

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İNŞAAT MÜHENDİSLİĞİ ANABİLİM DALI**

**SU KALİTESİ İNDEKSİNİN TAHMİNİ İÇİN MAKİNE
ÖĞRENME YÖNTEMLERİNİN KARŞILAŞTIRILMASI:
KASKİ ÖRNEĞİ**

**Hazırlayan
Nur ORHAN**

**Danışman
Prof. Dr. A. Alper ÖNER**

Yüksek Lisans Tezi

**Ağustos 2024
KAYSERİ**

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İNŞAAT MÜHENDİSLİĞİ ANABİLİM DALI**

**SU KALİTESİ İNDEKSİNİN TAHMİNİ İÇİN MAKİNE
ÖĞRENME YÖNTEMLERİNİN KARŞILAŞTIRILMASI:
KASKİ ÖRNEĞİ
(Yüksek Lisans Tezi)**

**Hazırlayan
Nur ORHAN**

**Danışman
Prof. Dr. A. Alper ÖNER**

**Bu çalışma, Erciyes Üniversitesi Bilimsel Araştırma Projeleri Birimi
tarafından FYL-2024-13494 kodlu proje ile desteklenmiştir.**

**Ağustos 2024
KAYSERİ**

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim.

Nur ORHAN

“Su Kalitesi İndeksinin Tahmini İçin Makine Öğrenme Yöntemlerinin Karşılaştırılması: KASKİ Örneği” adlı Yüksek Lisans Tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi’ ne uygun olarak hazırlanmıştır.

Hazırlayan

Nur ORHAN

Danışman

Prof. Dr. A. Alper ÖNER

İnşaat Mühendisliği ABD Başkanı

Prof. Dr. Zülküf KAYA

Prof. Dr. A. Alper ÖNER danışmanlığında **Nur ORHAN** tarafından hazırlanan “**Su Kalitesi İndeksinin Tahmini İçin Makine Öğrenme Yöntemlerinin Karşılaştırılması: KASKİ Örneği**” adlı bu çalışma jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü **İnşaat Mühendisliği** Anabilim Dalında **Yüksek Lisans** tezi olarak kabul edilmiştir.

/ 08 /2024

JÜRİ:

Danışman : Prof. Dr. A. Alper ÖNER

Üye : Prof. Dr. Hatice ÇITAKOĞLU

Üye : Doç. Dr. Oğuz ŞİMŞEK

ONAY:

Bu tezin kabulü Enstitü Yönetim Kurulunun tarih ve sayılı kararı ile onaylanmıştır.

..... /...../

Prof. Dr. Emel KIZILKAYA AYDOĞAN

Enstitü Müdürü

TEŞEKKÜR

Bana Yüksek lisans çalışmam süresince her türlü yardımı ve fedakârlığı sağlayan, bilgi ve deneyimleriyle bana yol gösteren ve üzerimden hiçbir emeğine esirgemeyen sevgili danışmanım Prof. Dr. A. Alper ÖNER hocama sonsuz teşekkür ediyorum.

Çalışmalarım esnasında benimle tüm bilgi ve deneyimlerini paylaşarak tez yazma sürecimde bana desteğini ve hiçbir fedakârlığını esirgemeyerek çalışma umudumu diri tutan, beni motive eden ve bana yol gösteren güzel kalpli kıymetli hocam sayın Prof. Dr. Hatice ÇITAKOĞLU'na ve üzerimde emeği olan tüm hocalarıma teşekkürlerimi bir borç bilirim.

Bana Kayseri Su ve Kanalizasyon İdaresi (KASKİ)'de staj imkânı sağlayan, tez çalışma konum için kurumsal verilerini paylaşarak iş ve yaşam tecrübelerini benimle paylaşan 2019-2013 yılları arasında Genel Müdür olan sayın Prof. Dr. Özgür ÖZDEMİR hocama çok teşekkür ederim. Ayrıca KASKİ'de görev yapan ve bana bu tecrübeyi yaşamamda yardımcı oldukları için teşekkür ederim.

Bu tez çalışmasına maddi destek sağlayan Erciyes Üniversitesi Bilimsel Araştırma Projeleri Birimi'ne (Proje No: FYL-2024-13494) teşekkür ederim.

Beni bugünlere kadar yetiştirip büyüten ve her zaman yanımda olan üzerimden maddi ve manevi desteklerini esirgemeyen bana her zaman güvenen sevgili aileme minnetlerimi sunarım.

Nur ORHAN

Ağustos 2024, KAYSERİ

SU KALİTESİ İNDEKSİNİN TAHMİNİ İÇİN MAKİNE ÖĞRENME YÖNTEMLERİNİN KARŞILAŞTIRILMASI: KASKİ ÖRNEĞİ

Nur ORHAN

Erciyes Üniversite, Mühendislik Fakültesi İnşaat Mühendisliği
Yüksek Lisans, Ağustos 2024
Tez Danışmanı: Prof. Dr. A. Alper ÖNER

ÖZET

Bu tez çalışmasında, Kayseri Su ve Kanalizasyon İdaresi (KASKİ) tarafından 2015-2022 yılları arasındaki içme suyu verileri kullanarak su kalitesinin tahmin edilmesi araştırılmıştır. Yapılan tahmin çalışmasında giriş verisi olarak pH, iletkenlik, toplam siyanür, sülfat, nitrat, sodyum ve serbest klor kullanılırken çıkış verisi ise su kalitesi indeksi (WQI) kullanılmıştır. Modelleme yapılırken, yapay zekâ teknikleri olarak uyarlamalı ağ tabanlı bulanık çıkarım sistemi, Yapay Sinir Ağları, Gauss Süreç Regresyon, Mars Gezgin Model, Destek Vektör Makinesi, En Küçük Kareler Destek Vektör Regresyonu, Araştırılmış Regresyon Topluluğu Yöntemi, Yakın Komşu Sınıflandırılması Modeli, LSSVM ve M5 Model Ağacı Yöntemi kullanılmıştır. Modellerin tahmin başarıları performans kriterlerine göre doğruluğu test edilmiştir. Performans kriteri olarak, Determinasyon Katsayısı, Ortalama Mutlak Hata, Ortalama Karesel Hata, Kling-Gupta Verimliliği, Bağlı Kök Ortalama Hata, Nash-Sutcliffe Etkinliği, Overall Index ve Willmott Rafine İndeksi kullanılmıştır. Model sonuçlarına göre, karşılaştırma için kullanılan yöntemlerin tatmin edici sonuçlar verdiği görülmüştür. En iyi tahmin modeli ise LSSVM'dir. Önerilen yöntemlerin etkinliğini ise Kruskal-Wallis test ile doğruluğu desteklenmiştir. Bu araştırmanın amacı, WQI' belirlemek için doğrulanmış makina modeli sunarak, zaman alıcı, yoğun emek gerektiren, hesaplama hatalarına yatkın olan maliyetli ve geleneksel yöntemlere alternatif çözüm üretmektir.

Anahtar Kelimeler: Su Kalitesi İndeksi, Yapay Sinir Ağı Modelleri, Performans Kriterleri, Kruskal Wallis Testi

COMPARISON OF MACHINE LEARNING METHODS FOR WATER QUALITY INDEX ESTIMATION: KASKI EXAMPLE

Nur ORHAN

**Erciyes University, Graduate School of Natural and Applied Sciences
Master Thesis, August 2024
Supervisor: Prof. Dr. A. Alper ÖNER**

ABSTRACT

In this thesis, the estimation of water quality was investigated by Kayseri Water and Sewerage Administration (KASKİ) using drinking water data between 2015 and 2022. In the estimation study, pH, conductivity, total cyanide, sulfate, nitrate, sodium and free chlorine were used as input data, while water quality index (WQI) was used as output data. While modeling, adaptive network based fuzzy inference system, Artificial Neural Networks, Gaussian Process Regression, Mars Rover Model, Support Vector Machine, Least Squares Support Vector Regression, Researched Regression Ensemble Method, Nearest Neighbor Classification Model, LSSVM and M5 Model Tree Method were used as artificial intelligence techniques. The accuracy of the models' prediction success was tested according to performance criteria. Coefficient of Determination, Mean Absolute Error, Mean Square Error, Kling-Gupta Efficiency, Relative Root Mean Error, Nash-Sutcliffe Efficiency, Overall Index and Willmott Refined Index were used as performance criteria. According to the model results, it was seen that the methods used for comparison gave satisfactory results. The best estimation model is LSSVM. The effectiveness of the proposed methods was supported by Kruskal-Wallis test. The aim of this research is to produce an alternative solution to the traditional methods that are time-consuming, labor-intensive, costly and prone to calculation errors by presenting a validated machine model for determining WQI.

