t) \quad (3.21)$$

Burada $g_i = \partial_i g_i^{t-1} l(\hat{y}_i, \hat{y}_i^{(t-1)})$, $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(\hat{y}_i, \hat{y}_i^{(t-1)})$ ikinci dereceden istatistiklerdir "kayıp fonksiyonuna" aittir.

Daha basit olan Eşitlik (3.21)'de aşağıdaki fonksiyonu elde etmek için sabit fonksiyonlar da kaldırılabilir.

$$L^{(t)} = \sum_{i=1}^n \left(g_i f_t(X_i) + \frac{1}{2} h_i f_t^2(X_i) \right) + \Omega(f_t) \quad (3.22)$$

XGBoost yönteminin algoritması aşağıdaki adımlarda açıklanabilir:

- a) Amaç fonksiyonu, $\text{Obj}(\theta) = \text{TL}(\theta) + \text{R}(\theta)$ Burada $\text{TL}(\theta)$ eğitim kaybı ve $\text{R}(\theta)$ düzenlemedir.
- b) Taylor teoremi kullanılarak yeni amaç fonksiyonu oluşturulur.

$$\text{Obj}^t = \sum_{i=1}^n \left[m_i f_t(p_i) + \frac{1}{2} C_i f_t^2(p_i) \right] + \text{R}(f_t) \quad (3.23)$$

m_i ve C_i girdilerdir. Amaç fonksiyonunun sonucu, kendisini modele ekleyen bir ağaçtır.

- Ağacın karmaşıklığını düzenlemek için, $f_t(p) = w_{q(p)}$, $w \in R^L$, $q : R^d \rightarrow \{1, 2, 3, \dots, L\}$ L yaprak sayısıdır, q yaprakları ilgili veri noktalarına atar.

$$\text{R}(f) = \alpha \left(\alpha L + \frac{1}{2} \beta \sum_{j=1}^L w_j^2 \right) \quad (3.24)$$

Yukarıdaki denklem (Eşitlik 3.24), XGBoost sınıflandırıcısının karmaşıklığını verir.

- Her bir yaprak üzerinde ikinci dereceden gradyan istatistikleri ile gradyan eklenmesi üzerine puanlama formülünün uygulanması yardımı ile kalite skoru elde edilebilir (Bhati ve diğerleri, 2021).

LogitBoost (Logistic Boosting (LogitBoost)) sınıflandırma modeli

İlk olarak Schapire ve Singer (1998) tarafından tanıtılan LogitBoost, son yıllarda geliştirilen boosting algoritmalarından birisidir. Boosting, başlangıçta, sınıflandırma performansını iyileştirmek için birkaç zayıf sınıflandırıcıyı birleştirmek için önerilmiştir. Daha sonra,

Freund ve Schapire (1997) tarafından “AdaBoost” olarak adlandırılan daha yetenekli ve pratik bir artırma algoritması önerilmiştir. Adaptive Boosting’in kısaltması olan AdaBoost, bir meta (MA)-öğrenme algoritmasıdır. AdaBoost önceki zayıf sınıflandırıcıların performansına göre yinelemeli olarak diğerleri üzerinde zayıf bir sınıflandırıcı oluşturmaya çalışır. Buna göre AdaBoost, önceki zayıf sınıflandırıcılarla doğru bir şekilde sınıflandırılmayan örnekler daha fazla ağırlık vererek zor örnekler odaklanmaya yönlendirilir. AdaBoost, zayıf sınıflandırıcılar kullanarak eğitim hatalarını azaltmaktadır. AdaBoost’un çok iyi bir genelleme (yeni verileri sınıflandırma yeteneği) olduğu gözlemlenmektedir. Ancak, diğer sınıflandırıcıların çoğu gibi, AdaBoost da çok gürültülü verilerle uğraşırken aşırı uyum problemi konusunda sıkıntılıdır. Bu durumla başa çıkmak için Friedman ve diğerleri (2000), LogitBoost kullanmanın eğitim hatalarını doğrusal olarak azaltabileceğini ve dolayısıyla daha iyi genelleme sağlayabileceğini tespit etmiştir. AdaBoost, monoton ve pürüzsüz olan ve etkin bir şekilde çözülebilen üstel bir LF (Eşitlik 3.25) beklentisini en aza indirmek için ek bir LGR modeline (Eşitlik 3.26) uygun olarak kabul edilebilir.

$$F(\vec{X}) = \sum_{t=1}^T \alpha_t f_t(\vec{X}) \quad (3.25)$$

$$ELOSS(F) = E(e^{-yF(\vec{X})}) \quad (3.26)$$

Üstel kayıp işlevi sınıflandırma hatasıyla katlanarak değişir ve gürültülü verileri işlerken AdaBoost algoritmasını savunmasız hale getirmektedir. Problemi çözmek için Friedman ve diğerleri (2000), sınıflandırma hatası ile doğrusal olarak değişen ve gürültü ve aykırı değerlere karşı daha az duyarlı olduğu ortaya çıkan bir binom log-olasılık kaybı fonksiyonu (Log-Probability Loss Function (LLOSS)) (Eşitlik 3.27) önermiştir.

$$LLOSS(F) = E \left[-\log(1 + e^{-yF(\vec{X})}) \right] \quad (3.27)$$

Optimizasyon, eklemeli simetrik bir lojistik modele uyması için Newton adımları kullanılarak gerçekleştirilebilir (Cai ve diğerleri, 2006; Tehrany ve diğerleri, 2019).

3.3.2. ANN ve MLP regresyon (Tahmin) modeli

İnsan beyninin öğrenme modelini taklit eden ANN'ler, mantıklı kararlar vermek için karmaşık bilgileri işlemek için üretilmiştir. ANN'ler genel olarak üç katmandan (giriş katmanı, gizli katman, çıktı katmanı) oluşmaktadır (Şekil 3.6). Bu katmanlar, insan beyninin nöron sistemini taklit eden düğümler aracılığıyla birbirine bağlıdır. Düğümler arasındaki bağlantı, ağırlık adı verilen bir sayı ile ilgilidir. Bu sayıya ağırlık denir. Bu yapıda veri, giriş katmanından (Input Layer) ağa girer, gizli katmanda (Hidden Layer) veriler işlenir ve mantıklı bir karar alındıktan sonra çıkış katmanında (Output layer) ağdan çıkar (Hopfield, 1988). Bazen de bir ANNs'de birden fazla gizli katman olmaktadır. İki gizli katmana sahip ağlara örnek verecek olursak, herhangi bir karmaşıklıkta karar sınırlarını oluşturabilen ağlar olarak tanımlanabilir. İlk gizli katman, özellik alanını birkaç bölgeye ayırırken, ikinci gizli katman farklı bölgeleri, her bir sınıf için dış bükey bölgeler oluşturmak üzere bir araya getirir. Gizli katmanın çıktısı Eşitlik 3.28 ile tanımlanır. Çıktı katmanı, rastgele ayırık bölgeler oluşturmak için önceki katmanlar ile birlikte çıktı verir. Öğrenme sırasında, ağlar bir kavram temsili bir çok birime dağıtılabilir veya nöronları bireysel "alt görevlere" atayabilir (Dutton ve Conroy, 1997).

$$h_i = \sigma(\sum_{j=1}^n V_{ij}X_j + T_i^{hid}) \quad (3.28)$$

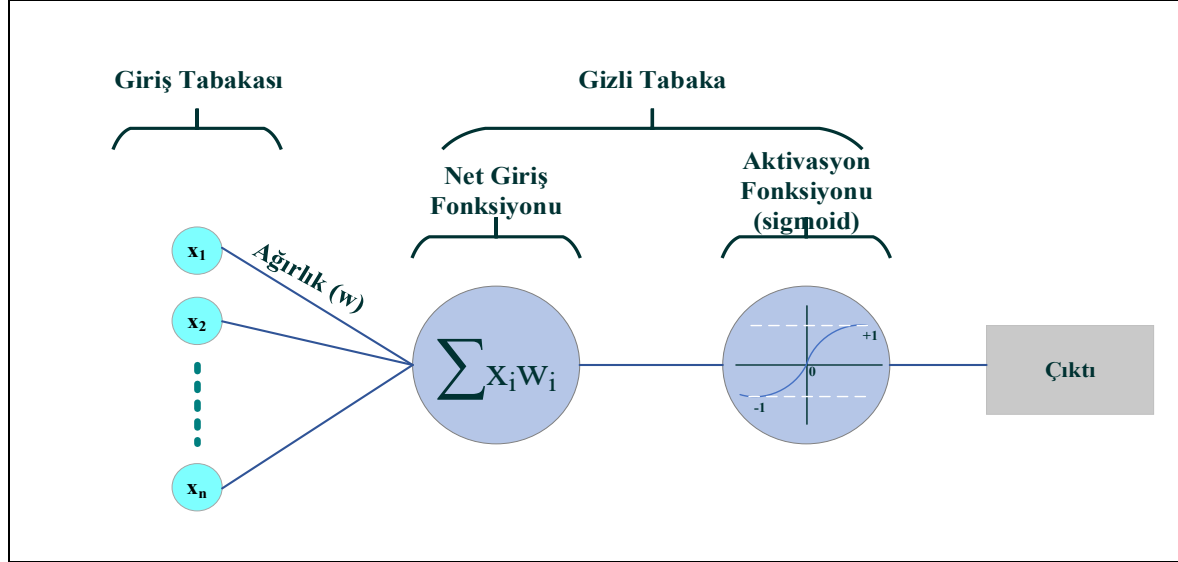
Burada, $\sigma()$, N , V_{ij} , X_j , and T_i^{hid} sırasıyla aktivasyon fonksiyonu, girdi sayısı, ağırlıklar, girdi düğümlerine girdiler ve eşiklerdir.

Temel olarak, sistemin girdi ve çıktısı (aktivasyon fonksiyonu veya transfer fonksiyonu) arasındaki ilişki sigmoiddır (Eşitlik 3.29).

$$\sigma(v) = \frac{1}{1+\exp(-v)} \quad (3.29)$$

Verilere ve çalışmanın amacına bağlı olarak bu işlevler değişiklik gösterebilir (örneğin, ark tanjantı, hiperbolik tanjant vb.). Ağa doğrusal olmayanlığı tanıtmak için yalnızca aktivasyon işlevi kullanılmaz, aynı zamanda nöronların değerini de birbirine bağlar. Bu nedenle, ağı farklı nöronlar tarafından felç edilmesi sorunu çözülmüş olur. Ayrıca, ANNs bir mantıksal

kararı diğerine besleyerek, düğümler arasındaki bağlantıyı ileri beslemeli yapmış olur (Hopfield, 1988; Wang, 2003).



Şekil 3.6. Yapay sinir ağı çalışma akış şeması.

3.4. Sınıflandırma Çalışmaları İçin Değerlendirme

Çok sınıflı bir karışıklık matrisi (Confusion Matrix (CM)), çok sınıflı veri kümelerine uygulanan sınıflandırıcıların performans değerlendirme için bir araçtır. Başka bir ifade ile, bir sınıflandırıcının performansı, çok sınıflı bir CM ile değerlendirilir. CM'den, pozitif (P) veya negatif (N) koşullardan kaç tanesinin doğru şekilde sınıflandırıldığı (yani, True Positive (TP), True Negative (TN)) veya tersinin (yani, False Positive (FP) olduğu algılanır) False Negative (FN)) anlaşılır (Fatourehci ve diğerleri, 2008). Bu bilgiyi çok sınıflı karışıklık matrisinden çıkararak, bir sınıflandırıcının performansını tanımlamak için çeşitli metrikler (yani, Doğruluk (Accuracy), Kesinlik (Precision), Geri Çağırma (Recall), F1 puanı, Yaygınlık (Prevalance), phi katsayısı vb.) hesaplanabilir (Vujović, 2021; Danaei Mehr ve Polat, 2022). Çizelge 3.6. şematik birçok sınıflı CM'yi göstermektedir. Burada matrisin birincil köşegenindeki (X_{11}, X_{ij}, X_{nn}) elemanlar, bir sınıflandırma sürecinde doğru sınıflandırılan örnekleri (yani, TP & TN) gösterir (Dalianis ve Dalianis, 2018). Bu çalışmada, sınıflandırıcıların performansını değerlendirmek için üç metrik kullanılmıştır. Eşitlik 3.30 ile tanımlanan sınıflandırmanın doğruluğu, pozitif olarak sınıflandırılan tüm örneklerin kesridir. Ayrıca, Eşitlik 3.31 ve Eşitlik 3.32 ile tanımlanan sırayla kesinlik ve geri çağırma (Mehr ve Polat, 2019), bu çalışmada sınıflandırmanın CM'den doğrulukla birlikte

hesaplanan diğer metriklerdir. Geri çağırma (Hassasiyet) su kalite sınıfı etiketlerinin ne kadarının doğru tahmin edildiğini gösterir. Kesinlik öngörülen su kalitesi sınıfı etiketlerinin ne kadarının doğru tahmin edildiğini gösterir.

Çizelge 3.6. Şematik birçok sınıflı CM.

	Tahmin Edilen			
Gerçek		V ₁	V _j	V _n
	V ₁	X ₁₁	X _{1j}	X _{1n}
	V _i	X _{i1}	X _{ij}	X _{in}
	V _n	X _{n1}	X _{nj}	X _{nn}

$$\text{Doğruluk} = \frac{\sum_{i=1}^n X_{ii}}{\sum_{i=1}^n \sum_{j=1}^n X_{ij}} \quad (3.30)$$

$$\text{Kesinlik} = \frac{X_{ii}}{\sum_{k=1}^n X_{ki}} \quad (3.31)$$

$$\text{Geri Çağırma} = \frac{X_{ii}}{\sum_{k=1}^n X_{ik}} \quad (3.32)$$

3.5. Tahmin Çalışmaları İçin Değerlendirme Metrikleri

Girdi değerlerinin bağımlı değerler üzerindeki etki göstergesi olan Pearson korelasyon katsayısı, çoklu değişkenler arasındaki korelasyon derecelerini ölçer. Korelasyon derecesi değeri [-1,1] arasındadır. Burada 1, değişkenler arasında güçlü bir korelasyonu, -1 ise zayıf bir korelasyonu gösterir. Sonuç olarak, korelasyon derecesinin mutlak değeri ne kadar düşükse, korelasyon o kadar zayıftır ve bunun tersi de geçerlidir. Pearson korelasyon katsayısı formülü Eşitlik 3.33 ile tanımlanır, burada, E(X), X ve cov (X, Y)'nin beklenen değerini belirler ve $\sigma_X \sigma_Y$, sırasıyla X, Y'nin kovaryansını ve X ve Y'nin standart sapmasını tanımlar (Lord ve diğerleri, 2021).

$$p_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\sqrt{E(Y^2)-E^2(Y)}} \quad (3.33)$$

Ortalama mutlak hata (Mean Absolute Error (MAE)), tahmin edilen ve tahmin edilen değerlerin gerçek değerlere ne kadar yakın olduğunu ölçen bir vektörden vektöre regresyon değerlendirme yöntemidir. Dolayısıyla, MAE değeri ne kadar düşükse, tahminler ile gerçek değerler arasındaki farklar o kadar düşük olur ve bunun tersi de geçerlidir (Eşitlik 3.34). Eşitlik 3.34’de, n , testlerin veya eğitim setlerinin sayısını gösterir ve A_t ve F_t , gerçek ve tahmin edilen değerlerdir (Qiao ve diğerleri, 2019).

$$MAE = \frac{1}{n} \sum_{t=1}^n |A_t - F_t| \quad (3.34)$$

Tahminlerin performansını değerlendirmek için kök ortalama standart sapma olarak bilinen Kök Ortalama Kare Hatası (Root Mean Square Error(RMSE)) kullanılır. Bu nedenle, tahminler ile gerçek değerler arasındaki mesafeyi (artık) ölçer, artıkların ortalama değerini hesaplar ve ortalama değerlerin karekökünü hesaplar. RMSE, Eşitlik 3.35 ile tanımlanır. Bu eşitlikte, N , değerlerin sayısını temsil eder, ve y_i ve $\hat{y}_i$ i’inci değerini ve bununla ilgili tahmini gösterir (Rustam ve diğerleri, 2020).

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \|y_i - \hat{y}_i\|^2}{N}} \quad (3.35)$$

3.6. Eğitim, Doğrulama ve Deneme

ML çalışmalarında eğitim, doğrulama ve deneme veri setleri ayırımı çok önemli rol oynamaktadır. Örneklem dışı teknik olarak da adlandırılan CVAL, bir veri bilimi projesinin temel bir unsurudur. Bu unsur ML modellerini değerlendirmek ve bağımsız bir test veri kümesi için modelin nasıl performans göstereceğine erişmek için kullanılan bir yeniden örneklem prosedürüdür. Bir ML modeli performansını değerlendirmek için kullanılan sekiz eğitim, doğrulama ve deneme kümeleri ayırımı yöntemleri mevcuttur. Bunlar, p ’yi dışta bırakma çapraz doğrulama (Leave p out cross-validation), tek örnek dışta bırakma çapraz doğrulama (Leave one out cross-validation), bekletme çapraz doğrulama (Holdout cross-validation (HCV)), tekrarlanan rastgele alt örneklem doğrulama (Repeated random subsampling validation), k -kat çapraz doğrulama (k -fold cross-validation (K-FCV)), tabakalı k -kat çapraz doğrulama (Stratified k -fold cross-validation), zaman Serisi çapraz

doğrulama (Time Series cross-validation), iç içe çapraz doğrulama (Nested cross-validation) olarak sayılabilir (Berrar, 2018).

ML modeli geliştirmekte genellikle veri kümesi rastgele olarak eğitim ve deneme kümelerine bölünür, ve eğitim verileri, ML modelini eğitmek için kullanılır. Aynı model, modelin performansını değerlendirmek için bağımsız test verileri üzerinde test edilir. Bölünmenin rastgele durumundaki değişiklik ile modelin doğruluğu da değişir, bu nedenle model için sabit bir doğruluk elde edilemez. Deneme verileri, veri sızıntısı olmaması için eğitim verilerinden bağımsız tutulmalıdır. Eğitim verilerini kullanarak bir ML modelinin geliştirilmesi sırasında model performansının değerlendirilmesi gerekir. İşte CVAL verilerinin önemi burada devreye girmektedir. Yani, eğitim kümesi model geliştirmek için kullanılırken, doğrulama kümesi aynı modelin performansını doğrulamak için kullanılır. CVAL, verilerimizi daha da iyi kullanmamızı sağlamaktadır (Stone, 1978; Berrar, 2018).

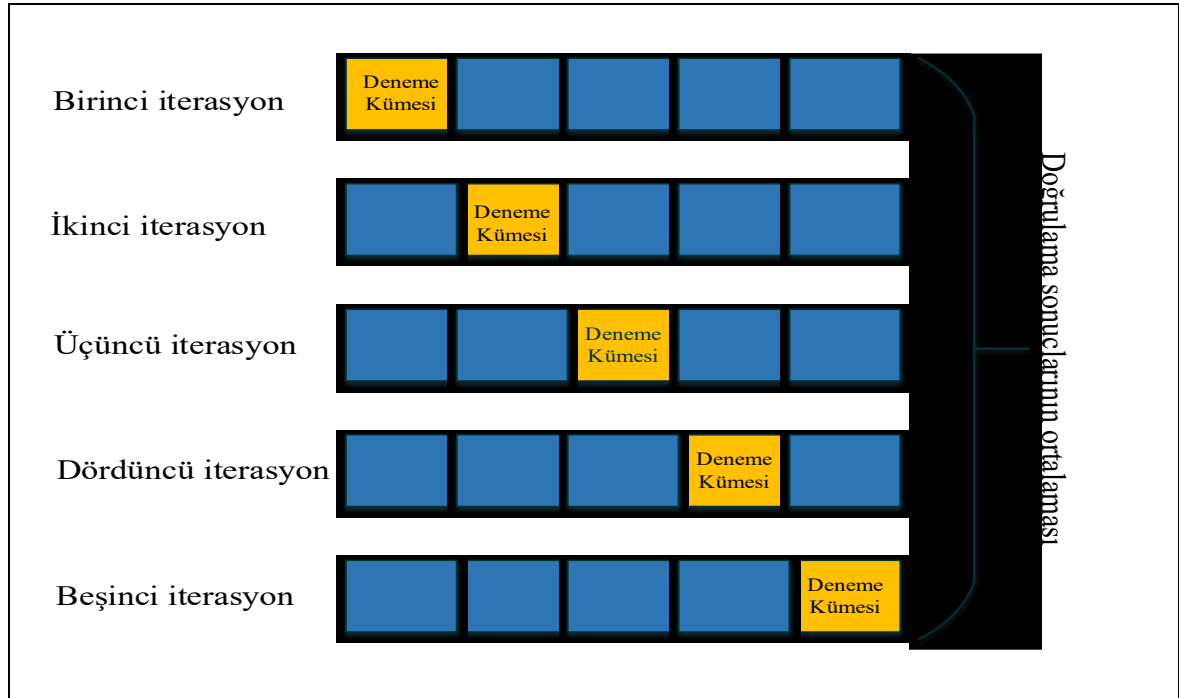
Bu tez çalışmasında, yukarıda isimleri geçen doğrulama yöntemleri arasından HCV ve K-FCV yöntemleri modellerin performansını doğrulamak ve değerlendirmek için kullanılmıştır. Bu seçim ismi geçen yöntemlerin diğerlerine göre avantajları ve uygulanabilirlikleri göz önünde bulundurularak yapılmıştır.

3.6.1. HCV doğrulama tekniği

HCV tekniği, veri analizine bağlı olarak veri kümesini rastgele olarak eğitim ve deneme verilerine ayıran kapsamlı bir CVAL yöntemidir. HCV’de, veri kümesi rastgele eğitim ve doğrulama verilerine bölünür. Genel olarak, eğitim verilerinin yüzdesi, deneme verilerinden daha fazladır. Eğitim verileri modeli başlatmak için kullanılır ve doğrulama verileri modelin performansını değerlendirir. Modeli eğitmek için ne kadar çok veri kullanılırsa daha iyi sonuçlar veren model ortaya konulacaktır. Fakat eğitim kümesi yüzdesi aşırıda olursa, ezberci model oluşumuna yol açar. HCV yöntemi için, eğitimden uygun oranda veri izole edilir (Browne, 2000). Veri bilimi uzmanları bu yöntemle eğitim ve doğrulama veri kümelerinin %80-20’yi geçmemesini önerirler. Bu tez çalışmasında farklı oranlarda eğitim-doğrulama kümeleri oluşumu (%70-30, %75-25, %80-20) yapıldı ve modellerin performansı üzerinde bunların etkisi ayrıca değerlendirilmiştir.

3.6.2. K-FCV tekniği

K-FCV’de, orijinal veri kümesi eşit olarak k alt parçaya bölünür. Her bir iterasyon için k -kat veya gruplardan bir grup doğrulama verisi olarak, geri kalan $(k-1)$ gruplar ise eğitim verisi olarak seçilir. İşlem, her grup doğrulama ve geri kalanı eğitim verisi olarak kabul edilene kadar k kez tekrarlanır. Modelin nihai doğruluğu, k -modelleri doğrulama verilerinin ortalama doğruluğu alınarak hesaplanır (Şekil 3.7). Bu tekniğin avantajları modelin önyargısının düşük olması, düşük zaman karmaşıklığı ve tüm veri seti hem eğitim hem de doğrulama için kullanılması söylenebilir (Berrar, 2018). Bu tez çalışmasında aynı bekletme tekniğinde de uygulandığı gibi farklı bölme oranları (5-FCV ve 10-FCV) denenmiş ve modelleri performansı üzerinde ayrıca etkileri değerlendirilmiştir.



Şekil 3.7. K-FCV tekniğinin çalışma şeması.

3.7. Ön Eğitim Çalışması

Basit bir ifadeyle, bir ML modelini önceden eğitmek, önce bir modelin bir görev veya veri kümesi üzerinde eğitilmesini ifade eder. Deneme çalışması kapsamında 214 satır ve üç parametreden (boy, ağırlık ve yaş) oluşan balık verilerinden oluşan veri kümesi kullanılmıştır. Bu çalışmada, üç makine öğrenme algoritmasının (NB, J48 DT, RF) etkinliği

değerlendirilmiştir. Sonuç olarak, her üç algoritma da ANNs'den daha iyi performans gösterdiği tespit edilmiştir. Gürültülü ve doğrusal olmayan veri kümeleriyle karşılaşma durumunda alternatif bir yaklaşım olabileceği değerlendirilmektedir. Ayrıca, bu üç algoritma arasında J48 DT ve RF, belirli balık yaş grupları için verilerin az olduğu durumlarda olağanüstü performans gösterdiği tespit edilmiştir (Benzer ve diğerleri, 2022).

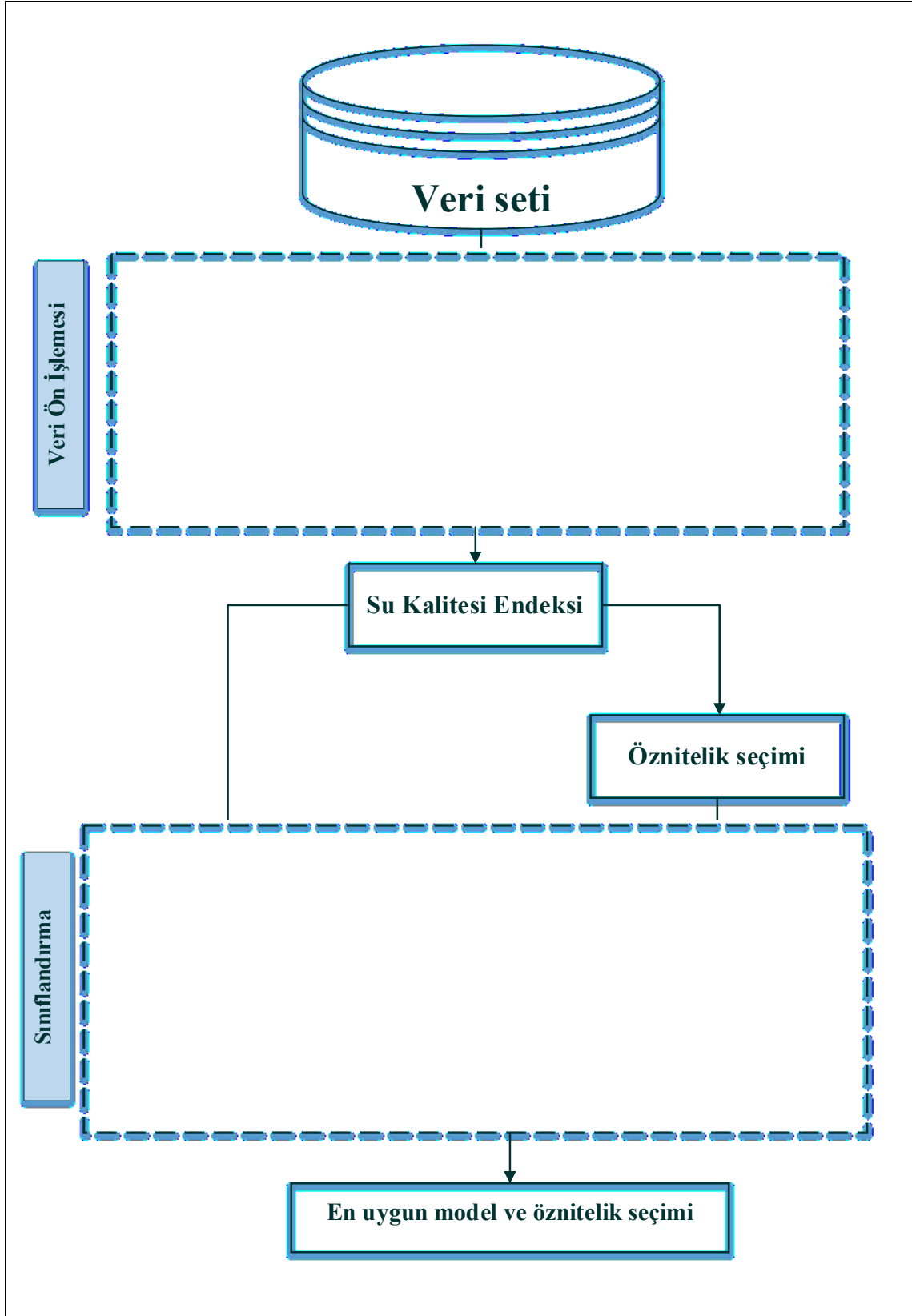
3.8. Sınıflandırma Çalışması Prosedürü

Şekil 3.8., Büyük Menderes Havzası için EL yöntemleriyle su kalitesi sınıflandırma modellemesi çalışma akış şemasını göstermektedir. Şekilde görüldüğü gibi bu çalışma üç ana aşamadan oluşmaktadır.

- Veri ön işlemesi, çalışmanın birinci aşamasında önce bölümlerde anlatıldığı üzere Büyük Menderes Havzası için DSİ tarafından sağlanan su kalitesi verileri modelleme aşamasında kullanılabilmek için yeni veri seti şemasına entegre edilmiştir (Çizelge 3.2). Veriler yeni düzenlenmiş olan verisetine aktarıldıktan sonra kayıp veriler tespit edilmiştir. Kayıp veriler tespit edildikten sonra uygun kayıp veri işlemesine karar verilmiştir ve ML tahmin modellemesi yönteminden yardım almak üzere kayıp veriler tahmin edilerek yeni veriseti inşa edilmiştir. Yeni oluşan veriseti üzerinde veri normalleştirme çalışmaları yapılmıştır. Veri normalleştirilmesi aykırı değerlerin verisetinden çıkarılması ve veri ölçeklemesi ile gerçekleştirilmiştir.
- Çalışmanın hedefi ML modelleri yardımıyla su kalitesi sınıflandırmasıdır. Modelleri eğitim sırasında denetlemek için (SUPL yöntemi) belli bir yıl, ay, ve istasyona ait verileri WQI yöntemiyle sınıflandırılması gerekmektedir. Bunun için kullanılan bir çok WQI yöntemi bulunmaktadır. Bir su kütlesinin İnsani Tüketim açısından uygun olup olmadığı değerlendirmek için en uygun WQI yöntemi seçimine karar verilmiştir. Araştırma kapsamında WAWQI yöntemi seçilmiş ve belli yıl, ay, ve istasyona ait veriler göz önünde bulundurularak su kalitesi sınıfları belirlenmiştir.
- Bu çalışmanın neredeyse en önemli ana hattı pratik, gerçekçi ve performansı yüksek olan bir ML modeli seçimidir. Araştırmada son yıllarda sınıflandırma çalışmalarında çok iyi performans sergileyen dört EL modeli kullanılmasına karar verilmiştir. Modeller önceden endekslenmiş olan su kalitesi verileri ile eğitilmiş ve sonunda doğrulama aşamasından çıkan sonuçlar kaydedilmiştir. Bu aşamada modeller Çizelge 3.2'de anlatıldığı üzere tüm parametreler ile eğitilmiştir. Fakat bir modeli verimli kılmak için

sadece performansının yükseltmesi değil, hesaplamaların azaltılmasıyla birlikte sonuç verme zamanı açısından da verimli kılmak gerekmektedir. Aykırı hesaplamaların azaltılması için yöntemlerinin en önemlisi öznitelik seçimiyle veri kümesi boyutsallığını düşürmek olduğu söylenebilir. Bunun için üç farklı öznitelik seçimi yöntemiyle yeni veri kümeleri oluşturulmuştur. Modeller tekrar yeni veri kümelerinin her birisiyle eğitilerek ve doğrulama aşaması sonrası sonuçlar çalışmanın başlangıcında tüm verisiyle eğitilerek modellerin sonuçlarıyla kıyaslanmıştır. Sonuçların değerlendirilmesiyle en uygun boyuttaki veri kümesi ve en verimli modelin hangisi olduğuna karar verilmiştir.

Bu çalışma kapsamında kullanılan ML ve öznitelik seçimi yöntemleri deneme amaçlı olarak 214 adet (boy, ağırlık ve yaş verisi) balık verisiyle ele alınmıştır. NB, J48 DT ve RF modelleri 214 adet üç parametreden oluşan balık verilerinden oluşan veri kümesi ile eğitilmiştir. Bu üç algoritmanın (J48 DT ve RF) belirli balık yaş grupları için performansları karşılaştırılmıştır.



Şekil 3.8. Büyük Menderes Havzası su kalitesi sınıflandırma çalışması akış şeması

3.9. ÇO Tahmin Çalışması Prosedürü

ÇO miktarının sürekli izlenmesi ve modellenmesi havza yönetimi çalışmalarında önemli rol oynamaktadır. Ayrıca suyun kalitesini yansıtan temel parametrelerden biri olan ÇO seviyelerinin izlenmesi, çevre bilimcilerin su sistemlerinin sağlığı hakkında karar vermeleri için büyük önem taşımaktadır. Bu nedenle bu tez çalışmasının başka bir boyutu olarak Büyük Menderes Havzası'nın 2004-2014 yılları arasında ÇO tahmini doğrultusunda bir etkili regresyon modeli geliştirilmeye çalışılmıştır.

Bu doğrultuda aynı sınıflandırma çalışmasında anlatıldığı gibi modelin performansı ve zaman-maliyet açısından verimini artırmak için ayrıca öznelik seçimi çalışmasıyla boyutu düşürülmüş altküme oluşturulmaya çalışılmış ve modellerin performansı hem tüm veriseti hem de boyutu düşürülmüş altküme ile eğitilerek yapılmıştır.

Öznelik seçimi çalışması sonuçlarından önce bu tahmin çalışmasının bazı farklarını ortaya konulması önemlidir. Birinci olarak bu tahmin çalışmasında kullanılan veriseti sınıflandırma çalışmasında kullanılan veriseti ile parametre açısından farklı olarak 13 parametreden oluşmaktadır (Yıl, Ay, İstasyon, Sıcaklık, pH, İletkenlik, KOİ, BOİ-5, Amonyum, Nitrit, Nitrat, Fosfat, ve ÇO). Böyle bir parametre azaltma çalışmasının ardında iki mantık göz önünde bulundurulmuş ve böylece 19 parametre arasından tahmin çalışmasında kullanılmak üzere 13 parametre seçilmiştir;

- a) Çalışmanın birinci aşamasında modeller tüm parametrelerden oluşan (19 parametre) veri setiyle eğitilmiş ve doğrulama sonuçları değerlendirilmiştir. Sonuçlar bir regresyon modellemesi çalışması sonuçlarının kabul edilemeyecek kadar düşük olduğu için nedenleri araştırılmıştır. Bu nedenlerden birisi tahmin çalışmalarının verisetinin gürültüsünden etkilenmesidir. Veriseti gürültüsü azaltma yöntemlerinden birisi veriseti boyutu düşürülmesidir. Bu nedenle gürültüye neden olan ve bir ÇO konsantrasyonu tahmini çalışmasında ilgisiz olan parametreler tespit edildi ve verisetinden çıkarılmıştır.
- b) İkinci olarak ÇO miktarını ya etkilemeyecek yada etki oranı oldukça düşük olan parametreleri verisetinden çıkarılması çalışmasıdır. Bu nedenle besin madde kontaminasyonunu destekleyen ve doğrudan ÇO miktarıyla ilgili olan parametrelerden oluşan yeni bir alt küme oluşturulmuştur.

Bu nedenlerle 19 parametreden oluşan verisetini boyutu düşürülerek 13 parametreden oluşan yeni bir veriseti ÇO tahmin çalışmasında kullanılmıştır.

Bundan ilave olarak tahmin modellerinin gürültüye oldukça fazla duyarlılığı nedeniyle başka bir gürültü azaltma çalışmasıda yapılmıştır. Bu araştırma sonucunda iki istasyondan elde edilen veriler gürültü oluşturma için tahmin modellemesi çalışmasına uygun bulunmadığı için veri setinden çıkarılmıştır. Bu çalışmada 8. ve 10. istasyon verileri tüm parametreleri ile birlikte verisetinden çıkarılmıştır. Dolayısıyla tahmin modellemesi çalışmasında kullanılan veri kümesi 660 veri yerine 558 veriden oluşmaktadır.

Ayrıca bu çalışmada da tıpkı sınıflandırma çalışması gibi doğrulama yöntemlerinin etkisi ÇO tahmin modelleri üzerinde araştırılmıştır. Bu doğrultuda, aynı sınıflandırma çalışmasında olduğu gibi iki ana doğrulama yöntemi yani K-FCV ve HCV farklı bölme teknikleriyle (5-FCV, 10-FCV, %70-30 HCV, %75-25 HCV, ve %80-20 HCV) uygulanmıştır.

Çalışmanın bu aşamasında farklı öznelik yöntemleri ÇO tahmini çalışmasında kullanılmak üzere boyutu azaltılmış altkümeler oluşturulması için denemeler yapılmıştır. Modellerin arasından BFE-WRAPM yöntemi en mükemmel ve kabul edilebilir sonucu verdiği için bu çalışma kapsamında kullanılmak üzere seçilmiştir.

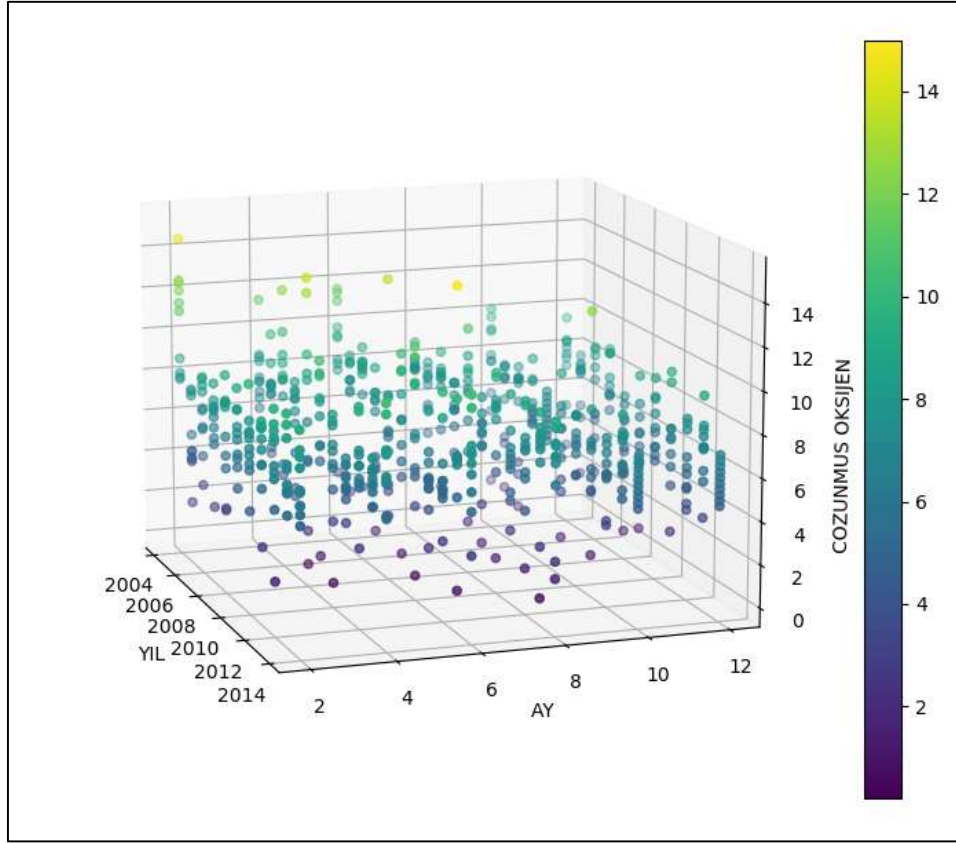
4. BULGULAR

4.1. Büyük Menderes Havzası'nın Genel Su Kalitesi Durumu

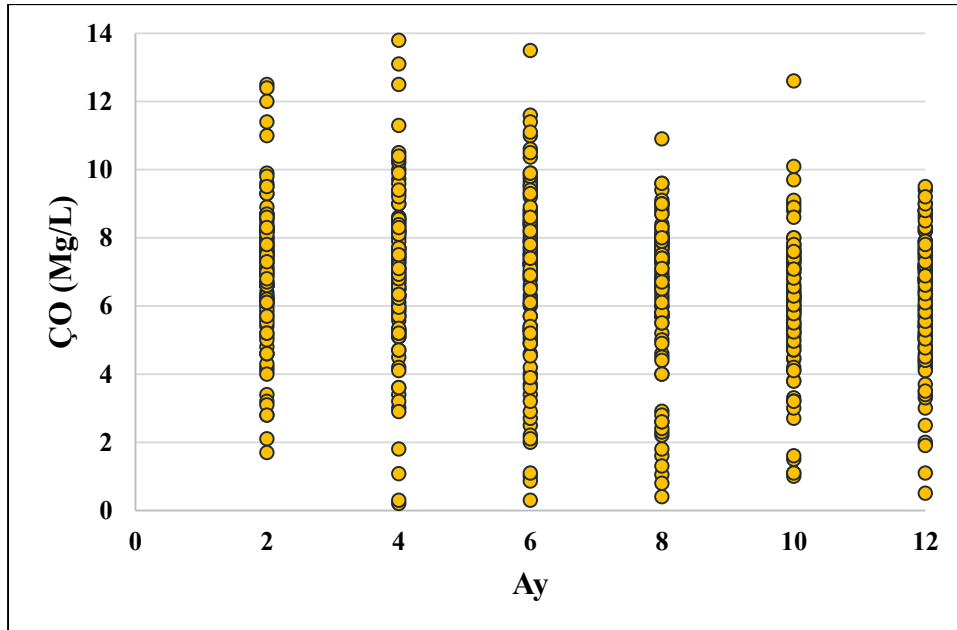
Büyük Menderes Havzasının 2004-2014 yılları arasında aylık ÇO durumu Şekil 4.1'de görülmektedir. Bu şekile bakıldığında Nisan ayından Ağustos ayına kadar ÇO miktarının düşük olduğu görülmektedir. Öncelikle Büyük Menderes Havzası'nın ÇO durumuna bakıldığında (Şekil 4.1 ve Şekil 4.2), genel olarak Şubat ayında ÇO miktarının Nisan ve Haziran aylarına göre yüksek olduğu görülmektedir. Ağustos ayında ÇO miktarı en düşük değere ulaşmakta, Ekim ayında Ağustos ayına nazaran yükselmekte, Aralık ayında tekrar azalmaktadır. Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'ne göre ÇO değeri yüksek kaliteli sularda 8 mg/L'den fazla olmalıdır. Yani ÇO doygunu bir suyun ÇO değeri bu miktardan fazla olmalıdır. ÇO miktarı 6 mg/L'den düşük olduğunda artık su kütlesi ÇO açısından kalitesini kaybetmekte ayrıca bu değer bir kirlilik göstergesi olarakda tabir edilebilir (Orman ve Su İşleri Bakanlığı, 2015).

Çözünmüş Oksijenin değişimlerinde açıklamak için iki faktörü göz önünde bulundurmak gerekebilir. Bunlar sıcaklık ve besin madde zenginleşmesidir. Sıcaklık havada bulunan oksijenin suda çözülmesini etkiler. Buda suyun ÇO derişimi dalgaları şeklinde yansır. Soğuk su sıcak suya göre daha çok ÇO'yu kendinde tutar. Şubat ayında sıcaklığın daha düşük olması sebebiyle suda ÇO miktarının yükselmesi beklenebilir (Şekil 4.1, Şekil 4.2 ve Şekil 4.3).

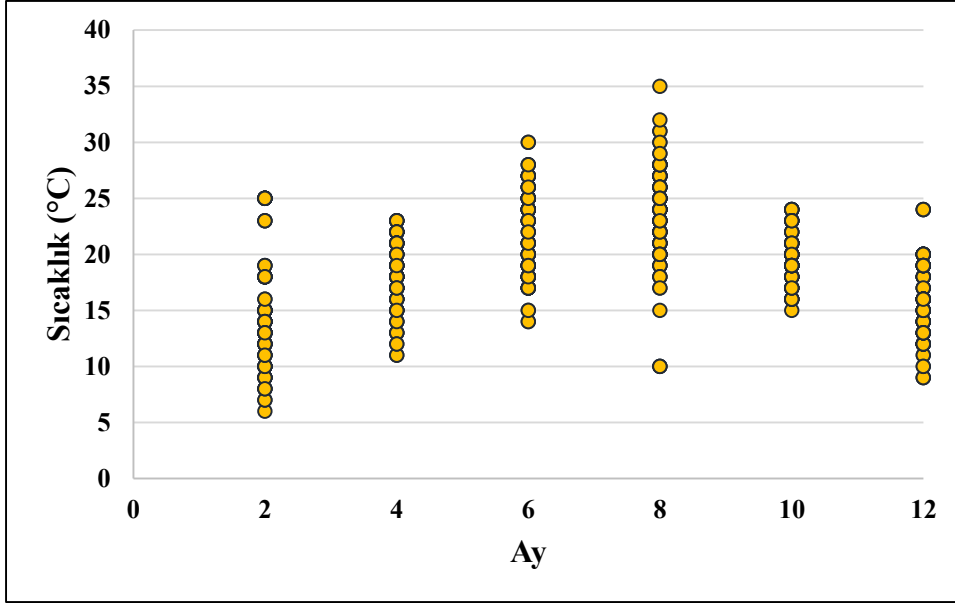
Araştırmada elde edilen verilere göre Nisan ayından itibaren Büyük Menderes Nehri'nin sıcaklığının hızlı bir şekilde yükselmeye başladığı ve Ağustos ayında en yüksek sıcaklığa ulaştığı görülmektedir. Şekil 4.2 ve Şekil 4.3 incelendiğinde sıcaklığın ÇO derişimini nasıl etkilediği görülmektedir. Ağustos ayında düşük ÇO miktarının en düşük derişimlerde olmasının sıcaklığın yüksek olmasından kaynaklandığı düşünülmektedir.



Şekil 4.1. Büyük Menderes Havzası'nın Çözünmüş Oksijen durumu



Şekil 4.2. Büyük Menderes Havzası'nın aylar bazında Çözünmüş Oksijen durumu

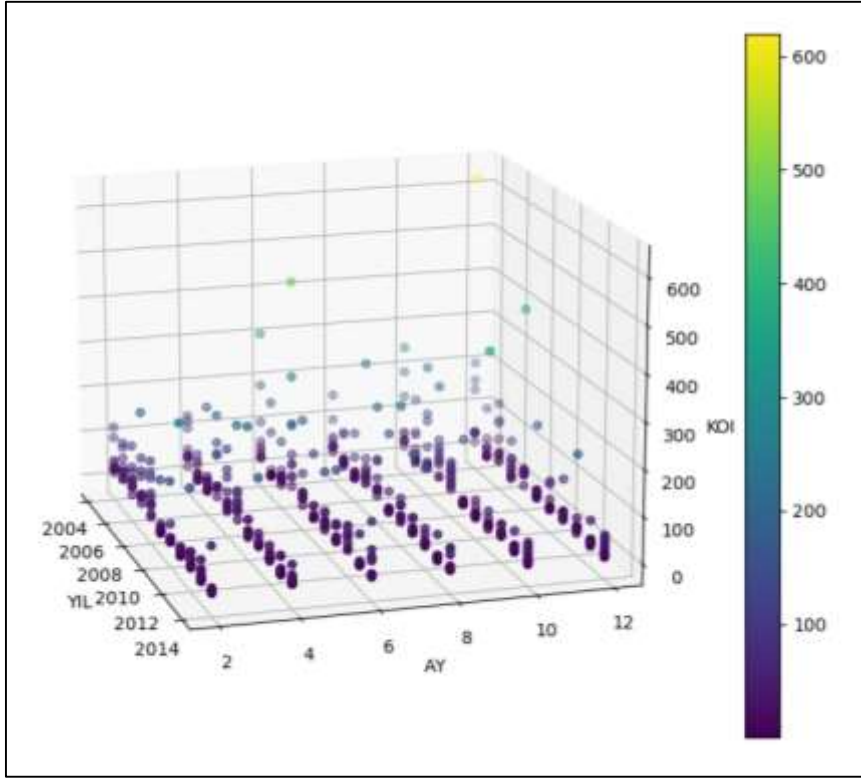


Şekil 4.3. Büyük Menderes Havzası'nın aylar bazında sıcaklık durumu

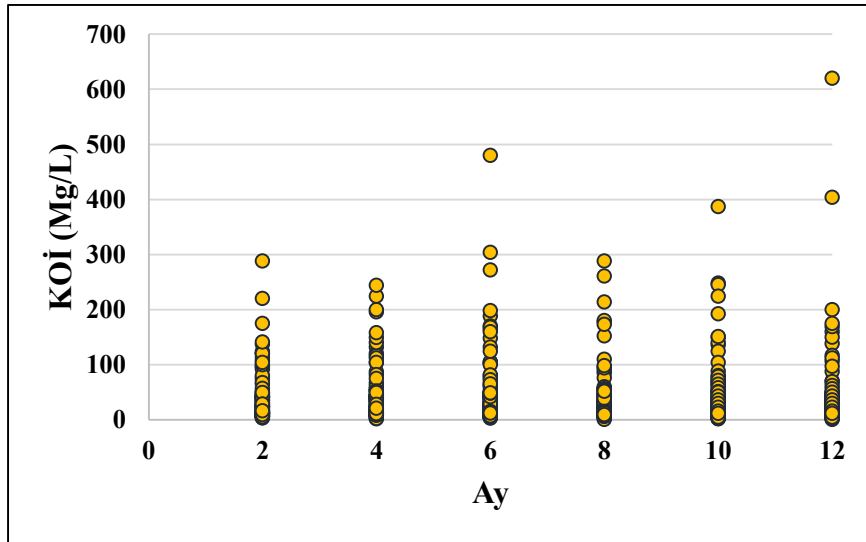
Büyük Menderes Havzası'nın KOİ ve BOİ-5 durumuna bakıldığında (Şekil 4.4 – Şekil 4.7) Şubat ayından Ağustos ayına kadar bir besin madde zenginleşmesi söz konusu olduğu görülmektedir.

Şekil 4.4 ve Şekil 4.5 incelendiğinde Büyük Menderes Havzası'nın KOİ durumunun standart değerlerin üstünde olduğu görülmektedir. Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliğinde yüksek kaliteli bir suyun KOİ değerinin 25 mg/L'den az olması gerektiğini bildirilmektedir.

Ayrıca KOİ miktarının kritik değeri ismi geçen yönetmelikte 50 mg/L olduğunda belirtilmiştir (Orman ve Su İşleri Bakanlığı, 2015). Bir su kütlesinin KOİ miktarı 50 mg/L'den fazla olursa su kütlesinde bir kirlilik ortaya çıktığı sonucuna varılabilir. Bu durumda, Şekil 4.4 ve Şekil 4.5'e göre, Büyük Menderes Havzası'nda genellikle KOİ değeri Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği standart değerlerin üstünde olduğu görülmektedir.



Şekil 4.4. Büyük Menderes Havzası'nın KOİ durumu

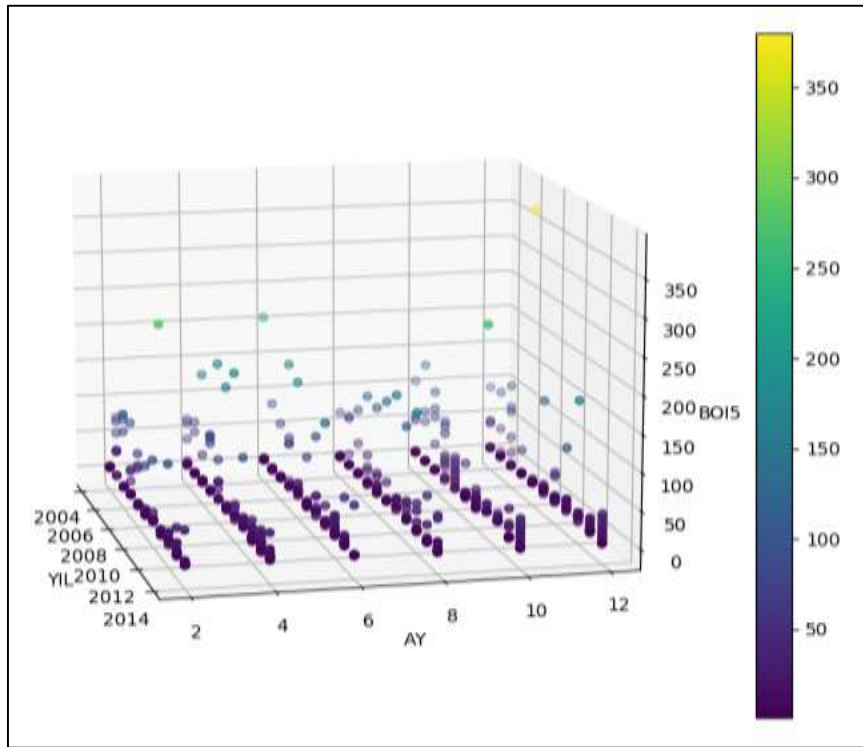


Şekil 4.5. Büyük Menderes Havzası'nın aylar bazında KOİ durumu

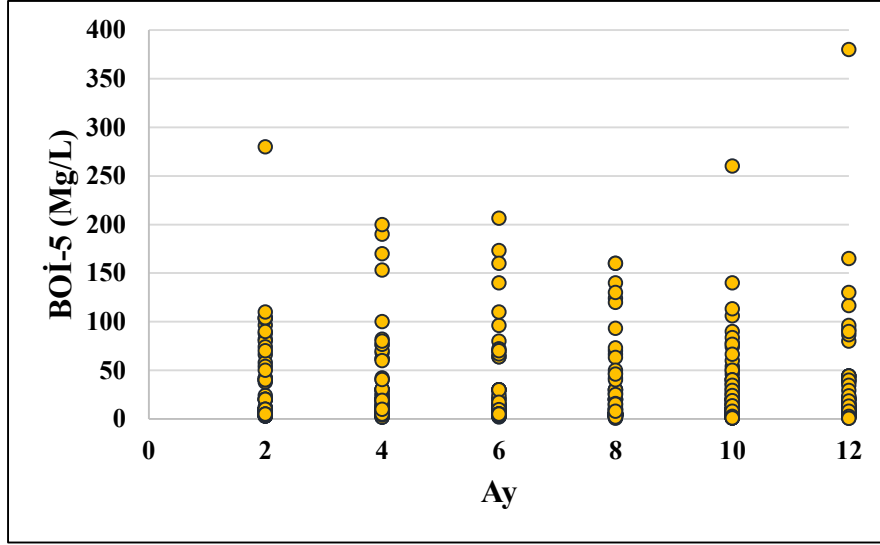
Su sistemlerinde KOİ ve BOİ-5 parametrelerinin birbiriyle çok yakın ilişkide olduğu bilinmektedir. Bir suyun KOİ değeri yükseldiğinde BOİ-5 değerinde de dolaylı bir şekilde yükselmesi ön görünür. Şekil 4.6 ve Şekil 4.7 Büyük Menderes Havzası'nın BOİ-5 açısından durumunu göstermektedir. Araştırmada elde edilen verilere göre yinede Nisan ayından itibaren Büyük Menderes Havzası'nda BOİ-5 miktarı hızlı bir şekilde artmaktadır. Çalışma

alanında BOİ-5 değerinin Haziran ve Ağustos aylarında en yüksek miktarına ulaştığı belirlenmiştir. BOİ-5 miktarının Haziran ve Ağustos aylarındaki bu artışı KOİ miktarının aynı aylarda artışıyla (Şekil 4.4 ve Şekil 4.5) uyumakta olduğu görülmektedir.

BOİ-5 miktarının standart değerleri açısından incelendiğinde, Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetmeliği'ne göre yüksek kaliteli bir suyun BOİ-5 değerinin 4 mg/L'den az olmalıdır (Orman ve Su İşleri Bakanlığı, 2015). Ayrıca BOİ-5 miktarının kritik değeri ismi geçen yönetmelikte 8 mg/L olduğuda belirtilmiştir. Şekil 4.6 ve Şekil 4.7'e bakıldığında Büyük Menderes Havzası'nın BOİ-5 miktarı genellikle Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetmeliği standart değerlerin oldukça üstünde olduğu görülmektedir. Buda bu havzanın KOİ miktarıyla uyumakla birlikte bir kirlilik göstergesi olarak değerlendirilebilir.

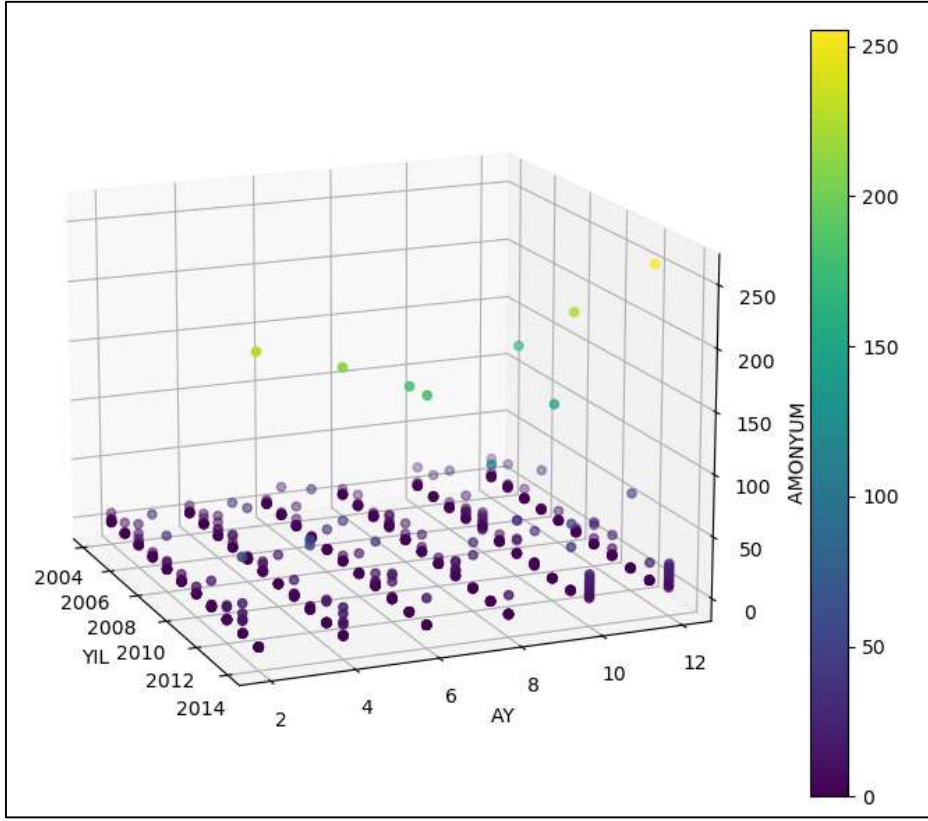


Şekil 4.6. Büyük Menderes Havzası'nın BOİ-5 durumu

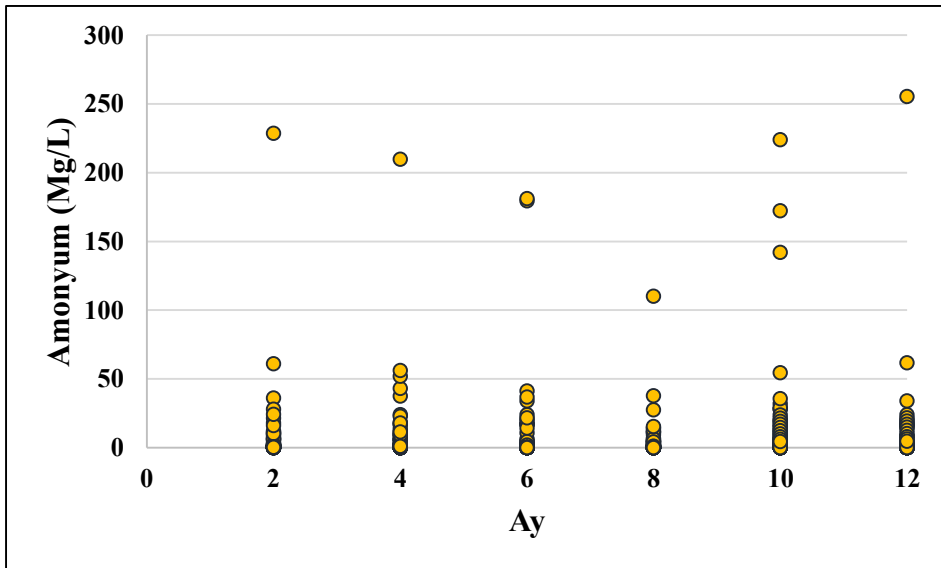


Şekil 4.7. Büyük Menderes Havzası'nın aylar bazında BOİ-5 durumu

Şekil 4.8 ve Şekil 4.9 Büyük Menderes Havzası'nın Amonyum miktarı durumunu göstermektedir. Besin zenginleşmesiyle ilgili olarak Nisan ayından itibaren bu havzada Amonyum miktarının artışı görülmektedir. Amonyum'un standart değerlerine baktığımızda TS-266'ya göre standart değeri 0,50 mg/L olmalıdır (TSE, 2005). Ayrıca Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'ne göre yüksek kaliteli bir suyun Amonyum miktarı 0,20 mg/L'den az olması gerekmektedir. Bu yönetmeliğe göre Amonyumun kritik miktarı 1,00 mg/L belirlenmiştir (Orman ve Su İşleri Bakanlığı, 2015). Şekil 4.8 ve Şekil 4.9 standart değerler açısından incelendiğinde Büyük Menderes Havzası'nda Amonyum miktarı genellikle Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetmeliği ve TS-266 standart değerlerinin oldukça üstünde olduğu görülmektedir.



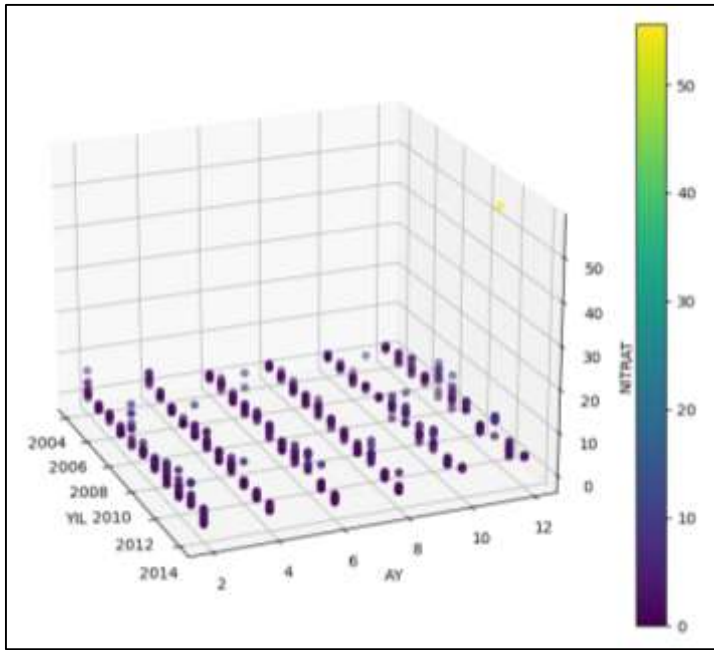
Şekil 4.8. Büyük Menderes Havzası'nın Amonyum durumu



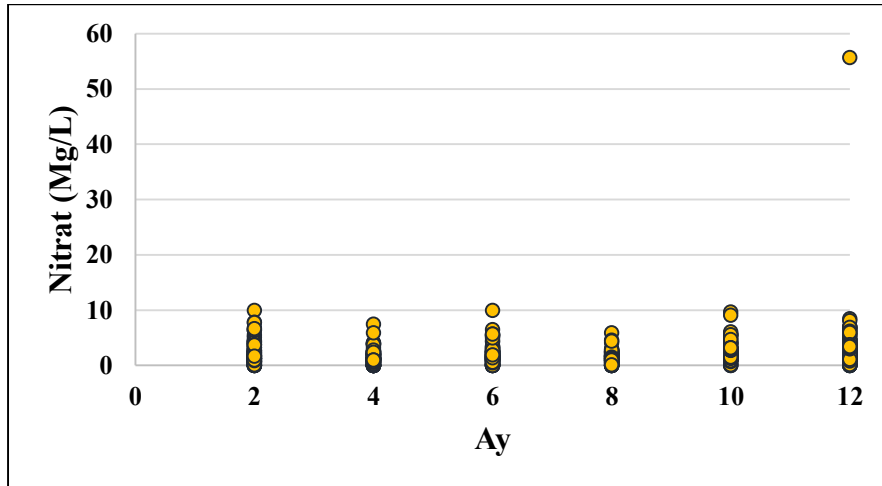
Şekil 4.9. Büyük Menderes Havzası'nın aylar bazında Amonyum durumu

Şekil 4.10 ve Şekil 4.12 Büyük Menderes Havzası'nın Nitrat ve Nitrit durumunu göstermektedir. Araştırmada elde edilen verilere göre Şekil 4.11 ve Şekil 4.13 Büyük Menderes Havzası'nın Nitrat ve Nitrit durumu aylar bazında görülmektedir. WHO'ya göre Nitrat için standart değer 50 mg/L belirlenmiştir (WHO, 2006). Orman ve Su İşleri Bakanlığı

tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'ne göre yüksek kaliteli bir suyun Nitrat miktarı 5 mg/L'den az alması gerekmektedir. Bu yönetmeliğe göre Nitratın kritik miktarı 10 mg/L belirlenmiştir (Orman ve Su İşleri Bakanlığı, 2015). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği standart değerleri ile Büyük Menderes Havzası'nın Nitrat değerleri karşılaştırıldığında Nitrat'ın genellikle Türkiye standart değerlerinin üstünde olduğu görülmektedir. Büyük Menderes Havzası'nda Şubat ayından Ekim ayına kadar Nitrat miktarının yükseldiği görülmektedir.

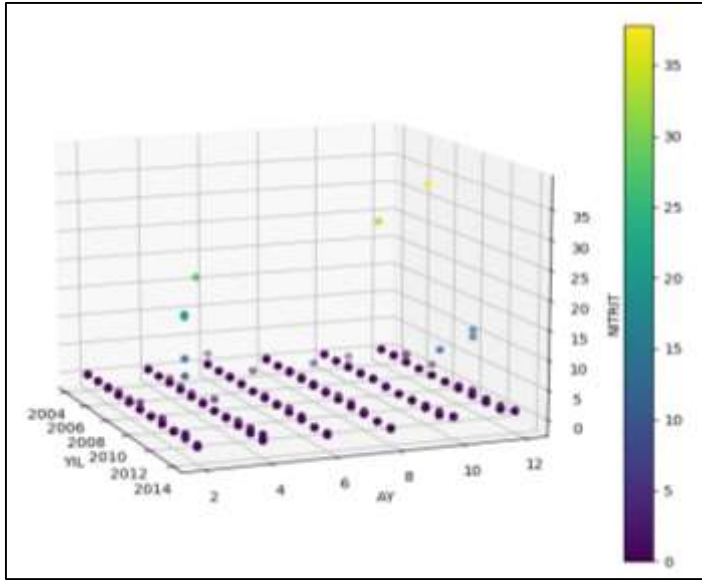


Şekil 4.10. Büyük Menderes Havzası'nın Nitrat durumu

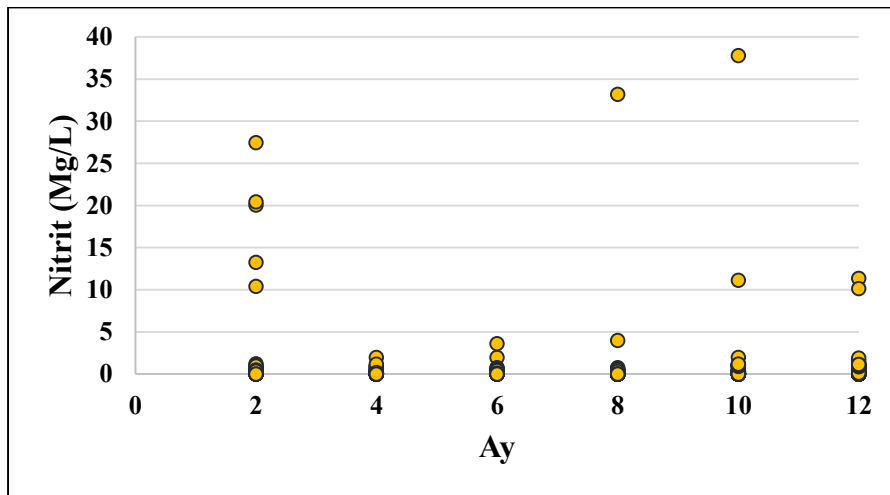


Şekil 4.11. Büyük Menderes Havzası'nın aylar bazında Nitrat durumu

Şekil 4.12 ve Şekil 4.13 Büyük Menderes Havzası'nın Nitrit durumunu göstermektedir. Nitrit için WHO, 3 mg/L'yi standart değer olarak belirtmiştir (TSE, 2006). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'ne göre yüksek kaliteli bir suyun Nitrit miktarı 0,01 mg/L'den az olması gerekmektedir. Bu yönetmeliğe göre Nitrat'ın kritik miktarı 0,06 mg/L belirlenmiştir (Orman ve Su İşleri Bakanlığı, 2015). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği standart değerleri açısından Büyük Menderes Havzası'nın Nitrit miktarı karşılaştırıldığında Nitrit'in genellikle Türkiye için standart değerlerin üstünde olduğu görülmektedir.

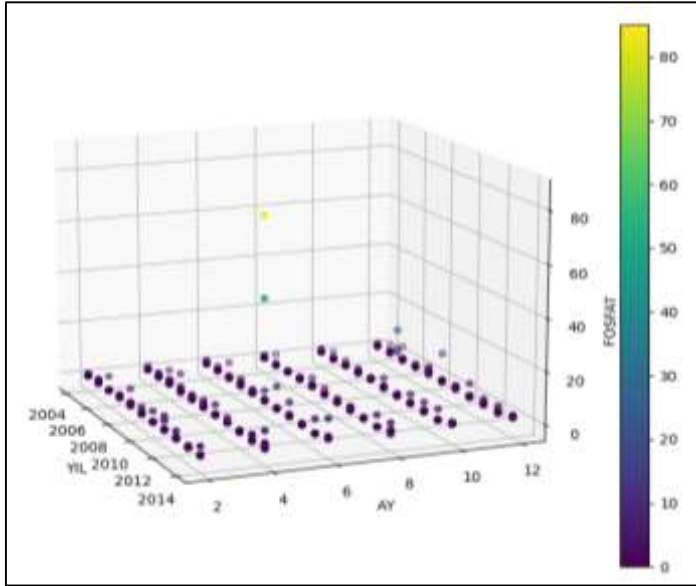


Şekil 4.12. Büyük Menderes Havzası'nın Nitrit durumu

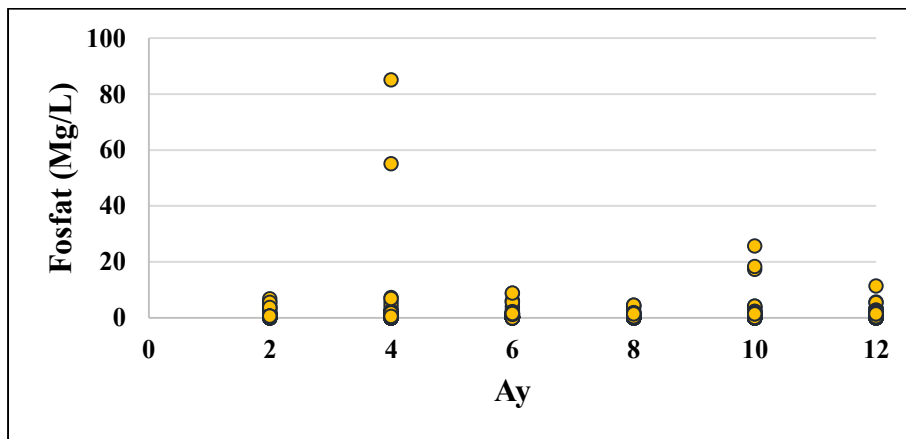


Şekil 4.13. Büyük Menderes Havzası'nın aylar bazında Nitrit durumu

Şekil 4.14 ve Şekil 4.15, Büyük Menderes Havzası'nın Fosfat durumu göstermektedir. Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'ne göre yüksek kaliteli bir suyun Fosfat miktarı 0,05 mg/L'den az olması gerekmektedir. Bu yönetmeliğe göre Nitrat'ın kritik miktarı 0,16 mg/L belirlenmiştir (Orman ve Su İşleri Bakanlığı, 2015). TS-266'ya göre Fosfat'ın standart değeri 0,4 mg/L dir (TSE, 2005). Büyük Menderes Havzası'nda Fosfat miktarı TS-266 ve Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği standart değerler açısından incelendiğinde Fosfat'ın genellikle standart değerlerin üstünde olduğu görülmektedir. Şekil 4.15'de görüldüğü gibi Fosfat'ın Haziran ayında en yüksek miktarda olduğu tespit edilmiştir.