Keywords: Water Quality Index, Artificial Neural Network Models, Performance Criteria, Kruskal Wallis Test,

İÇİNDEKİLER

KASKİ İÇME SUYU ŞEBEKE İŞLETME VERİLERİNİN ANFİS VE YSA MODELLERİ KULLANARAK SU KALİTESİNİN TAHMİN EDİLMESİ

BİLİMSEL ETİĞE UYGUNLUK	ii
YÖNERGEYE UYGUNLUK.....	iii
KABUL VE ONAY	iv
TEŞEKKÜR.....	v
ÖZET	vi
ABSTRACT.....	vii
İÇİNDEKİLER	viii
KISALTMALAR	xii
TABLolar LİSTESİ.....	xiv
ŞEKİLLER LİSTESİ	xv
GİRİŞ	1

1. BÖLÜM

GENEL BİLGİLER ve LİTERATÜR ÇALIŞMASI

1.1. Problem Durumu	4
1.2. Araştırmanın Amacı	5
1.3. Araştırmanın Önemi.....	6
1.4. Literatür Taraması	7

2. BÖLÜM

YÖNTEM VE MATERYAL

2.1. Kayseri Su ve Kanalizasyon İdaresi Genel Müdürlüğü	17
2.1.1. KASKİ Su Kalitesi Verilerinin Analiz Süreci	18
2.2. Su Kalitesi Parametreleri	19

2.2.1. Fiziksel Parametreler	21
2.2.1.1. Bulanıklık.....	21
2.2.1.2. Sıcaklık.....	21
2.2.1.3.Tat ve Koku	21
2.2.1.4. Renk.....	22
2.2.1.5. Askıda Katı Madde (AKM).....	22
2.2.1.6.Toplam Tuz Konsantrasyonu ve Elektriksel İletkenlik (Eİ).....	22
2.2.2. Kimyasal Parametreler	23
2.2.2.1. Çözünmüş Oksijen (ÇO)	23
2.2.2.2.Biyolojik Oksijen İhtiyacı (BOİ).....	23
2.2.2.3.Kimyasal Oksijen İhtiyacı (KOİ).....	24
2.2.2.4. pH Değeri	24
2.2.2.5. Nitrat	24
2.2.2.6. Serbest Klor	24
2.2.2.7. Sodyum Adsorbsiyon Oranı (SAO).....	25
2.2.3.Biyolojik Parametreler	25
2.2.3.1.Fekalkoliform	25
2.2.3.2.Klorofil-a.....	25
2.2.4.Ağır Metaller	25
2.2.4.1.Siyanür (CN ⁻).....	25
2.2.4.2.Civa (Hg).....	26
2.2.4.3.Arsenik (As)	26
2.3. KASKİ'den Alınan Veriler.....	26
2.4. Yapay Sinir Ağları (ANN)	27
2.4.1. Yapay Sinir Ağının Tarihçesi ve Kullanım Alanları	28
2.4.2. Yapay Sinir Ağlarının Avantaj ve Dezavantajları.....	30
2.4.3. Yapay Sinir Ağlarının Yapısı	31

2.4.3.1. Hücre Gövdesi	32
2.4.3.2. Dendrit	33
2.4.3.3. Akson.....	33
2.4.3.4. Sinaps	33
2.4.4. Yapay Sinir Ağı Mimarileri	38
2.4.4.1. Tek Katmanlı Ağlar	38
2.4.4.2. Çok Katmanlı Ağlar.....	39
2.4.5. Yapılarına Göre Yapay Sinir Ağları.....	40
2.4.5.1. İleri Beslemeli Yapay Sinir Ağları.....	40
2.4.5.2. Geri Beslemeli Yapay Sinir Ağları	41
2.4.6. Öğrenme Algoritmalarına Göre Yapay Sinir Ağları.....	42
2.4.6.1. Denetimli Öğrenme	42
2.4.6.2. Denetimsiz Öğrenme	43
2.4.6.3. Takviyeli Öğrenme.....	43
2.5. Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS)	43
2.6. GAUSS Süreç Regresyonu Modeli (GPR)	46
2.7. Mars Gezgin Model (MARS)	49
2.8. Destek Vektör Makinesi Modeli (SVM).....	50
2.9. En Küçük Kareler Artırma Modeli (LSBoost).....	52
2.10. K-En Yakın Komşu Modeli (KNN)	53
2.11. Radyal Tabanlı M5 Model Ağacı (M5TREE)	55
2.12. En Küçük Kareler Destek Vektör Makinesi (LSSVM)	56
2.13. Performans Kriterleri.....	58
2.13.1. Ortalama Mutlak Yüzde Hata (MAPE)	58
2.13.2. Bağlı Kök Ortalama Karesel Hata (RRMSE).....	59
2.13.3. Ortalama Mutlak Hata (MAE).....	60
2.13.4. Determinasyon Katsayısı (R^2).....	60

2.13.5. Nash-Sutcliffe Etkinliği (NSE).....	61
2.13.6. Kling-Gupta Verimliliği (KGE).....	61
2.13.7. Overall Index (OI)	62
2.13.8. Willmott Rafine İndeksi (WI).....	62
2.14. Su Kalite İndeksi (WQI).....	63

3. BÖLÜM

BULGULAR

3.1. Model Gelişimi.....	65
3.2. Makine Öğrenmesi Modellerinin Elde Edilmesi	66
3.2.1. ANFIS Model Gelişimi	66
3.2.2. ANN Model Gelişimi.....	67
3.2.3. GPR Model Gelişimi.....	69
3.2.4. SVM Model Gelişimi	71
3.2.5. M5TREE Model Gelişimi.....	72
3.2.6. KNN Model Gelişimi	73
3.2.7. MARS Model Gelişimi	73
3.2.8. LSSVM Model Gelişimi	73
3.2.9. LSBoost Model Gelişimi.....	74
3.3. Analiz Sonuçları	74

4. BÖLÜM

TARTIŞMA-SONUÇ ve ÖNERİLER

4.1. Tartışma	86
4.2. Sonuç ve Öneriler	90
KAYNAKÇA	92
ÖZGEÇMİŞ.....	97

KISALTMALAR

AB	: Avrupa Birliđi
AKM	: Askıda Katı Madde
AN	: Amonyok Azotu
ANFIS	: Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi
ANN	: Yapay Sinir Ađı
As	: Arsenik
BOİ	: Biyolojik Oksijen İhtiyacı
Ca ⁺²	: Kalsiyum
CN ⁻	: Siyanür
ÇO	: Çözünmüş Oksijen
DÖ	: Derin Öğrenme
DSİ	: Devlet Su İşleri
DT	: Karar Ađacı
Eİ	: Elektriksel İletkenlik
GPR	: Gauss Süreç Regresyon
HCO ₃ ⁻¹	: Bikarbonat
Hg	: Civa
KASKİ	: Kayseri Su ve Kanalizasyon İdaresi
KGE	: Kling-Gupta Verimliliđi
KNN	: Yakın Komşu Sınıflandırılması Modeli
KOİ	: Kimyasal Oksijen İhtiyacı
KW	: Kruskall Wallis
LSBooST	: Araştırılmış Regresyon Topluluđu Yöntemi
LSSVM	: En Küçük Kareler Destek Vektör Makinesi
LSTM	: Uzun Kısa Süreli Bellek
M5TREE	: M5 Model Ađacı Yöntemi
MAE	: Ortalama Mutlak Hata
MAPE	: Ortalama Mutlak Yüzde Hata
MARS	: Mars Gezgin Model
MLP-ANN	: Çok Katmanlı Yapay Sinir Ađı

Na^{+1}	: Sodyum
NO_2^{-1} :	Nitrit
NSE	: Nash-Sutcliffe Etkinliđi
OI	: Overall Index
R^2	: Determinasyon Katsayısı
RBF-ANN	: Radyal Tabanlı Fonksiyon Sinir Ağları
RRMSE	: Bağlı Kök Ortalama Karesel Hata
SAO	: Sodyum Adsorbsiyon Oranı
SMO	: Sıralı Minimum Optimizasyon
SO_4^{-2}	: Sülfat
SVM	: Destek Vektör Regresyon
WDT-ANFIS	: Nöro Bulanık Çıkarım Sistemi
WHO	: Dünya Sağlık Örgütü
WI	: Willmott Rafine Index
WQI	: Su Kalite İndeksi

TABLOLAR LİSTESİ

Tablo 1. İçme suyu kalitesi parametre değerlerinin kabul edilebilir maksimum değerleri	20
Tablo 2. Yapay sinir ağı sisteminin biyolojik sinir sistemindeki karşılığı.....	34
Tablo 3. Toplama fonksiyonları	35
Tablo 4. Aktivasyon fonksiyonları.....	37
Tablo 5. MAPE hata kriterlerinin sınıflandırılması	59
Tablo 6. RRMSE hata kriterleri sınıflandırılması	59
Tablo 7. R^2 hata kriterlerinin sınıflandırılması	60
Tablo 8. NSE hata kriterlerinin sınıflandırılması	61
Tablo 9. KGE hata kriterlerinin sınıflandırılması	62
Tablo 10. WQI modeli su kalitesi sınıfları	64
Tablo 11. ANFIS tahmin modeli parametreleri	67
Tablo 12. YSA tahmin modeli parametreleri	68
Tablo 13. GPR tahmin modeli parametreleri	70
Tablo 14. SVM tahmin modeli parametreleri	72
Tablo 15. WQI tahmini model performanslarının karşılaştırılması	77
Tablo 16. KW testinin sonuçları	85

ŞEKİLLER LİSTESİ

Şekil 1. Biyolojik nöron yapısı ve elemanları.....	32
Şekil 2. Yapay sinir hücresinin yapısı ve elemanları	34
Şekil 3. Tek katmanlı iki girdi ve bir çıktıdan oluşan yapay sinir ağı modeli	38
Şekil 4. Tek çıkışlı çok katmanlı yapay sinir ağı modeli	40
Şekil 5. İleri beslemeli yapay sinir ağı model gösterimi	41
Şekil 6. Geri beslemeli yapay sinir ağı modeli	42
Şekil 7. İki girdi ve bir çıktıya sahip ANFIS yapısı.....	44
Şekil 8. SVM model gösterimi.....	51
Şekil 9. LSBoost modeli şematik gösterimi	53
Şekil 10. K-en yakın komşu modeli örneği.....	55
Şekil 11. WQI değeri için kullanılan su kalitesi parametreleri	64
Şekil 12. WQI modellerine ait saçılma diyagramları ve zaman serisi grafikleri	81
Şekil 13. WQI tahmini için kullanılan modellere ait taylor diyagramı.....	82
Şekil 14. WQI tahmini için kullanılan modellere ait violin diyagramları.....	83
Şekil 15. WQI tahmini için kullanılan modellere ait hata violin diyagramları.....	84

GİRİŞ

Uzaydan bakıldığında gezegenimizin yaklaşık %71 su ile kaplı olduğu görülmektedir. Sıvı haldeki su Dünya'yı diğer gezegenlerden ve Güneş sisteminden ayıran en önemli maddedir. Dünya'daki suyun %96,5'i okyanuslarda geri kalan kısmı ise nehirlerde, göllerde, buzullarda ve su buharı şeklinde atmosferde bulunmaktadır. Fakat canlıların hayatını devam ettirebilmesi için tatlı sulara ihtiyaçları vardır. Dünya'daki suların ise sadece %2,5'ni tatlı sular oluşturmaktadır. Ülkemizde ise bu rakam 95 km³ yüzey suyunu oluştururken bizler ise potansiyelimizin sadece 27,5 km³ lük kısmından yararlanabilmekteyiz. Bu kaynakların ise yaklaşık %76'sını tarımsal faaliyetlerde, %14'ünü içme suyunda, %10'unu ise endüstriyel faaliyetlerde kullanılmaktadır. Yeryüzünün de bu kadar kısıtlı miktarda bulunan tatlı sular insan yaşamı için hayati önem taşımakta olup çeşitli çalışmalara da araştırma konusu olmuştur [1].