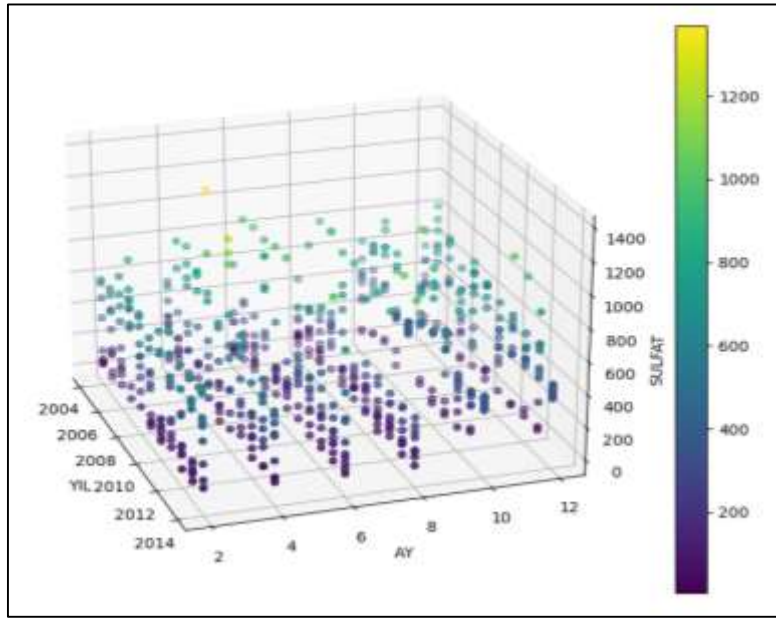


Şekil 4.14. Büyük Menderes Havzası'nın Fosfat durumu

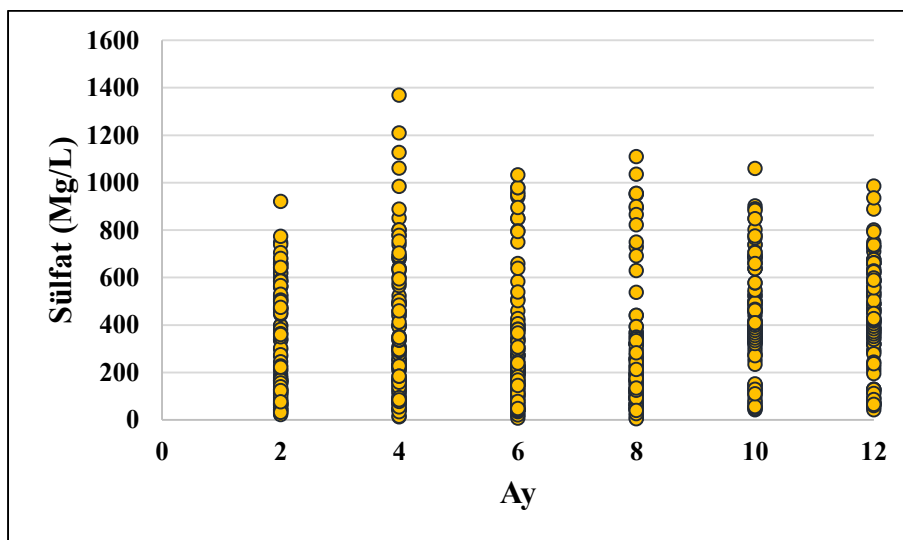


Şekil 4.15. Büyük Menderes Havzası'nın aylar bazında Fosfat durumu

Şekil 4.16 ve Şekil 4.17 Büyük Menderes Havzası'nın Sülfat durumunu göstermektedir. Bu iki şekile bakıldığında Sülfatın Nisan ayı itibarıyla artışı görülmektedir. WHO Sülfat için 500 mg/L'yi standart değer olarak belirtmiştir (WHO, 2006). Bu değer TS-266'ya göre 250 mg/L dir (TSE, 2005). Büyük Menderes Havzası'nın Sülfat miktarını standartlar açısından incelendiğinde Sülfat miktarı genellikle hem WHO standart değeri hemde TS-266'nın önerdiği Türkiye için standart değerden fazla olduğu tespit edilmiştir.

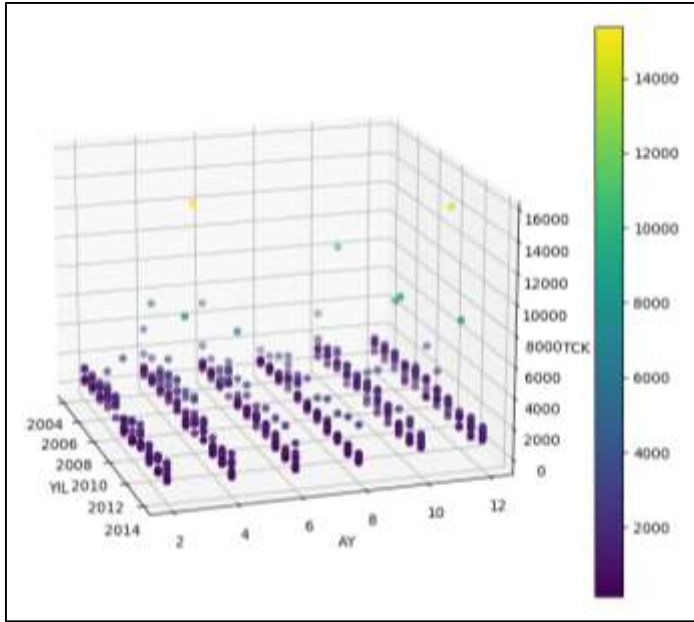


Şekil 4.16. Büyük Menderes Havzası'nın Sülfat durumu

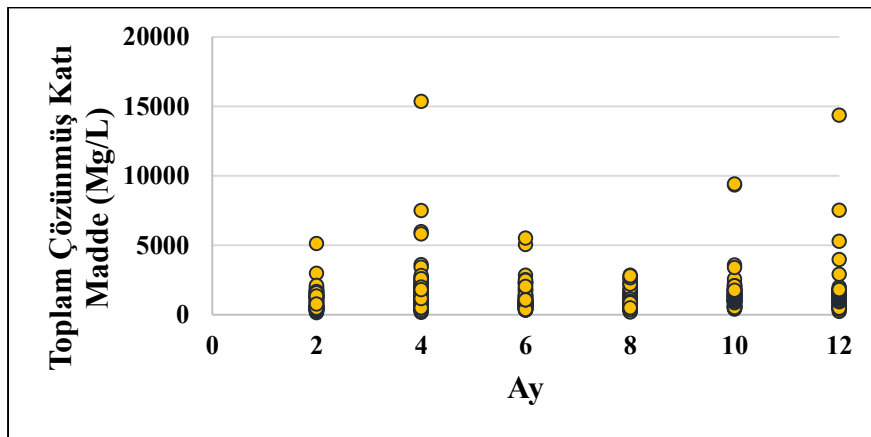


Şekil 4.17. Büyük Menderes Havzası'nın aylar bazında Sülfat durumu

Şekil 4.18 ve Şekil 4.19, Büyük Menderes Havzası'nın TÇK durumunu göstermektedir. Büyük Menderes Havzası'nda Nisan ayından itibaren TÇK'nın artışı görülmektedir. Nisan ve Ağustos ayları arasında ayrıca bu parametrenin en yüksek miktarlarda olduğu görülmektedir. Türkiye'de bulunan yönetmelikler ve standartlarda TÇK için her hangi bir standart değer belirtilmemiştir. Fakat WHO, TÇK için 1000 mg/L'yi standart değer olarak belirtmiştir (WHO, 2006). Şekil 4.18 ve Şekil 4.19 standart değer açısından incelendiğinde, Büyük Menderes Havzası'nda genellikle TÇK değerleri WHO ve Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetmeliği standart değerlerin üstünde olduğu görülmektedir.



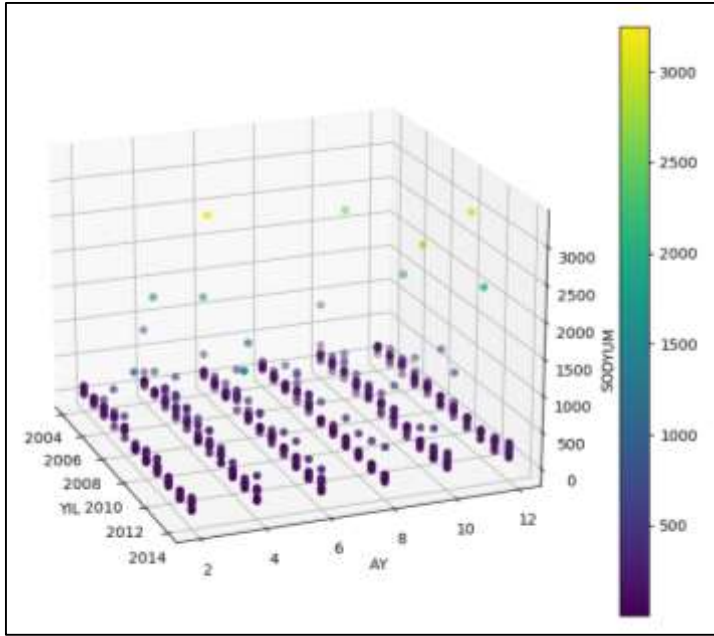
Şekil 4.18. Büyük Menderes Havzası'nın Toplam Çözünmüş Katı durumu



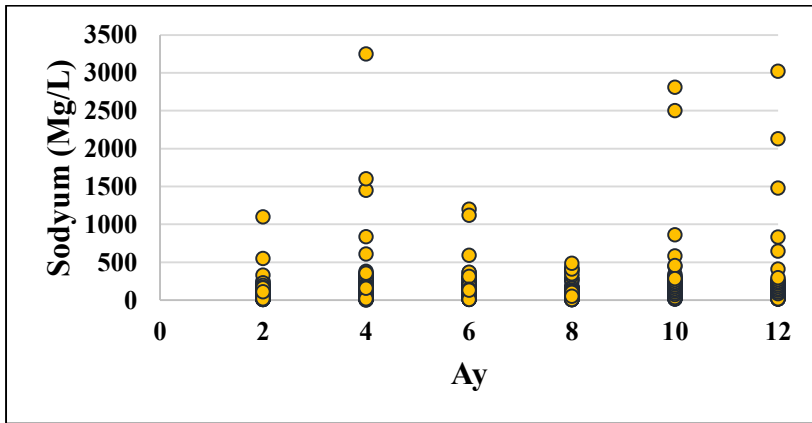
Şekil 4.19. Büyük Menderes Havzası'nın aylar bazında TÇK durumu

İnorganik tuzlar (esas olarak Kalsiyum, Magnezyum, Potasyum, Sodyum, bikarbonatlar, Klorürler ve Sülfatlar) TÇK ile ilişkidir.

Şekil 4.20 ve Şekil 4.21 Büyük Menderes Havzası'nın Sodyum durumunu göstermektedir. Bu iki şekilde görüldüğü gibi Nisan ayından itibaren Büyük Menderes Havzası'nda Sodyum miktarı artmaktadır. Sodyum için WHO ve TS-266, 200 mg/L'yi standart değer olarak belirtmiştir (WHO, 2006; TSE, 2005). Şekil 4.20 ve Şekil 4.21 incelendiğinde Büyük Menderes Havzası'nda genellikle Sodyum değerleri WHO ve TS-266 standart değerlerin üstünde olduğu görülmektedir.

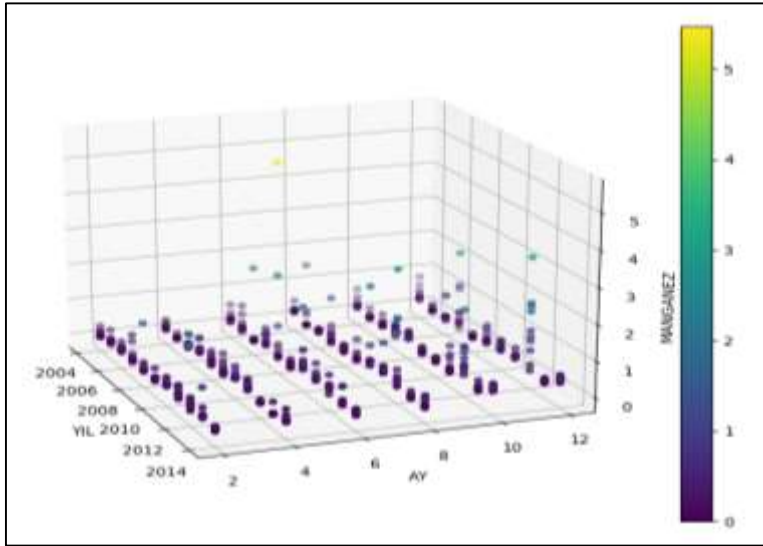


Şekil 4.20. Büyük Menderes Havzası'nın Sodyum durumu

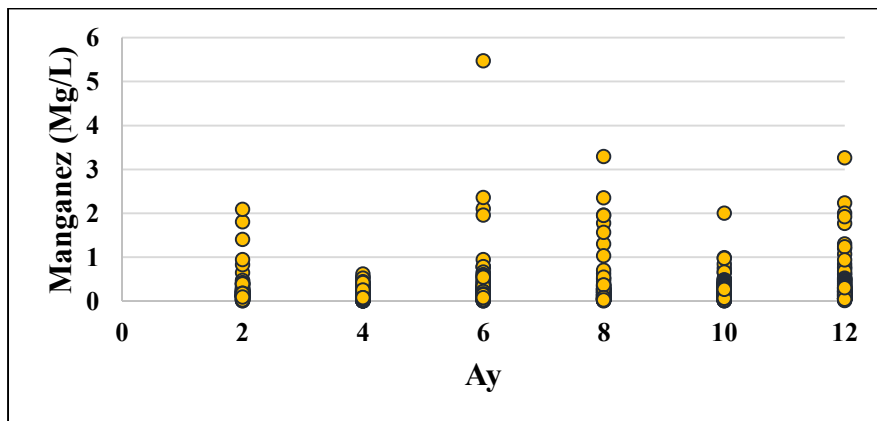


Şekil 4.21. Büyük Menderes Havzası'nın aylar bazında Sodyum durumu

Şekil 4.22 ve Şekil 4.23 Büyük Menderes Havzası'nın Manganez durumunu göstermektedir. Manganezin Haziran, Ağustos, ve Aralık aylarında çok yüksek miktarlarda olduğu görülmektedir. WHO Manganez için 0,4 mg/L'yi standart değer olarak belirtmiştir. Bu değer TS-266'ya göre 0,05 mg/L'dir (TSE, 2005). Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetmeliği'ne göre yüksek kaliteli bir suyun Manganez miktarı 0,1 mg/L'den az alması gerekmektedir. Bu yönetmeliğe göre Manganez'in kritik miktarı 0,5 mg/L belirlenmiştir (Orman ve Su İşleri Bakanlığı, 2015). Şekiller standart değerler açısından incelendiğinde genellikle Büyük Menderes Havzası Manganez değerleri TS-266 ve Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetmeliği standart değerlerinin üstünde olduğu görülmektedir.

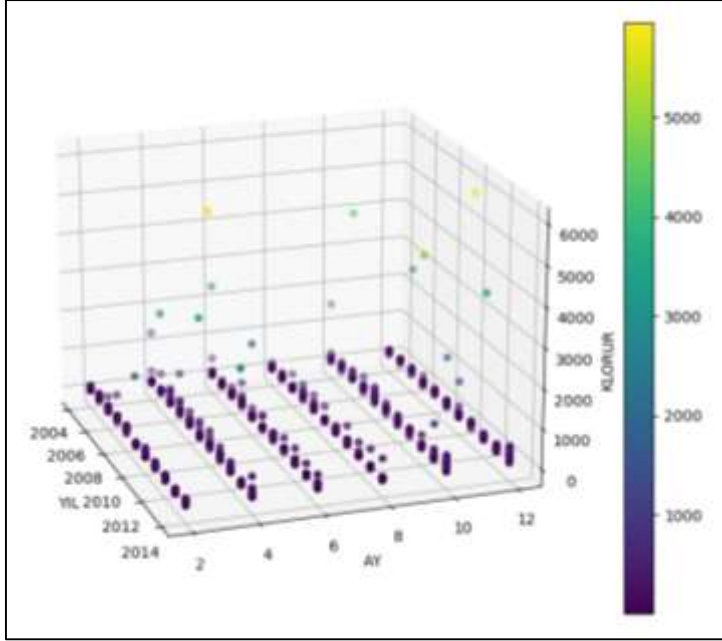


Şekil 4.22. Büyük Menderes Havzası'nın Manganez durumu

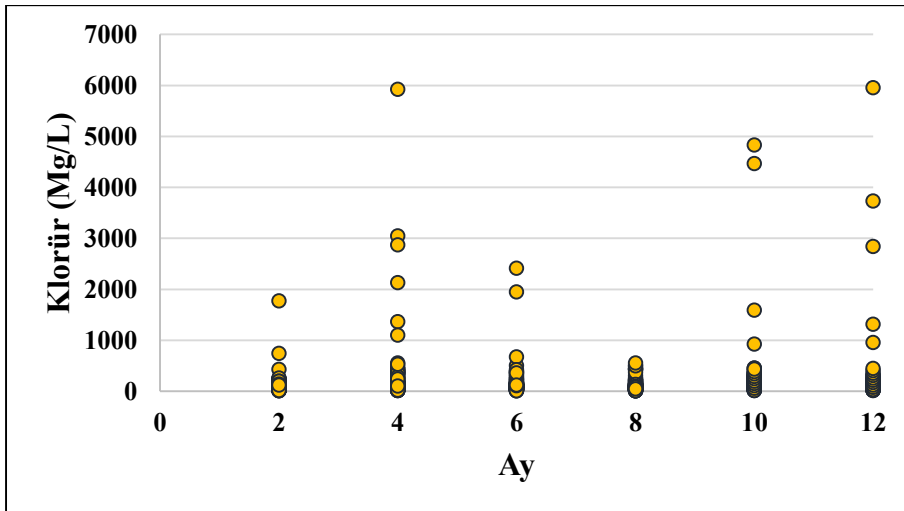


Şekil 4.23. Büyük Menderes Havzası'nın aylar bazında Manganez durumu

Şekil 4.24 ve Şekil 4.25 Büyük Menderes Havzası'nın Klorür durumunu göstermektedir. WHO ve TS-266 Klorür için 250 mg/L'yi standart değer olarak belirtmişlerdir (TSE, 2005; WHO, 2006). Standart değerler açısından incelendiğinde genellikle Büyük Menderes Havzası'nda Klorür miktarının WHO ve TS-266 standart değerlerin üstünde olduğu görülmektedir.



Şekil 4.24. Büyük Menderes Havzası'nın Klorür durumu



Şekil 4.25. Büyük Menderes Havzası'nın aylar bazında Klorür durumu

4.2. Sınıflandırma Çalışması Sonuçları

Bu bölümde Büyük Menderes Havzası için EL yöntemleriyle su kalitesi sınıflandırma modellemesi değerlendirilmiştir.

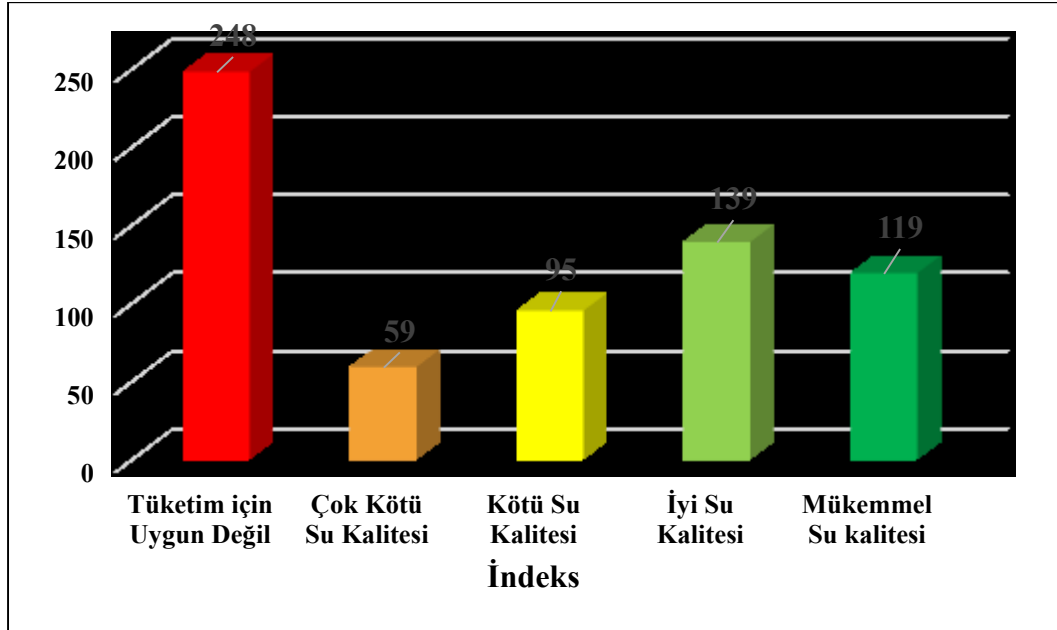
4.2.1. WQI sınıflandırması sonuçları

WQI toplu bilgiyi azaltmak ve suyun genel kullanıma yönelik kalitesini tanımlamak için tek bir mantıksal ifade oluşturmaktadır. Su sistemlerinin genel durumunu tanımlamak için dünya çapında ulusal ve uluslararası kuruluşlar tarafından geliştirilen fonksiyonlardan olan WQI kullanılmaktadır. Bu amaç doğrultusunda WAWQI yöntemi seçilerek ve belli yıl, ay ve istasyona ait veriler göz önünde bulundurularak su kalitesi sınıfları belirlenmiştir (Çizelge 4.1). 2004-2014 yılları arasında toplam 660 veriden oluşan Büyük Menderes Havzası su kalitesi genel durumu WQI yöntemi ile sınıflara ayrılmıştır (Şekil 4.26). 2004-2014 Büyük Menderes Havzasında belli yıl, ay ve istasyona ait toplam 248 tüketim için “Uygun Olmayan Su Kalitesi” vakası olduğu tespit edilmiştir.

Çizelge 4.1. WQI'leri belirtilmiş verilerin şematik veri seti çizelgesi (MSK: Mükemmel Su Kalitesi; İSK: İyi Su Kalitesi; KSK: Kötü Su Kalitesi; ÇKSK: Çok Kötü Su Kalitesi; TUD: Tüketim İçin Uygun Değil)

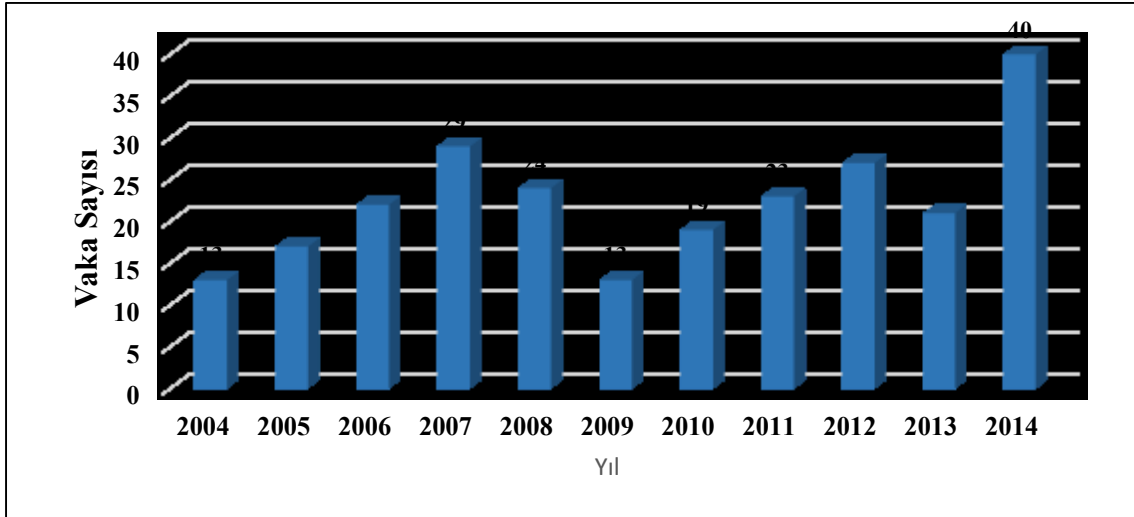
Sayı				İzleme Bilgileri							Su Kalitesi Parametreleri											
	Yıl	Ay	İst.	T	pH	EC	ÇO	KOI	BOİ	NH ₄	NO ₃	NO ₂	PO ₄	Cl ⁻	Fe	Mn	Na ⁺	SO ₄	TDS	Smf		
1	2004	2	1	15	8	900	7.8	120	6.4	8	1.25	0.32	1.14	70.9	2.983	0.329	107.2	68	580	TUD		
2	2004	4	1	18	8.2	740	6	132	68	6.4	3.25	0.29	1.71	49.6	0.27	0.06	46	43.2	470	TUD		
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
269	2008	10	5	20	7.94	1903	6.28	50.99	24.8	6.26	2.286	0.28	0.948	212	0.462	0.307	172.6	373.4	1308	TUD		
270	2008	12	5	14	7.7	2110	5.2	49	6	0.43	4.7	0	1.22	124.1	0.278	0.06	149.4	584	1580	KSK		
271	2008	2	6	7	8.4	1730	8.2	39.14	4.8	0.18	4.1	0	0.82	131.2	0.01	0.09	119.4	398.6	1300	KSK		
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
658	2014	8	10	29	8.3	870	9	9.3	7.85	0.05	0.1	0	1.47	46.53	0.04	0.027	52.69	135.2	510	ÇKS K		
659	2014	10	10	19	8.04	2504	7.08	11.35	0.87	4.37	3.177	1.17	1.468	431.5	0.064	0.269	282.8	411	1769	TUD		
660	2014	12	10	19	8.01	2599	6.88	11.45	0.76	4.60	3.44	1.15	1.433	448.1	0.517	0.3	295.3	428.1	1829	TUD		

Mevcut veriler ile değerlendirildiğinde 2004-2014 yılları arası Büyük Menderes Havzası'nda toplam 402 defa “Kötü Su Kalitesi” vakası olduğu (Kötü Su Kalitesi, Çok Kötü Su Kalitesi ve Tüketim İçin Uygun Olmayan Su Kalitesi) ve suyun insani tüketim açısından kullanılmaz hale geldiği belirlenmiştir. Buda Büyük Menderes Havzası'nın 2004-2014 yılları arasında ne kadar kirliliğe maruz kaldığı hakkında bilgi vermektedir.



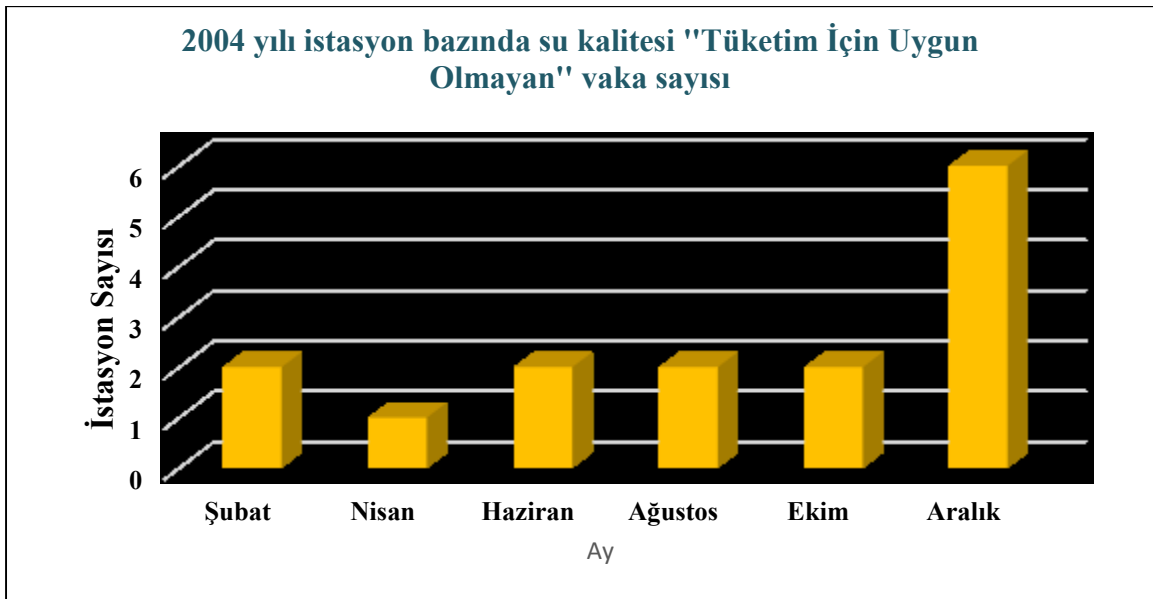
Şekil 4.26. Büyük Menderes Havzası su kalitesi genel durumu

Şekil 4.26’de görüldüğü gibi “Tüketim İçin Uygun Olmayan” su kalitesi Büyük Menderes Havzası’nın baskın su kalitesidir. Bunu tartışmak için Büyük Menderes Havzası’nda uygun olmayan su kalitesi vakalarını yıllar bazında tartışılması gerekmektedir. Şekil 4.27, yıllar bazlı Büyük Menderes Havzası’nda tüketim için “Tüketim İçin Uygun Olmayan” su kalitesi vaka sayısını göstermektedir. Şekil 4.27’de görüldüğü gibi 2014 yılı Büyük Menderes Havzası’nın su kalitesi açısından “Tüketim İçin Uygun Olmayan” 40 su kalitesi vakasıyla en kirli yıl olduğu belirlenmiştir. 2014 yılının ardından 2007 ve 2012 yılları sırayla 29 ve 27 “Uygun Olmayan Su Kalitesi” vakasıyla en kirli yıllar olarak belirlenmiştir. Fakat diğer yıllarda da vaka sayısı göz ardı edilebilecek kadar az olmadığı ve bunun da Büyük Menderes Havzası’nın önceki bölümlerde nedenleri tartışıldığı üzere besin madde kirliliğine maruz kalmasından kaynaklı olabileceği değerlendirilmektedir.



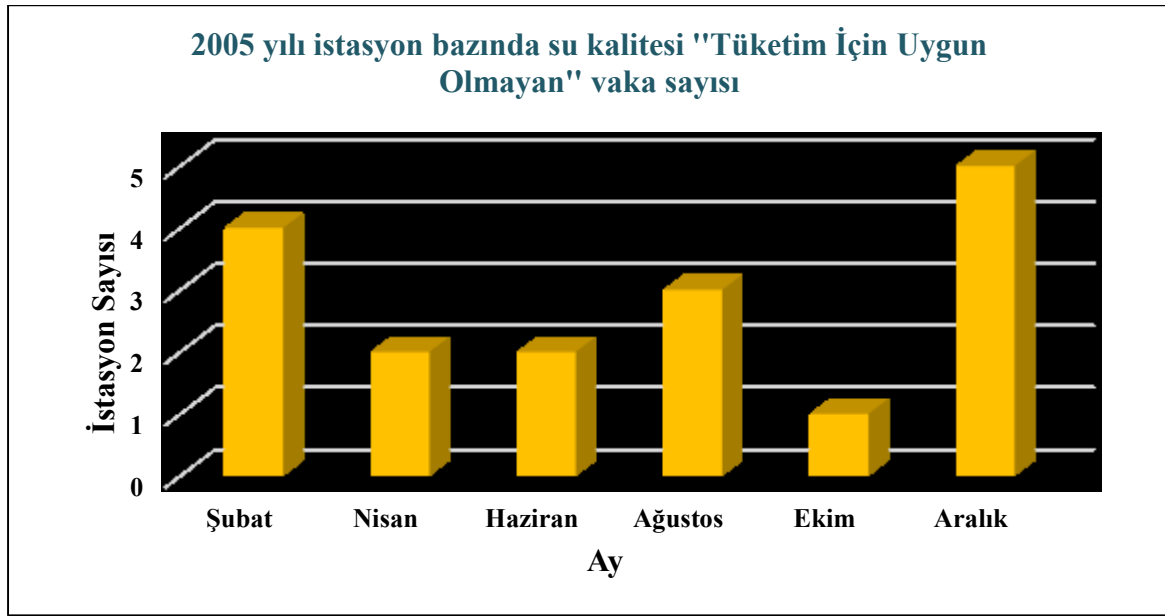
Şekil 4.27. Yıllar bazlı Büyük Menderes Havzası’nda “Tüketim İçin Uygun Olmayan” su kalitesi vaka sayısı

Büyük Menderes Havzası için aylar bazında tüketim için “Tüketim İçin Uygun Olmayan” su kalitesi vakaları incelendiğinde (Şekil 4.28 - Şekil 4.38), 2004 Aralık ayının Büyük Menderes Havzası için tüketim için “Tüketim İçin Uygun Olmayan” su kalitesi vakası açısından en kirli ay olduğunu görülmektedir (Şekil 4.28). Bu çalışma kapsamında su kalitesi dikkate alındığında Aralık ayında toplam 10 istasyonun 6’sında “Tüketim İçin Uygun Olmayan” su kalitesi vakasının varlığı tespit edilmiştir.



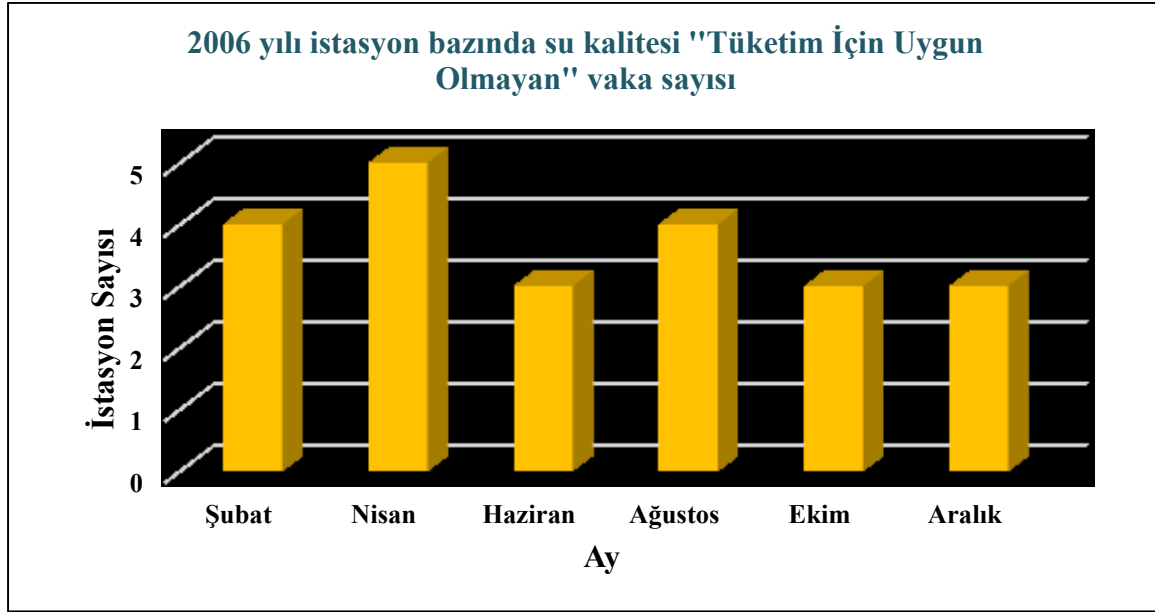
Şekil 4.28. 2004 yılı için Büyük Menderes Havzası’nda aylar bazında “Tüketim İçin Uygun Olmayan” su kalitesi vakaları

Aynı şekilde 2004 yılında olduğu gibi, 2005 yılı Aralık ayı Büyük Menderes Havzası'nın su kalitesi açısından en kirli ay olduğu görülmektedir. (Şekil 4.29). Bu ayda 5 istasyonda "Tüketim İçin Uygun Olmayan" su kalitesi vakası tespit edilmiştir. Ayrıca 2005 yılı Şubat ayında aynen Aralık ayı gibi su kalitesi açısından kirlilik problemlerinin olduğu belirlenmiştir. Şubat ayında da 4 istasyonda "Tüketim İçin Uygun Olmayan" su kalitesi vakası tespit edilmiştir.



Şekil 4.29. 2005 yılı için Büyük Menderes Havzası'nda aylar bazında "Tüketim İçin Uygun Olmayan" su kalitesi vakaları

Büyük Menderes Havzası'nın 2006 yılı "Tüketim İçin Uygun Olmayan" su kalitesi vakalarına bakıldığında Şubat, Nisan ve Ağustos aylarında sırayla 4, 5 ve 4 istasyonda "Tüketim İçin Uygun Olmayan" su kalitesi vakasıyla en kirli aylar olduğu görülmektedir (Şekil 4.30). Fakat verilerin incelenmesi sonucunda diğer aylarda da "Tüketim İçin Uygun Olmayan" su kalitesi vakası açısından problemler olduğu görülmektedir.