Yaşam kaynağı olan su insanlar için vazgeçilmez bir madde olmuştur. İnsan yaşamını devam ettirebilmesi için hayatında su olmazsa olmaz bir faktördür. Fakat son zamanlarda ciddi bir şekilde değişim gösteren iklim değişikliği ve beraberinde getirdiği küresel ısınma insan yaşamını son derece olumsuz etkilemektedir. Mevsim normallerinin üzerinde artış gösteren sıcaklık ve kuraklık etkisi ile sınırlı sayıda bulunan su kaynaklarımıza ulaşım günden güne azalmaktadır. Ayrıca her geçen gün artan nüfus, beraberinde getirdiği günlük su tüketim miktarının artması, evsel ve endüstriyel faaliyetlerin hız kazanması, atık sular ve tarımsal faaliyetlerde kullanımın yaygınlaşmasına bağlı olarak su kaynaklarımız kirlenmekte ve giderek tükenmektedir.

Diğer yanda su kaynaklarının yeryüzünde eşit dağılmamasının sonucu olarak dünya ülkelerinin üçte birinin suya erişimi çok kısıtlı miktarda bulunmaktadır. Durum böyle iken dünya nüfusunun % 18'i temiz suya ulaşmakta güçlük çekmektedir. Su kaynaklarımızın azalması ve kirlenmesi sonucu olarak salgın hastalıkların artması ve

göç gibi durumların yaşanması insan sağlığını ve su kalitesini olumsuz yönde etkilemektedir.

Su kaynaklarının korunması için su kalitesinin takip edilmesi ve izlenilmesi gerekmektedir. Bundan dolayı su kalitesini etkileyen parametrelerin düzenli olarak ölçülmesi ve kayıt altına alınması gerekmektedir. Bu kapsamda Avrupa Birliği (AB) su politikaları gelişimini üç aşamaya ayırmış olup bunların en sonuncusu ve en kapsamlı olanı 23 Ekim 2000 yılında yayınlanan 2000/60/EC sayılı ‘Su Çerçeve Direktifi’ olmuştur. Yer altı ve yer üstü suları olmak üzere bütün su kaynaklarının korunmasını sağlamak, su kaynaklarının kirlenmesine sebebiyet veren faktörlerin kontrol altına almak ve su kullanımı ile ilgili ücretlendirilmeleri belirlemek bu direktifin temel amaçlarının bir kısmı arasında yer almaktadır [2].

Türkiye’de ise 28 Temmuz 2016 tarihinde yayınlanan ‘İnsani Tüketim Amaçlı Sular Hakkında Yönetmelik’ sayesinde içme ve kullanım amaçlı suların hijyen şartlarına uygunluğunun test edilmesi, su kalite parametrelerinin ölçülmesi ve standartlara uygunluğunun kontrol edilmesi, kaynak ve içme sularının elde edilmesi ve üretiminin sağlanması, etiket ve ambalajlanmasının yapılması ve denetlenmesi gibi esasları düzenlemek amacıyla oluşturulmuştur. Bu yönetmeliğe göre büyükşehir belediyelerine bağlı su ve kanalizasyon idareleri mevcut olup ilgili belediyeler alt yapı hizmetlerini üstlenerek bakım, onarım, inşaat ve işletme işlerini yerine getirmektedir.

Yapılan çalışmalar ve deneyler sonucunda elde edilen veriler toplanıp bilgiye erişim kolaylığını kolaylaştırmaktadır. Öte yandan çok sayıda veriyi kullanıp aynı anda tüm verileri birleştirerek ilerideki durumu hakkında tahminde bulunmak gelişen teknoloji ile daha kolay hale gelmiştir.

Geçmiş yıllarda yapılan yorucu ve uzun zamanlar alan deneysel faaliyetler sonucunda elde edilen çok kapsamlı veri kümelerinin yapay zekâ teknikleriyle kullanarak modellemeleri ve tahminlerini yapmak mümkün hale gelmiştir.

Yapay zekâ teknikleri geçmişte ve günümüzde elde edilen verilerin toplu kullanılmasını sağlayarak gelecekte nasıl sonuçlar doğuracağını tahmin etmek amacıyla da kullanılmaktadır. Su kaynakları yönetimini de kapsayan su kalite tahmininin yapılması bu çalışmalardan bir tanesi olarak örnek gösterilmektedir. Böylece su kalitesini etkileyen

parametrelerin düzenli olarak ölçülmesi, süreklilik göstermesi ve verilerin saklanması yapılacak olan çalışmalarda araştırmacıların işlerine önemli derecede kolaylıklar sağlayacaktır.

Bu tez kapsamında Kayseri Büyükşehir Belediyesi Su ve Kanalizasyon İdaresi'nden temin edilen 2015-2022 yıllarına ait ölçümü yapılmış merkez su kalite kontrol raporlarına uygun veriler ile 'Water Quality Index' (WQI) değerinin modellenmesi yapılmıştır. Modelleme giriş verileri: pH, toplam siyanür, sülfat, nitrat, sodyum, serbest klor ve iletkenlik iken çıkış verisi olarak su kalitesi indeksi (WQI) değerleri kullanılmıştır. Modelleme yapılırken ise yapay zekâ teknikleri olarak Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS), Yapay Sinir Ağları (ANN), Gauss Süreç Regresyon (GPR), Mars Gezgin Model (MARS), Destek Vektör Makinesi (SVM), Araştırılmış Regresyon Topluluğu Yöntemi (LSBooST),Yakın Komşu Sınıflandırılması Modeli (KNN), En Küçük Kareler Destek Vektör Makinesi (LSSVM) ve M5 Model Ağacı Yöntemi (M5TREE) kullanılmıştır. Kullanılan yöntemler ile su kalitesi parametrelerin tahmini üzerinde durulup sonuçlar değerlendirilmiştir.

1. BÖLÜM

GENEL BİLGİLER VE LİTERATÜR ÇALIŞMASI

1.1. Problem Durumu

Nüfusun artmasıyla beraber hızla gelişen sanayileşme çevre kirliliğini de beraberinde getirmektedir. Artan sanayileşme çeşitli atık suları içeren kirleticiler oluşturmakta ve doğal su ortamlarını önemli ölçüde kirletmektedir. Doğrudan veya dolaylı bir şekilde insan tüketimine sunulan bu sular insanlarda salgın hastalıklara yakalanmasına zemin hazırlarken sucul alanlarda yaşayan canlıların doğal dengesinin bozulma riskini de artırmaktadır. Öte yandan küresel ısınmanın da etkisiyle üst seviyelere çıkan sıcaklık değerleri mevsimlerin kurak geçmesine neden olmuş su kaynaklarımızın azalmasını tetiklemektedir.

Türk Standartları Enstitüsü tarafından Nisan 1965’de çıkarılan ‘Türk İçme Suyu Standartları’ (TS 266), 1997 yılında revize edilerek daha kapsamlı hale gelmiştir. Giderek azalan ve kirlenen su kaynaklarını kontrol altına almak amacıyla içme suyu standartlarına göre tespit edilen parametrelerin devamlı olarak izlenilip ölçümleri yapılmaktadır. Ölçümleri yapılan parametrelerin su kalitesine uygunluğu tespit edilerek güvenli bir şekilde kullanımı sağlanmaktadır [3].

Günümüzde su kaynakları çeşitli kullanımlar için tüketilmektedir. Bunlardan bazıları sanayileşme, kentsel atıklar, tarımsal faaliyetler, taşımacılık ve insan aktiviteleri gibi faaliyetler su tüketim alanlarına örnek gösterilmektedir. Tüketim oranı artan su kaynaklarının bilinçsizce kullanılmasının yanı sıra su kaynakları üzerinde kirlilik oluşumunu da artırmakta ve su kalitesini etkilemektedir. Bu faktörler su kalitesi üzerinde olumsuz etkiler oluşturmaktadır. Su kalitesinde meydana gelen bozulmalar dünya genelinde ciddi sorun haline gelmektedir.

Araştırmacılar, su kalitesi hakkında daha detaylı ve güvenilir bilgiyi bulmak ve su kalitesinde önem arz eden parametreleri yakından takip etmek için deneysel, gözlemsel, bilgisayar destekli veri analizleri gibi çeşitli yöntemler kullanmaktadırlar.

1.2. Araştırmanın Amacı

Sucul sistemlerde su kalitesi parametrelerinin modellenmesi son derece önem arz etmektedir. Su kalitesini, suyun fiziksel, kimyasal ve biyolojik özellikleri dikkate alınarak değerlendirilmektedir. Değerlendirilen bu

Dünya üzerinde düzensiz dağılan su kaynakları bölgelere göre değişkenlik ve farklılık göstermektedir. Bu durum saha verilerinde sapmalar yaşanmasına ve araştırmacıların işini zorlaştırmasına neden olmaktadır. Bu yüzden veri işlemek için kullanılan geleneksel yöntemler, hem çalışmaların bütünlüğünün sağlanması hem de kısa sürede sonuçlara ulaşmanın gecikmesini zorlaştırdığından dolayı su kalitesini belirlemek için yeterince verimli olmayan yöntemler olarak değerlendirilmektedir. Su kalitesiyle ilgili sorunların çözülmesi için daha farklı yöntemler ve daha fazla çabalar gösterilmelidir. Kullanılan ek çabalar bulguların tutarlılığının iyileşmesine yardımcı olacaktır. Böylelikle düşük kaliteli su tüketiminin önüne geçilmesi beklenmektedir.