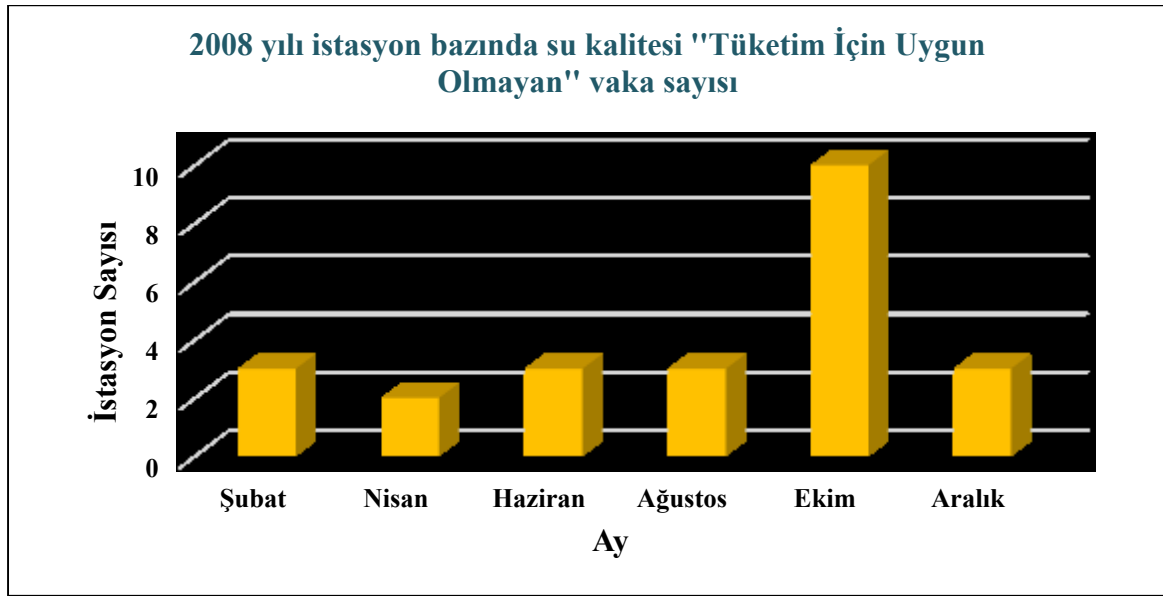
Şekil 4.30. 2006 yılı için Büyük Menderes Havzası'nda aylar bazında "Tüketim İçin Uygun Olmayan" su kalitesi vakaları

2007 yılında ise (Şekil 4.31) Şubat, Ağustos ve Ekim aylarında 6 istasyonda "Tüketim İçin Uygun Olmayan" su kalitesi vakası tespit edilerek en kirli aylar tanınmıştır. Fakat diğer aylarda (Nisan, Haziran, Aralık) oldukça ciddi sayıda vakalar olduğu söylenebilir.



Şekil 4.31. 2007 yılı için Büyük Menderes Havzası'nda aylar bazında "Tüketim İçin Uygun Olmayan" su kalitesi vakaları

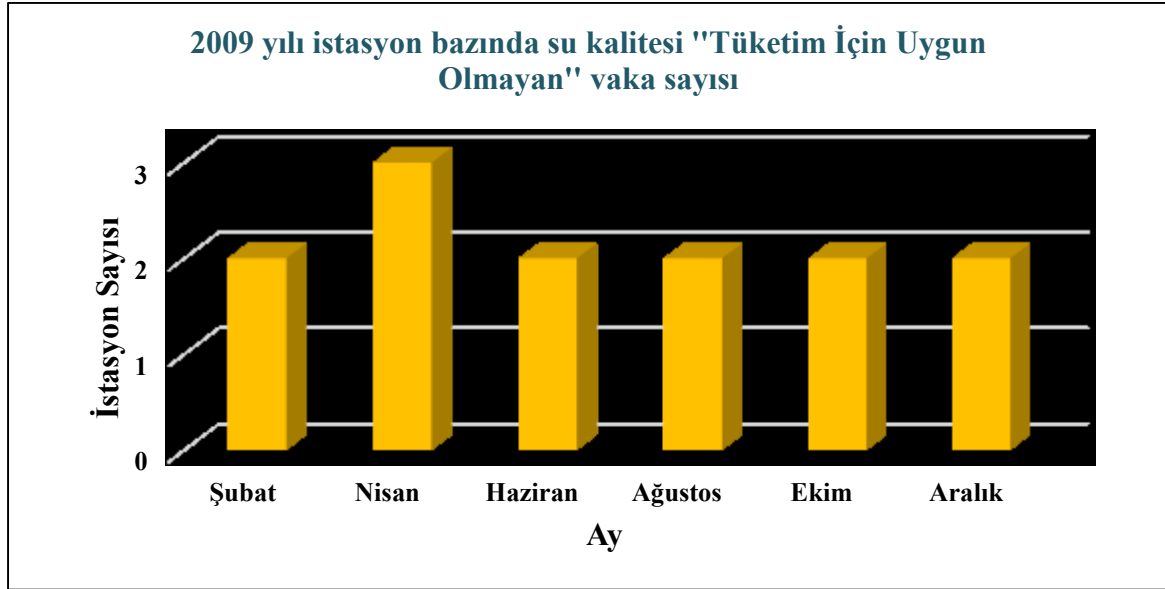
Büyük Menderes Havzası 2008 yılında Ekim ayına kadar “Tüketim İçin Uygun Olmayan” su kalitesi açısından oldukça az vakalı aylar olduğu görülmektedir (Şekil 4.32). Fakat Ekim ayında su kalitesi verileri göz önünde bulundurulan her 10 istasyonda da “Tüketim İçin Uygun Olmayan” su kalitesi vakası tespit edilmiştir. Önceki yıllar ile kıyaslandığında 2008 Ekim ayına Büyük Menderes Havzası’nın su kalitesi bakımından en kötü dönem olduğu söylenebilir.



Şekil 4.32. 2008 yılı için Büyük Menderes Havzası’nda aylar bazında “Tüketim İçin Uygun Olmayan” su kalitesi vakaları

Büyük Menderes Havzası 2009 yılı “Tüketim İçin Uygun Olmayan” su kalitesi vakalarına bakıldığında Nisan ayı 3 istasyonda vaka tespiti ile en kirli ay olduğu görülmektedir (Şekil 4.33).

2010 yılı Ekim ayında Büyük Menderes Havzası 7 istasyonda “Tüketim İçin Uygun Olmayan” su kalitesi tespitiyle yinede bir kirlilik patlaması yaşamıştır (Şekil 4.34).

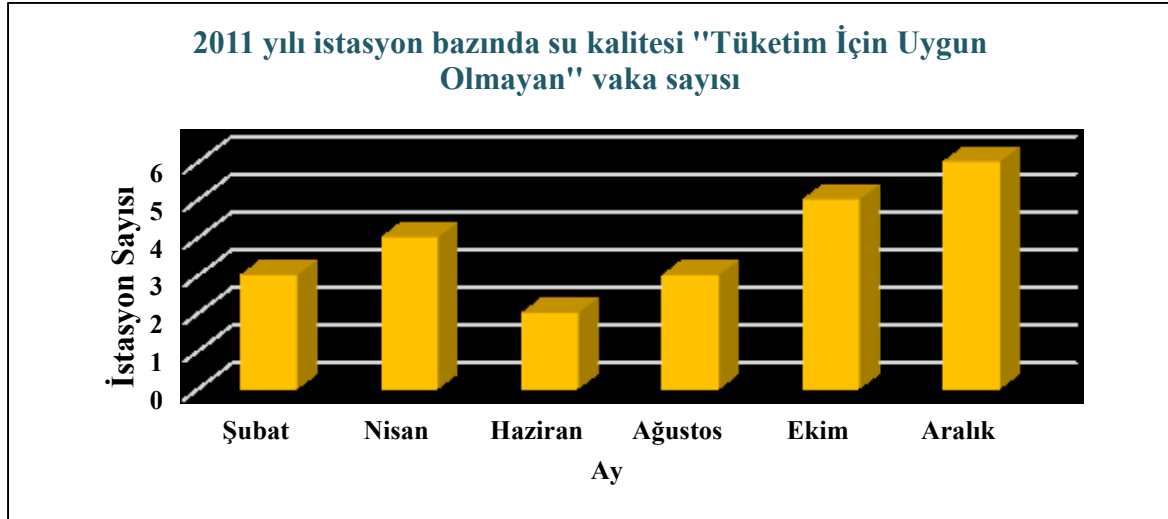


Şekil 4.33. 2009 yılı için Büyük Menderes Havzası'nda aylar bazında "Tüketim İçin Uygun Olmayan" su kalitesi vakaları



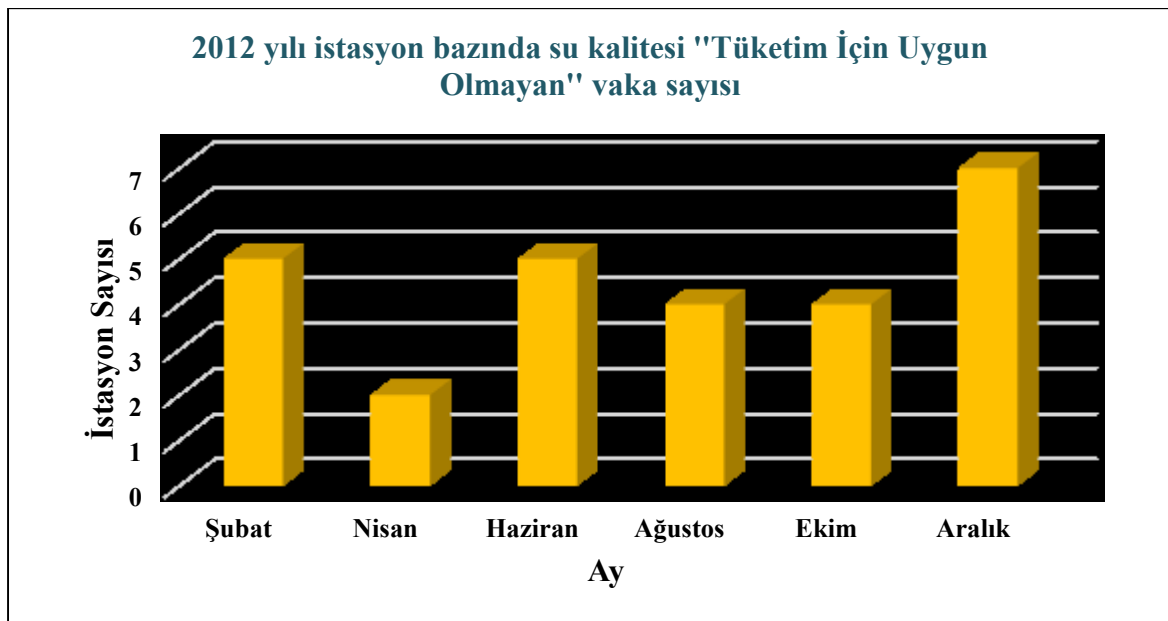
Şekil 4.34. 2010 yılı için Büyük Menderes Havzası'nda aylar bazında "Tüketim İçin Uygun Olmayan" su kalitesi vakaları

2011 yılında ise tüm yıl boyunca "Tüketim İçin Uygun Olmayan" su kalitesi vakaları gözlenmemektedir (Şekil 4.35). 2011 yılının Nisan, Ekim ve Aralık ayları sırayla 4, 5 ve 6 istasyonda "Tüketim İçin Uygun Olmayan" su kalitesi vakaları tespiti ile en kirli ayları olarak tespit edilmiştir.



Şekil 4.35. 2011 yılı için Büyük Menderes Havzası’nda aylar bazında “Tüketim İçin Uygun Olmayan” su kalitesi vakaları

2012 yılı Büyük Menderes Havzası’nın en kirli yıllarından birisi olduğu görülmektedir (Şekil 4.36). 2012’de tüm yıl boyunca ciddi “Tüketim İçin Uygun Olmayan” su kalitesi vakaları tespit edilmiştir. Aralık ayı 7 istasyonda “Tüketim İçin Uygun Olmayan” su kalitesi vakasıyla 2012 yılının en kirli ayı olarak tespit edilmiştir. Aralık ayının ardından Şubat ve Haziran ayları 5 istasyonda “Tüketim İçin Uygun Olmayan” su kalitesi vakalarıyla bu yılın diğer kirli ayları olarak tespit edilmiştir.



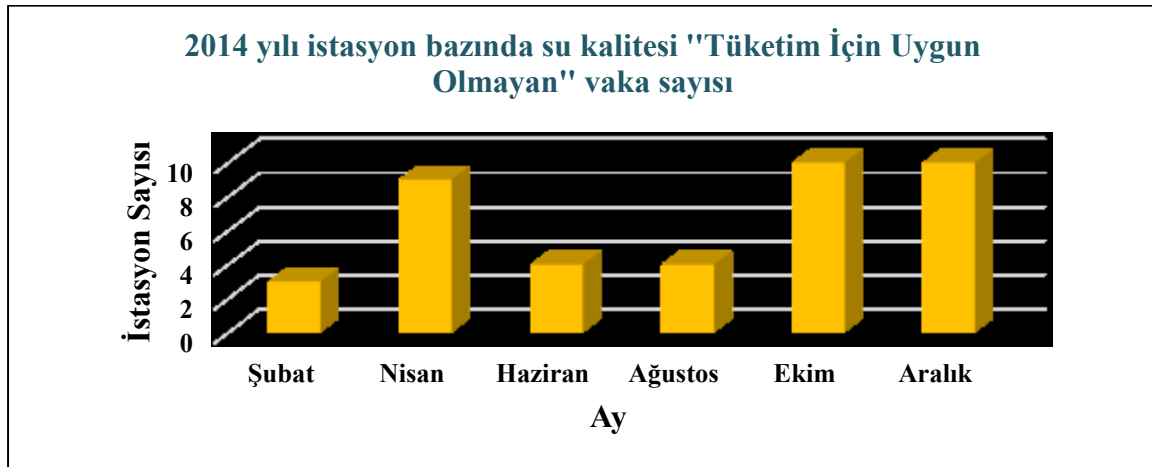
Şekil 4.36. 2012 yılı için Büyük Menderes Havzası’nda aylar bazında “Tüketim İçin Uygun Olmayan” su kalitesi vakaları

2013 yılı Şubat ayı dışında Büyük Menderes Havzası çokda ciddi bir kirlilik patlaması yaşamadığı görülmektedir (Şekil 4.37). Fakat Şubat 2013'e 8 istasyonda "Tüketim İçin Uygun Degil" su kalitesi vakası tespitiyle Büyük Menderes Havzası'nın su kalitesi sorunu yaşandığı söylenebilir.



Şekil 4.37. 2013 yılı için Büyük Menderes Havzası'nda aylar bazında "Tüketim İçin Uygun Olmayan" su kalitesi vakaları

2014 yılı Büyük Menderes Havzası'nda su kalitesi sorunlarının yaşandığı yıl olarak değerlendirilebilir (Şekil 4.38). Bu yılın Nisan, Ekim ve Aralık aylarında sırayla 8, 10 ve 10 istasyonda "Tüketim İçin Uygun Olmayan" su kalitesi tespitiyle su kalitesi sorunlarının yaşandığı görülmektedir.



Şekil 4.38. 2014 yılı için Büyük Menderes Havzası'nda aylar bazında "Tüketim İçin Uygun Olmayan" su kalitesi vakaları

4.2.3. Öznitelik seçimi sonuçları

ML sınıflandırma ve tahmin çalışmalarında en önemli konulardan birisi modelin verimli hale getirilmesidir. Bir model zaman-maliyet açısından verimli olması için aykırı hesaplamalar düşürülmelidir. Bu sonuca varmanın en önemli ve yaygın kullanılan yöntemlerinden birisi öznitelik seçimidir. Öznitelik seçimi yöntemleri, ilgisiz özelliklerin ortadan kaldırılmasına katkıda bulunan özelliklerin boyutlarını küçültmekle kalmaz, aynı zamanda ana sınıf/sınıflarla ilgili en etkili özellik alt kümesini de seçmektedir. Bu nedenle, öznitelik seçimi, veri kümesinin boyutunun küçülmesine, eğitim süresinin kısalmasına, sınıflandırmanın doğruluğunun artmasına ve aşırı uyum sorununun azalmasına neden olmaktadır. Öznitelik seçiminden kaynaklanan küçük bir altküme, sınıflandırıcıyı yanıltabileceğinden, en önemli özniteliklerin seçilmesi önemli bir konudur. FLTM, WRAPM ve EMBM yöntemleri, özellik seçme yöntemlerinde kullanılan kategorilerdir. Bu tez çalışmasında Büyük Menderes Havzası su kalitesi verileri ML modellerini üç kategori öznitelik seçimi yöntemlerinin en etkin özniteliklerden oluşan altkümeler kurulması için uygulanmıştır. Çizelge 4.2. Büyük Menderes Havzası veri seti için uygulanan öznitelik seçimi yöntemlerinin sonuçları gösterilmektedir. Büyük Menderes Havzası ana veri seti 19 parametreden oluşmaktadır. Bu çalışmada DSİ tarafından sağlanan veriler incelenerek düzenli ölçüm ve havzanın ana kirlilik yükü göz önünde bulundurularak sınıflandırma çalışmalarında kullanılmak üzere 19 parametreden oluşan veri seti oluşturulmuştur. Bu 19 parametre içinde Büyük Menderes Havzası'nın ana kirlilik kaynağı olan tarım ve hayvancılık faaliyetlerinden gelen besin madde ile ilişkili olan parametreler (KOİ, BOİ-5, NH₄, NO₃, NO₂, PO₄, SO₄) bulunmaktadır. Bunun yanı sıra bir havzanın su kalitesini oldukça negatif yönde etkileyen TÇK ile ilgili olan parametrelerde (Cl⁻, Fe, Mn, Na⁺, SO₄, TÇK) ele alınmıştır. Ayrıca su kalitesi izlenimi ve değerlendirmelerinde önemli rol oynayan Sıcaklık, pH, İletkenlik gibi parametrelerde izleme bilgileri arasında (Yıl, Ay, İstasyon) göz önünde bulundurulmuştur. Çizelge 4.2'e bakıldığında BFE-WRAPM öznitelik seçimi yöntemi 19 (yıl, ay ve istasyon dahil) parametre arasından 7 parametreyi su kalitesi sınıflandırmasında en etkin parametreler olarak önermiş ve bu doğrultuda daha sonra sınıflandırıcıların performansı üzerinde etkisini değerlendirmek için bu 7 parametreden oluşan altküme oluşturulmuştur. BFE-WRAPM yöntem izleme bilgilerinden İstasyon, İzleme Yılı ve Ay'ına göre ağırlık vermiştir. Bu yöntem su kalitesi genel özellikleri arasında pH'ı su kalitesi sınıflandırmasında daha etkin parametre bulmuş ve ÇO ve iletkenliği veri setinden çıkarmıştır. Besin madde zenginleşmesi ile ilişkili parametrelerden ise Amonyum,

Nitrit ve Fosfatı seçmiş Sülfat ve Nitrat'ı çıkarmıştır. Ayrıca TÇK madde konsantrasyonunu etkileyen mineraller ve bileşimler arasından sadece Manganez ve Sodyum'a ağırlık vermiştir.

Çizelge 4.2. Büyük Menderes Havzası veri seti için uygulanan öznitelik seçimi yöntemlerinin sonuçları

Veri seti	Öznitelik seçici	Parametre sayısı	Seçilen parametreler
1	Tüm veri seti	19	Yıl, Ay, İstasyon, Sıcaklık, pH, İletkenlik, ÇO, KOİ, BOİ-5, NH ₄ , NO ₃ , NO ₂ , PO ₄ , Cl ⁻ , Fe, Mn, Na ⁺ , SO ₄ , TÇK
2	BFE-WRAPM	7	İstasyon, pH, NH ₄ , NO ₂ , PO ₄ , Mn, Na ⁺
3	MI-FLTM	9	İstasyon, ÇO, KOİ, BOİ-5, NH ₄ , NO ₂ , PO ₄ , Mn, SO ₄
4	RF-EMBM	14	İletkenlik, ÇO, KOİ, BOİ-5, NH ₄ , NO ₃ , NO ₂ , PO ₄ , Cl ⁻ , Fe, Mn, Na ⁺ , SO ₄ , TÇK

MI-FLTM yöntemi aynı BFE-WRAPM yöntemi gibi izleme bilgilerinden İstasyon bilgilerini önemli bulmuş ve Yıl ve Ay'ı çıkarmıştır. Bu yöntem genel su kalitesi parametreleri arasından ÇO seçmiş, pH ve iletkenliği veri setinden çıkarmıştır. Besin madde kirlenmesiyle ilişkili olan parametrelere baktığımızda MI-FLTM yöntemi aynı BFE-WRAPM yöntemi gibi Amonyum, Nitrit ve Fosfat'ı seçmiş ve Nitrat'ı çıkarmıştır. Fakat bu parametrelerin yanında su kalitesi sınıflandırması için kullanılması ve dikkate alınması gereken KOİ, BOİ-5 ve Sülfat'ın göz önünde bulundurulması gereğini önermiştir. TÇK ile ilişkili olan parametreler arasında MI-FLTM yöntemi sadece Manganez'i seçmiş, Klorür ve Sodyum'u veri setinden çıkarmıştır. MI-FLTM yöntemi uygulaması sonucunda 9 parametreden oluşan altküme oluşturmuştur.

RF algoritmasıyla EMBM yöntemi 19 parametre arasından 14 parametre seçmiştir. RF-EMBM izleme bilgilerini önemsiz bulmuş ve hepsini çıkarmıştır. Bu yöntem diğer yöntemlere nazaran genel su kalitesi parametrelerine biraz daha ağırlık vermiş, iletkenlik ve ÇO su kalitesi sınıflandırmasında göz önünde bulundurulması gereğini önermiştir. Bu yöntem besin madde kirlenmesi ile ilişkili olan tüm parametrelerin tutulmasını önermiştir. TÇK ile ilişkili olan tüm parametreler bu yöntemin seçimi üzere tutulmuştur.

Genel bir değerlendirme yapıldığında her üç yöntem besin madde kirliliği ile ilgili olan Amonyum, Nitrit ve Fosfat'ın her zaman veri setinde bulundurulması gerektiği önerilmiştir.

Her üç yöntemde Manganezi TÇK ile ilişkili olan en önemli parametre bulmuş ve her zaman veri setinde bulundurulması önerilmiştir.

4.2.4. Sınıflandırma çalışması sonuçları

Tez çalışması kapsamında su kalitesi sınıflandırma çalışmasında iki yol izlenmiştir. Bunlardan birisi öznitelik seçimi yöntemleri uygulanarak oluşturulan veri seti boyutsallığı düşürülmüş altkümelerin etkisi ile sınıflandırıcıların performansı incelenmiştir. Ayrıca doğrulama yöntemlerinin her bir sınıflandırıcı bazında performansı da değerlendirilmiştir.

Birinci aşamada modeller 19 parametreden oluşan veri seti ile eğitilmiştir. Çizelge 4.3, modellerin tüm veri seti ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçları gösterilmektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%94,5454), kesinlik (%95,00) ve hassasiyet (%95,00)) XGBoost modelinin en iyi sonucu verdiği tespit edilmiştir. LogitBoost modeli ise elde edilen metrik değerlere (doğruluk (%94,3939), kesinlik (%94,40) ve hassasiyet (%94,40)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve RF modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%51,0606), kesinlik (%66,00) ve hassasiyet (%51,00); doğruluk (%79,3939), kesinlik (%78,90) ve hassasiyet (%79,40) olarak bulunmuştur. 5- FCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu XGBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.3. Tüm veri seti ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	94,3939	79,3939	51,0606	94,5454
Kesinlik (%)	94,40	78,90	66,00	95,00
Hassasiyet (%)	94,40	79,40	51,00	95,00

Çizelge 4.4, modellerin tüm veri seti ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçlarını göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%96,9696), kesinlik (%97,00), ve hassasiyet (%97,00)) XGBoost modelinin en iyi sonucu verdiği tespit edilmiştir. LogitBoost modeli ise elde edilen metrik değerlere ((doğruluk (%95,303), kesinlik (%95,30) ve hassasiyet (%95,30)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve RF modellerinin elde

ettiği metrik değerler sırasıyla doğruluk (%59,8484), kesinlik (%81,00) ve hassasiyet (%60,00); doğruluk (%80,6061), kesinlik (%79,80) ve hassasiyet (%80,60) olarak bulunmuştur. 10-FCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu XGBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.4. Tüm veri seti ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	95,303	80,6061	59,8484	96,9696
Kesinlik (%)	95,30	79,80	81,00	97,00
Hassasiyet (%)	95,30	80,60	60,00	97,00

Çizelge 4.5, modellerin tüm veri seti ve %70-30 eğitim-doğrulama oranıyla HCV yöntemiyle eğitildiği doğrulama sonuçlarını göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%96,9696), kesinlik (%92,60) ve hassasiyet (%90,90)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%79,2929), kesinlik (%77,80) ve hassasiyet (%79,30)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%53,0303), kesinlik (%56,00) ve hassasiyet (%53,00); doğruluk (%77,7777), kesinlik (%69,00) ve hassasiyet (%72,00) olarak bulunmuştur. %70-30 eğitim-doğrulama oranıyla HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.5. Modellerin tüm veri seti ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	90,9091	79,2929	53,0303	77,7777
Kesinlik (%)	92,60	77,80	56,00	69,00
Hassasiyet (%)	90,90	79,30	53,00	72,00

Çizelge 4.6, modellerin tüm veri seti ve %75-25 eğitim-doğrulama oranıyla HCV yöntemiyle eğitildiği doğrulama sonuçlarını göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%92,7273), kesinlik (%93,60) ve hassasiyet (%92,70)) LogitBoost modelinin en iyi sonucu verdiği tespit

edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%80,00), kesinlik (%80,40) ve hassasiyet (%80,00)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%54,5454), kesinlik (%65,00) ve hassasiyet (%55,00); doğruluk (%73,9393), kesinlik (%75,00) ve hassasiyet (%74,00) olarak bulunmuştur. %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.6. Tüm versi seti ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	92,7273	80,00	54,5454	73,9393
Kesinlik (%)	93,60	80,40	65,00	75,00
Hasasiyet (%)	92,70	80,00	55,00	74,00

Çizelge 4.7, modellerin tüm veri seti ve %80-20 eğitim-doğrulama oranıyla HCV yöntemiyle eğitildiği doğrulama sonuçlarını göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%91,6667), kesinlik (%93,20) ve hassasiyet (%91,70)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%84,0909), kesinlik (%83,70) ve hassasiyet (%84,10)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%63,6363), kesinlik (%62,00) ve hassasiyet (%64,00); doğruluk (%77,2727), kesinlik (%75,00) ve hassasiyet (%74,00) olarak bulunmuştur. %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.7. Tüm veri seti ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlar

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	91,6667	84,0909	63,6363	77,2727
Kesinlik (%)	93,20	83,70	62,00	75,00
Hassasiyet (%)	91,70	84,10	64,00	74,00

Çizelge 4.5-4.7, modellerin tüm veri seti ve farklı oranlarda (70-30, 75-25, 80-20) eğitim-doğrulama HCV yöntemiyle eğitildiği doğrulama sonuçlarını göstermektedir. Çalışmanın bu noktasına kadar modellerin en iyi doğrulama sonuçlarını 10-FCV yöntemiyle elde edildiğini göz önünde bulundurulduğunda, HCV yönteminin genel olarak verimsiz olduğu görülmektedir.

Çizelge 4.8, modelleri BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçlarını göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%93,6364), kesinlik (%93,60) ve hassasiyet (%93,60)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. XGBoost modeli ise elde edilen metrik değerlere ((doğruluk (%93,3333), kesinlik (%94,00) ve hassasiyet (%93,00)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve Random Forest modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%51,0606), kesinlik (%66,00) ve hassasiyet (%51,00); doğruluk (%84,0909), kesinlik (%83,70) ve hassasiyet (%84,10) olarak bulunmuştur. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.8. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	93,6364	84,0909	51,0606	93,3333
Kesinlik (%)	93,60	83,70	66,00	94,00
Hassasiyet (%)	93,60	84,10	51,00	93,00

Çizelge 4.9, BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%95,6060), kesinlik (%96,00) ve hassasiyet (%96,00)) XGBoost modelinin en iyi sonucu verdiği tespit edilmiştir. LogitBoost modeli ise elde edilen metrik değerlere ((doğruluk (%93,0303), kesinlik (%93,10) ve hassasiyet (%93,00)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve Random Forest modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%59,8484), kesinlik (%81,00) ve hassasiyet (%60,00); doğruluk

(%85,7576), kesinlik (%85,30) ve hassasiyet (%85,80) olarak bulunmuştur. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu XGBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.9. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	93,0303	85,7576	59,8484	95,6060
Kesinlik (%)	93,10	85,30	81,00	96,00
Hassasiyet (%)	93,00	85,80	60,00	96,00

Çizelge 4.10, BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%92,9293), kesinlik (%92,90) ve hassasiyet (%92,90)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%86,3636), kesinlik (%86,00) ve hassasiyet (%86,40)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%58,0808), kesinlik (%64,00) ve hassasiyet (%58,00); doğruluk (%80,3030), kesinlik (%80,00) ve hassasiyet (%80,00) olarak bulunmuştur. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.10. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	92,9293	86,3636	58,0808	80,3030
Kesinlik (%)	92,90	86,00	64,00	80,00
Hassasiyet (%)	92,90	86,40	58,00	80,00

Çizelge 4.11, BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında

(doğruluk (%93,3333), kesinlik (%93,40) ve hassasiyet (%93,30)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%88,4848), kesinlik (%88,80) ve hassasiyet (%88,50)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%60,00), kesinlik (%60,00) ve hassasiyet (%60,00); doğruluk (%81,8181), kesinlik (%82,00) ve hassasiyet (%82,00) olarak bulunmuştur. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.11. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	93,3333	88,4848	60,00	81,8181
Kesinlik (%)	93,40	88,80	60,00	82,00
Hassasiyet (%)	93,30	88,50	60,00	82,00

Çizelge 4.12, BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%92,4242), kesinlik (%92,80) ve hassasiyet (%92,40)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. XGBoost modeli ise elde edilen metrik değerlere ((doğruluk (%84,8484), kesinlik (%85,00) ve hassasiyet (%85,00)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve Random Forest modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%63,6363), kesinlik (%63,00) ve hassasiyet (%64,00); doğruluk (%84,0909), kesinlik (%84,80) ve hassasiyet (%84,10) olarak bulunmuştur. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.12. BFE-WRAPM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	92,4242	84,0909	63,6363	84,8484
Kesinlik (%)	92,80	84,80	63,00	85,00
Hassasiyet (%)	92,40	84,10	64,00	85,00

Çizelge 4.13, MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%93,4848), kesinlik (%94,00) ve hassasiyet (%94,00)) XGBoost modelinin en iyi sonucu verdiği tespit edilmiştir. LogitBoost modeli ise elde edilen metrik değerlere ((doğruluk (%93,33), kesinlik (%93,40) ve hassasiyet (%93,30)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve Random Forest modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%51,0606), kesinlik (%66,00) ve hassasiyet (%51,00); doğruluk (%83,1818), kesinlik (%82,50) ve hassasiyet (%83,20) olarak bulunmuştur. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu XGBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.13. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	93,33	83,1818	51,0606	93,4848
Kesinlik (%)	93,40	82,50	66,00	94,00
Hassasiyet (%)	93,30	83,20	51,00	94,00

Çizelge 4.14, MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 10-FCV yöntemiyle eğitilen doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%95,1515), kesinlik (%96,00) ve hassasiyet (%95,00)) XGBoost modelinin en iyi sonucu verdiği tespit edilmiştir. LogitBoost modeli ise elde edilen metrik değerlere ((doğruluk (%93,6364), kesinlik (%93,70) ve hassasiyet (%93,60)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve Random Forest modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%59,8484), kesinlik (%81,00) ve hassasiyet (%60,00); doğruluk

(%82,2727), kesinlik (%81,70) ve hassasiyet (%82,30) olarak bulunmuştur. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu XGBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.14. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 10-FCV yöntemiyle eğitilen doğrulama sonuçları

	Modeller			
	LogitBoost	Ranom Forest	AdaBoost	XGBoost
Doğruluk (%)	93,6364	82,2727	59,8484	95,1515
Kesinlik (%)	93,70	81,70	81,00	96,00
Hassasiyet (%)	93,60	82,30	60,00	95,00

Çizelge 4.15, MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%93,4343), kesinlik (%93,80) ve hassasiyet (%93,40)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%83,3333), kesinlik (%82,50) ve hassasiyet (%83,30)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%62,6262), kesinlik (%61,00) ve hassasiyet (%63,00); doğruluk (%80,8080), kesinlik (%82,00), ve hassasiyet (%81,00) olarak bulunmuştur. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.15. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	93,4343	83,3333	62,6262	80,8080
Kesinlik (%)	93,80	82,50	61,00	82,00
Hassasiyet (%)	93,40	83,30	63,00	81,00

Çizelge 4.16, MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk

(%92,7273), kesinlik (%93,00) ve hassasiyet (%92,70)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%82,4242), kesinlik (%81,90) ve hassasiyet (%82,40)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%62,4242), kesinlik (%67,00) ve hassasiyet (%62,00); doğruluk (%81,2121), kesinlik (%83,00) ve hassasiyet (%81,00) olarak bulunmuştur. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.16. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	92,7273	82,4242	62,4242	81,2121
Kesinlik (%)	93,00	81,90	67,00	83,00
Hassasiyet (%)	92,70	82,40	62,00	81,00

Çizelge 4.17, MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%91,6667), kesinlik (%92,40) ve hassasiyet (%91,70)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%82,5758), kesinlik (%82,40) ve hassasiyet (%82,60)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%71,9696), kesinlik (%70,00) ve hassasiyet (%72,00); doğruluk (%78,7878), kesinlik (%82,00) ve hassasiyet (%79,00) olarak bulunmuştur. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.17. MI-FLTM Yöntemi'nin önerdiği parametrelerden oluşan altküne ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	91,6667	82,5758	71,9696	78,7878
Kesinlik (%)	92,40	82,40	70,00	82,00
Hassasiyet (%)	91,70	82,60	72,00	79,00

XGBoost modeli için Çizelge 4.13-4.17 sonuçlarını kıyasladığımızda bu modelin HCV yöntemi uygulanmasından negatif etkilenmiştir.

Çizelge 4.18, RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküne ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%95,7576), kesinlik (%95,90) ve hassasiyet (%95,80)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. XGBoost modeli ise elde edilen metrik değerlere ((doğruluk (%94,6969), kesinlik (%95,00) ve hassasiyet (%95,00)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve Random Forest modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%51,06), kesinlik (%66,00) ve hassasiyet (%51,00); doğruluk (%81,2121), kesinlik (%80,40) ve hassasiyet (%81,20) olarak bulunmuştur. RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküne ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.18. RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküne ve 5-FCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	95,7576	81,2121	51,06	94,6969
Kesinlik (%)	95,90	80,40	66,00	95,00
Hassasiyet (%)	95,80	81,20	51,00	95,00

Çizelge 4.19, RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküne ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%96,2121), kesinlik (%97,00) ve hassasiyet (%96,00)) XGBoost modelinin en iyi sonucu verdiği tespit edilmiştir. LogitBoost modeli ise elde edilen metrik değerlere ((doğruluk (%95,1515),

kesinlik (%95,20) ve hassasiyet (%95,20)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve Random Forest modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%59,8484), kesinlik (%81,00) ve hassasiyet (%60,00); doğruluk (%81,9697), kesinlik (%81,30) ve hassasiyet (%81,20) olarak bulunmuştur. RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu XGBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.19. RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve 10-FCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	95,1515	81,9697	59,8484	96,2121
Kesinlik (%)	95,20	81,30	81,00	97,00
Hassasiyet (%)	95,20	81,20	60,00	96,00

Çizelge 4.20, RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%93,4343), kesinlik (%94,40) ve hassasiyet (%93,40)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%80,8081), kesinlik (%79,10) ve hassasiyet (%80,80)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%58,5858), kesinlik (%61,00) ve hassasiyet (%59,00); doğruluk (%79,7979), kesinlik (%79,00) ve hassasiyet (%80,00) olarak bulunmuştur. RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.20. RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %70-30 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	93,4343	80,8081	58,5858	79,7979
Kesinlik (%)	94,40	79,10	61,00	79,00
Hassasiyet (%)	93,40	80,80	59,00	80,00

Çizelge 4.21, RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%91,5152), kesinlik (%92,90) ve hassasiyet (%91,50)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen metrik değerlere ((doğruluk (%83,6364), kesinlik (%83,50) ve hassasiyet (%83,60)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%67,2727), kesinlik (%65,00) ve hassasiyet (%67,00); doğruluk (%81,8181), kesinlik (%81,00) ve hassasiyet (%82,00) olarak bulunmuştur. RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.21. RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %75-25 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	91,5152	83,6364	67,2727	81,8181
Kesinlik (%)	92,90	83,50	65,00	81,00
Hassasiyet (%)	91,50	83,60	67,00	82,00

Çizelge 4.22, RF-EMBM öznitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları göstermektedir. Ele alınan modellerde, değerlendirme metriklerini göz önünde alındığında (doğruluk (%91,6667), kesinlik (%93,80) ve hassasiyet (%91,70)) LogitBoost modelinin en iyi sonucu verdiği tespit edilmiştir. Random Forest modeli ise elde edilen

metrik değerlere ((doğruluk (%87,1212), kesinlik (%87,20) ve hassasiyet (%87,10)) göre en iyi ikinci sonucu vermiştir. Ele alınan modeller arasında AdaBoost ve XGBoost modellerinin elde ettiği metrik değerler sırasıyla doğruluk (%71,9696), kesinlik (%69,00) ve hassasiyet (%72,00); doğruluk (%81,8181), kesinlik (%81,00) ve hassasiyet (%82,00) olarak bulunmuştur. RF-EMBM öz nitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçlarına göre en iyi sonucu LogitBoost, en kötü sonucu ise AdaBoost modelinin verdiği tespit edilmiştir.