Sahadan alınan numuneler deneysel yöntemlerle ölçümleri yapıldıktan sonra elde edilen veriler istatistiksel modellemeleri yapılarak genel kurallar çıkarılmaktadır. Analiz tekniklerinin ve varsayımların doğrulanmasının yanı sıra istatistiksel modelleme ve değerlendirme titiz bir seçim gerektirmektedir. Bu tür modellemelerin çoğu karmaşık yapıya sahiptir ve analizini gerçekleştirmek için önemli miktarda saha verisi gerektirmektedir. Buna ek olarak, su kalitesi parametreleri çeşitli etkenlere bağlı olarak etkilenmektedir. Veri için kullanılan geleneksel yöntemler kullanmak artık bu sorunları çözmede yetersiz kalmaktadır. Son zamanlarda gelişen teknolojinin de beraberinde getirdiği yapay zekâ teknikleri istatistiksel yaklaşımlar, yöntemleri uzaktan algılama, sayısal verilerin kullanımı ve tahmin edilmesi gibi yöntemleri kullanmak için çeşitli algoritmalarından oluşmaktadır. Doğrusal olmayan veya eksik verilerin modellenmesi için etkili bir yöntem olarak kullanılmaktadır. Böylece büyük çapta problemleri çözerek insanlar tarafından eksik kalan veri miktarlarının karar verme sürecini kolaylaştırmaktadır.

Su kalitesi ölçüm yöntemi, verilerin değerlendirilmesi ve sentezlenmesi zor olan büyük veri kümelerinin toplanması ve analiz edilmesi gerekmektedir. Bu nedenle su kalitesini değerlendirmek için çeşitli teknikler oluşturulmuştur. Su Kalitesi İndeksi (WQI) modeli bu tekniklerden biridir. Su Kalitesi İndeksi (WQI), ölçümü yapılmış tüm su kalitesi parametreleri kullanarak birimsiz bir değere dönüştürülmesi için kullanıcılara bilgi veren matematiksel hesap yöntemidir. Çeşitli kimyasal ve organik değişkenlere dayalı olarak su kütlelerinin kalitesinin değerlendirilmesi ve sınıflandırılması için bir gösterge olarak kullanılmaktadır.

Araştırmanın temel amacı Kayseri Büyükşehir Belediyesi Su ve Kanalizasyon İdaresi (KASKİ) Genel Müdürlüğü tarafından temin edilen içme suyu kalitesi parametreleri, modelleme verisi olarak kullanılmıştır. Modellemede kullanılan insani tüketime uygun içme suyu verileri 2015-2022 yıllarına ait olup aylık dönemlerde gözlemi yapılmıştır. İçme suyu amacına yönelik değerlendirilen su kalitesi parametreleri yönetmelik ve standartlarla kullanım alanlarına göre sınır değerleri tespit edilmiştir. Yılın bazı aylarında eksiklik görülmesi ve su kalitesinin insan yaşamı için önemli bir yere sahip olmasından dolayı gelecekle ilgili tahmin edilmesi öngörülmüştür. Yapılan işlemler su kalite indeksi (WQI) ve su kalite parametrelerinin (ph, toplam siyanür, sülfat, nitrat, sodyum, serbest klor) tahmin edilmesinde ‘Bulanık Çıkarım Sistemi’ (ANFIS) ve çeşitli yapay zekâ modellemeleri kullanılmıştır. Bu işlemler ABD merkezli MathWorks firması tarafından geliştirilen pozitif bilim hesaplamaları ve mühendislik uygulamaları için kullanılan çok paradigmatlı bir dili olan MATLAB programı vasıtasıyla modellenmiştir. Modellenen veriler sonraki yıllarda içme suyu kalitesine uygunluğunun tahmin edilmesi amaçlanmaktadır.

1.3. Araştırmanın Önemi

Su kalitesi değerlendirme ölçütleri olarak suyun fiziksel, kimyasal, biyolojik doğası, insan etkileri üzerindeki kullanım esasları, çevresel etkilenme ve suyun kirlilik durumunun belirlenmesi gibi insan sağlığı üzerindeki etkisine göre analiz sonuçlarının uygunluğuna karar verilmektedir. Su kalitesinin değerlendirilmesi, suyun mevcut durumunu tanımlamak, kalitesini etkileyecek parametrelerin birbiri arasında sebep-sonuç ilişkileri kurarak bilgilerin izlenimi kolaylaştırmaktadır. İzlenimi yapılan parametreler yönetmelik esaslarına göre raporlanması gerçekleştirilmektedir.

Birçok ülkede içme suyu kalitesi parametrelerinin sınır değerleri Dünya Sağlık Örgütü (WHO)'nın standartlarına göre saptanmıştır. (WHO,1984,1993) Ülkemizde ise TS266 standartlarına göre su kalitesi parametrelerinin sınır değerleri belirlenmektedir. Bu tez kapsamında elde edilen veriler Türkiye'nin Kayseri ili merkez idari sınırları içerisinde ölçümü yapılmış gözlemler ve deneyler İnsani Tüketim Amaçlı Sular Hakkında Yönetmelik esaslarına dayanmaktadır. Yönetmeliğe göre içme suyu jeolojik koşullara uygun kendi kaynağında doğal olarak oluşan bir çıkış noktası bulup oradan sürekli olarak akan veya teknik usullerle çıkarılan ve bağlı olduğu kurum esaslarına göre uygun görülen dezenfeksiyon, saflaştırma, çöktürme ve filtrasyon gibi işlemler uygulayarak parametrelerin elde edilmesini ambalajlanarak piyasaya sunması esasına dayanmaktadır.

Sahadaki çalışmaları kolaylaştırmak, hayatımıza giren artan teknolojinin etkisi ile beraber mühendislik çalışmalarında büyük hız kazanan yapay zekâ tekniklerini de kullanarak çeşitli iyileştirmeler ve tahminler yapılmaktadır. Bununla birlikte su kalitesi parametrelerinin karmaşık ve büyük ölçekli sonuçlarının incelenmesinin yetersiz kaldığı verileri başarılı bir şekilde elde edebilmek amacıyla araştırmacıların işini kolaylaştıran çeşitli makine öğrenimi yöntemi kullanılmıştır. Yapay zekânın bir dalı olan makine öğreniminde ki algoritmalar verileri analiz etmek ve verilerdeki potansiyel modellemeleri ortaya çıkarmak için tercih edilmektedir. Su kalitesi analizi, değerlendirilmesi ve tahmini için elde edilmiş somut verileri kullanılarak makine öğrenimi sayesinde karmaşıklıkların giderilmesi amaçlanmaktadır.

1.4. Literatür Taraması

Geçmiş dönemdeki bazı çalışmalar bu tezin yazılmasında, veri analizlerinin nasıl ve hangi yöntemler kullanarak modellenmesinde fikir olmuştur. Bu çalışmalardan bazıları ise aşağıda verilmiştir.

Tarımsal ve endüstriyel ihtiyaçları karşılamak için kullanılan su kaynaklarından birisi de nehirler olarak kabul edilmektedir. Nehirler kolay ulaşılabilir olmasından kaynaklanan, dinamik yapısı ve atıkları bertaraf etmesinden dolayı kirlenmektedir. Bu amaçla Johor Nehri'nde çeşitli kalkınma ve insan faaliyetleri nedeniyle yüzey suyu kalitesinin değerlendirilmesi ve tahmin edilmesi ve buna ek olarak nehir havzalarının etkin yönetiminde gerekli olan kirlilik seviyesini korumak için su kalitesi

parametrelerinin modellenmesi Najah Ahmed A, Binti Othman F, Abdulmohsin Afan H et al. (2019) makine öğrenimi yoluyla modellemesi yapılmıştır. Yaptıkları çalışma kapsamında Bulanık Çıkarım Sistemi (ANFIS), Radyal Tabanlı Fonksiyon Sinir Ağları (RBF-ANN) ve Çok Katmanlı Sinir Ağları (MLP-ANN) gibi farklı modelleme yaklaşımlarını kullanmışlardır. Fakat deneylerden ve izleme istasyonlarından elde edilen veriler gürültünün varlığı nedeniyle sistematik hatalar vermiş doğru tahmin yapılmasını güçleştirmiştir. Bu durumu önlemek Nöro-Bulanık Çıkarım Sistemi (WDT-ANFIS) tabanlı sinir ağı kullanımını artırmışlardır. Böylelikle su kalitesinin geçmiş verilerine dayanan dalgacık gürültü giderme tekniği işe yaramıştır. Modelleme verilerinde ise su kalitesi parametresi olarak pH, amonyak azotu (AN) ve askıda katı madde (AKM) kullanmışlardır. Model üzerindeki etkileri değerlendirmek için ise üç değerlendirme süreci kullanmışlardır. İlk değerlendirme sürecinde her girdi parametresinin önemini belirleyen sinir sisteminin bölünmesine yardımcı olan ağ bağlantı ağırlıkları üzerinde durulmuşlardır. İkinci ve üçüncü değerlendirme süreçlerinde ise birinci senaryo olarak su kalitesi için tahmin modeli oluşturmaktadırlar. Tüm su kalitesi parametrelerinde %0,5-%5 aralığında iyileşme olduğunu gözlemlemişlerdir. ($R^2 \geq 0,9$) İkinci senaryo olarak ise aynı parametrelerin her istasyondaki değerlerine dayalı olarak bir tahmin modelini geliştirmişlerdir. Senaryo ikinin senaryo üçe göre daha yeterli olduğunu gözlemlemişlerdir. Böylece önerilen modelin doğrulanması sonucunda tüm verileri tatmin edici bir şekilde tahmin ettiğini görmüşlerdir [4].