Çizelge 4.22. RF-EMBM öz nitelik seçimi Yöntemi'nin önerdiği parametrelerden oluşan altküme ve %80-20 eğitim-doğrulama oranlı HCV yöntemiyle eğitildiği doğrulama sonuçları

	Modeller			
	LogitBoost	Random Forest	AdaBoost	XGBoost
Doğruluk (%)	91,6667	87,1212	71,9696	81,8181
Kesinlik (%)	93,80	87,20	69,00	81,00
Hassasiyet (%)	91,70	87,10	72,00	82,00

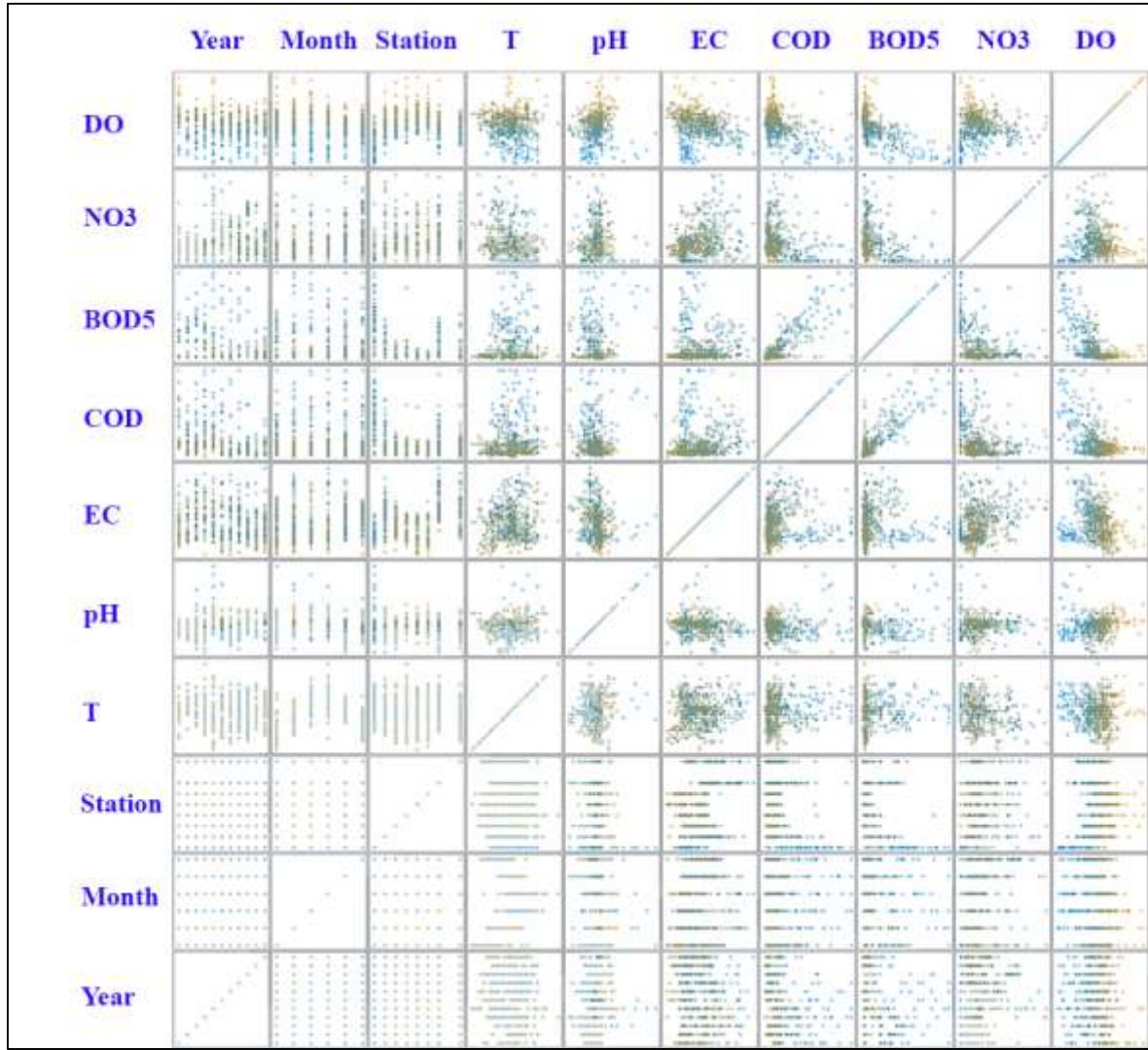
4.3. Çözünmüş Oksijen (ÇO) Konsantrasyonu Tahmin Modellemesi Çalışması Sonuçları

ÇO tahmin çalışması uygulama prosedüründe, bu uygulamalardan sonra RF modeli için en iyi sonuçlar hem tüm veri seti (13 parametreden oluşan veri seti) hemde boyutsallığı düşürülmüş veri setiyle 10-FCV doğrulama yöntemiyle elde edildiği için bu modelin 10-FCV doğrulamayla tahmin sonuçları grafikler ve çizelgeler ile görselleştirilmiştir. Aynı konu MLP tahmin modeli içinde geçerli olarak, bu model hem tüm veri seti hemde boyutsallığı düşürülmüş veri setleri ve %70-30 eğitim-deneme oranı HCV yöntemiyle en üstün sonuçları verdiği için, MLP modelinin ÇO tahmin çalışmasında sonuçları bu koşullar ile Çizelgelere yansıtıldı ve grafikler ile görselleştirilmiştir.

4.3.1. Öz nitelik seçimi sonuçları

Bu çalışmada, ÇO tahmininde ML algoritmalarına yardımcı olmak için seçilen 13 ilgili parametre arasından, veri setinin boyutunu azaltmak için BFE-WRAPM yöntem ile 9 (Yıl, Ay, İstasyon, Sıcaklık, pH, iletkenlik, KOİ, BOİ-5, Nitrat) parametre seçilmiştir. Seçilen değişkenler arasındaki korelasyon çizim matrisi, Şekil 4.39'de görülmektedir. Elenen

değişkenler ÇO konsantrasyonu ile yakın ilişki içinde olmasına rağmen, BFE-WRAPM öznitelik seçici, bunları gürültünün nedeni olarak bulmuştur (Chandrashekar ve Sahin, 2014). İlgisiz değişkeni ortadan kaldıran ve en belirgin değişkenleri tutan bir öznitelik seçici algoritmasının faydası, sınıflandırıcının veya öngörücünün performansını artırmak için veri kümesinin boyutsallığını ve dolayısıyla gürültüsünü azaltmaktır. Kullanılan WRAPM öznitelik seçicinin sonucu, algoritmanın tahmin edicilere yardımcı olmak için yalnızca bir adet Azot kaynağı tutma eğiliminde olduğunu tespit edilmiştir. Özellik seçicinin, oluşturulan özellik alt kümesinde tutmak için Nitrit yerine Nitratı seçmesi dikkat çeken sonuç olarak tespit edilmiştir. Ayrıca suda besin kirliliğinin ana kaynağı olan Fosfor, ilgilenilen değişken setinden çıkarılmıştır. Bir su sisteminde Fosfor ve ÇO konsantrasyonu arasındaki ilişki; Fosfor bir su sistemine salındığında besin zenginleşmesine ve ardından alg patlamasına neden olur ve sonuç olarak ÇO konsantrasyonu önemli ölçüde azaltır (Varol ve Şen, 2012; Motevalli ve diğerleri, 2019). Ancak uygulanan WRAPM öznitelik seçici, ÇO konsantrasyonunu tahmin etmek için Fosforu özelliklerin alt kümesinde seçmemiştir. Ayrıca, gürültüyü ve boyutsallığı azaltmak için özdeş değişkenleri ortadan kaldırma prosedürüne rağmen, KOİ ve BOİ-5 birlikte WRAPM önerdiği alt kümede tutulmuştur.



Şekil 4.39. BFE-WRAPM öznitelik seçici tarafından seçilen değişkenler arasındaki korelasyon çizim matrisi

Gözlem bilgilerinin (Yıl, Ay ve İstasyon) yanı sıra, T ve pH gibi tabakalaşma göstergeleri ve iletkenlik gibi mineral bütçe göstergelerinin (Ouma ve diğerleri, 2020) ÇO konsantrasyonunu tahmin etmede temel parametreler olarak bulunmuştur. WRAPM öznitelik seçici tarafından seçilen değişkenleri doğrulayan birçok araştırmacı, herhangi bir özellik seçme yöntemi kullanmadan uzmanlıklarına dayalı olarak çalışmalarında giriş suyu kalitesi değişkenleri olarak seçmişlerdir (Heddami ve Kisi, 2017; Keshtegar ve Heddami, 2018; Csábrági ve diğerleri, 2019; Ouma ve diğerleri, 2020; Zhu ve Heddami, 2020; Dehghani ve diğerleri, 2022). Bazı araştırmacılar seçilen değişkenler (gözlem bilgisi ve tabakalaşma göstergeleri) dışında KOİ gibi parametreleri de dahil etmişlerdir (Zhu ve Heddami, 2020). Ancak su sistemlerinde ÇO seviyeleri ile yakın ilişki içinde olan NO₃ ve

BOİ-5 gibi parametrelerin etkisinin bu tür çalışmalarda daha fazla ele alınması gerektiği değerlendirilmektedir.

4.3.2. Çözünmüş Oksijen miktarı tahmin çalışması sonuçları

Çizelge 4.23’de Büyük Mendere Havzası’ndaki ÇO konsantrasyonunu tahmin etmek için uygulanan her bir ML modelinin performansını değerlendirmek için kullanılan metriklerin sonuçları görülmektedir.

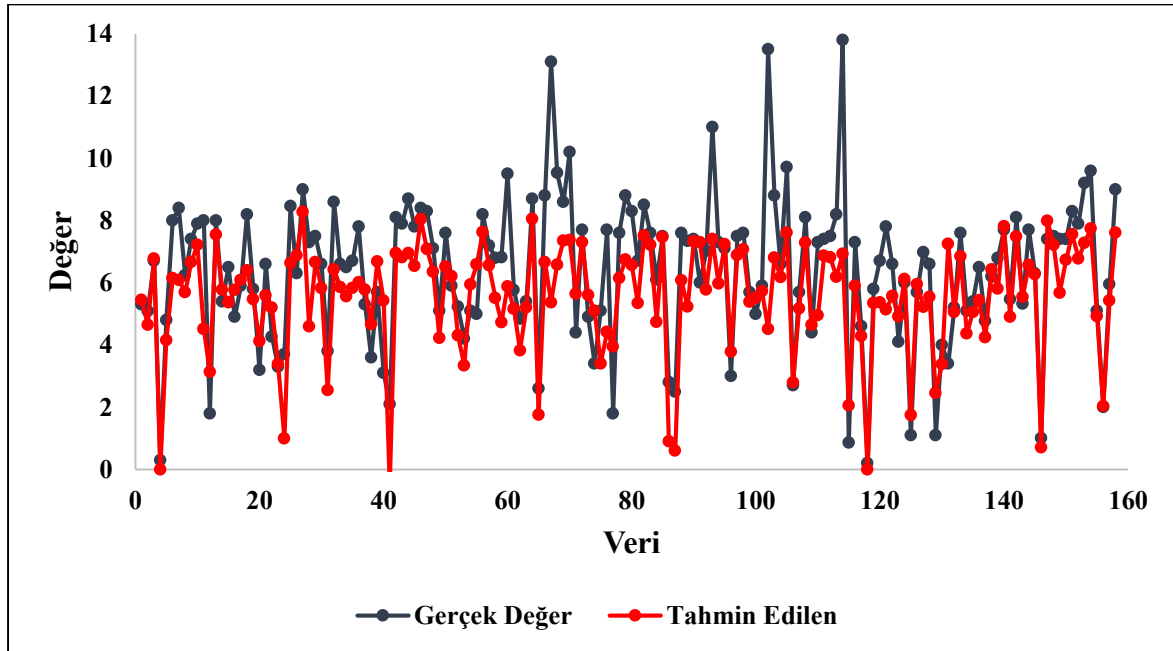
Çizelge 4.23. Tahmin modellerinin performans değerlendirme sonuçları

Model	Alt Küme	Pearson Correlation Coefficient (PCC)	MAE	RMSE
Random Forest	Tüm veri seti	0,7958	0,9188	1,3093
MLP	Tüm veri seti	0,7586	1,2424	1,7291
Random Forest	Boyutu Düşürülmüş Alt Küme	0,8052	0,8911	1,2805
MLP	Boyutu Düşürülmüş Alt Küme	0,7367	1,2495	1,7525

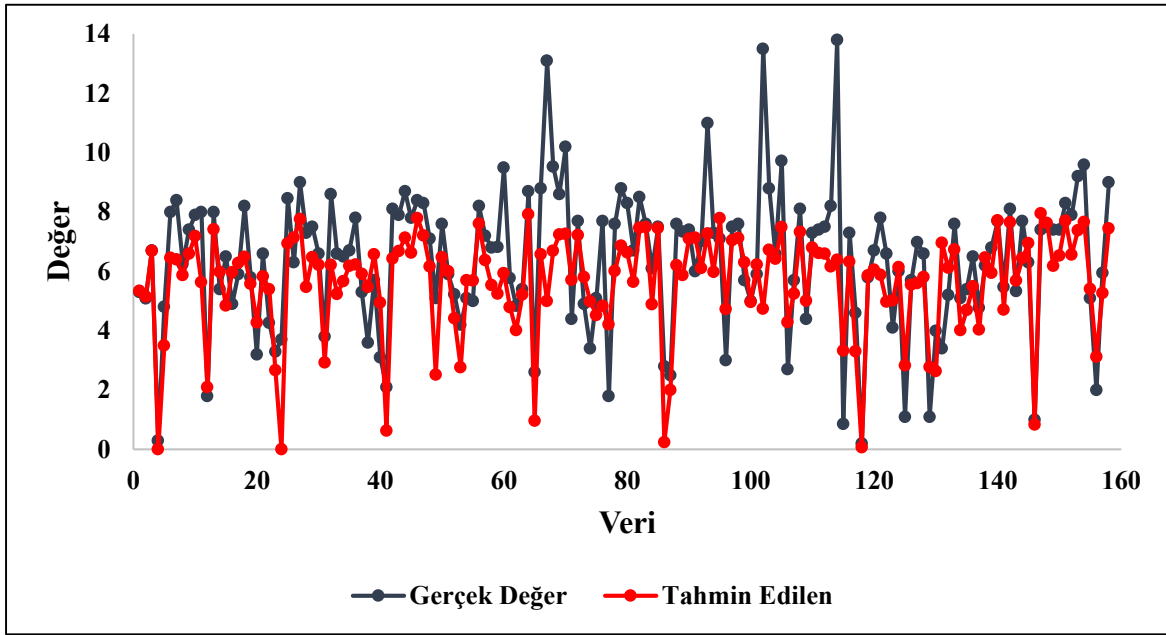
RF tahmin modeli tüm metrikler açısından MLP'ye göre daha iyi performans göstermiş, ancak azaltılmış değişkenlerin tahmin performansı üzerindeki etkisi dikkate alınması gerekmektedir. Tahmin sürecine 9 değişken dahil edildiğinde RF tahmin modelinin performansı bir miktar artarken, özellikleri azaltılmış alt kümesi dikkate alındığında MLP'nin performansı önemli ölçüde azalmaktadır. Bu sonuç, ANNs algoritmalarında özellik açısından boyutu düşürülmüş alt kümesi kullanılması olumsuz etki gösterebilir. Bir veri kümesinin içinde ne kadar çok parametre bulunursa ANNs o kadar iyi performans sunar ayrıca ilgili değişkenlerin ANNs üzerinde olumsuz etkisi olmamaktadır. Başka bir deyişle, ANNs modellerinin RF gibi ağaç tabanlı modellere göre gürültüye daha az eğilimli olduğu sonucuna varılabilir. Csábrágyi ve diğerleri (2019), çalışmalarında farklı şekilde optimize

edilmiş ANNs modellerinin performansını karşılaştırdıklarını ve ANNs modelinin performansının daha büyük bir veri kümesiyle önemli ölçüde arttığını bildirmişlerdir. Ouma ve diğerleri (2020)'da bir ANNs modelinin performansını bir istatistiksel modelle karşılaştırdıkları çalışmalarında, ÇO tahmini sürecine ilgili tüm değişkenler dahil edildiğinde ANNs modelinin performansının arttığını tespit etmişlerdir.

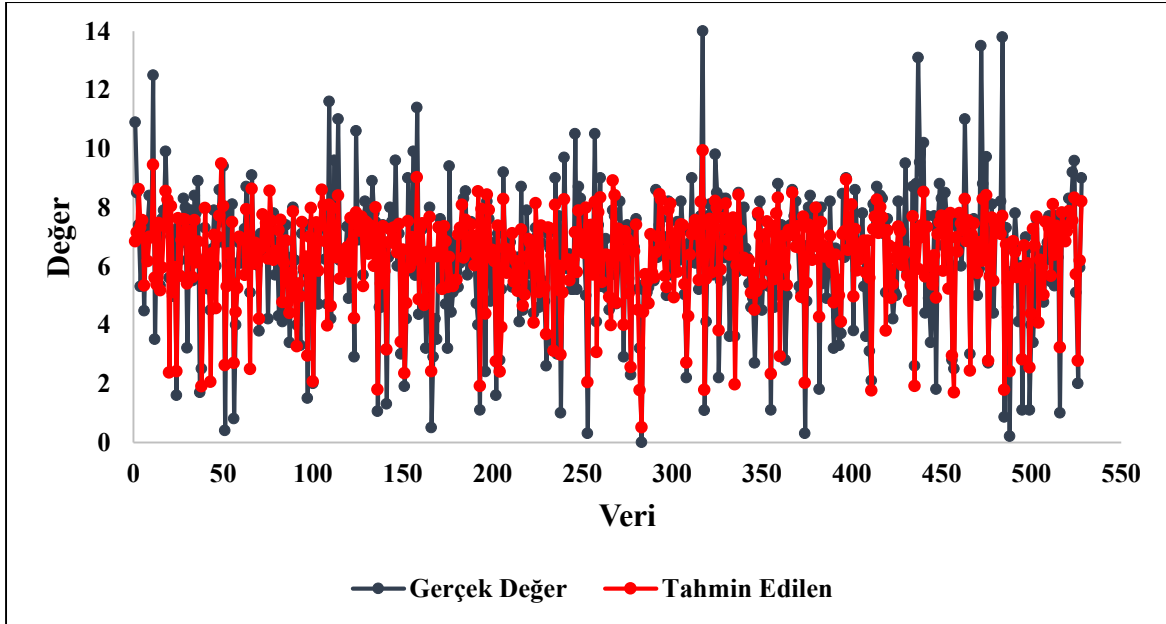
Çizelge 4.23. incelendiğinde, MLP tahmin modelinin boyutsallığı düşürülmüş veri seti dahil edilince PCC metriğinin değeri yaklaşık olarak yüzde iki (%2,00) oranında düşmüş olup, tahminin yakınlığını gösteren MAE artmamıştır. Bu nedenle, tahminlerin gerçek ÇO konsantrasyonlarından uzaklaşması sonucuna varılabilir (Şekil 4.40 ve Şekil 4.41). Şekil 4.41'de görüldüğü gibi Şekil 4.40'e nazaran tahmin edilen değerlerin gerçek değerlerden uzaklaştığı tespit edilmiştir. Boyutsallığı azaltılmış alt kümesiyle RF tahmin modeli PCC'deki artışa rağmen, MAE ve RMSE önemli ölçüde azaldığı görülmektedir (Çizelge 4.22). Şekil 4.42 ve Şekil 4.43'de görüldüğü gibi boyutsallık azalmasıyla birlikte tahmin edilen değerler azda olsa gerçek değerlere yaklaşmaktadır. Bu, gürültü ve boyutsallık düştüğünde algoritma tarafından tahminlerin performansının arttığını göstermektedir.



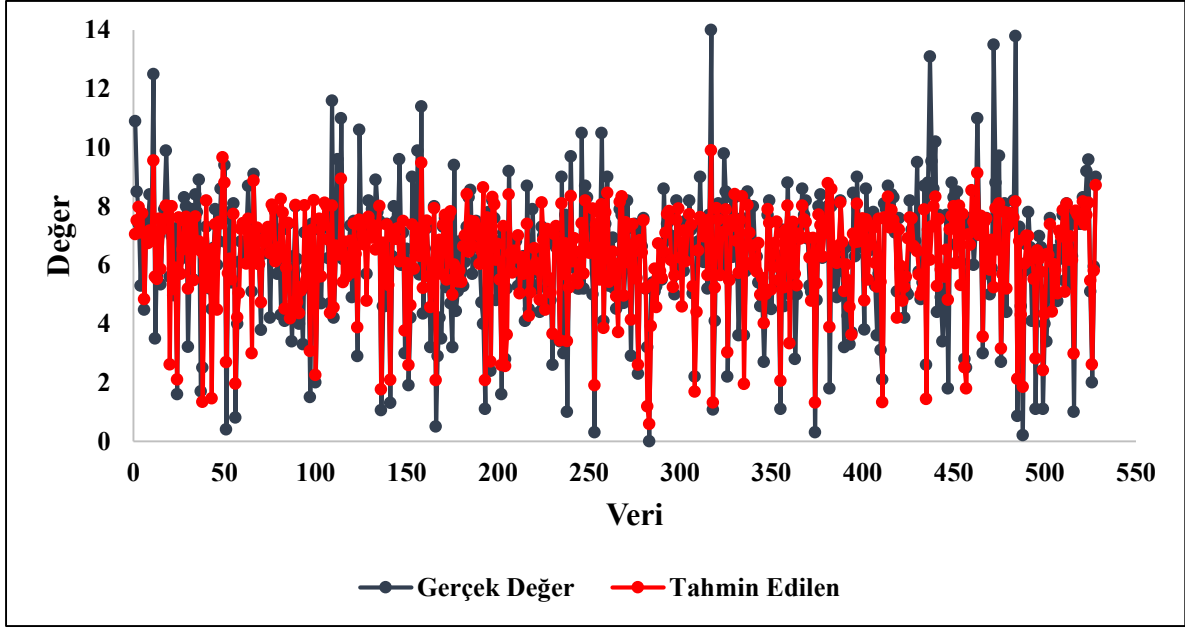
Şekil 4.40. Tüm veri seti ve %70-30 eğitim-deneme oranlı HCV doğrulama metodu uygulanmasıyla MLP modelinin gerçek ve tahmin edilen değerleri



Şekil 4.41. Boyutsallığı düşürülmüş veri seti ve %70-30 eğitim-deneme oranlı HCV doğrulama metodu uygulanmasıyla MLP modelinin gerçek ve tahmin edilen değerleri



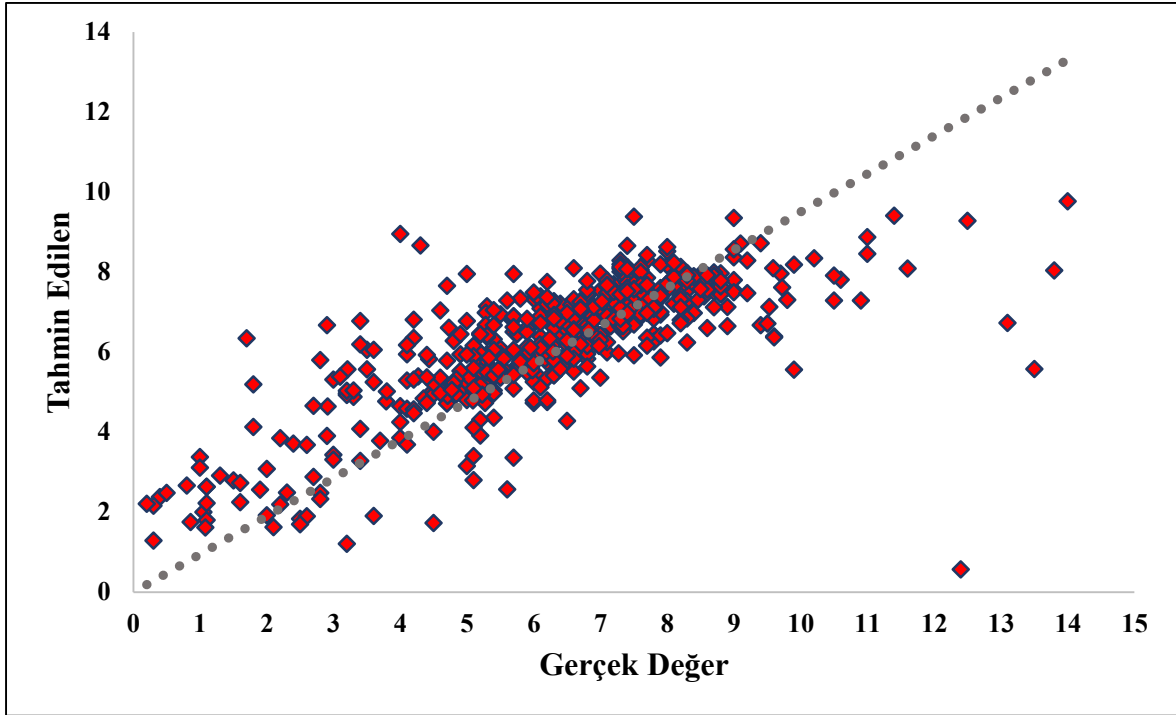
Şekil 4.42. Tüm veri seti ve 10-FCV doğrulama metodu uygulanmasıyla RF modelinin gerçek ve tahmin edilen değerleri



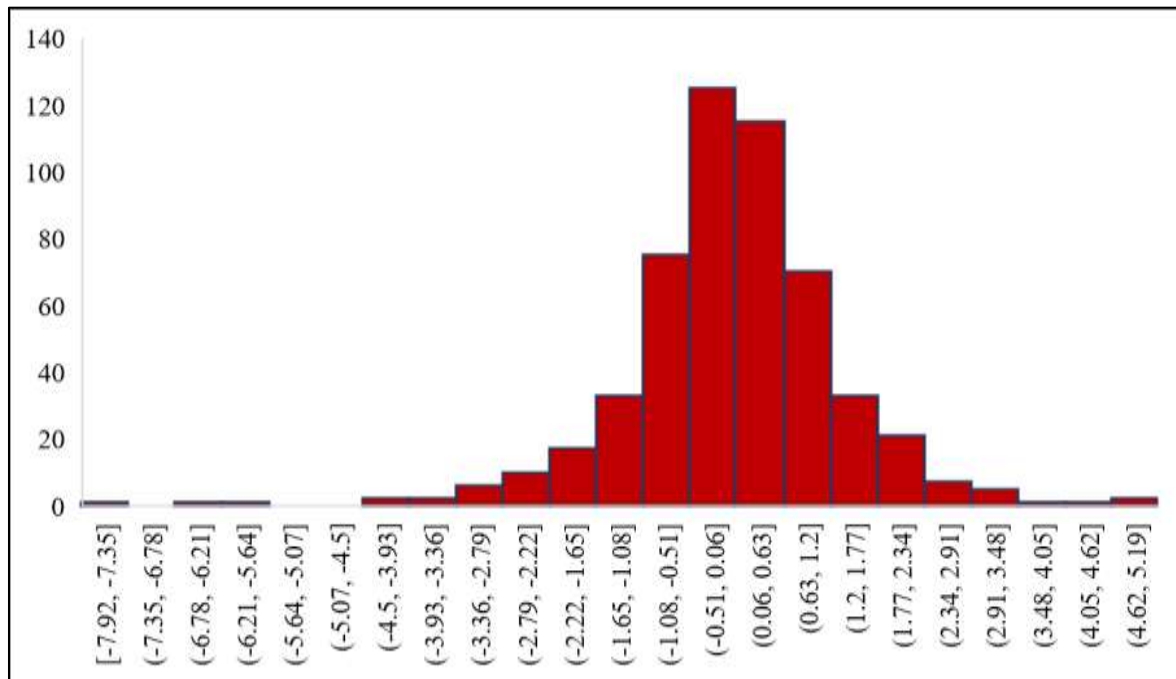
Şekil 4.43. Boyutsallığı düşürülmüş veri seti ve 10-FCV doğrulama metodu uygulanmasıyla RF modelinin gerçek ve tahmin edilen değerleri

Şekil 4.44, ilgili tüm değişkenleri göz önünde bulundurarak RF tahmin modeli tarafından yapılan ÇO'nin gerçek değerlerine tahmin edilen değerlerin uyumluluğunu göstermektedir. Şekil 4.44'de görüldüğü gibi, 5 ile 9 mg/L arasındaki ÇO konsantrasyonları için, algoritma tarafından tahmin edilen değerler ÇO'nun gerçek değerlerine yakın olduğu tespit edilmiştir. Ayrıca, bu aralığın altında ve üstündeki değerler algoritma tarafından tam olarak tahmin edilememiştir.

RMSE'yi analiz edersek, tüm değişkenlerle eğitilmiş RF tahmin modelinin (Şekil 4.45) Tahmin Hatası Sapması (Estimation Error Deviation (EED)) daha düşüktür (-1 ile 1.3 arasında). Test aşamasında tahmin edilen değerlerin yaklaşık %72,91'i bu hata aralığında tespit edilmiştir. Kabul edilebilir hata aralığı dışında kalan değerlerin (%27,09) EED'si bu aralığı aşmasına rağmen, tahmin edilen değerlerin yalnızca %7,38'i kabul edilemez EED aralığında ($2,34 < EED < -2,22$) kaldığı belirlenmiştir.



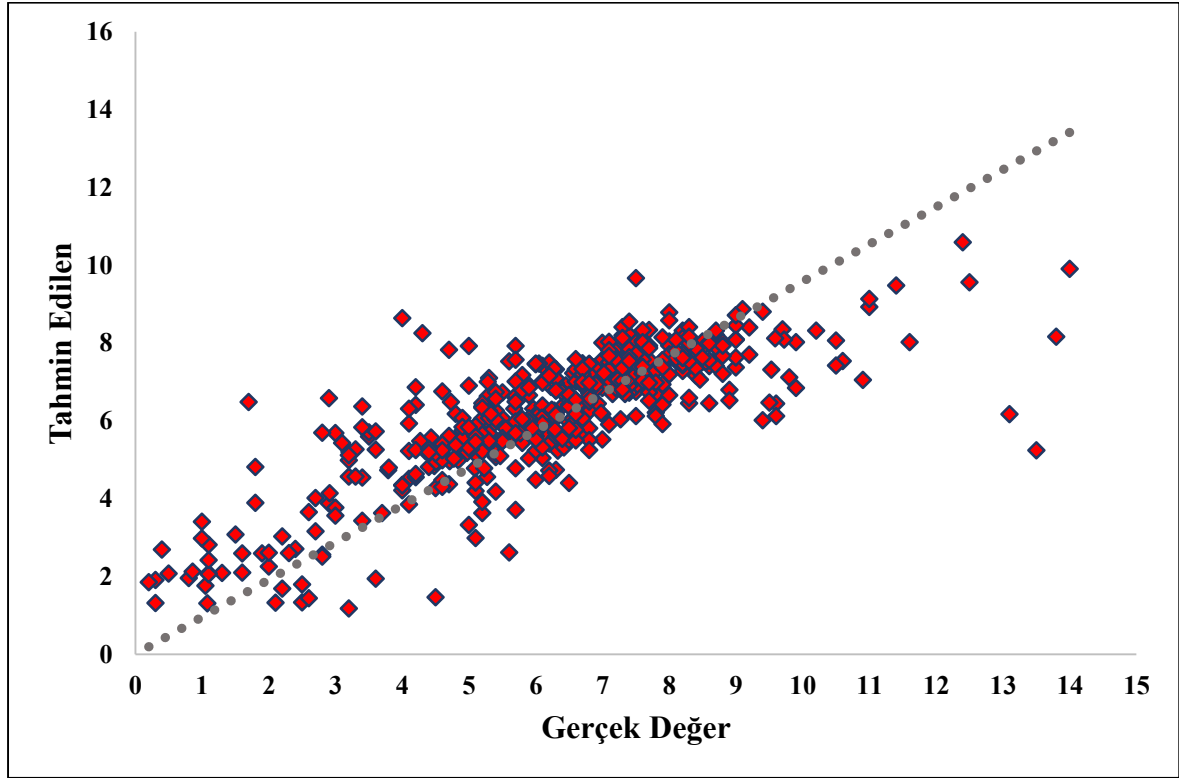
Şekil 4.44. Tüm veri seti ile RF tahmin modelinin gerçek ÇO değerine karşı tahmin edilen ÇO konsantrasyonu.



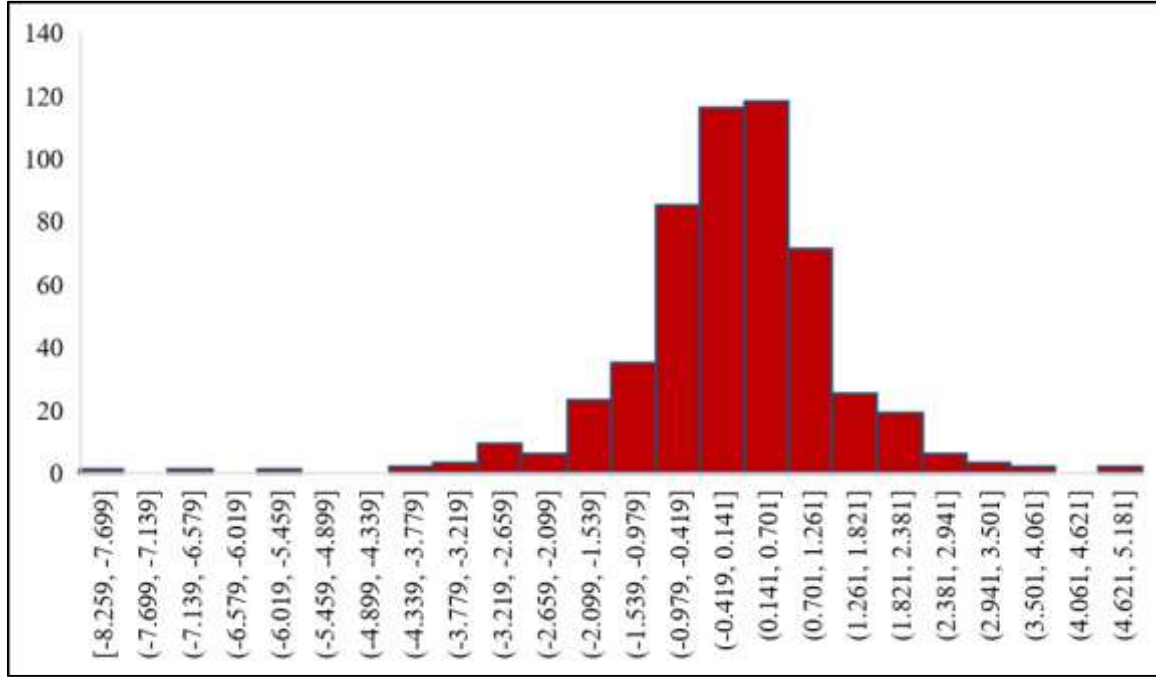
Şekil 4.45. Tüm değişkenler ile eğitilen RF modelinin test aşamasındaki EED'sı.

RF tahmin modelinin gerçek ve tahmin edilen dağılım grafiğinin uyumu karşılaştırıldığında, boyutsallığı düşürülmüş alt küme ile ÇO'nin tahmin edilen değerleri gerçek değerlere daha uyumlu olduğu görülmektedir (Şekil 4.46). Şekil 4.44'deki sonuçlara benzer şekilde, RF

tahmin modeli, test aşamasında boyutsallığı düşürülmüş alt küme ile denetlendiğinde, 5 ile 9 mg/L konsantrasyonları arasındaki ÇO değerlerini en ideal şekilde tahmin edebilmiştir. Azaltılmış özellik alt kümesine sahip modelin tahmin EED histogram grafiği (Şekil 4.47), veri kümesinin gürültü giderme ve boyutsallık azaltmanın model tahmin yeteneği üzerindeki olumlu etkisini doğrulamaktadır. Burada, tahmin modeli, -1 ile 1,26 arasında mükemmel bir EED aralığı ile örneklerin %73,86'sını tahmin edebilmiştir. Örneklerin yalnızca %6,81'i kabul edilemez bir EED aralığında tahmin edilmiştir ($2,38 < \text{EED} < -2,1$).