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H. & Ye, L. 2022, yaptıkları çalışma kapsamında artan nüfusun beraberinde getirdiği doğal su kaynaklarındaki kirliliği önlemek amacıyla su kalitesi analizi ve değerlendirmesi yapmışlardır. Bu çalışmada, içme suyu, atık su, yeraltı suyu, yüzey suyu, deniz ve tatlı sulardaki su kalitesi farklı özelliklere sahip oldukları için verilerin derlenmesinde zorlukların gözlenmektedir. Bu durumu önlemek amacıyla makine öğrenmesini kullanarak farklı algoritmaların karşılaştırılmasını yapmışlardır. Kullanılan algoritmalar ise şu şekilde özetlenebilir: SVM, Destek Vektör Makinesi; RF, Rastgele Oran; YSA, Yapay Sinir Ağı; PCA, Temel Bileşen Analizi; XGBoost, Aşırı Gradyan Artırma; SOM, Kendi Kendine Organize Hareket Eden Harita; DT, Karar Ağacı; DO, Çözünmüş Oksijen; LSTM, Uzun Kısa Süreli Bellek; WQI, Su Kalitesi İndeksi; KOİ, Kimyasal Oksijen İhtiyacı gibi çeşitli algoritmaları kullanarak su kirliliği kontrolü, su kalitesinde iyileştirmenin yapılması, su

arıtma ve inşası, izlenmesi, değerlendirilmesi süreçlerinde su ortamlarının gelecekteki durumları hakkında tahmin edilmesine olanak sağlamışlardır [5].

Zamili, H., Bakan, G., Zubaidi, S. L., & Alawsi, M. A., 2023 yılında yazmış olduğu yüksek lisans tezi kapsamında Türkiye'nin Çorum ili sınırları içinde olan Yeşilırmak nehrindeki su kalitesi parametreleri ile Su Kalite İndeksi (WQI) tahmin etmek ve buna ek olarak çevresel felaketleri önceden önlemek amacıyla erken uyarıların saptanması için makine öğrenimi modelleri oluşturmuştur. Model parametresi olarak su kalitesinin on bir fizikokimyasal değerleri kullanmıştır. Kullanılan parametreler pH, biyokimyasal oksijen ihtiyacı (BOI), çözünmüş oksijen (ÇO), elektriksel iletkenlik (EI), kalsiyum (Ca^{+2}), sodyum (Na^{+1}), magnezyum (Mg^{+2}), klorür (Cl^{-1}), nitrit (NO_2^{-1}), sülfat (SO_4^{-2}) ve toplam çözünmüş katılar (TDS) olup su kalite indeksi (WQI) değerini hesaplamak için kullanılmıştır. Ayrıca kısıtlama katsayısına dayalı parçacık sürüsü optimizasyonu ve kaotik yerçekimi arama algoritması (CPSOCSGA), parçacık sürüsü optimizasyonu (PSO) ve deniz avcısının optimizasyon algoritması (MPA) Levenberg-Marquardt üç katmanlı bir yapay sinir ağı (YSA) ile birleştirilerek bazı yerlerdeki içme suyu için uygun olmayan noktaların optimizasyonunu yapmıştır. Yapılan modellemelerde istatistiksel doğrulama değerleri olarak $\text{MAE}=0,01627$, $\text{RMSE}=0,0187$ ve $\text{R}^2=0,965$ tespit edilmiş olup modellemede CPSOCSGA-ANN performansının MPA-ANN ve PSO-ANN performanslarına göre daha üstün olduğunu ortaya koymuştur [6].

Shamsuddin, I. I. S., Othman, Z., & Sani, N. S., 2022, son yıllarda artan su kirliliği su kalitesine zarar verdiği için su kalitesi parametrelerinin izlenmesi daha hassas ve pratik ölçüm yapmayı gerektirmektedir bundan dolayı da maliyeti artırmaktadır. Maliyeti azaltmak, su kirliliğini kontrol etmek, etkili su kalitesi için doğru bir şekilde tahminin edilmesini sağlamak, su kalitesi tahmini için çeşitli modeller oluşturmak ve kötü su kalitesinin tespiti durumunda tüketicileri bilgilendirmek amacıyla makine öğreniminden faydalanmışlardır. Langat nehir havzası suyu verilerini sınıflandırmak için yapay sinir ağı (YSA), Karar Ağaçları (DT) ve Destek Vektör Makineleri (DVM) kullanarak üç makine öğrenimi modeli geliştirmişlerdir. Bu üç model arasında karşılaştırmalı performans analizine göre DVM'nin en uygun olduğunu göstermişlerdir. DVM, DT, YSA modelleri arasında ise 0,05 güven düzeyinde performans farkı olduğunu göstermişlerdir [7].

Ay M., 2014, yapmış olduğu çalışmada karmaşık ve doğrusal olmayan, farklı akarsulardan ölçülen su kalitesi ve debi parametrelerinin modellenmesi için klasik denklemlerin aksine yapay zeka tekniklerini kullanarak bu konu üzerindeki yaklaşımını değerlendirmiştir. Bu çalışma kapsamında radyal tabanlı yapay sinir ağı, çok katmanlı yapay sinir ağı, uyarlamalı bulanık sinir ağı çıkarım sistemi, c- ortalamalar ve k- ortalamalar yöntemlerini kullanmıştır. Her bir yöntemi ayrı ayrı bütünleştirerek çözünmüş oksijen, kimyasal oksijen ihtiyacı ve toplam tuz yoğunlaşma değerlerini modellemiştir. Elde edilen sonuçları klasik istatistik yönteminden olan çoklu lineer regresyon ve otoregresif hareketli ortalamalar yöntemleri ile karşılaştırmıştır. Karşılaştırma sonucuna göre yapay zekâ tekniklerinin klasik yöntemlere göre su kalitesi parametrelerinin modellenmesinde daha başarılı olduğunu ortaya koymuştur. Buna ek olarak bütünleştirilmiş modellerin diğer modellere göre tahmin performansının daha güçlü olduğunu belirlemiştir [8].

Haghiabi, A. H., Nasrolahi, A. H., & Parsaie, A., 2018, İran'ın güneybatısında yer alan çok yıllık ana akarsulardan birisi olan Tireh Nehri'nin aşama deşarjını belirlemek için su kalitesi bileşenlerini izlemek ve ilişkilendirmek için Yapay Sinir Ağı (YSA), Grup Veri İşleme Yöntemi (GMDH) ve Destek Vektör Makinesi (DVM) tekniklerini kullanmışlardır. Ölçüm parametreleri olarak sıcaklık (T), özgül iletkenlik (EC), pH, klorür (Cl), toplam çözünmüş katılar (TDS), sodyum (Na^{+1}), kalsiyum (Ca^{+2}), magnezyum (Mg^{+2}), sülfat (SO_4^{-2}) ve bikarbonat (HCO_3^{-1}) kullanmışlardır. YSA ve DVM farklı tipte transfer ve çekirdek fonksiyonları ile test edildiğinde sonuçların her iki modelde de tahmin için uygun performansa sahip olduğu görülmüştür. Kullanılan teknikler karşılaştırıldığında YSA'nın tansig modelinin test edilen transfer fonksiyonunun en iyi performans değerini verdiğini görürken GMDH modellerinin en düşük doğruluk oranı olduğunu göstermişlerdir [9].

Özeren Y., 2022, yapmış olduğu yüksek lisans kapsamında Sakarya Havzasına ait 1995-2014 yılları arasındaki beş ay ölçümü yapılmış (Şubat, Nisan, Haziran, Ağustos ve Kasım) su kalitesine ait parametreler ile çözünmüş oksijen (ÇO) değerinin Yapay Sinir Ağları (YSA), Destek Vektör Makinesi (DVM), Derin Öğrenme (DÖ) ve Gauss Süreç Regresyon (GSR) yöntemleri ile modellemesini yapmıştır. Giriş verisi olarak beş parametre kullanmıştır bunlar ise aylar, pH, sıcaklık (T), biyolojik oksijen ihtiyacı (BOİ) ve elektriksel iletkenlik (Eİ)'dir. Çıkış verisi olarak ise çözünmüş oksijen (ÇO) kullanmıştır.

Oluşturulan modeller ile Taylor ve Violin diyagramları yoluyla ölçülmüş veriler modellerden elde edilen değerler ile karşılaştırılarak birbirine ne kadar yakın olduğunu incelemiştir. Buna ek olarak kullanılan yöntemlerin etkinliği Kruskal-Wallis (KW) testi ile desteklenmiştir. Modellerin tahmin performanslarını ölçmek için ise Ortalama Mutlak Hata (OMH), Kök Ortalama Karesel Hata (KOKH) ve Nash-Sutcliffe verimlilik katsayısı (NSE) ve Determinasyon Katsayısı (R^2) değerleri göz önünde bulundurulmuştur. Yapılan modellerde en iyi sonucu YSA modeli vermiştir [10].