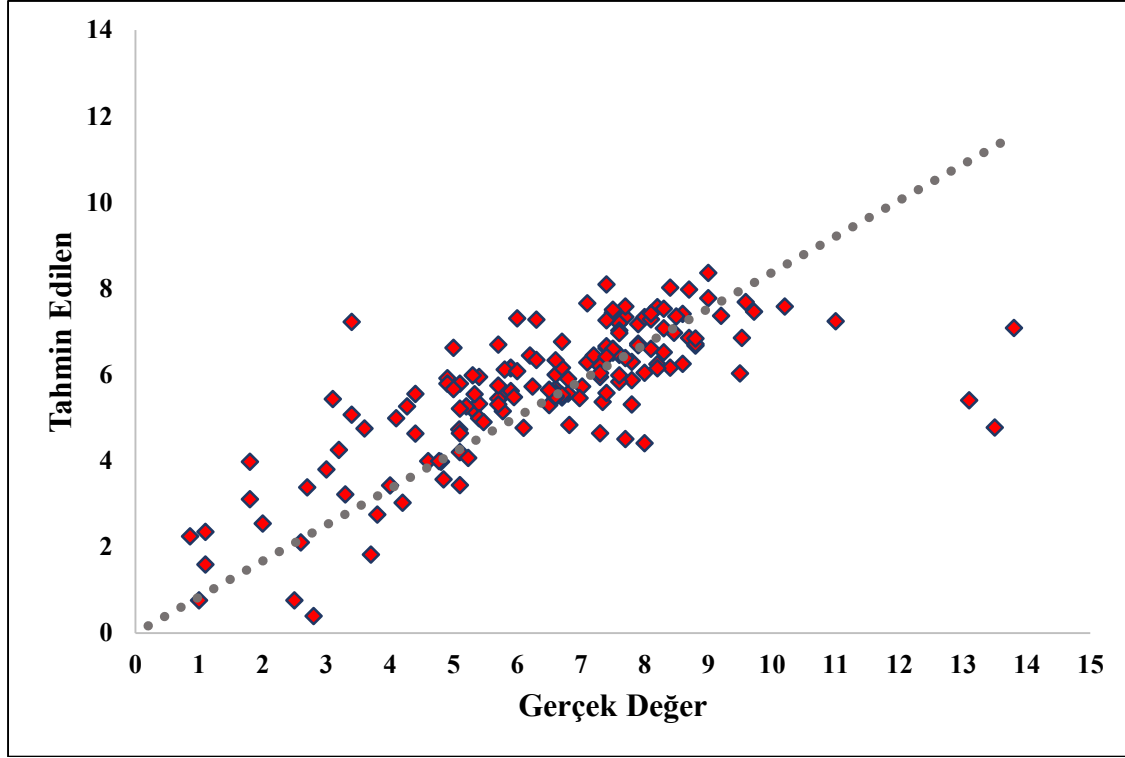
Şekil 4.46. Boyutsallığı düşürülmüş alt küme ile RF tahmin modelinin gerçek ÇO değerine karşı tahmin edilen ÇO konsantrasyonu.



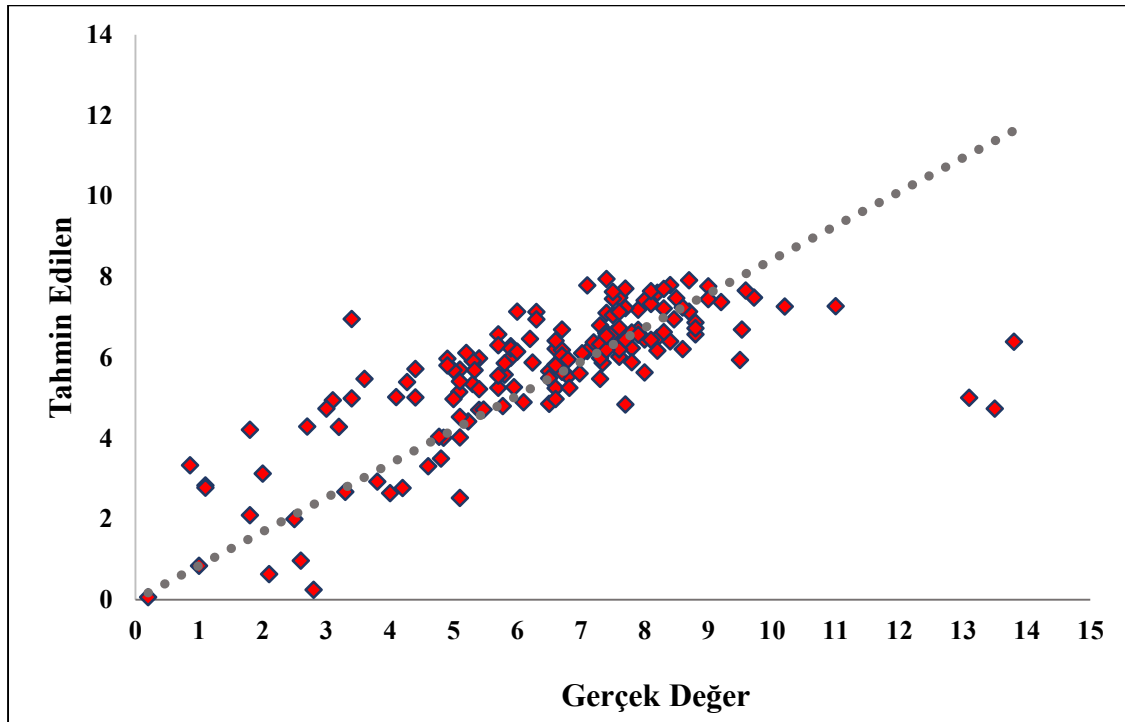
Şekil 4.47. Boyutsallığı düşürülmüş alt küme ile eğitilen RF modelinin test aşamasındaki EED'sı.

Elde edilen sonuçlara göre MLP algoritması, tahmin yeteneğinin büyük ölçüde eğitim aşamasındaki değişkenlerin sayısına bağlı olduğunu göstermektedir (Çizelge 4.23). Başka bir deyişle, eğitim aşamasına ne kadar çok değişken dahil edilirse, MLP algoritması test aşamasındaki değerleri o kadar doğru tahmin edebilmektedir.

Şekil 4.48 ve Şekil 4.49, ilgili tüm değişkenler ve boyutsallığı düşürülmüş alt küme alt ile MLP'nin gerçek ve tahmin edilen ÇO değerlerinin dağılım grafiğini göstermektedir. Karşılaştırmalı olarak ilgili tüm değişkenler eğitim aşamasına dahil edildiğinde çizim daha uygun görülmektedir. Şekil 4.48 ve Şekil 4.49 kıyaslandığında, MLP algoritmasının sınırlı sayıda özellik yerine daha fazla kaynaktan eğitime bağımlılığını göstermektedir. MLP algoritması, bir hedef parametre değerinin tahmini için çok sayıda özelliğe duyarlıdır. Bu ağaç tabanlı algoritmaların mekanizmasına ve sonuçlarına aykırıdır (Kohavi ve John, 1997). Tıpkı RF tahmin modelinin sonuçları gibi, MLP algoritmalarıyla en doğru ÇO değeri tahmini 5 ile 9 mg/L arasındadır.

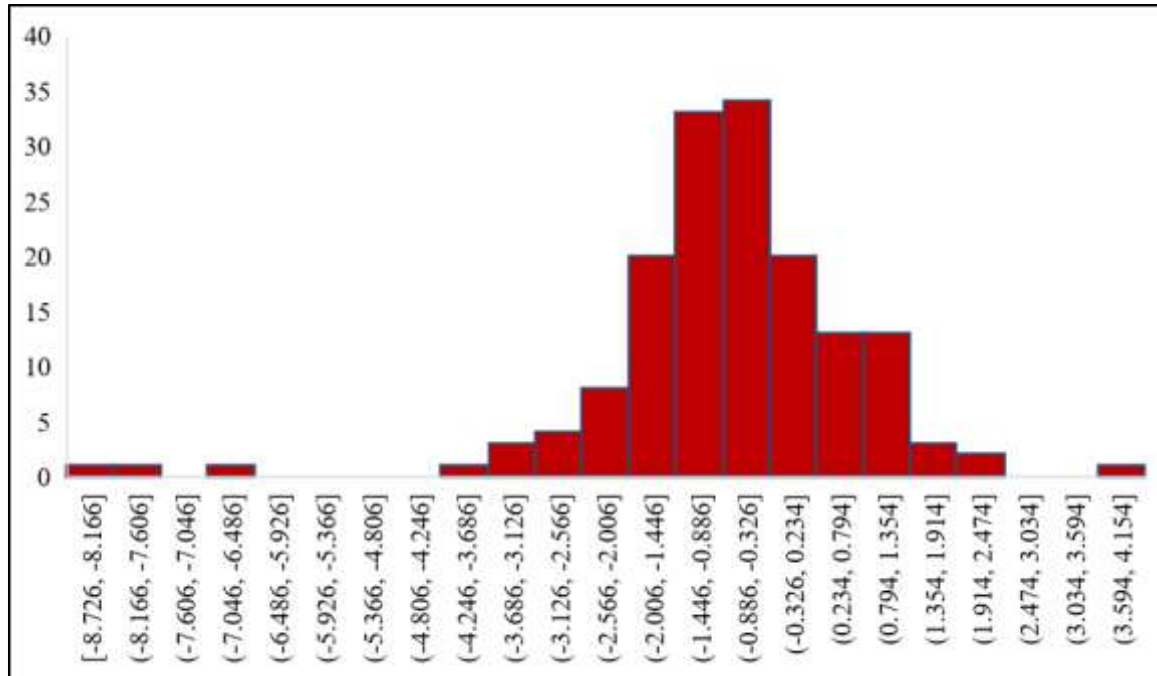


Şekil 4.48. Tüm veri seti ile MLP tahmin modelinin gerçek ÇO değerine karşı tahmin edilen ÇO konsantrasyonu.

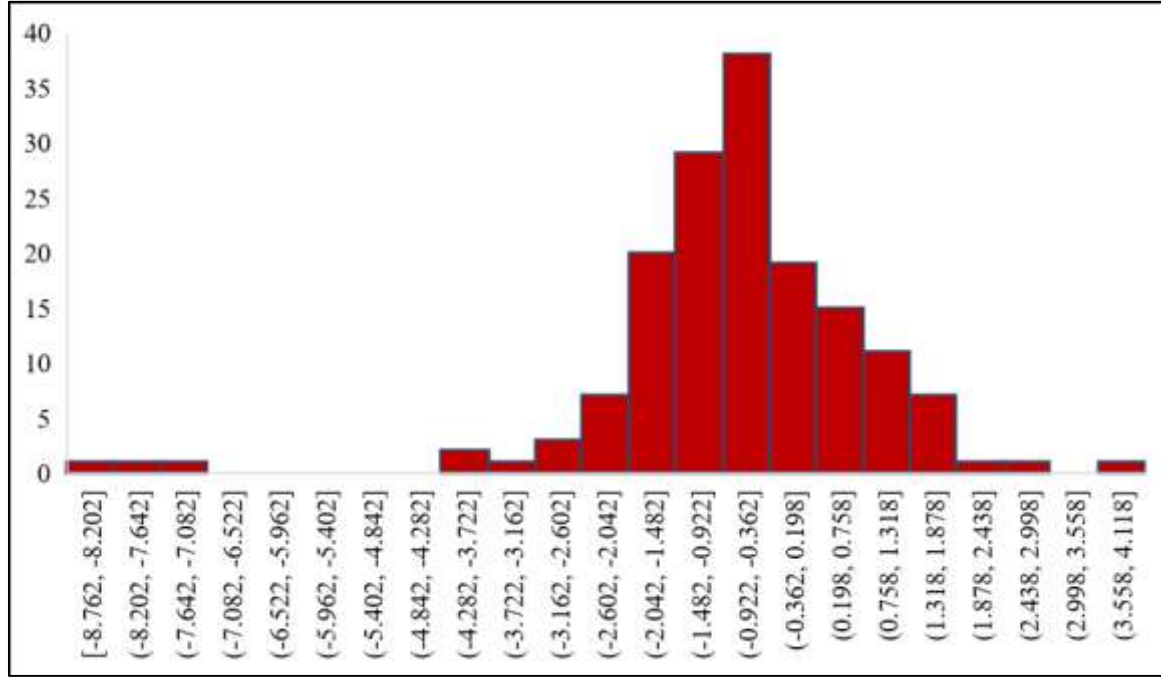


Şekil 4.49. Boyutsallığı düşürülmüş alt küme ile MLP tahmin modelinin gerçek ÇO değerine karşı tahmin edilen ÇO konsantrasyonu.

MLP Modelinin EED'si incelendiğinde, tüm veri seti ile (Şekil 4.50) tahmin edilen ÇO değerlerinin %71,51'i kabul edilebilir (-1,44 ile 1,35) EED aralığında olduğu, oysa algoritma tarafından tahmin edilen ÇO değerlerinin %77,21'i, boyutsallığı düşürülmüş alt kümesi ile bu aralıkta olduğu tespit edilmiştir (Şekil 4.51). Her ne kadarda veri seti boyutsallığı düşürülmesi MLP modeli üzerinde MAE ve RMSE metriklerini göz önünde bulundurarak negatif etki yaratsada, EED açısından incelendiğinde boyutsallık azaltma tahmin hatasını azaltma açısından pozitif etkisi görülmektedir. Her iki tahmin modelinin tahmin doğruluğu üzerinde gürültü engellemenin yanı sıra boyutsallık azaltmanın dikkate değer etkisini kanıtlamaktadır.



Şekil 4.50. Tüm veri seti ile eğitilen MLP modelinin test aşamasındaki EED'si.



Şekil 4.51. Boyutsallığı düşürülmüş alt küme ile eğitilen MLP modelinin test aşamasındaki EED'sı.

5. TARTIŞMA

5.1. Büyük Menderes Havzası Su Kalitesi Durumu Tartışması

Bir havzanın su kalitesi durumunu değerlendirmek için ilk olarak incelenmesi gereken parametre ÇO olmalıdır (Makubura ve diğerleri, 2022). Çözünmüş Oksijene bakıldığında bir su kütlesinin sağlığı hakkında bilgi elde edinebilir. İçsullarda yaşayan canlılar 4-15 mg/L arasında ÇO'ya ihtiyaç duymaktadırlar (Nacar ve diğerleri, 2020). ÇO seviyesi bir su kütlesinde düşerse o su sisteminde aşırı oksijen tüketimini gösterir. Bu aşırı oksijen tüketimi yada başka bir deyişle ÇO düşüşü bir çok faktöre bağlıdır. Bunlardan birisi suda besin maddenin çoğalmasıyla birlikte aerobik mikrobiyal aktivitenin çoğalması ve ardından ÇO konsantrasyonunun düşüşüdür. ÇO dalgalanmalarını açıklamak için iki faktörü göz önünde bulundurmak gerekebilir. Bunlar sıcaklık ve besin madde zenginleşmesidir. Sıcaklık havada bulunan oksijenin suda çözülmesini etkilemektedir. Soğuk su sıcak suya göre daha çok ÇO'yu kendinde tutar (Karpuzcu, 2007). Büyük Menderes Havzası'nda Şubat ayında sıcaklığın daha düşük olduğu ve bundan dolayı suda ÇO miktarının yükselmesinin beklenebilir olduğu söylenebilir. Nisan ayından itibaren sıcaklık arttığı için başka faktörler ile birlikte (Örneğin besin madde faktörü) ÇO seviyesinin düşüşü ön görülmektedir . Havzanın su kalitesi durumunu gösteren verilerin değerlendirilmesi sonucu elde edilen bulgular bu durumu desteklemektedir. Büyük Menderes Havzası'nın ÇO ve sıcaklık durumu incelendiğinde ÇO derişiminin sıcaklık ile değişimi görülmektedir. Ağustos ayında düşük ÇO miktarının en düşük derişimlerde olmasının sıcaklığın yüksek olmasından (Nacar ve diğerleri, 2020) kaynaklandığı düşünülmektedir. Ayrıca Büyük Menderes Havzası'nın ÇO durumu göz önünde bulundurulduğunda, ÇO derişimi dağılımı genelde 2 ile 8 mg/L aralığında olduğu görülmektedir. Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde içme suyu olarak kullanılabilecek yüksek kaliteli suların ÇO seviyesi en az 6 mg/L ve en uygun ÇO seviyesi ise 8 mg/L'den fazla olması gerekliliği göz önünde bulundurulduğunda, Büyük Menderes Havzası ÇO açısından kritik durumda olduğu görülmektedir. Bu olayı sıcaklık faktörü açısından tartışmak için Sütgibi (2015) Büyük Menderes havzası'nın sıcaklık eğilimlerini saptamaya çalışarak havza genelinde sıcaklık artışı olduğunu belirtmiştir. Sıcaklığın yanı sıra tarım ve hayvancılık faaliyetlerine bağlı olan ÇO derişimi dalgalanmalarının göz önünde bulundurulması gerekmektedir. Büyük Menderes Havzası'nın baskın sanayileri tarım ve

hayvancılık olduğu için (Demirci, 2019) Şubat ayından tarımcılık faaliyetlerinin artması nın göz önünde bulundurulmasıyla birlikte Nisan ayından Ağustos ayına kadar bu havzada ÇO miktarının düşüşü görülmektedir. Tarım-Orman Bakanlığına göre Mart ayından ekim işlemleri başlar. Nisan ayından itibaren meyve fidanlarının dikimi yapılır. Bu ayda arpa, Referans ekle tarım orman bakanlığına atıf yaptın buğday gibi hububatın yazlık ekimleri tamamlanır. Kışlık hububatta Nitrat'lı gübreleme yapılır. Mayıs ayında artık tarımcılık başlamış olur ve Azot'lu gübreyle gübreleme yapılır. Tarımın yanı sıra hayvancılıkta başlamış olur ve artık hayvanlar meralarda otlamaya başlar. Haziran ayında hala Azot'lu gübre ile gübreleme devam eder. Hayvanlar ise merada otlamaya devam eder. Mayıs ayının ortalarından alabalık dışındaki su ürünlerini Ağustos'un sonuna doğru sürecekle avlanma yasağı getirilir. Ağustos ayında kışlık sebzelerin ekimi ve fidelerinin dikimi devam eder ve mera hayvancılığına devam edilir. Eylül ayı itibariyle tarım faaliyetleri azalır ve meralar kurumaya başlar ve dolayısıyla hayvanların mera otlamasına yavaş yavaş son verilir. Bunların hepsini göz önünde bulundurduğumuzda Büyük Menderes Havzası'nda Nisan ayından Ağustos ayına kadar sudaki besin maddesi artışı ve dolayısıyla ÇO miktarının düşüşü açıklanır (Tarım ve Orman Bakanlığı, 2022). Bilim insanları tarafından Büyük Menderes Havzası su kalitesi ile ilgili araştırmalar bu havzada besin madde kirlenmesi söz konusu olduğunu onaylamaktadır. Dügel ve Kazancı (2004) araştırmalarında Büyük Menderes Havzası'nın besin madde açısından kirli bir havza olduğunu belirtmişler. Boyacıoğlu ve diğerleri (2005) tarımsal kirleticiler Büyük Menderes Havzası'nın en önemli kirleticiler olduğunu belirtmişler. Boyacıoğlu ve Boyacıoğlu (2017) Büyük Menderes Havzası'nın organik kirlenmeye maruz kaldığını ortaya konulmuşlar.

Bir su kütlesinde besin madde zenginleşmesi durumunda aerobik biyolojik aktivite yükselir ve dolayısıyla sucul canlılar çoğalmak için oksijen tükettiklerinde ÇO miktarı azalır ve biyolojik oksijen ihtiyacı yükselir (Bütünoğlu, 2018). Bunun yanında mikrobiyal faaliyet yükseldiği için mikrobiyal kimyasal reaksiyonlarda çoğalır ve dolayısıyla KOİ'de dolaylı bir şekilde yükselir. Besin zenginleşmesi ise tarım ve hayvancılık faaliyetlerinin yaygın olduğu aylarda gübreleme faaliyetleri ve hayvan dışkılarının suya bırakılmasından kaynaklıdır (Yılmaz, 2013). Büyük Menderes Havzası'nın Amonyum durumuna bakıldığında Şubat ayında gübreleme işlemlerinin başladığını belirtebiliriz. Ayrıca Koç (2010) Büyük Menderes Havzası su kalitesi ile ilgili çalışması sonucunda bu havzanın genelinde su kalitesi BOİ, KOİ, amonyak, Nitrit ve Nitrat açısından III. ve IV. su kalitesi sınıfında yer aldığını belirtmektedir. Fosfat ve Sülfat gübrelerde bulunan iki önemli

bileşimdir (Sönmez ve diğerleri, 2008). Bitkilerin büyüme için Azot'tan ziyade Fosfor'a ihtiyaçları vardır (Soylu ve diğerleri, 2016). Ayrıca Sülfat, gübrelerde bir karşı anyon olabilir veya yavaş salınan bir kaplama olarak kullanılır. Gübreleme işlemleri sırasında Amonyum, Nitrit ve Nitrat ile birlikte Fosfat ve Sülfat'da suya bırakılır (Sönmez ve diğerleri, 2008). Fosfat ve Sülfat birlikte suda besin zenginleşmesine neden olurlar (Bütünoğlu, 2018). Büyük Menderes Havzası'nda Sülfat miktarı neredeyse her zaman yüksektir. Buradan da anlaşılacağı gibi tarımcılık ve gübreleme faaliyetleri dışında farklı bir faktör Sülfat miktarını etkilemektedir. Sülfat birçok mineralde doğal olarak bulunur ve ticari olarak, özellikle kimya endüstrisinde kullanılır (Güven, 2004). Endüstriyel atıklarda ve atmosferik çökeltme yoluyla suya atılır. Ancak, en yüksek seviyeler genellikle doğal kaynaklardan gelen yeraltı sularında meydana gelir (Aras ve Fındık, 2018). Büyük Menderes Havzası'nda yüksek Sülfat miktarının yukarıda belirtilen kaynaklar ve tarımcılık faaliyetlerinden oluştuğu değerlendirilmektedir. Somay ve Gemici (2012) araştırmalarında Büyük Menderes Havzası iyon kompozisyonunda Sülfat miktarının çok olduğunu ve doğal kaynaklardan ortaya çıktığını belirtmişler. Büyük Menderes Havzası'nın su kalitesi durumu incelenmesi sonucu Nisan ayından Ağustos ayına kadar en yüksek Sülfat miktarı gözlenmektedir. Buda tarım faaliyetlerinin bu aylarda baskın olduğu ve başka kaynaklar ile birleştiğinde bu aylarda yüksek Sülfat miktarını açıklamaktadır. TÇK, inorganik tuzları (esas olarak Kalsiyum, Magnezyum, Potasyum, Sodyum, bikarbonatlar, Klorürler ve Sülfat'lar) ve suda çözünen az miktarda organik madde içerir. İçme suyundaki TÇK, doğal kaynaklardan, kanalizasyondan, kentsel akış ve endüstriyel atık sulardan gelir (Ünlü ve diğerleri, 2008). TÇK ve dolayısıyla inorganik tuzlar farklı kaynaklardan su sistemine döküldüğü için Büyük Menderes Havzası'nın TÇK ve diğer inorganik tuzlar miktarının dalgalanmalarını açıklamak pekde mümkün görülmemektedir. Ayrıca bu kaynaklar genelde noktasal olmayan kaynaklar olduğu için her hangi bir ÇED raporundan bu maddelerin Büyük Menderes Havzası için kirlilik yükünü bulmakta mümkün görülmemektedir.

5.2. Sınıflandırma Sonuçlarının Tartışması

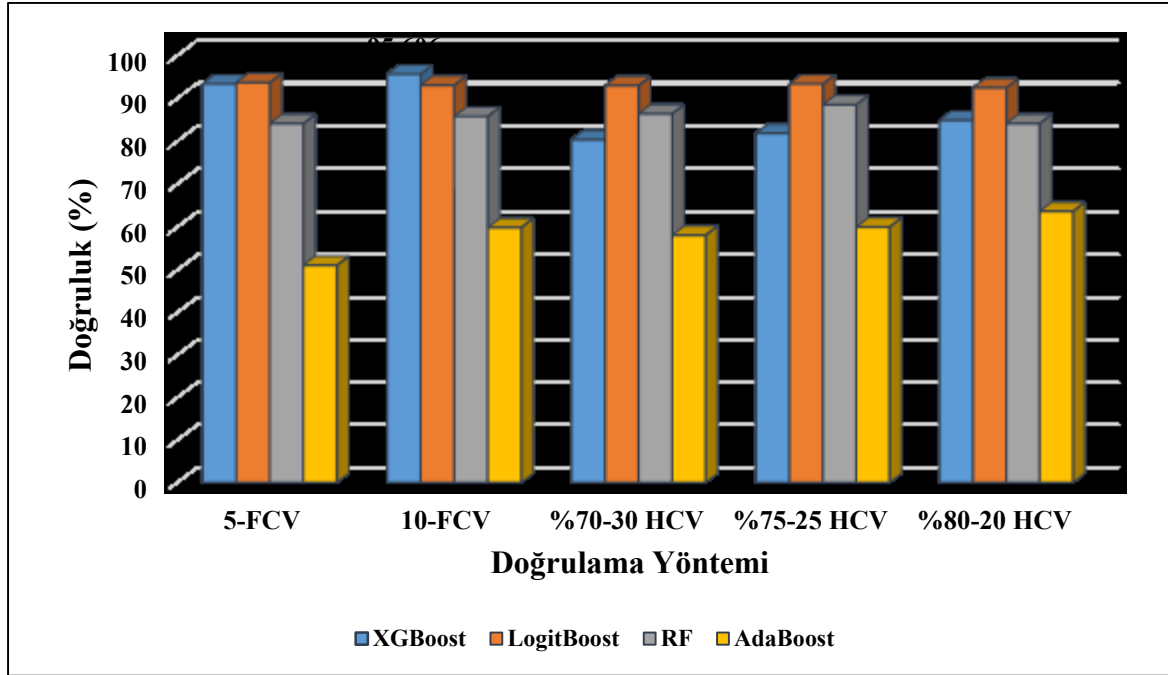
Büyük Menderes Havzası su kalitesi sınıflandırma çalışması sonuçlarına bakıldığında uygulanan AdaBoost ve RF modellerinin uygulanan altküme ve doğrulama metodu farketmeksizin performans açısından etkisiz modeller olduğu tespit edilmiştir (Şekil 5.1-Şekil 5.4).

XGBoost modeli su kalitesi sınıflandırması için en verimli model olarak tespit edilmiştir. Fakat XGBoost modeli doğrulama yönteminden etkilenmekte olup performansı özellikle HCV yöntemi uygulanmasıyla düşmektedir. Sonuç olarak, bu model K-FCV Yöntemi'ne daha duyarlı olup XGBoost modelinin performansını pozitif yönde etkilemektedir.

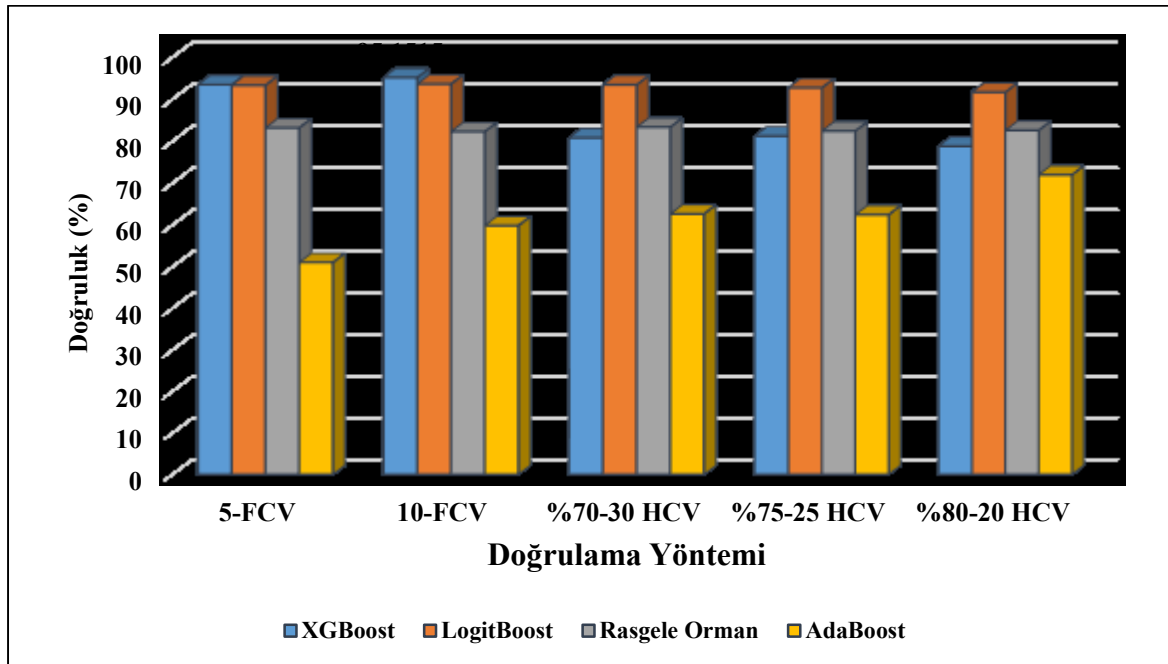
LogitBoost modeli performansına bakıldığında , bu modelin doğrulama yönteminden en az etkilendiği görülmektedir. LogitBoost modelinin doğrulama yönteminden hiç etkilenmediği anlamına gelmemekte olup doğrulama yöntemi değiştiğinde modelin performansında XGBoost modeli gibi ciddi iniş ve çıkışlar görülmemektedir.

RF modeli de aynı LogitBoost modeli gibi doğrulama yöntemine az duyarlıdır. Fakat AdaBoost modeli performansı doğrulama yöntemine duyarlıdır.

Tüm modellerin performansı veri seti boyutundan etkilenmektedir. Veri kümesinde ne kadar çok parametre varsa daha iyi eğitilmiş olmaktadır. Veri seti boyutu etkisi performans cinsinden modellerin doğruluğu, kesinliği ve hassasiyetinin artması anlamını taşımaktadır. Küme boyutunun etkisi XGBoost performansı artışında diğer modellerde olduğu gibi çok olmamaktadır. Bu sonuçlar doğrultusunda, XGBoost modelinin çok verimli bir model olduğu kanıtlanmış olacaktır. XGBoost modelinin en iyi performansı, MI-FLTM öznitelik seçimi yönteminin önerdiği 9 parametre ile eğitildiğinde değil, BFE-WRAPM öznitelik seçimi yönteminin önerdiği 7 parametrelilik alt küme ile eğitildiğinde elde edildiği tespit edilmiştir (Şekil 5.2). XGBoost modeli daha az parametre ile daha iyi eğitildiği dolayısıyla performansının daha iyi olduğu tespit edilmiştir.



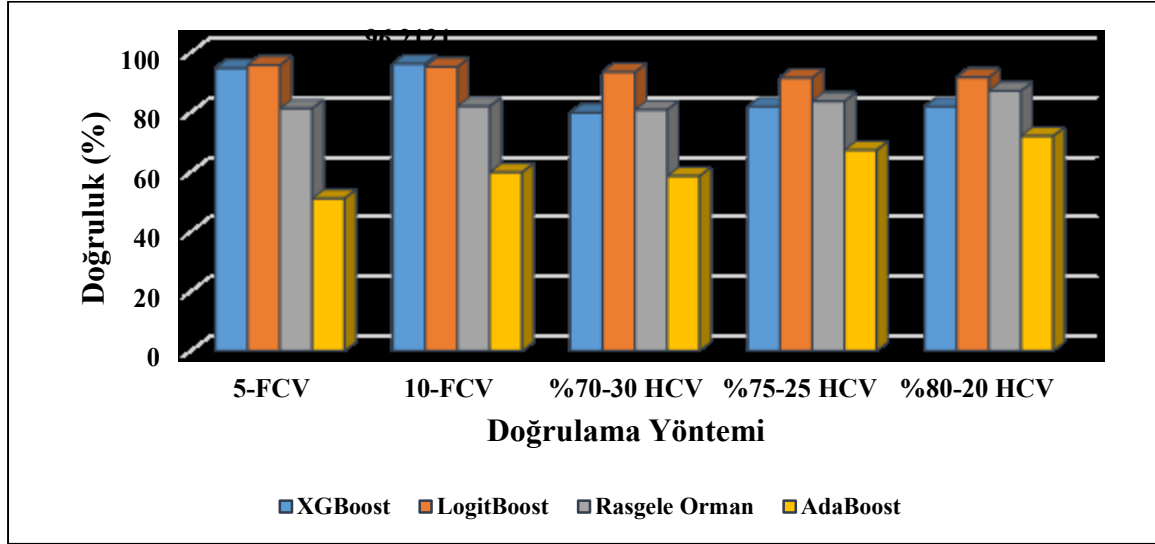
Şekil 5.1. Farklı doğrulama yöntemleri ve BFE-WRAPM öznitelik seçiminin önerdiği altküme ile modellerin doğruluk kıyaslaması



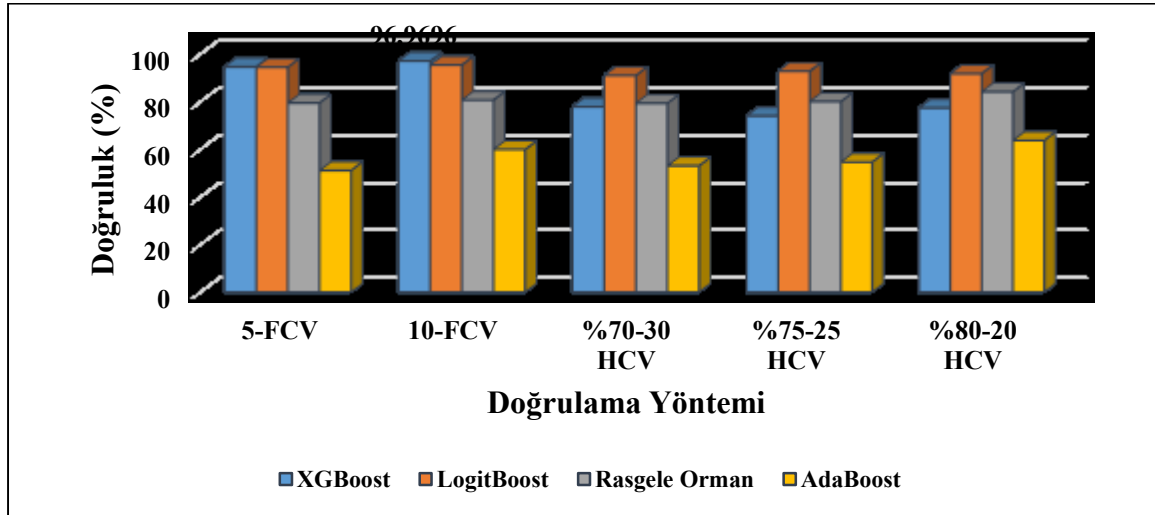
Şekil 5.2. Farklı doğrulama yöntemleri ve MI-FLTM öznitelik seçiminin önerdiği altküme ile modellerin doğruluk kıyaslaması

XGBoost modelinin en iyi performansı RF-EMBM öznitelik seçimi yönteminin önerdiği 14 parametreden oluşan alt küme, 19 parametreden oluşan tüm veri seti ve 10-FCV ile görülmektedir (Şekil 5.3 ve Şekil 5.4). XGBoost modelinin en iyi performanslarına yakın

olarak LogitBoost modeli de en iyi performansını RF-EMBM öznitelik seçimi yönteminin önerdiği 14 parametreden oluşan alt küme, 19 parametreden oluşan tüm veri seti ve 10-FCV uygulanmasıyla göstermiştir. Buradan elde edilen sonuçlara göre, modellerin veri kümesi boyutundan ne kadar çok etkilendiği anlaşılmaktadır.



Şekil 5.3. Farklı doğrulama yöntemleri ve RF-EMBM öznitelik seçiminin önerdiği altküme ile modellerin doğruluk kıyaslaması



Şekil 5.4. Farklı doğrulama yöntemleri ve tüm veri seti ile modellerin doğruluk kıyaslaması

Çalışma kapsamında elde edilen bulgulara göre XGBoost modeli Büyük Menderes Havzası su kalitesi sınıflandırmasında kullanılan veri kümesi farketmeksizin ve 10-FCV sonucunda diğer modellere göre en üstün performansla en verimli model olduğu tespit edilmiştir. Bu

modeli sınıflandırma performansı açısından inceleyecek olursak, sınıflandırma çalışması sonucu elde edilen her senaryo için CM incelenmesi gerekmektedir.

Çizelge 5.1, BFE-WRAPM öznitelik seçici tarafından önerilen veri kümesi ve 10-FCV ile en iyi sonuçları elde eden XGBoost sınıflandırıcısının CM tablosudur. XGBoost sınıflandırıcısının, BFE-WRAPM öznitelik seçici yöntemi tarafından seçilen özelliklerin alt kümesi ile sınıflandırma sürecindeki performansı, sırasıyla %99,15 ve %100'lük maksimum sınıflandırma doğruluğuyla “Mükemmel” ve “Tüketim İçin Uygun Değil” sınıf etiketlerinin kullanılabileceği anlaşılmaktadır. Ancak “Çok Kötü” sınıf etiketinin sınıflandırma doğruluğu söz konusu olduğunda, XGBoost sınıflandırıcısının performansı %72,88 doğrulukla önemli ölçüde düşmüştür. Modelin “İyi” sınıf etiketi sınıflandırmasındaki performansı doğruluğu %97,84'dır. Son olarak, XGBoost sınıflandırıcısının BFE-WRAPM öznitelik seçici tarafından önerilen özniteliklerin veri seti ile genel sınıflandırma doğruluğu 660 örnekten 631'ini doğru bir şekilde sınıflandırdığı için %95,606 olarak hesaplanmıştır.