Altuntaş M., 2018, Devlet Su İşleri'nden (DSİ) alınan Yeşilirmak Havzasına ait 1995-2015 yılları arasındaki yirmi farklı gözlem istasyonunun su kalitesi parametrelerinin modellemesini yapmıştır. Yılların şubat, nisan, haziran, ağustos ve kasım aylarına ait verileriyle beraber çözünmüş oksijen yoğunluğunu modellerde kullanmıştır. Modellerde yapay zekâ tekniği olan, Uyarlamalı Bulanık Sinir Ağı Çıkarım Sistemi (ANFIS) Ve Çok Katmanlı Yapay Sinir Ağı (ÇKYSA) kullanmıştır. Modelleme verisi olarak aylar, pH, sıcaklık (T), bulanıklık (TH), elektriksel iletkenlik (Eİ) ve çözünmüş oksijen (ÇO) parametrelerini kullanmıştır. Her iki modellemede beş değişkene kadar birçok kombinasyon denemesi yapmıştır. Yapılan analizler kıyaslandığında ikisinin de sonuçlarının çok yakın olduğu gözlemlenmiş birbirlerine karşı üstün olmadığını sonucuna varmıştır. Ayrıca ANFIS metodunda en iyi analiz sonucunu, aylar- sıcaklık (T) - elektriksel iletkenlik (Eİ)- bulanıklık (TH) parametreleri verirken ÇKYSA metodunda ise en iyi analiz sonucunu, aylar- sıcaklık (T)- elektriksel iletkenlik (Eİ) bağımsız değişkenlerinde ortaya çıkarmıştır [11].

Dede, Ö. T., & Sezer, M., 2017 Giresun iline bağlı içme suyu kaynağı olan, Aksu Çayı üzerinde bulunan üç farklı istasyon noktasından gözlemlenen 1 yıllık su kalitesi parametresi değerini incelemiştir. Trabzon Devlet Su İşleri Genel Müdürlüğü Su Analiz Laboratuvarında incelenen 36 adet su kalitesi parametresi analizi sırasında çözünmüş oksijen, T.coli, E.coli ve fenol değerlerinin su kalitesi standartlarının uygunluğu dışında kalıp ana kirleticiler olarak saptanmıştır. Buna ek olarak renk, bulanıklık, mangan, demir, alüminyum, cıva ve amonyak azotunun da sınır değerlerini aştığı görülmüştür. Bu amaçla verilerin anlaşılabilirliğinin sağlanması ve yorumlanması, su kalitesini etkileyen parametrelerin belirlenmesi için Kanada Su Kalitesi İndeks Modeli (CWQI) uygulamıştır. İstasyon değerleri için CWQI değerleri zayıf kategorisinde

değerlendirilmiş hesaplanan indeks değerinin mantıklı sonuçlar verdiğini saptamışlardır [12].

Özen S., 2018, Kayseri Su ve Kanalizasyon İdaresi (KASKİ) merkez ileri biyolojik atık su arıtma tesisine ait 2011-2014 yılları arasındaki kimyasal oksijen oranı (KOİ) ve biyolojik oksijen ihtiyacı (BOİ) değişkenleri kullanılarak atık su arıtma verilerinin tahmin edilmesi çalışmasını yürütmüştür. KOİ değişkeni tahmininde ölçülen pH, akış debisi (Q), BOİ, azot, askıda katı maddeler (AKM) değerleri kullanılırken BOİ değişkeni tahmininde pH, fosfor (P), akış debisi (Q) , KOİ, askıda katı maddeler (AKM) bağımsız değişken olarak kullanmıştır. KOİ ve BOİ tahmininde Yapay Sinir Ağı (YSA) ve Adaptif Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS) yöntemlerini kullanmıştır. Yapay zekâ metotları ile analiz edilmeden önce elde olan verileri eğitim ve test datası şeklinde iki ayrı küme şeklinde gruplandırmıştır. Toplamda 724 verinin %75'i yani 543 adet eğitim verisi olarak kullanılırken kalan %25'i yani 181 adeti test verisi şeklinde kullanmıştır. Karakök Ortalama Karesel Hata (KOKH), En Düşük Ortalama Karesel Hata (OMH) istatistiksel doğrulama yöntemleri kullanılarak BOİ değişkeninin tahminde bu verileri karşılaştırmıştır. Tahmin sonucuna göre YSA ve ANFIS modellerinin sonuçlarının birbirine çok yakın çıktığını gözlemlemiştir. Sonuçların determinasyon katsayısına göre değerlendirildiğinde istenilen durumda olmadığına sonucuna varmış bu amaçla da ikinci aşama olarak ise KOİ, BOİ ve AKM değişkenlerin birbirleri arasındaki korelasyonun 0,5 den diğer parametreler arasındaki korelasyon 0,5'den daha düşük çıkmıştır. Bu durumu çözmek için de Excel- çözücü yardımıyla matematiksel eşitlik geliştirmiştir. Buna göre zaman ve hata kriterlerine göre yakın sonuçlar vermiştir. Uygulanabilirlik açısından da OMH ve KOKH kriterlerine göre KOİ ve BOİ değerlerinin tahmin edilmesini tercih etmiştir [13].

Najah Ahmed A, Binti Othman F, Abdulmohsin Afan H et al. 2019, Yapılan çalışma kapsamında Malezya'nın üçüncü büyük bölgesi olarak kabul edilen Johor eyaletinde bulunan beş ana nehirde sadece Sungai nehir havasından alınan su kalitesine etki eden veriler kullanılmıştır. Johor nehri ve kollarının su kaynağı olarak önemli bir rol oynamasının yanı sıra çeşitli kalkınma ve insan faaliyetlerinden dolayı nehirde bozulmalar olmuştur. Geleneksel modelleme yöntemlerinin büyük miktarda bilinmeyen ve eksik kalmış veri setlerinin tamamlanmasında yetersiz kaldığı gözlemlenmiştir. Bu amaçla sucul sistemlerde su kalitesi parametrelerinin modellenmesi büyük önem

taşımaktadır. Su kaynaklarının daha iyi yönetimi için su kalitesi parametreleri yapay zekâ aracılığıyla modellenmiştir. Farklı modelleme yaklaşımlarında ise, Adaptif Nöro-Bulanık Çıkarım Sistemi (ANFIS), Radyal Tabanlı Fonksiyon Sinir Ağları (RBF-ANN) ve Çok Katmanlı Perceptron Sinir Ağları (MLP-ANN) kullanmışlardır. Yapay Sinir Ağı (YSA) modellemesi için sıcaklık (Temp), elektrik iletkenlik (COND), tuzluluk (SAL), nitrat (NO_3), bulanıklık (TURB), fosfat (PO_4), klorür (Cl), potasyum (K), sodyum (Na), magnezyum (Mg), demir (Fe) ve Escherichia coli (E-coli) olmak üzere toplamda on iki su kalitesi parametrelerini kullanmışlardır. Temel istatistiksel değerlerinden ortalama, minimum, maksimum standart sapma (S.D.) ve girdi-çıkı parametrelerinin varyasyon katsayısı (CV) verilere ilişkin dağılım için bir ölçü olarak hesaplanmıştır. Fakat hem giriş hem de çıkış parametrelerinin konsantrasyon seviyelerinde büyük değişiklikler görülmüş yüksek varyasyon katsayısı değerine ulaşılmıştır. (Amonyak Azotu için %254,94 ve E-coli için %325,96) Farklılıklarının yaşanmasının nedeni olarak nehir havzasının geniş coğrafi alanı kaplaması ve birçok kollarından arıtılmamış atık suların nehir suyuna akıtan kanallar olması ve iklime ilişkin mevsimsel değişimlerin yaşanması etkilemiştir. Bu durumda su kalitesinin geçmiş verilerine dayanarak dalgacık gürültü giderme tekniği önermişlerdir. Nöro – Bulanık Çıkarım sistemlerinde kullanılan teknikler (WDT-ANFIS) geçmiş su kalitesi parametre verileri (2009-2010) ile karşılaştırmışlardır. Karşılaştırılan parametreler pH, askıda katı madde (AKM) ve amonyak azotunu (AN) içerirken üç kısımda değerlendirilmiştir. İlk değerlendirme süreci sinir sisteminin bölünmesine bağlı iken ikinci ve üçüncü değerlendirme süreçlerinde ise tek bir parametre ve bir parametre birleşimlerini oluşturarak en etkili girdiyi tespit etmek için kullanmışlardır. Bu işlemler sırasında iki senaryo oluşturmuşlardır. Senaryo 1, su kalitesi için bir tahmin modeli oluştururken Senaryo 2 de ise aynı parametrelerin her istasyondaki değerlerine dayalı bir tahmin modelini geliştirmek amacıyla kullanmışlardır. Her iki senaryoda on iki girdi parametrelerini kullanmışlardır. Senaryo 1'e göre tüm su kalitesi parametreleri için %0,5 - %5 aralığında önemli bir iyileşme olduğu gözlemlenmiş olup önerilen modelin tüm veriler için tatmin edici bir şekilde su kalitesi parametrelerinin tahmin edilmesini ortaya koymuşlardır. (su kalitesi parametreleri R^2 değerleri 0,9'a eşit veya daha büyüktür.) [14].

Acar, R., & Saplıoğlu, K, 2022, çalışmalarında Fırat Havzası üzerinde yer alan 2119 numaralı Kemahboğazı akım gözlem istasyonundan elde edilen 1995-2002 yıllarına ait su sertliği verilerini kullanmışlardır. Su sertliğinin tahmin edilmesi amacı ile sıcaklık (T),

pH, elektriksel iletkenlik (EC), sodyum (Na^+), potasyum (K^+), karbonat (CO_3)⁻², sülfat (SO_4)⁻², bikarbonat (HCO_3)⁻, klor (Cl)⁻ ve su miktarı (SM) modellemelerde girdi parametreleri olarak kullanılırken su sertliği (SS) çıktı parametresi olarak kullanılmışlardır. Temin edilen bu parametrelerle ANFIS modelleri oluşturulması amaçlanmaktadır fakat girdi parametrelerinin sayıca çok olması model kombinasyonu oluşumunun sayısını artırmakta ve en iyi modelin tespit edilmesini zorlaştırmıştır. Bu zorluğa çözüm üretebilmek amacıyla ANFIS modelinde etkili parametrelerin tespit edilebilmesi için çoklu regresyon modelini kullanmışlardır. Kullandıkları çoklu regresyon modelinde ise klor (Cl)⁻, elektriksel iletkenlik (EC), bikarbonat (HCO_3)⁻ ve sülfat (SO_4)⁻² parametreleri girdi parametreleri olarak kullanılmış Düzeltilmiş R^2 değerindeki değişimler incelenerek farklı kombinasyon kümeleri ANFIS modellemeleri ile oluşturulmuştur. Elde edilen eğitim ve test dataları için Wilcoxon testi, R^2 ve ağırlıklı karesel hata değerleri kullanılarak su sertliğinin tahmin edilmesinde elde edilen modellemeleri göstermişlerdir [15].

Şemsüddin, I. I. S., Osman, Z. ve Sani, N. S., 2022, geleneksel yöntemlerle su kalitesi değerlendirilmesi uzun gözlemler sonucu ölçüldüğünden ve pahalı laboratuvar deneyleri gerektirdiğinden dolayı hem su kirliliğini kontrol altına almak hem de zamandan ve maliyetten tasarruf sağlamak amacıyla gelişen makine öğrenimi yöntemini kullanarak su kalitesi modellemesini yapmışlardır. Böylece kullanım kalitesi düşük suların kalitesini tespit edip tüketicileri bilgilendirerek etkili su yönetim çalışmasını yürütmeyi amaçlamışlardır. Makine öğrenimi yönteminde kullanılacak parametreleri Malezya Çevre Bakanlığı Langat Nehir Havzası su kalitesi verilerinden temin etmişlerdir. Veri kümesi on dört izlemem istasyonundan alınan toplam 560 kayıt ve 46 özniteliğe sahip olup Ocak 2012'den Aralık 2016'ya kadar toplam beş yıllık veri kümesini almışlardır. Aldıkları veri setlerini Yapay Sinir Ağı (YSA), Destek Vektör Makineleri (DVM) ve Karar Ağaçları (DT) modellerini geliştirilerek üç makine öğrenimi yönteminde kullanmışlardır. Model girdisi olarak altı su kalitesi parametresi (pH, AN, SS, COD, BOD ve DO), çıktı parametresi olarak Su Kalite İndeks değeri (WQI) kullanmışlardır. Bu modellerin performans analizleri karşılaştırıldığında ise 0,05 güven düzeyinde performans farkı elde edilerek en uygun modelin DVM olduğu sonucuna varmışlardır. Yapılan analizler sonucunda YSA diğer geleneksel yöntemlere kıyasla daha yüksek

performanslı doğruluk oranı verirken, DVM modelinin YSA ve DT'ye göre daha iyi parametreleri sınıflandırmada etkili bir şekilde kullanılabileceğini göstermişlerdir [16].

Haghiabi, A. H., Nasrolahi, A. H. ve Parsaie, A., 2018, İran'ın güneybatısında yer alan Tireh Nehri'nin su kalitesi parametrelerini tahmin etmek için Yapay Sinir Ağı (YSA), Destek Vektör Makinesi (DVM) ve Grup Veri İşleme Yöntemi (GMDH) tekniklerini kullanarak performans analizlerini yapmışlardır. YSA modelinin tasarımında kullandıkları işlem adımları sırasıyla şu şekilde olmaktadır. Veri kümesinin %80'i eğitim %20'si test için bölündükten sonra ilk adımda gizli bir katman oluşturularak giriş özelliklerine göre eşit sayıda nöronlar kullanmışlardır. Farklı performans parametre sonuçları değerlendirilerek transfer fonksiyonlarından en iyilerini seçmişlerdir. YSA ve DVM modellerinde farklı tipte transfer ve çekirdek fonksiyonları test edilerek her iki model içinde uygun tahmin performansına sahip olduğunu göstermişlerdir. Transfer ve Kernel fonksiyonlarında tansig ve RBF test edilen modeller arasında en iyi performansa sahip olmuştur. YSA'nın en iyi performansı (0,92 ve 0,84), Kök Ortalama Kare Hatası (0,238 ve 0,295) olup eğitim ve test aşamalarında tansig fonksiyonu en iyi transfer fonksiyonu olmuştur. GMDH modelinin analiz sonuçlarında ise su kalitesi bileşenlerini tahmin etmede kabul edilebilir bir performansa sahip iken doğruluk oranı YSA ve SVM'ye göre daha az vermektedir. Uygulanan modellerin doğruluğunu değerlendirmek için hata indekslerine bakıldığında SVM'nin en doğru model olduğunu göstermiştir [17].

Uncuoglu, E., Citakoglu, H., Latifoglu, L., Bayram, S., Laman, M., Ilkentapar, M., & Oner, A. A., 2022, inşaat sektöründe hidroloji mühendisliği ve geoteknik mühendisliğine ait inşaat mühendisliği alanlarında karşılaşılan problemleri analiz etmek için makine öğrenmesi (ML) yöntemlerinden yararlanarak nasıl performans gösterdiğini araştırmışlardır. Yapılan çalışmada girdi parametrelerinin çıktı parametreleri üzerindeki etkinliği incelenmiştir. Geoteknik mühendisliğinde giriş parametreleri kazık genişliği (B), minimum sıcaklık (Tmin), dönme derecesi (R) ve akış akışı (Q) olarak girilmiştir. Problemin temsili özelliğinin belirlenmesi için veri setlerinin homojen bir şekilde dağılması test aşamasında elde edilen hata değerlerini etkilenmesine neden olmaktadır. Ele alınan problemlerin (proje zamanında tamamlanmalı, proje bütçesi aşılmamalı, öngörülen kalite sağlanmalı) MGGP ve M5Tree modellemesi yapılarak her birinin çıktılarını hesaplamaya yönelik denklemler oluşturmuşlardır. Oluşturulan denklemlere göre geoteknik mühendisliğinde en iyi sonucu YSA modeli vermiştir. Hidroloji

mühendisliği alanında yapılan çalışmalarda ise Türkiye’de 2010-2020 yılları arasında kullanılmış farklı illere ait 496 (160’ı su isale hattı, 111’i sulama sistemi, 80’i tedavi hizmetleri, 78’i baraj ve 40’ı besleme kanalı) su yapısı verilerini Devlet Su İşleri Genel Müdürlüğü (DSİ) tarafından temin etmişlerdir. Veriler Yapay Sinir Ağı (ANN), Gauss Süreci Regresyonu (GPR), Destek Vektör Makinesi (SVMR), Uzun Kısa Süreli Hafıza (LSTM), Çoklu Gen Genetik Programlama (MGGP) ve M5 Model Ağacı (M5Tree) makine öğrenimi yöntemlerini kullanmışlardır. Oluşturulan modellerin performanslarını değerlendirmek için genel Performans İndeksi (OI) hesaplanarak makine öğrenimi modelinin göreceli performanslarını ise Taylor diyagramı, Keman diyagramı ve Tek kuyruklu Wilcoxon İşaretili Sıra testleri kullanılarak değerlendirilmişlerdir. Sonuç olarak doğal ortamda MGGP modeli en iyi tahminleri verirken YSA modelinde ise yöntemsel ve deneysel veri kümeleri kullanıldığında en iyi tahminleri ürettiğini ortaya koymuşlardır. Taylor ve Keman diyagramlarına göre daha geniş açıdan bakıldığında ANN, MGGP ve GPR modelleri daha başarılı sonuçlar vermiştir. Doğal durumlar için de MGGP, M5 Ağacı ve YSA modelleri en iyi tahminleri üretmiştir. Ayrıca tüm modeller Kruskal - Wallis testini sorunlar açısından geçmiş bulunmaktadır [18].

2. BÖLÜM

YÖNTEM VE MATERYAL

İnsan hayatının birçok yerinde var olan su yaşamın devam ettirilebilmesi için son derece önemli bir madde olmuştur. Fakat sınırlı olan su kaynaklarının kötü bir şekilde yönetilmesi, artan nüfusla beraber günlük su tüketiminin artması, endüstriyel faaliyetlerin hız kazanması, tarımsal faaliyetlerde içme suyu kullanımının yaygınlaşması, atık suların su kaynaklarına karışması ve iyi arıtılmaması gibi nedenlerden dolayı su kaynaklarımız giderek kirlenmektedir.

Son zamanlarda yaşanan iklim değişikliği ve kuraklık etkisi küresel ısınmanın göstergesi olarak belirlenirken su kaynaklarımızın tükenmesine sebebiyet vermektedir. Böylece yakın gelecekte ülkemizde su kıtlığının yaşanması öngörülmektedir. Durumun bu denli ciddi olması ve doğrudan su tüketimiyle beraber insan vücuduna giren çeşitli kirletici maddelerin salgın hastalıklara yakalanma riskinin artmasına neden olacaktır. Tüm bu yaşanacak durumları önlemek amacıyla mevcut su kaynaklarının korunması ve su kalitesine etkileyen parametrelerin düzenli bir şekilde ölçülmesi, ölçülen verilerin saklanması bu alanda çalışmalar yürütecek olan araştırmacıların işini kolaylaştırmaktadır.

2.1. Kayseri Su ve Kanalizasyon İdaresi Genel Müdürlüğü

Kayseri Büyükşehir Belediyesi bünyesine bağlı su ve kanalizasyon hizmetlerini üstlenerek içme, kullanma ve endüstri gibi alanlarında kullanılacak olan su ihtiyaçlarını karşılamak için, atık suların toplanmasını ve arıtılmasında öncü bir yol izleyerek kent dışı kalan su kaynaklarını korumasını ve muhafaza edilmesi amacıyla yapılan hizmetlerin devamı niteliğinde gerekli olan tesisleri kurmak ve kurulu olan tesisleri devralarak tek elden işletilmesinde görev alarak Bakanlar Kurulunun 89/14886 sayılı kararı ile 1989 yılında Kayseri Su ve Kanalizasyon İdaresi Genel Müdürlüğü (KASKİ) kurulmuştur.