Çizelge 5.1. BFE-WRAPM öznitelik seçimi yönteminin önerdiği alt küme ve 10-FCV uygulanmasıyla XGBoost modelinin CM tablosu

		Gerçek				
		Mükemmel Su kalitesi	İyi Su Kalitesi	Kötü Su Kalitesi	Çok Kötü Su Kalitesi	Tüketim için Uygun Değil
Tahmin Edilen	Mükemmel Su kalitesi	118	0	0	0	0
	İyi Su Kalitesi	1	136	9	0	0
	Kötü Su Kalitesi	0	3	86	6	0
	Çok Kötü Su Kalitesi	0	0	0	43	0
	Tüketim için Uygun Değil	0	0	0	10	248

“İyi” ve “Kötü” sınıf etiketlerinin sınıflandırılması performansı iki farklı altküme (MI-FLTM ve BFE-WRAPM) ile kıyaslandığında, BFE-WRAPM yönteminin önerdiği parametreler oluşan altküme ile eğitilen XGBoost modeli bu sınıf etiketlerinin sınıflandırılmasında daha iyi performans göstermiştir.

Bununla birlikte, “Mükemmel”, “Çok Kötü” ve “Tüketim İçin Uygun Değil” sınıf etiketleri için MI-FLTM yöntemiyle seçilen özelliklerin alt kümesiyle bu modelin sınıflandırma doğruluğu, BFE-WRAPM öznitelik yöntemi tarafından seçilen özelliklerin alt kümesi ile bu sınıflandırıcı tarafından elde edilen sonuçların aynı olduğu tespit edilmiştir.

Çizelge 5.2. MI-FLTM Öznitelik Seçimi Yöntemi’nin önerdiği alt küme ve 10-FCV uygulanmasıyla XGBoost modelinin CM tablosu

		Gerçek				
		Mükemmel Su kalitesi	İyi Su Kalitesi	Kötü Su Kalitesi	Çok Kötü Su Kalitesi	Tüketim için Uygun Değil
Tahmin Edilen	Mükemmel Su kalitesi	118	1	0	0	0
	İyi Su Kalitesi	1	134	10	0	0
	Kötü Su Kalitesi	0	4	85	3	0
	Çok Kötü Su Kalitesi	0	0	0	43	0
	Tüketim için Uygun Değil	0	0	0	13	248

“Kötü” ve “Çok Kötü” sınıf etiketlerinin en iyi sınıflandırması, XGBoost modeli tarafından RF-EMBM öznitelik seçim yöntemi tarafından seçilen özelliklerin alt kümesi ile eğitildiğinde elde edilmiştir (Çizelge 5.3). Burada modelim sınıf etiketinin sınıflandırılma doğruluğu sırasıyla %92.63 ve %76.27’ olarak bulunmuştur. XGBoost modelinin daha önce tartışılan diğer özellik alt kümeler ile eğitilen model tarafından yapılan tüm sınıflandırmalara göre daha iyi sonuç verdiği görülmektedir.

Çizelge 5.3. RF-EMBM Öz nitelik Seçimi Yöntemi'nin önerdiği alt küme ve 10-FCV uygulanmasıyla XGBoost modelinin CM tablosu

		Gerçek				
		Mükemmel Su kalitesi	İyi Su Kalitesi	Kötü Su Kalitesi	Çok Kötü Su Kalitesi	Tüketim için Uygun Değil
Tahmin Edilen	Mükemmel Su kalitesi	118	0	0	0	0
	İyi Su Kalitesi	1	136	7	0	0
	Kötü Su Kalitesi	0	3	88	3	0
	Çok Kötü Su Kalitesi	0	0	0	45	0
	Tüketim için Uygun Değil	0	0	0	11	248

Çizelge 5.4'e göre, XGBoost modeli tarafından “İyi”, “Kötü” ve “Çok Kötü” sınıf etiketlerinin sınıflandırılmasında, tüm özellikler sınıflandırma sürecine dahil edildiğinde en yüksek performans elde edilmiştir. Bu üç sınıf, aynı sınıflandırıcının sınıflandırma işlemine boyutsallığı düşürülmüş alt kümeler dahil edildiğinde en yüksek sınıflandırma hatasına sebep vermektedir.

Daha yüksek su kalitesi sınıflandırma doğruluğu için, parametre sayısını azaltmaktansa daha fazla parametreyi dahil etmenin daha iyi olabileceği görülmektedir. Bunun nedeni, ML algoritmalarının, sınırlı sayıda parametre yerine çeşitli parametrelerden öğrendiklerinde çok sınıflı su kalitesini tahmini doğrultusunda daha iyi performans göstermesidir. Ancak, iki nedenden dolayı daha fazla parametreyi dahil etmek de mantıklı değildir. Bu nedenlerden birincisi, bir ML algoritması tarafından yapılan ekstra hesaplamaları azaltmak ve ardından maliyet etkin hale getirmek için, daha düşük boyutta alt küme kullanmak gereklidir. İkinci olarak, bir havzanın su kalitesinin izlenmesi ve değerlendirilmesini ekonomik hale getirmek için, bazı ölçümlerin pahalı olduğu kadar zaman alıcı da olması, araştırmacının sınırlı sayıda etkin parametreleri seçmesi gerektiğini ortaya koymaktadır.

Çizelge 5.4. Tüm veri seti ve 10-FCV uygulanmasıyla XGBoost modelinin CM tablosu

		Gerçek				
		Mükemmel Su kalitesi	İyi Su Kalitesi	Kötü Su Kalitesi	Çok Kötü Su Kalitesi	Tüketim için Uygun Değil
Tahmin Edilen	Mükemmel Su kalitesi	118	0	0	0	0
	İyi Su Kalitesi	1	136	4	0	0
	Kötü Su Kalitesi	0	3	91	0	0
	Çok Kötü Su Kalitesi	0	0	0	47	0
	Tüketim için Uygun Değil	0	0	0	12	248

Çalışmanın başında seçilen tüm ilgili öznitelikler dahil edildiğinde en iyi performans XGBoost sınıflandırıcı tarafından elde edilmiş olsa da, XGBoost sınıflandırıcıların performansının diğer alt kümelerle birlikte değerlendirilmesi daha doğru olacaktır. Bu sınıflandırıcı tarafından elde edilen en iyi ikinci performans, RF-EMBM öznitelik seçici tarafından seçilen 14 parametre ile yapılmasına rağmen, elenen parametrelerin (örn. Yıl, Ay, İstasyon, Sıcaklık, pH) ekstra bir maliyeti yoktur. Bu nedenle, ilgili tüm parametreleri dahil etmek ve performanstan ödün vermemek anlamlı olacaktır. Bununla birlikte, üçüncü en iyi performansa sahip olan BFE-WRAPM özelliği seçici tarafından seçilen 7 parametreyi içeren XGBoost sınıflandırıcısının performansı incelendiğinde, bu özellik kombinasyonunun seçilmesi ve performansın küçük bir yüzdesinin göz ardı edilmesi önerilebilir. Bu durumun, model daha fazla örnek sayısından oluşan büyük veri kümeleriyle eğitildiğinde telafi edilebileceği değerlendirilmektedir.

Çalışma kapsamında, sınıflandırma çalışması sonuçlarını literatürde bulunan XGBoost bir su kalitesi sınıflandırması için kullanılan bir model olarak kıyasladığımızda, çalışmanın doğruluğunu ve bulgularını doğrulamaktadır. Yusri ve diğerleri (2022) XGBoost ve SVM modellerinin performansını su kalitesi sınıflandırmasında değerlendirmişlerdir. Çalışmanın

genel sonucunda 7 parametreden oluşan veri kümesi ile XGBoost modeli su kalitesi sınıflandırması için %94,00 doğrulukla en üstün model olarak bulunmuştur. 5-FCV uygulanması sonucunda XGBoost modeli için elde edilen %94,00 doğruluk, bu çalışmada elde ettiğimiz bulgular ile örtüşmektedir. Bu çalışmadaki BFE-WRAPM öznitelik seçimi yönteminin önerdiği 7 parametreden oluşan alt küme kullanılması ve 5-FCV yöntemi uygulanmasıyla XGBoost modeli için %93,3333 sınıflandırma doğruluğu Yusri ve diğerleri (2022)'nin çalışması ile benzerlik taşımaktadır.

Xin ve Mou (2022) su kalitesi sınıflandırması için 12 ML modeli performansını değerlendirmişlerdir. Çalışmanın sonucunda XGBoost modeli en üstün üç model arasında yer almıştır. Performans açısından XGBoost modelini geçen diğer iki modelde yine XGBoost mantığına dayalı çalışan modellerdir. Bu sonuçların çalışmanın bulgularıyla benzerlik gösterdiği görülmektedir.

Nasir ve diğerleri (2022) su kalitesi sınıflandırma modeli oluşturmak için 7 ML modelinin performansını değerlendirmişlerdir. Yazarlar çalışmada GBoost tekniğiyle çalışan iki model kullanmışlardır. Bunlardan birisinin bu çalışmada kullanıldığı gibi XGBoost modeli, diğerinin ise CATBoost modeli olduğunu belirtmektedirler. Nasir ve diğerleri (2022) nin çalışmalarında kullanılan veri seti 7 parametreden oluşmaktadır. Bunun yanı sıra 10-FCV yöntemi modellerin performansını değerlendirmek için kullanılmıştır. Bu çalışmada CATBoost modeli %94,51 doğrulukla en verimli model bulunmuştur. XGBoost modelinin doğruluğu %88,10 olarak bulunmuştur.

Literatürde su kalitesi sınıflandırması için XGBoost modelinin performansı değerlendirilen çalışma bulunmamaktadır. XGBoost modelinin sınıflandırma performansı bir çok farklı disiplininde değerlendirilmiştir . Ancak bu çalışmalar bu tez çalışmasının ana hatlarıyla aynı çizgide olmadığı için bu tez çalışmasında değerlendirilme imkanı bulunamamıştır.

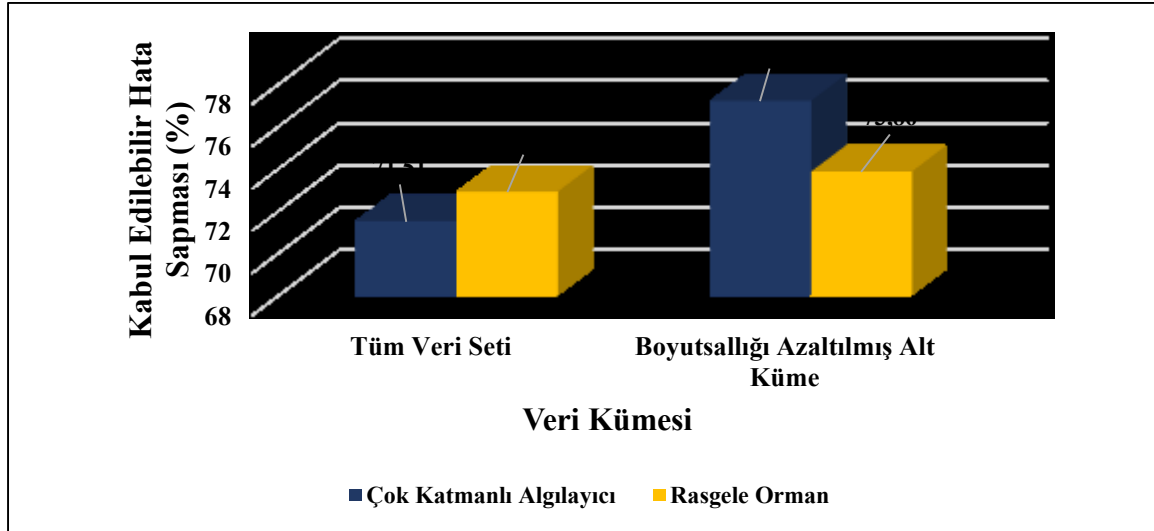
5.3. Tahmin Çalışması Sonuçlarının Tartışılması

ANNs modellerinin ve EL teknikleriyle çalışan performansı yüksek ağaç tabanlı algoritma olan RF dahil olmak üzere ağaç tabanlı modellerin performansı, sınıflandırma ve tahmin modelleri problemlerini çözmek için verimli modeller önermek üzere birçok araştırma incelenmiştir.

Çalışmanın sonuçları, RF modelinin su kalitesi parametrelerinin azaltılmış alt kümesini kullanarak ÇO konsantrasyonlarını tahmin etmede güçlü bir tahmin modeli olduğunu göstermiştir. Elde edilen kabul edilebilir sonuçlara rağmen, MLP algoritması RF modeline karşı daha zayıf performans göstermiştir. Ayrıca, öznetelik seçiminin tahmin modelleri performansı üzerindeki etkisi de bir gerçektir.

Veri kümesi boyutsallığının azaltılmasının etkisini modellerin performansı üzerinde incelendiğinde, MLP modelinin EED'si daha fazla etkilenmektedir. Boyutsallık azaltılmasıyla birlikte MLP modelinin EED'si RF modeline göre daha düşük olduğu tespit edilmiştir.

Şekil 5.5, tahmin modellerinin farklı kümeler kullanılmasıyla elde edilen tahmin örneklerin yüzde kaçının kabul edilebilir EED aralığında olduğunu göstermektedir. Şekil 5.5, veri kümesi boyutunun etkisini EED'sinin azaltılması üzerinde göstermek için kullanılmıştır. Boyutsallık azaltması genel olarak RF modelinin performansını yükselttiği ve MLP modelinin tahmin performansı üzerinde negatif etkisi olduğu belirtilmiştir. Şekil 5.5 incelendiğinde boyutsallık azalmasının her ne kadar da MLP modelinin genel performansı üzerinde negatif etkisi olsada, modelin tahmin EED'sini oldukça düşürmüş olduğu tespit edilmiştir. Bu açıdan bakıldığında tüm veri setinin kullanılmasıyla RF modelinin tahmin ettiği örneklerin %72,91'ni kabul edilebilir EED aralığında olduğu, boyutsallığı azaltılmış olan alt küme kullanılmasıyla bu modelin tahmin ettiği örneklerin %73,86'sının aynı aralıkta olduğu görülmektedir. RF modelinin EED azaltma performansının boyutsallığı azaltılmış olan alt küme kullanılmasıyla %0,95 yükselmiş olduğu görülmektedir. MLP modelinin EED açısından performansını değerlendirildiğinde, tüm veri setinin kullanılmasıyla tahmin ettiği örneklerin %71,51'i kabul edilebilir EED aralığında olduğu, boyutsallığı azaltılmış olan alt küme kullanıldığında modelin tahmin ettiği örneklerin %77,21'nin aynı aralıkta olduğu tespit edilmiştir. EED performansının , MLP modelinin boyutsallığı azaltılmış olan bir alt küme kullanılarak %5,7 oranında artırıldığı tespit edilmiştir. Veri setinin boyutsallığının azaltılması etkisinin, MLP modelinin EED azaltması performansı üzerinde daha fazla olduğu görülmektedir. Buradan her iki model için boyutsallık ve veri setinin gürültüsünün azaltılması, dolayısıyla tahmin performansının arttığı sonucuna varılabilir. Ancak gürültü azalması etkisinin MLP modelin tahmin hatası performansı üzerinde daha fazla olduğu görülmektedir.



Şekil 5.1. Şekil 5.5. Tahmin edilen örneklerin kabul edilebilir EED aralığı

Literatürde, RF ve türevlerinin regresyon problemlerinin çözümünde birçok ML modeline göre üstünlüğü görülmektedir. Örneğin, Singh ve diğerleri (2013), hava kalitesi parametrelerini tahmin etmede RF algoritmasının basit bir karar ağacı ve SVM modeline göre üstünlüğünü bildirmişlerdir. Wang ve diğerleri (2016), bir regresyon problemini çözmede RF modelinin ve orijinal RF kavramından türetilen bir yöntem olan Rastgele Bit Ormanı'nın (Random Bit Forest (RBF)) diğer yedi ML algoritmasıyla performansını değerlendirmişlerdir. Sonuç olarak RF ve RBF modellerinin diğerlerine göre üstünlüğü ortaya konulmuştur (Wang ve diğerleri, 2016). Maheshwari ve Lamba (2019) bir hava kalitesi regresyon problemini çözmede RF ve MLP modelleri dahil 6 regresörün performansını araştırmış ve bunun sonucunda, MLP modelinin olağanüstü performansına rağmen, RF regresörünün doğruluk açısından diğer dört ML regresöründen daha iyi performans gösterdiğini bulmuşlardır. Bu çalışmanın sonuçlarından MLP regresörünün bu tür regresyon problemlerini çözmek için de sağlam bir algoritma olduğu sonucuna varılabilir. Başka bir çalışmada, Ghorbani ve diğerleri (2016) nehir akışını tahmin etmede MLP modelinin SVR modeline göre tahmin performansının mükemmelliğini bildirmişlerdir. Heidari ve diğerleri (2016) MLP algoritmasının nano akışkanların viskozitesinin tahmininde regresyon kabiliyetini araştırmışlar ve MLP algoritmasını böyle bir regresyon problemini çözmede uygulayarak olağanüstü bir sonuç elde ettiğini bildirmişlerdir. Bu da MLP modelinin tahmin çalışmalarında olağanüstü yeteneğini ortaya koymaktadır. MLP modeli ele alındığında diğer regresörlerin zayıf olduğu anlamına gelmemektedir. Örneğin, Amid ve Gundoshmian (2017) kazanların çıkış enerjilerinin tahmininde MLP modeli dahil üç regresörün performansını araştırmışlar ve sonuç olarak MLP regresörü üç regresör grubu

arasında en zayıf performansa sahip olduğunu bulmuşlardır. Literatürde MLP modelinin regresyon problemlerini çözmek için etkili bir algoritma olmadığı sonucuna varan başka örnekler de bulunmaktadır (Ghritlahre ve Prasad, 2018). Bu nedenle, öncelikle bu algoritmanın popülaritesi ve tabii ki uygulama kolaylığı nedeniyle, regresyon performansının farklı senaryolar üzerinden daha fazla araştırılma yapılması ihtiyacını ortaya koymaktadır.

Ahmed ve Lin (2021) bir RF modeli olan Nicel Regresyon Ormanı'nın (Quantitative Regression Forest (QRF)) ÇO konsantrasyonunu tahmin etme performansının MLP ile karşılaştırıldığı bir çalışma yapmışlardır. QRF ÇO'yu daha az hatayla tahmin etmede daha iyi performans gösterdiğini belirtmişlerdir. Elde edilen sonuçlar, çalışmanın bulgularını doğrulamaktadır (Ahmed ve Lin, 2021).

RF regresörünün performansının MLP dahil diğer ML modellerine göre ağırlıklandırıldığı farklı çalışmalar da çalışmanın sonuçlarını doğrulamaktadır. Singh ve diğerleri (2022), buğdayın Azot durumunu tahmin etmek için farklı regresörlerin performansı karşılaştırılmışlardır. Sonuç olarak, RF modelinin performansının MLP modeline göre daha iyi sonuçlar verdiği belirtilmiştir.

Son zamanlarda, öznetelik seçimi, regresyon ve/veya sınıflandırma çalışmasını uygun maliyetli hale getirme gayretleri nedeniyle farklı araştırma alanlarında birçok araştırmacının dikkatini çekmiştir. Öznetelik seçiminde hedef yalnızca ekstra hesaplamaları azaltmak, veri kümesinin boyutsallığını düşürmek , aynı zamanda tahmin performansını iyileştirmek için ana veri kümesinden azaltılmış bir özellik alt kümesi oluşturmaktır. Her zaman parametre sayısı azaltılmış alt küme performansı artırmaz. Bazı ML tahmin modellerinin performansını olumsuz etkileyebilir. Masmoudi ve diğerleri (2020) veri kümesinin boyutunu azaltmak için öznetelik seçimi uygulandıklarında, regresörün tahmininde önemli bir gelişme olduğunu bildirmişlerdir. Castangia ve diğerleri (2021) özellik sayısı azaltılmış alt kümesinin, RF regresörü dahil beş ML tahmin modelleri performansı üzerindeki olumlu etkisini bildirmişlerdir. Bu çalışmanın sonuçları, ÇO konsantrasyonu tahmin etme doğrultusunda özellik sayısı azaltılmış alt kümesinin olumlu etkisini regresörlerin performansı üzerinde gösterilmiştir. Bununla birlikte, literatürde özelliklerin azaltılmış alt kümesinin MLP tahmin modelleri performansı üzerindeki etkisinin araştırıldığı çalışmalar bulunmaktadır. Hossain ve diğerleri (2013) özellik sayısı azaltılmış alt kümesi kullanılarak güneş enerjisinin tahmini

için MLP regresörünün doğruluğunda olumlu bir gelişme olduğunu bildirmişlerdir. Literatürde özelliklerin azaltılmış alt kümesinin MLP modelinin performansı üzerinde olumsuz bir etkiye sahip olduğunu doğrulayan çalışmalar da mevcuttur (Kapadia ve Jariwala, 2022).

Özellik seçiminin faydaları ve öznitelik sayısı azaltılmış alt kümesi göz önüne alındığında, önemli ölçüde azalan hesaplamalar için doğruluktan ödün vermenin doğru bir yaklaşım olmadığı değerlendirilmektedir.

Fakat, RF gibi güçlü bir regresör, özellik seçimi olsun veya olmasın çok iyi sonuçlar elde ettiğinde, ÇO konsantrasyonunun tahmini için etkisiz bir algoritmayı kullanmak doğru olmayacaktır. MLP tahmin modelinin sağlamlığı ve doğruluğu, farklı senaryolar ve farklı veri kümeleri aracılığıyla araştırılabileceği öngörülmektedir.

6. SONUÇ VE ÖNERİLER

Su yeryüzündeki yaşamın en önemli gereksinimlerinden birisidir. İnsanoğlu sağlık açısından, su kalitesi, hem su kaynaklı hastalık salgınları yoluyla hem de arka plandaki hastalık oranlarına katkıda bulunarak sağlık üzerinde büyük bir etkiye sahip olabilir. Bu nedenle insanlar ve hatta hayvanların tüketimi için yüzey su kaynaklarının kalitesini sürekli olarak izlenmesi gerekmektedir. Bu sadece insan sağlığı açısından çok büyük bir öneme sahip olmakla kalmaz, aynı zamanda Azot ve Fosfor gibi besinlerin ve/veya bunların türevleri bilerek veya bilmeyerek sucul sistemlere salınırsa, yüzey suyu rezervuarlarının ötrofikasyonu gibi istenmeyen sonuçlar meydana getirebilir. Ötrofikasyon ardından su sistemlerinde ÇO seviyelerini azalması su yaşamının yok olmasına neden olacaktır. Sonuç olarak, suda yaşayan organizmaların yetiştirilmesi için uygun ortamın engellenmesi ve kısıtlanması ortaya çıkacaktır.

Tez çalışması kapsamında, su kalitesini değerlendirmek, su kalitesi sınıfı belirlemek ve bir model oluşturmak üzere Türkiye'nin 25 su havzası arasından Büyük Menderes Havzası seçilmiştir. Büyük Menderes Havzası Türkiye'nin en kalabalık yedinci nehir havzasıdır. Bu havzanın en başta gelen su kullanıcıları tarım, mesken, ve endüstriyel tesislerdir. Büyük Menderes Havzası'nın en büyük arazi kullanıcısı tarım olduğu bilinmektedir.

Büyük Menderes Havzası'nın genel durumu incelenmiş olup 2004-2014 yılları arasında 10 istasyonda 248 ‘‘Tüketim İçin Uygun Olmayan’’ su kalitesi vaka sayısı ile kirli bir havza olduğu tespit edilmiştir. Bu havzada araştırılan yıllar arasında 2014 yılı 40 ‘‘Tüketim İçin Uygun Olmayan’’ su kalitesi vaka sayısı ile en kirli yılı olarak tespit edilmiştir. Araştırılan yıllar arası ve havza genelinde Ekim ve/veya Aralık ayları ‘‘Tüketim İçin Uygun Olmayan’’ su kalitesi vakası açısından en kirli aylar olarak tespit edilmiştir. Büyük Menderes Havzası'nın genel durumu su kalitesi parametreleri açısından incelendiğinde, bu araştırma için kullanılan su kalitesi parametreleri değerleri genelde TS-266 ve Orman ve Su İşleri Bakanlığı tarafından 2015 yılında yayınlanan Yüzeysel Su Kalitesi Yönetimi Yönetmeliği'nde Türkiye için belirlenen standart değerlerin üstünde olduğu tespit edilmiştir. Bu araştırmanın sonuçlarına göre 2004-2014 yılları arasında Büyük Menderes Havzası'nın su kalitesi açısından kirli bir havza olduğu sonucuna varılmıştır.

Büyük Menderes Havzası'nın önemli bir sanayi bölgesinde olması ve tarım faaliyetlerinin çok yapılması yüzey sularının kirlilik kaynağının yüksek olabileceği tahminini vermektedir. Tarım sanayisinden gelen en önemli kirlilik besin zenginleşmesi ve su sistemlerinin ötrofikasyona uğraması bilinen bir gerçektir. Bu nedenle bu tez çalışmasında değerlendirilen su kalitesi parametreleri besin zenginleşmesi kirliliğiyle ilgili parametreler göz önünde bulundurulmuştur.

Bu tez çalışmasında izlenilen hedef iki aşamadan oluşmaktadır. Bunlardan birisi Büyük Menderes Havzası ve Büyük Menderes Nehri'nin mevcut yıllara ait (2004-2014) su kalitesi verileri göz önünde bulundurularak WQI yöntemiyle su kalitesi sınıflarına etiketlenmesi ve daha sonra bu etiketlenilmiş veri kümesi kullanılarak ML'ye dayalı etkin ve verimli bir su kalitesi sınıflandırma modeli önerilmesidir. Çalışmanın bu hedefi doğrultusunda, bir sınıflandırma modelini verimli kılmak için kullanılan yöntemler arasında öznitelik seçimiyle veri kümesinin boyutsallığını düşürmektir. Veri kümesi boyutsallığı düşmesi bir modeli hem zaman-maliyet açısından verimli yapmakta hemde bazı durumlarda modelin karmaşıklığını azaltmakla birlikte modelin performansını arttıracaktır. Bu nedenle performansı değerlendirilen sınıflandırıcıların yanında yinede ML'ye dayalı üç ana öznitelik seçimi yöntemi ve önerdikleri boyutsallığı düşürülmüş veri kümelerinin etkileri değerlendirilmiştir.

Ayrıca çalışmanın bu aşamasında EL mantığına dayalı AdaBoost, LogitBoost, XGBoost ve RF adlı dört sınıflandırıcının performansı değerlendirilmiştir. Çalışmanın bu aşaması modellerin performansı tüm veri seti (19 parametrede oluşan) ve boyutsallığı azaltılmamış veri kümesi kullanarak farklı doğrulama yöntemleri (farklı K-FCV ve farklı oranlarda HCV doğrulama yöntemleri) kullanılarak değerlendirilmiştir. Bu aşamanın sonuçlarına göre AdaBoost ve RF modelleri doğrulama yöntemi farketmezin performansı çok düşük olan en etkisiz modeller olarak tespit edilmiştir. Fakat XGBoost ve LogitBoost modelleri oldukça kabul edilebilir ve değerlendirilebilir modeller olarak tespit edilmiştir. Hatta XGBoost modeli 10-FCV yöntemi uygulanmasıyla çok iyi performans göstermiştir. XGBoost modelinin sergilediği performans (Tüm veri seti ve 10-FCV ile) literatürde bulunan benzer çalışmalara göre çok iyi sonuçlar vermiştir. Ayrıca doğrulama yönteminin modellerin performansı üzerinde ne kadar önemli etkisi olduğu da tespit edilmiştir.

İkinci aşamada üç öznitelik seçimi yönteminin önerdikleri boyutsallık düşürülmüş veri kümeleri kullanılarak modellerin performansı üzerinde etkisi değerlendirilmiştir. Veri kümesinin boyutsallığının modellerin performansı üzerinde önemli etkisi olduğu bulgusuna ulaşılmıştır. Fakat çalışmanın bu aşamasının bulgularını daha önce uygulanan ve değerlendirilen modellerin performansı (tüm veri seti kullanılarak) ile kıyasladığında yinede AdaBoost ve RF Modelleri performans açısından en etkisiz ve verimsiz modeller bulunmasıyla birlikte LogitBoost modeli XGBoost modeline nazaran daha düşük performans gösterdiği görülmüştür. Çalışmanın sınıflandırma aşamasının bulguları değerlendirildiğinde, XGBoost modelinin su kalitesi sınıflandırmasında çok etkili bir model olduğu sonucuna varılmıştır. XGBoost modeli tüm veri kümelerinin etkisi (tüm veri seti yada boyutsallığı düşürülmüş alt kümeler) değerlendirilerek 10-FCV Yöntemiyle literatürde bulunan bulguların çok üstünde performans sergilediği tespit edilmiştir. Ancak boyutsallık düşürülmesi negatif etkisi hatta XGBoost gibi güçlü bir model üzerinde göz ardı edilemeyecek bir gerçek olduğu ortaya konulmuştur. Bu tez çalışmasının sınıflandırma çalışması aşaması bulgularına dayanarak XGBoost modelinin BFE-WRAPM öznitelik seçimi yönteminin önerdiği 7 parametreden oluşan ve 10-FCV yönteminin uygulanmasının doğru bir yaklaşım olacağı değerlendirilmektedir. Bu koşullarda elde edilen XGBoost modelinin performansı (Doğruluk, Kesinlik, ve Hassasiyet) aynı koşullarda fakat tüm veri seti kullanılarak elde edilen performans kıyasladığında göz ardı edilebilir bir düşüş izlenmiştir. Performansdaki bu düşüşün daha sonra modeli aynı yedi parametreden oluşan fakat veri sayısı daha çok veri kümesinin eğitilmesi ile telafi edilebileceği değerlendirilmektedir. Daha büyük ve sağlıklı bir veri seti ile XGBoost modelin performansının daha büyük veri kümesiyle (veri sayısı açısından) eğitilip sonuçları bu çalışmanın sonuçlarıyla kıyaslanması önerilmektedir.

Çalışmasının ikinci aşaması karar mercilerine ileri zamanlarda bir havzanın su kirliliği açısından fikir vererek su kalitesini iyileştirmek için hızlı tedbirler alınması açısından önemlidir.

Bir su sisteminde özellikle besin kirlenmesi ortaya çıktığında suda bulunan ÇO miktarı dolaylı bir şekilde azalmaya başlar. Bu da sucul canlılarının yaşamını kısıtlar hatta ilerideki aşamalarda sucul canlıların ve en başta balıkların ölümüne neden olabilir . Bu durum sucul ekosistemin dengesinin bozulması olarak sonuçlanabilir. Bu nedenle suda bulnan ÇO

miktarı sürekli izlenmeli ve hatta bu çalışmanın devamında ileri zamanlarda ÇO miktarı tahmini doğrultusunda bir model oluşturulmalıdır.

ANNs tabanlı MLP adlı ve EL tabanlı RF adlı regresyon modellerinin performansı ÇO tahmininde değerlendirilmiştir. Zaman-maliyet verimliliği hedefini takip ederek veri kümesinin boyutsallığı azalmasının modellerin performansı üzerinde etkisi araştırılmıştır. Bunun için literatürde denenen BFE-WRAPM öznitelik seçimi yöntemi mantıklı ve verimli veri kümesi önermesi ve başka yöntemlerin denenmesi ve yüksek performanslı sonuç elde edilememesi sebebiyle boyutsallığı azaltılmış 9 parametreden oluşan alt küme veri kümesi (BFE-WRAPM Yöntemi) kullanılmıştır.

Çalışmanın bulgularına göre RF modeli ÇO tahmini doğrultusunda hangi küme kullanılması farketmez MLP tahmin modeline göre daha üstün performans sergilemiştir. Ayrıca boyutsallık düşürme modellerin performansını etkilediği tespit edilmiştir. Genel performansda MLP modeli boyutsallık azalması ile negatif olarak etkilense de EED açısından boyutsallık azaldıktan sonra tahmin hatasının kabul edilebilir aralığa girdiği tespit edilmiştir. Genel performans değerlendirmesinde bu modelin performansının düşmesiyle veri kümesinde ne kadar fazla parametre varsa bu model o kadar doğru tahminler yürütebilir sonucuna varılmaktadır. Ancak yine de RF modeli ÇO tahmini doğrultusunda daha iyi performans gösterdiği için bu modelin ilerideki çalışmalarda güvenilir bir model olarak kullanılabileceği değerlendirilmektedir. Bu çalışmada, 9 parametreden oluşan BFE-WRAPM öznitelik seçimi yönteminin önerdiği veri kümesi zaman-maliyet hatta performans açısından RF modelinin tahmin performansını yükseltmesi sebebiyle benzer çalışmalarda göz önünde bulundurulabileceği değerlendirilmektedir.

Genel bir yaklaşımla sınıflandırma ve tahmin çalışmalarından elde edilen sonuçlar literatürde bulunan benzer çalışmaların sonuçlarının üstünde olup daha fazla araştırmayla uygulanabilir hale gelebileceği değerlendirilmektedir.

KAYNAKLAR

- 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.
- Ahmed, M. H., and Lin, L. S. (2021). Dissolved oxygen concentration predictions for running waters with different land use land cover using a quantile regression forest machine learning technique. *Journal of Hydrology*. 597, 126-213.
- Alexakis, D. E. (2022). Applying Factor Analysis and the CCME Water Quality Index for Assessing Groundwater Quality of an Aegean Island (Rhodes, Greece). *Geosciences*, 12(10), 384.
- Aldhyani, T. H., Al-Yaari, M., Alkahtani, H., Maashi, M. (2020). Water quality prediction using artificial intelligence algorithms. *Applied Bionics and Biomechanics*. 2020, 6659314.
- Akyildiz, G. K., Duran, M. (2021) Evaluation of the impact of heterogeneous environmental pollutants on benthic macroinvertebrates and water quality by long-term monitoring of the buyuk menderes river basin. *Environmental Monitoring and Assessment*, 193(2021), 1-23.
- Amid, S., and Mesri Gundoshmian, T. (2017). Prediction of output energies for broiler production using linear regression, ANN (MLP, RBF), and ANFIS models. *Environmental Progress & Sustainable Energy*. 36(2), 577-585.
- Asadollah, S. B. H. S., Sharafati, A., Motta, D., Yaseen, Z. M. (2021). River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *Journal of Environmental Chemical Engineering*. 9(1), 104599.
- Arabgol, R., Sartaj, M., Asghari, K. (2016). Predicting Nitrate Concentration and Its Spatial Distribution in Groundwater Resources Using Support Vector Machines (SVMs) Model. *Environmental Modeling and Assessment*. 21, 71–82.
- Aras, S. Fındık, Ö. (2018). Nevşehir ili için Kızılırmak Nehri'nin içme suyu potansiyelinin araştırılması. *Nevşehir Bilim ve Teknoloji Dergisi*, 7(2), 214-222.
- Avila, R., Horn, B., Moriarty, E., Hodson, R., Moltchanova, E. (2018). Evaluating statistical model performance in water quality prediction. *Journal of Environmental Management*. 206, 910-919.
- Barral, N., Maleki, M., Madani, N., Cánovas, M., Husillos, R., Castillo, E. (2022). Spatio-temporal geostatistical modelling of sulphate concentration in the area of the Reocín Mine (Spain) as an indicator of water quality. *Environmental Science and Pollution Research*. 29(57), 86077-86091.
- Barron, A. R. (1987). *Are Bayes rules consistent in information?*. In Open Problems in Communication and Computation. New York, NY., Springer, 85-91.

- Bauder, J. W., Bauder, T. A., Waskom, R. M., & Scherer, T. F. (2008). Assessing the suitability of water (Quality) for irrigation-Salinity and Sodium. *Western Fertilizer Handbook*.
- Benzer, S., Garabaghi, F. H., Benzer, R., Mehr, H. D. (2022). Investigation of some machine learning algorithms in fish age classification. *Fisheries Research*. 245, 106151.
- Berrar D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*. 1, 542-545.
- Bhati, B. S., Chugh, G., Al-Turjman, F., Bhati, N. S. (2021). An improved ensemble based intrusion detection technique using XGBoost. *Transactions on Emerging Telecommunications Technologies*. 32(6), e4076.
- Bilgin, A. (2018). Evaluation of surface water quality by using Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) method and discriminant analysis method: a case study Coruh River Basin. *Environmental Monitoring and Assessment*. 190(9), 1-11.
- Breiman, L. (2001). Random forests. *Machine Learning*. 45(1), 5-32.
- Bolón-Canedo, V., Sánchez-Marño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. *Knowledge and Information Systems*, 34, 483-519.
- Boyacioglu, H., Boyacioglu, H., Gunduz, O. (2005) Application of Factor Analysis in the Assessment of Surface Water Quality in Buyuk Menderes River Basin. *European Water Research Association*, 9(10): 43-49.
- Boyacioglu, H. (2006). Surface water quality assessment using factor analysis. *Water SA*, 32(3), 389-393.
- Boyacıoğlu, H., Boyacıoğlu, H. (2017). *Environmental Determinants of Surface Water Quality Based on Environmetric Methods*. 15th International Conference on Environmental Science and Technology. Rhodes, Greece.
- Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*. 44(1), 108-132.
- Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of the Total Environment*. 721, 137612.
- Bütünoğlu, A. (2018). Su kaynaklarında yüzer sulak alan ve sucul bitkiler ile nütrient gideriminin değerlendirilmesi. Uzmanlık Tezi. *TC Tarım ve Orman Bakanlığı Su Yönetimi Genel Müdürlüğü*, Ankara.
- Cai, Y. D., Feng, K. Y., Lu, W. C., Chou, K. C. (2006). Using LogitBoost classifier to predict protein structural classes. *Journal of Theoretical Biology*. 238(1), 172-176.

- Camejo, J., Pacheco, O., Guevara, M. (2013, January). *Classifier for drinking water quality in real time*. In 2013 International Conference on Computer Applications Technology (ICCAT). Sousse, Tunisia, 1-5.
- Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*. 8(3), 559-568.
- Castangia, M., Aliberti, A., Bottaccioli, L., Macii, E., Patti, E. (2021). A compound of feature selection techniques to improve solar radiation forecasting. *Expert Systems with Applications*. 178, 114979.
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*. 40(1), 16-28.
- Chen, H. (1995). Machine learning for information retrieval: Neural networks, symbolic learning, and genetic algorithms. *Journal of the American Society for Information Science*. 46(3), 194-216.
- Chen, C. W., Tsai, Y. H., Chang, F. R., Lin, W. C. (2020). Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. *Expert Systems*. 37(5), e12553.
- Csábrági, A., Molnár, S., Tanos, P., Kovács, J., Molnár, M., Szabó, I., Hatvani, I. G. (2019). Estimation of dissolved oxygen in riverine ecosystems: Comparison of differently optimized neural networks. *Ecological Engineering*. 138, 298-309.
- Çevre ve Şehircilik Bakanlığı (2016) Büyük Menderes Nehir Havzası Taslak Yönetim Planı SÇD Pilot Projesi; Stratejik Çevresel Değerlendirme Raporu. Ankara. 2016
- Dalai, C., Jha, R., Desai, V. R. (2015). Rice Husk and Sugarcane Bagasse Based Activated Carbon for Iron and Manganese Removal. *Aquatic Procedia*. 4, 1126 – 1133.
- Danades, A., Pratama, D., Anggraini, D., Anggriani, D. (2016, October). *Comparison of accuracy level K-nearest neighbor algorithm and support vector machine algorithm in classification water quality status*. In 2016 6th International Conference on System Engineering and Technology (ICSET). Bandung, Indonesia, 137-141.
- Danaei Mehr, H., and Polat, H. (2022). Diagnosis of polycystic ovary syndrome through different machine learning and feature selection techniques. *Health and Technology*. 12(1), 137-150.
- Davies-Colley, R. J., Vant, W. N., Smith, D. G. (2003). *Colour and clarity of natural waters*. Michigan: Blackburn Press.
- Day, N. E., and Kerridge, D. F. (1967). A general maximum likelihood discriminant. *Biometrics*. 23 (2), 313-323.

- Dayan, P., and Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Current Opinion in Neurobiology*. 18(2), 185-196.
- DeForest, D. K., Brix, K. V., Tear, L. M., Adams, W. J. (2018). Multiple linear regression models for predicting chronic aluminum toxicity to freshwater aquatic organisms and developing water quality guidelines. *Environmental Toxicology and Chemistry*. 37(1), 80-90.
- DeJong, G., Mooney, R. (1986). Explanation-based learning: An alternative view. *Machine Learning*, 1, 145-176.
- Dehghani, R., Torabi Poudeh, H., Izadi, Z. (2022). Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques. *Modeling Earth Systems and Environment*. 8(2), 2599-2613.
- Demirci, A. (2019). *Işıklı-Gökgöl (Denizli-Çivril) Sulak Alanları Ve Civarındaki Yüzey Ve Yeraltı Sularında Klorür, Nitrat Ve Nitrit Kirliliğinin Araştırılması*. Yayınlanmış Yüksek Lisans tezi, Pamukkale Üniversitesi, Pamukkale.
- De Vries, W., McLaughlin, M. J. (2013). Modeling the cadmium balance in Australian agricultural systems in view of potential impacts on food and water quality. *Science of the Total Environment*. 461, 240-257.
- Dezfooli, D., Hosseini-Moghari, S. M., Ebrahimi, K., Araghinejad, S. (2018). Classification of water quality status based on minimum quality parameters: application of machine learning techniques. *Modeling Earth Systems and Environment*. 4(1), 311-324.
- Divya, A. H., Solomon, P. A. (2016). Effects of some water quality parameters especially total coliform and fecal coliform in surface water of Chalakudy river. *Procedia Technology*. 24, 631-638.
- Dohare, D., Deshpande, S., Kotiya, A. (2014). Analysis of Ground Water Quality Parameters: A Review. *Research Journal of Engineering Sciences*. 3 (5), 26-31.
- Dong, X., Yu, Z., Cao, W., Shi, Y., Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*. 14(2), 241-258.
- Dutton, D. M., Conroy, G. V. (1997). A review of machine learning. *The Knowledge Engineering Review*. 12(4), 341-367.
- Dügel, M., and Kazancı, N. (2004). Assessment of water quality of the Büyük Menderes River (Turkey) by using ordination and classification of macroinvertebrates and environmental variables. *Journal of Freshwater Ecology*, 19(4), 605-612.
- Elphick, J. R. F., Bergh, K. D., Bailey, H. C. (2011). Chronic Toxicity of Chloride to Freshwater Species: Effects of Hardness and Implications for Water Quality Guidelines. *Environmental Toxicology and Chemistry*. 30 (1), 239–246.

- Faruk, D. Ö. (2010) A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*, 23(4), 586-594. <https://doi.org/10.1016/j.engappai.2009.09.015>
- Fewtrell, L. and Bartram, J. (2001). Water quality: guidelines, standards, and health: assessment of risk and risk management for water-related infectious diseases. World *Health Organization*.
- Freund, Y., and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*. 55(1), 119-139.
- Friedman, J., Hastie, T., Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*. 28(2), 337-407.
- Gardner, J. R., Yang, X., Topp, S. N., Ross, M. R., Altenau, E. H., & Pavelsky, T. M. (2021). The color of rivers. *Geophysical Research Letters*. 48(1), e2020GL088946.
- Gaya, M. S., Abba, S. I., Aliyu, M. A., Tukur, A. I., Saleh, M. A., Esmaili, P., Wahab, N. A. (2020). Estimation of water quality index using artificial intelligence approaches and multi-linear regression. *IAES International Journal of Artificial Intelligence*. 9(1), 126. <https://doi.org/10.11591/ijai.v9.i1.pp126-134>
- Geerdink, R. B., van den Hurk, R. S., Epema, O. J. (2017). Chemical oxygen demand: Historical perspectives and future challenges. *Analytica Chimica Acta*. 961, 1-11.
- Ghahramani, Z. (2004). *Advanced Lectures on Machine Learning: ML Summer Schools 2003*, Tübingen, Germany: Revised Lectures, 72-112.
- Ghorbani, M. A., Zadeh, H. A., Isazadeh, M., Terzi, O. (2016). A comparative study of artificial neural network (MLP, RBF) and support vector machine models for river flow prediction. *Environmental Earth Sciences*. 75(6), 1-14.
- Ghritlahre, H. K., and Prasad, R. K. (2018). Exergetic performance prediction of solar air heater using MLP, GRNN and RBF models of artificial neural network technique. *Journal of Environmental Management*. 223, 566-575.
- Güven, G. (2004). *Büyük Menderes nehri, Aydın yöresi yüzey sularında bazı kirleticilerin düzeylerinin araştırılması*. Yayınlanmış Doktora Tezi, Adnan Menderes Üniversitesi, Aydın.
- Haghiabi, A. H., Nasrolahi, A. H., Parsaie, A. (2018). Water quality prediction using machine learning methods. *Water Quality Research Journal*. 53(1), 3-13.
- Hameed, M., Sharqi, S. S., Yaseen, Z. M., Afan, H. A., Hussain, A., Elshafie, A. (2017). Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Computing and Applications*. 28(1), 893-905.

- Hamed, M. A. R. (2019). Application of surface water quality classification models using principal components analysis and cluster analysis. *Journal of Geoscience and Environment Protection*. 7, 26-41.
- Harper, D. M. (1992). *Eutrophication of freshwaters*. London: Chapman & Hall, 327.
- Heddam, S., and Kisi, O. (2017). Extreme learning machines: a new approach for modeling dissolved oxygen (DO) concentration with and without water quality variables as predictors. *Environmental Science and Pollution Research*. 24(20), 16702-16724.
- Heidari, E., Sobati, M. A., Movahedirad, S. (2016). Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN). *Chemometrics and Intelligent Laboratory Systems*. 155, 73-85.
- Hopfield, J. J. (1988). Artificial neural networks. *IEEE Circuits and Devices Magazine*. 4(5),
- Horton, R. K. (1965). An index number system for rating water quality. *J Water Pollut Control Fed*. 37(3), 300-306.
- Ho, J.Y., Afan, H.A., El-Shafie, A.H., Koting, S.B., Mohd, N.S., Jaafar, W.Z.B., Sai, H.L., Malek, M.A., Ahmed, A.N., Mohtar, W.H.M.W. and Elshorbagy, A. (2019). Towards a time and cost effective approach to water quality index class prediction. *Journal of Hydrology*. 575, 148-165.
- Hossain, M. R., Oo, A. M. T., Ali, A. B. M. S. (2013). The combined effect of applying feature selection and parameter optimization on machine learning techniques for solar Power prediction. *American Journal of Energy Research*. 1(1), 7-16.
- Hurley, T., Sadiq, R., Mazumder, A. (2012). Adaptation and evaluation of the Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) for use as an effective tool to characterize drinking source water quality. *Water Research*, 46(11), 3544-3552.
- Ibrahim, H., Kutty, A. A. (2013). Recreational stream assessment using Malaysia water quality index. *AIP Conference Proceedings*, 1571, 620–624.
- Ighalo, J. O., Adeniyi, A. G., Marques, G. (2020). Application of linear regression algorithm and stochastic gradient descent in a machine-learning environment for predicting biomass higher heating value. *Biofuels, Bioproducts and Biorefining*, 14(6), 1286-1295.
- Ilić, M., Srdjević, Z., Srdjević, B. (2022). Water quality prediction based on Naïve Bayes algorithm. *Water Science and Technology*. 85(4), 1027-1039.
- Internet: Kurtz, J., Griffith, M. B., Morrison, M. A., Hornig, C. E. (February 2021). Nutrients. U.S. Environmental Protection Agency (EPA), URL: <https://www.epa.gov/caddis-vol2/caddis-volume-2-sources-stressors-responses-nutrients>. Son Erişim Tarihi: 04.03.2022.

- Internet: TC. Tarım ve Orman Bakanlığı. Denizli Tarım Takvimi, URL: <https://denizli.tarimorman.gov.tr/Menu/32/Tarim-Takvimi>. Son Erişim Tarihi: 16.11.2022.
- Internet: United States Environmental Protection Agency. (2012). Water: Monitoring & Assessment/Conductivity. EPA, URL: <https://archive.epa.gov/water/archive/web/html/vms59.html> Son Erişim Tarihi: 06.03.2022.
- Internet: United States Environmental Protection Agency. (March 2012). Dissolved Oxygen and Biochemical Oxygen Demand. EPA, URL: <https://archive.epa.gov/water/archive/web/html/vms52.html>. Son Erişim Tarihi: 04.03.2022.
- Internet: USGS. (2018) Dissolved Oxygen and Water. USGS, URL: <https://www.usgs.gov/special-topics/water-science-school/science/dissolved-oxygen-and-water>. Son Erişim Tarihi: 16.01.2023.
- Internet: Suter, G., Cormier, S., Schofield, K., Gilliam, J., Barbour, C. pH. URL: <https://www.epa.gov/caddis-vol2/caddis-volume-2-sources-stressors-responses-ph>. Son Erişim Tarihi: 03. 03. 2022.
- Internet: United States Environmental Protection Agency. Sulfate in Drinking Water. EPA, URL: <https://archive.epa.gov/water/archive/web/html/sulfate.html>., Son Erişim Tarihi: 02.03. 2022.
- Isiyaka, H. A., Mustapha, A., Juahir, H., Phil-Eze, P. (2019). Water quality modelling using artificial neural network and multivariate statistical techniques. *Modeling Earth Systems and Environment*. 5(2), 583-593.
- Jarvis, P., Quay, K., Macadam, J., Edwards, M., Smith, M. (2018). Intake of lead (Pb) from tap water of homes with leaded and low lead plumbing systems. *Science of the Total Environment*, 644, 1346-1356.
- Johnston, C. A. (1991). Sediment and nutrient retention by freshwater wetlands: effects on surface water quality. *Critical Reviews in Environmental Science and Technology*. 21(5-6), 491-565.
- Johnson, A. A., Ott, M. Q., Dogucu, M. (2022). *Bayes rules!: An Introduction to Applied Bayesian Modeling*. Florida: CRC Press.
- Jović, A., Brkić, K., Bogunović, N. (2015). *A review of feature selection methods with applications*. In 2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO). Opatija, Croatia, 1200-1205.
- Kadam, A. K., Wagh, V. M., Muley, A. A., Umrikar, B. N., Sankhua, R. N. (2019). Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Modeling Earth Systems and Environment*. 5(3), 951-962.

- Karpuzcu, M. (2007). *Çevre Kirlenmesi ve Kontrolü*. İstanbul: Özal Yayınevi, 10-7.
- Katoch, S., Chauhan, S. S., Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*. 80(5), 8091-8126.
- Kapadia, D., and Jariwala, N. (2022). Prediction of tropospheric ozone using artificial neural network (ANN) and feature selection techniques. *Modeling Earth Systems and Environment*, 8(2), 2183-2192
- Keshtegar, B., and Heddami, S. (2018). Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. *Neural Computing and Applications*, 30(10), 2995-3006
- Khaire, U. M., and Dhanalakshmi, R. (2019). Stability of feature selection algorithm: A review. *Journal of King Saud University-Computer and Information Sciences*. 34 (4), 1060-1073
- Koc, C. (2010) A Study on the Pollution and Water Quality Modeling of the River Buyuk Menderes, Turkey. *Clean–Soil, Air, Water*, 38(12), 1169–1176.
- Kohavi, R., John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97, 273-324
- Kroupova, H., Machova, J., & Svobodova, Z. (2005). Nitrite influence on fish: a review. *Veterinarni Medicina*, 50(11), 461.
- Kumar, Z. M., and Manjula, R. (2012). Regression model approach to predict missing values in the Excel sheet databases. *IJCSET*, 3(4), 130-135.
- Kurita, T., Watanabe, K., Otsu, N. (2009, October). Logistic discriminant analysis. In 2009 IEEE International Conference on Systems, Man and Cybernetics, IEEE, San Antonio, 2167-2172.
- Kusnandar, D., Debataraja, N. N., Dewi, P. R. (2019). Classification of Water Quality in Pontianak City Using Multivariate Statistical Techniques. *Applied Mathematical Sciences*, 13(22), 1069-1075.
- Kükrer, S., Mutlu, E. (2019) Assessment of surface water quality using water quality index and multivariate statistical analyses in Saraydüzü Dam Lake, Turkey. *Environmental Monitoring and Assessment*, 191 (71).
- Lambora, A., Gupta, K., & Chopra, K. (2019, February). *Genetic algorithm-A literature review*. In 2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon), IEEE, Faridabad, 380-384.
- Law, M. H., Figueiredo, M. A., Jain, A. K. (2004). Simultaneous feature selection and clustering using mixture models. *IEEE transactions on pattern analysis and machine intelligence*. 26(9), 1154-1166.

- Learned-Miller, E. G. (2014). *Introduction to supervised learning*. Department of Computer Science, University of Massachusetts, Massachusetts, 3.
- Li, M., Liu, Z., Chen, Y., Hai, Y. (2016). Characteristics of iron corrosion scales and water quality variations in drinking water distribution systems of different pipe materials. *Water Research*, 106, 593-603.
- Li, B., Trueman, B. F., Rahman, M. S., Gao, Y., Park, Y., & Gagnon, G. A. (2019). Understanding the impacts of sodium silicate on water quality and iron oxide particles. *Environmental Science: Water Research & Technology*, 5(8), 1360-1370.
- Loague, K., Corwin, D. L. (2006). Point and nonpoint source pollution. *Encyclopedia of hydrological sciences*.
- Lord, D., Qin, X., Geedipally, S. R. (X. Q. S. R. G. Dominique Lord). (2021) Chapter 5: *Exploratory analyses of safety data, Highway Safety Analytics and Modeling*. Amsterdam: Elsevier, 135-177.
- Madalina, P., & Gabriela, B. I. (2014). Water quality index-an instrument for water resources management. *Aerul si Apa. Componente ale Mediului*, 2014, 391 – 398.
- Maheshwari, K., And Lamba, S. (2019, September). *Air quality Prediction Using Supervised Regression Model*. In 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), Ghaziabad, 1-7.
- Maldonado, S., Weber, R. (2009). A wrapper method for feature selection using support vector machines. *Information Sciences*. 179(13), 2208-2217.
- Malek, N. H. A., Wan Yaacob, W. F., Md Nasir, S. A., Shaadan, N. (2022). Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques. *Water*, 14(7), 1067.
- Makubura, R., Meddage, D. P. P., Azamathulla, H. M., Pandey, M., & Rathnayake, U. (2022). A Simplified Mathematical Formulation for Water Quality Index (WQI): A Case Study in the Kelani River Basin, Sri Lanka. *Fluids*, 7(5), 147.
- Marcy, S. M., Suter, G., Cormier, S. (September 2021). Dissolved Oxygen. EPA, URL: Available: <https://www.epa.gov/caddis-vol2/dissolved-oxygen>. Son Erişim Tarihi: 20.03. 2022.
- Masmoudi, S., Elghazel, H., Taieb, D., Yazar, O., Kallel, A. (2020). A machine-learning framework for predicting multiple air pollutants' concentrations via multi-target regression and feature selection. *Science of the Total Environment*, 715, 136991.
- Mazloomi, S., Dehghani, M. H., Norouzi, M., Davil, M. F., Amarluie, A., Tardast, A., & Karamitabar, Y. (2009). Physical and chemical water quality of ilam water treatment plant. *World Applied Sciences Journal*, 6(12), 1660-4.

- Mehr, H. D., & Polat, H. (2019, April). *Human Activity Recognition In Smart Home With Deep Learning Approach*. In 2019 7th International Istanbul Smart Grids and Cities Congress and Fair (ICSG), Istanbul, 149-153.
- Mekki, A. S., Yousif, S. M., Abed, M. M. J. (2023). Testing of chlorine dosage in drinking water using microstrip patch sensor. *TELKOMNIKA Telecommunication Computing Electronics and Control*, 21 (3), 528-534.
- Michie, D., Spiegelhalter, D.J., Taylor, C.C. (1994). *Machine Learning, Neural and Statistical Classification*. Cambridge, U.K.
- Mooney, R. J. (2000). Integrating abduction and induction in machine learning. *Abduction and Induction: essays on their relation and integration*, 181-191.
- Modaresi, F., and Araghinejad, S. (2014). A comparative assessment of support vector machines, probabilistic neural networks, and K-nearest neighbor algorithms for water quality classification. *Water resources management*, 28(12), 4095-4111.
- Mohammadpour, R., Shaharuddin, S., Chang, C. K., Zakaria, N. A., Ghani, A. A., Chan, N. W. (2015). Prediction of water quality index in constructed wetlands using support vector machine. *Environmental Science and Pollution Research*, 22(8), 6208-6219.
- Motevalli, A., Naghibi, S. A., Hashemi, H., Berndtsson, R. (2019). Inverse method using boosted regression tree and k-nearest neighbor to quantify effects of point and non-point source nitrate pollution in groundwater. *Journal of Cleaner Production*, 228, 1248-1263.
- Nacar, S., Mete, B., & Bayram, A. (2020). Günlük Çözünmüş Oksijen Konsantrasyonunun Çok Değişkenli Uyarlanabilir Regresyon Eğrileri İle Tahmin Edilmesi. *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 25(3), 1479-1498.
- 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.
- Nickson, R. T., McArthur, J. M., Shrestha, B., Kyaw-Myint, T. O., Lowry, D. (2005). Arsenic and other drinking water quality issues, Muzaffargarh District, Pakistan. *Applied Geochemistry*, 20(1), 55-68.
- Nyamunda, B. C. (2017). Review of the impact on water quality and treatment options of cyanide used in gold ore processing. *Water Quality*, 183, 225.
- Oğuz, T. C., Şiltu, E., Özdemir, O., Perçin, S., Çetin, T., Bahçeci, H., Turan, N., Yıldırım, E., Bakır, N., Koyun, S., Çakmakçı, M. (Y. K. Zeliha Selek, Dü). (2019). *Bölüm 3: Su Kalitesi, Ekosistem Esaslı Su Kalitesi Yönetimi*. Türkiye: Tarım ve Orman Bakanlığı, 39-95.
- Orman ve Su İşleri Bakanlığı (2015). Yüzeysel su kalitesi yönetimi yönetmeliğinde değişiklik yapılmasına dair yönetmeliği. *Orman ve Su İşleri Bakanlığı*. Ankara, Türkiye.

- Ostad-Ali-Askari, K., Shayannejad, M., Ghorbanizadeh-Kharazi, H. (2017). Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-rood River, Isfahan, Iran. *KSCE Journal of Civil Engineering*, 21, 134–140.
- Ouma, Y. O., Okuku, C. O., Njau, E. N. (2020). Use of artificial neural networks and multiple linear regression model for the prediction of dissolved oxygen in rivers: case study of hydrographic basin of River Nyando, Kenya. *Complexity*, 2020, 9570789.
- Pal, M., Samal, N. R., Roy, P. K., Roy, M. B. (2015). Electrical Conductivity of Lake Water as Environmental Monitoring – A Case Study of Rudrasagar Lake. *IOSR Journal of Environmental Science, Toxicology and Food Technology*, 9(3), 66-71.
- Pan, F., Converse, T., Ahn, D., Salvetti, F., Donato, G. (2009). *Feature selection for ranking using boosted trees*. In Proceedings of the 18th ACM conference on Information and knowledge management, 2025-2028.
- Parry, R. (1998). Agricultural phosphorus and water quality: A US Environmental Protection Agency perspective. *Journal of Environmental Quality*. 27(2), 258-261.
- Pérez-Suárez, A., Martínez-Trinidad, J. F., Carrasco-Ochoa, J. A. (2019). A review of conceptual clustering algorithms. *Artificial Intelligence Review*, 52, 1267-1296.
- Pérez, R., Martínez-Torrents, A., Martínez, M., Grau, S., Vinardell, L., Tomàs, R., Martínez-Lladó, X., Jubany, I. (2022). Chlorine Concentration Modelling and Supervision in Water Distribution Systems. *Sensors*, 22(15), 5578.
- Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, 4(2), 1883.
- Poonam, T., Tanushree, B., Sukalyan, C. (2013). Water quality indices-important tools for water quality assessment: a review. *International Journal of Advances in chemistry*, 1(1), 15-28.
- Press, S. J., and Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, 73(364), 699-705.
- Qiao, W., Tian, W., Tian, Y., Yang, Q., Wang, Y., Zhang, J. (2019). The forecasting of PM2.5 using a hybrid model based on wavelet transform and an improved deep learning algorithm. *IEEE Access*, 7, 142814-142825.
- Qin, Y., Alam, A. U., Pan, S., Howlader, M. M., Ghosh, R., Hu, N. X., ... & Deen, M. J. (2018). Integrated water quality monitoring system with pH, free chlorine, and temperature sensors. *Sensors and Actuators B: Chemical*, 255, 781-790.
- Rahman, M. S., Gagnon, G. A. (2014). Bench-scale evaluation of drinking water treatment parameters on iron particles and water quality. *Water research*, 48, 137-147.
- Rish, I. (2001, August). *An empirical study of the naive Bayes classifier*. In IJCAI 2001 workshop on empirical methods in artificial intelligence, New York, 3 (22), 41-46.

- Ritchie, J. C., Zimba, P. V., Everitt, J. H. (2003). Remote sensing techniques to assess water quality. *Photogrammetric Engineering & Remote Sensing*, 69(6), 695–704.
- Rodrigues, A., Jesus, F. T., Fernandes, M. A., Morgado, F., Soares, A. M., Abreu, S. N. (2013). Mercury toxicity to freshwater organisms: extrapolation using species sensitivity distribution. *Bulletin of Environmental Contamination and Toxicology*, 91(2), 191-196.
- Rown, R. M., McClelland, N. J., Deiniger, R. A., O'Connor, M. F. A. (1972). *Water quality index—crossing the physical barrier*. In Proceedings in International Conference on water pollution Research Jerusalem, Jerusalem, 6, 787-797.
- Roy, S., Edwards, M. A. (2019). Preventing another lead (Pb) in drinking water crisis: Lessons from the Washington DC and Flint MI contamination events. *Current Opinion in Environmental Science & Health*, 7, 34-44.
- Rozemeijer, J. C. and Broers, H. P. (2007). The groundwater contribution to surface water contamination in a region with intensive agricultural land use (Noord-Brabant, The Netherlands. *Environmental Pollution*, 148, 695-706.
- Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B. W., Aslam, W., Choi, G. S. (2020). COVID-19 future forecasting using supervised machine learning models. *IEEE Access*, 8, 101489-101499.
- Sagi, O., and Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8 (4), e1249.
- Samsudin, M. S., Azid, A., Khalit, S. I., Sani, M. S. A., Lananan, F. (2019). Comparison of prediction model using spatial discriminant analysis for marine water quality index in mangrove estuarine zones. *Marine Pollution Bulletin*, 141, 472-481.
- Sánchez, E., Colmenarejo, M. F., Vicente, J., Rubio, A., García, M. G., Travieso, L., & Borja, R. (2007). Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecological indicators*, 7(2), 315-328.
- Sappington, K., and Norton, S. B. (2022). Temperature. EPA, URL: https://www.epa.gov/caddis/authors-and-contributors-caddis#vol2_temp. Son Erişim Tarihi: 03.03. 2022.
- Schapire, R. E., Singer, Y. (1998). Improved boosting algorithms using confidence-rated predictions. *Machine Learning*, 37, 297–336.
- Scottish Development Department. (1975). *Towards cleaner water 1975*. Report of a second Rivers Pollution Survey of Scotland. Edinburgh.
- Sefidian, A. M., and Daneshpour, N. (2019). Missing value imputation using a novel grey based fuzzy c-means, mutual information based feature selection, and regression model. *Expert Systems with Applications*, 115, 68-94.

- Segal, M. R. (2004). Machine learning benchmarks and random *forest regression*. *Center for Bioinformatics and Molecular Biostatistics*. University of California, San Fransisco.
- Shallenberger, R.S. (1993). *Water and Inorganic Compounds*. In Taste Chemistry, Springer, Boston. https://doi.org/10.1007/978-1-4615-2666-7_4
- Shaw-Allen, P. and Suter, G. W. (February 2021). Metals. EPA, URL: <https://www.epa.gov/caddis-vol2/metals>. Son Erişim Tarihi: 07.03. 2022.
- Singh, H., Roy, A., Setia, R. K., Pateriya, B. (2022). Estimation of nitrogen content in wheat from proximal hyperspectral data using machine learning and explainable artificial intelligence (XAI) approach. *Modeling Earth Systems and Environment*, 8(2), 2505-2511.
- Singh, K. P., Gupta, S., Rai, P. (2013). Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmospheric Environmen*, 80, 426-437.
- Sim, J., Lee, J. S., Kwon, O. (2015). Missing values and optimal selection of an imputation method and classification algorithm to improve the accuracy of ubiquitous computing applications. *Mathematical Problems in Engineering*, 2015, 538613.
- Smith, V. H., Schindler, D. W. (2009). Eutrophication science: where do we go from here?. *Trends in ecology & evolution*, 24(4). 201-207.
- Singh, A., Thakur, N., Sharma, A. (2016). *A review of supervised machine learning algorithms*. In 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 1310-1315.
- Srivastava, S., Gupta, M. R., Frigyik, B. A. (2007). Bayesian quadratic discriminant analysis. *Journal of Machine Learning Research*, 8(6).
- Stone, M. (1978). Cross-validation: A review. *Statistics: A Journal of Theoretical and Applied Statistics*, 9(1), 127-139.
- Somay, M. A., and Gemici, U. (2012). Groundwater quality degradation in the Buyuk Menderes River coastal wetland. *Water, Air, & Soil Pollution*, 223, 15-27.
- Soylu, S., Sülü, S. M., & Bozkurt, İ. A. (2016). Bitki büyüme düzenleyici ve biyolojik mücadele etmeni olarak bakteriyel endofitler. *Mustafa Kemal Üniversitesi Ziraat Fakültesi Dergisi*, 21(1).
- Sulaiman, K., Ismail, L. H., Razi, M. A. M., Adnan, M. S., Ghazali, R. (2019, August). Water quality classification using an Artificial Neural Network (ANN). *IOP Conference Series: Materials Science and Engineering*, 601 (1), 012005.
- Sutton, R. S., Barto, A. G. (2018). *Reinforcement learning: An introduction*. Massachuset: MIT press.
- Sütgibi, S. (2015). Büyük Menderes Havzasının Sıcaklık, Yağış ve Akım Değerlerindeki Değişimler ve Eğilimler. *Marmara Coğrafya Dergisi*, (31), 398-414.

- Svinicki, M. D. (1998). A theoretical foundation for discovery learning. *Advances in physiology education*, 275(6), S4.
- Sönmez, İ., Kaplan, M., Sönmez, S. (2008). Kimyasal gübrelerin çevre kirliliği üzerine etkileri ve çözüm önerileri. *Derim*, 25(2), 24-34.
- Tan, J. F., Anastasi, A., Chandra, S. (2022). Electrochemical Detection of Nitrate, Nitrite and Ammonium for on-Site Water Quality Monitoring. *Current Opinion in Electrochemistry*, 32, 100926.
- Tarım ve Orman Bakanlığı. (2018). Büyük Menderes Nehir Havzası Yönetim Planı. *Havza Koruma Eylem Planlarının Nehir Havzası Yönetim Planlarına Dönüştürülmesi için Teknik Yardım*. Ankara.
- Tehrany, M. S., Jones, S., Shabani, F., Martínez-Álvarez, F., Tien Bui, D. (2019). A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using LogitBoost machine learning classifier and multi-source geospatial data. *Theoretical and Applied Climatology*, 137(1), 637-653.
- Tiwari, S., Babbar, R., Kaur, G. (2018). Performance evaluation of two ANFIS models for predicting water quality index of River Satluj (India). *Advances in Civil Engineering*, 2018, 8971079.
- Tyagi, S., Sharma, B., Singh, P., Dobhal, R. (2013). Water quality assessment in terms of water quality index. *American Journal of water resources*, 1(3), 34-38.
- TÜBİTAK (2013) Büyük Menderes Havzası. *Havza Koruma Eylem Planlarının Hazırlanması*. Ankara.
- Türkiye Standartları Enstitüsü (TSE). (2005). TS-266: *Water intended for human consumption*. Ankara: TSE.
- Uddin, M. G., Nash, S., Olbert, A. I. (2021). A review of water quality index models and their use for assessing surface water quality. *Ecological Indicators*, 122, 107218.
- Ustaoglu, F., Tepe, Y., Taş, B. (2020). Assessment of stream quality and health risk in a subtropical Turkey river system: A combined approach using statistical analysis and water quality index. *Ecological indicators*, 113, 105815.
- Van, N. H., Nguyen Viet, H., Truong Trung, K., Nguyen Hai, P., Nguyen Dang Giang, C. (2022). A comprehensive procedure to develop water quality index: A case study to the Huong river in Thua Thien Hue province, Central Vietnam. *Plos One*, 17(9), e0274673.
- Varol, M., Şen, B. (2012). Assessment of nutrient and heavy metal contamination in surface water and sediments of the upper Tigris River, Turkey. *CATENA*, 92, 1–10.
- Varol, M. (2020). Use of water quality index and multivariate statistical methods for the evaluation of water quality of a stream affected by multiple stressors: A case study. *Environmental Pollution*, 266, 115417.

- Vasanthavigar, M., Srinivasamoorthy, K., Vijayaragavan, K., Rajiv Ganthi, R., Chidambaram, S., Anandhan, P., Manivannan, R., Vasudevan, S. (2010). Application of water quality index for groundwater quality assessment: Thirumanimuttar sub-basin, Tamilnadu, India. *Environmental Monitoring and Assessment*. 171(1), 595-609.
- Vezhnevets, A., and Vezhnevets, V. (2005). Modest AdaBoost-teaching AdaBoost to generalize better. *Graphicon*, 12 (5), 987-997.
- Vujović, Ž. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, 12(6), 599-606.
- Walters, C. R., Somerset, V. S., Leaner, J. J., Nel, J. M. (2011). A review of mercury pollution in South Africa: current status. *Journal of Environmental Science and Health*, 46(10), 1129-1137.
- Wang, S. C. (2003). Artificial neural network. In *Interdisciplinary computing in java programming*. Springer, Boston, 81-100.
- Wang, H., Ma, C., & Zhou, L. (2009). *A brief review of machine learning and its application*. In 2009 international conference on information engineering and computer science. Wuhan, 1-4.
- Wang, Y., Li, Y., Pu, W., Wen, K., Shugart, Y. Y., Xiong, M., Jin, L. (2016). Random bits forest: a strong classifier/regressor for big data. *Scientific Reports*. 6(1), 1-8.
- Wang, Y., Zhou, J., Chen, K., Wang, Y., Liu, L. (2017). *Water quality prediction method based on LSTM neural network*. In 2017 12th international conference on intelligent systems and knowledge engineering (ISKE). Nanjing, 1-5.
- Wong, J. M. (1984). Chlorination-Filtration for Iron and Manganese Removal. *Journal AWWA Research and Technology*, 76(1), 76-79.
- World Health Organization. Water, Sanitation and Health Team. (2006). *Guidelines for drinking-water quality: incorporating first addendum*. Recommendations, 3rd ed. Geneva: World Health Organization.
- World Health Organization (2021). *Nickel in drinking water: background document for development of WHO Guidelines for drinking-water quality*. Geneva: World Health Organization.
- Xin, L., and Mou, T. (2022). Research on the Application of Multimodal-Based Machine Learning Algorithms to Water Quality Classification. *Wireless Communications and Mobile Computing*, 2022, 9555790.
- Xue, M., Zhu, C. (2009, April). *A study and application on machine learning of artificial intelligence*. In 2009 International Joint Conference on Artificial Intelligence, Hainan, 272-274.

- Yaseen, Z. M., Ramal, M. M., Diop, L., Jaafar, O., Demir, V., Kisi, O. (2018). Hybrid adaptive neuro-fuzzy models for water quality index estimation. *Water Resources Management*, 32(7), 2227-2245.
- Ye, J. (2007). *Least squares linear discriminant analysis*. In Proceedings of the 24th international conference on Machine learning. Oregon, 1087-1093.
- Yılmaz, B. (2013). *Terkos Gölü'ne dökülen Derelerde Mikrobiyolojik Ve Kimyasal Kirlilik düzeylerinin saptanması*. Yayınlanmış Doktora Tezi, Marmara Üniversitesi, İstanbul.
- Yılmaz, E., Koç, C., Gerasimov, I. (2020) A study on the evaluation of the water quality status for the Büyük Menderes River, Turkey. *Sustainable Water Resources Management*, 6, 100.
- Yozgatligil, C., Aslan, S., Iyigun, C., Batmaz, I. (2013). Comparison of missing value imputation methods in time series: the case of Turkish meteorological data. *Theoretical and Applied Climatology*. 112(1), 143-167.
- Yusri, H. I. H., Hassan, S. L. M., Halim, I. S. A., & Abdullah, N. E. (2022). *Water Quality Classification Using SVM And XGBoost Method*. In 2022 IEEE 13th Control and System Graduate Research Colloquium (ICSGRC). Shah Alam, 231-236.
- Zandbergen, P. A., Hall, K. J. (1998). Analysis of the British Columbia water quality index for watershed managers: a case study of two small watersheds. *Water Quality Research Journal*, 33(4), 519-550.
- Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *Journal of Applied Science and Technology Trends*, 1(2), 56-70.
- Zhou, Q., Zhou, H., Li, T. (2016). Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features. *Knowledge-based Systems*, 95, 1-11.
- Zhu, S., and Heddam, S. (2020). Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: extreme learning machines (ELM) versus artificial neural network (ANN). *Water Quality Research Journal*, 55(1), 106-118.
- Zhu, Y., Lu, Y., Hu, F. (2021, November). *The data analysis and smart discriminant method based on fisher model*. In 2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT). Chongqing, 843-846.



Gazili olmak ayrıcalıktır