Yaklaşık olarak 1 milyon 441 bin 523 kişiye çeşitli su yönetimi ve temini konusunda hizmet eden KASKİ toplamda 17 bin 043 km²'lik hizmet alanına sahip olup Kayseri ilinin günlük ortalama 370,000 m³/gün su ihtiyacını karşılamaktadır. Bu amaçla Kayseri iline kesintisiz suyun temin edilmesini sağlamak, kullanılan su kalitesinin içilebilir düzeyde olduğunu tespit etmek ve standartların üzerinde olup olmadığını test etmek için sahada çeşitli gözlem istasyonları kurulmuştur. Yapılan gözlemlerden alınan numuneler ise laboratuvar ortamında test edilerek Çevre ve Şehircilik Bakanlıktan yetkili ve Türkak'tan (Türk Akreditasyon Kurumu) akrediteli olarak analiz sonuçlarını şeffaf bir şekilde yayınlamak su kalite performanslarının uygunluğu belirlenmektedir [19].

Su hizmetlerinde abone memnuniyeti esas alındığında su kaynaklarını daha verimli kullanmak amacıyla gelişen teknolojiyi de beraberinde kullanarak daha kolay ve pratik çalışma alanı oluşturmak, insan ve çevre sağlığını da göz önünde bulundurmak KASKİ için bir görev haline dönüşmüştür. Bu amaçla artan nüfusa bağlı olarak büyüyen ve gelişen kentsel ihtiyaçları karşılayarak mevcut su kaynaklarının su kalitesini iyileştirmek ve alternatif su kaynakları oluşturmak için hazırlanan 2021-2025 stratejik plan kapsamına göre su kalitesi bakımından incelenen içme suyu kaynağı oranı 2020 yılı plan dönemi başlangıç değeri %25 iken 2025 yılı plan dönemi sonu hedeflenen değer %50 olması hedeflenmektedir.

2.1.1. KASKİ Su Kalitesi Verilerinin Analiz Süreci

Kayseri ili içme ve kullanım amaçlı suların hijyen kontrollerinin sağlanması, su kalite parametrelerinin ölçülmesi ve 28 Temmuz 2016 yılında yayınlanan 'İnsani Tüketim Amaçlı Sular Hakkında Yönetmelik' standartları baz alınarak kalite standartlarına göre uygunluğunun test edilmesi, içme ve kaynak sularının elde edilmesi gibi çalışmalar yönetmelik esaslarına göre belirlenmektedir. Bu yönetmeliğin ilkeleri doğrultusunda Kayseri Büyükşehir belediyesi himayesi altındaki KASKİ su ve kanalizasyon amaçlı alt yapı hizmetlerinin inşası, bakımı, onarımı ve işletilme süreçlerini yerine getirmektedir. Yapılan çalışmalar sonunda elde edilen veriler saklanmakta alınan numuneler laboratuvar ortamda deneysel incelemeler sonucunda incelenmektedir.

İçme suyu numune alımı sürecinde ise kuruma gelen şikâyetlerin değerlendirilmesi, özel taleplerin ve birimlerden gelen taleplerin toplanması ve rutin izleme programının

hazırlanması sonucu KASKİ portal üzerinden iş emri üretilerek numune alma programı başlatılmaktadır. Alınan numuneler hem kimyasal analize hem de mikrobiyolojik analize tabii tutulmaktadır. Sorasında serbest klor ölçüm süreci de tamamlandıktan sonra numuneler korunarak kayıt altında tutulmaktadır. Numunelerin laboratuvara teslim edildikten sonra deney numunelerine uygulanan işlemler sonucunda analiz raporları hazırlanmaktadır. Analiz rapor sonuçları incelendikten sonra sonuçların uygunluğu kalite standartlarına göre kontrol edilmektedir. Rutin izleme sonuçlarına uygunluğu ve şikâyet sonucu alınan numune sonuçlarının uygunluğu karşılaştırıldıktan sonra kurumsal zekâyaya veri girişleri yapılmaktadır.

Analiz raporu hazırlama sürecinde ise deney numunelerine uygulanan işlemler sürecinden elde edilen ham veriler derlenerek analiz sonuçları hesaplanmaktadır. Elde edilen sonuçlar sonuç girme sayfası kitlenerek numune verilerinin raporlanması sağlanmaktadır. Raporda görüş ve yorumlar, uygunluk beyanları belirtilerek raporun elektronik yolla ya da basılı olarak müşteriye iletimi sağlanmaktadır.

Deneyisel incelemelere başlamadan önce sahadan alınan numuneler laboratuvar personeline teslim edilerek numunenin deneye hazırlanması sağlanır. Deneyin laboratuvarda yapılmasından sonra doğruluk kontrolleri yapılmış olan analizleri laboratuvar barkod sistemi olan (GisLab) programına aktarılmaktadır. Bunun sonucunda ise deney numunelerine uygulanan işlemlere ilişkin sayısal veriler kurumsal zekâyaya girilerek verilerin uzun süre sistemde korunaklı bir şekilde kalması sağlanmaktadır.

2.2. Su Kalitesi Parametreleri

İnsan hayatını doğrudan etkileyen su; evsel, tarımsal, endüstriyel vb. olmak üzere çeşitli amaçlar için kullanılmaktadır. Sucul ortamda oluşan kirliliğin artması veya su kalitesinin düşmesi durumunda birçok alan ve insan sağlığı olumsuz yönde etkilenmektedir. Bu durumun önüne geçmek için su kalitesinin ve kullanılan suyun amacına uygunluğunun test edilmesi su kalitesinin belirlenmesin de büyük önem taşımaktadır. Su kalitesi parametreleri ise farklı kullanım alanlarına göre belirlenen yönetmelik ve standart değerlerine göre sınır değerleri taşımaktadır. Gözlem sonucu elde edilen veriler analiz edilmesi sonucunda sınır değerine uygunluğu tarafından karşılaştırılmaktadır. Buna göre

göre karşılaştırılan su kalitesi parametreleri fiziksel, kimyasal ve biyolojik parametreler olmak üzere sınıflandırılmaktadır.

Tablo 1. İçme suyu kalitesi parametre değerlerinin kabul edilebilir maksimum değerleri [20]

PARAMETRELER	Türk Standartları Enstitüsü (TSE266)	Avrupa Birliği (EC)	Dünya Sağlık Teşkilatı (WHO)
Fiziksel Parametreler			
Renk (Co-Pt birimi)	20	20	15
Bulanıklık (NTU birimi)	5	4	5
Tat	Duyusal	Duyusal	Duyusal
Koku	Duyusal	Duyusal	Duyusal
Kimyasal Parametreler			
Ph	6.5 - 9.5	6.5 - 9.5	6.5 - 9.5
İletkenlik ($\mu\text{S}/\text{cm}$)	2500	2500	2500
Nitrit ($\mu\text{g}/\text{L}$)	0.5	0.5	0.5
Nitrat (mg/L)	50	50	50
Demir ($\mu\text{g}/\text{L}$)	0.2	0.2	0.3
Magnezyum	50	50	50
Serbest Klor (mg/L)	0.5	0.5	5
Florür (mg/L)	1.5	1.5	1.5
Sodyum (mg/L)	175	200	200
Nikel ($\mu\text{g}/\text{L}$)	0.02	0.02	0.02
Bor ($\mu\text{g}/\text{L}$)	1	2	2
Sülfat (mg/L)	250	250	250
Mangan ($\mu\text{g}/\text{L}$)	0.05	0.05	0.1
Alüminyum (mg/L)	0.2	0.2	0.2
Amonyum (mg/L)	0.5	0.5	1.5
Biyolojik Parametreler EMS/100 mL			
Koliform	0	0	0
Esherichia Coli	0	0	0
C.perfringers	0	0	0
Pseudomonas Aeruginosa	0	0	0

2.2.1. Fiziksel Parametreler

2.2.1.1. Bulanıklık

Saydam bir yapıya sahip olan su ışık geçirgenliğine bağlı olarak bulanıklık parametresi ölçülmektedir. Yapılan ölçümler suda çözünmüş askıda katı maddelere bağlı olarak ışık geçirgenliğinin sonucu olarak tespit edilmektedir.

Suyun bulanık olması içerisinde çözünen maddelere bağlı olarak renginin bozulmasıdır. İçme sularının günlük hayatta kullanılabilmesi için berrak olması istenmektedir.

2.2.1.2. Sıcaklık

Suyun içilebilir durumda olabilmesi için istenilen sıcaklık değeri 10-12 °C olması beklenmektedir. Su sıcaklığının artması suyun içilebilir olmasını engellerken suyun lezzetsiz olmasına bağlı olarak tercih edilmemesine neden olacaktır.

Sıcaklık değerinin artması suda gerçekleşen biyolojik reaksiyonların gerçekleşme hızını artıracığı için sucul ortamda yaşayan mikroorganizma gelişim hızını artıracaktır. Bundan ötürü suda çözünmüş oksijen (ÇO) miktarının azaltacaktır. Su içindeki askıda katı maddelerin çökme hızı artarak su kalitesini olumsuz yönde etkileyecektir.

2.2.1.3. Tat ve Koku

Suyun içerisinde çözünmüş madde ve organizma sayısındaki artış suya farklı tat ve koku oluşmasına neden olmaktadır. Suda oluşan tat acı, ekşi ve tuzlu olurken suda oluşan koku ise baharatsı, otsu, balıksı, küfümsü gibi kötü kokuları oluşturmaktadır. İçme suyunun kokusuz ve insanların içimini seveceği lezzette olması için suyun kirleticilerden ayrıştırılması gerekmektedir. Bunun için suyun içerisinde bulunan mikroorganizmaların yok edilmesi gerekmektedir. Suyun temizlenmesi için bakır sülfat (CuSO_4) ile birlikte 0,3- 3 mg/ lt dozunda her organizma için farklı miktarda klor kullanılmaktadır. Bakır sülfat (CuSO_4)'ın mikroorganizmaları yok etmede yetersi kaldığı durumda klorlama işlemi takviye görevi görmektedir. Ayrıca klorlama işlemi oluşan kötü kokuların giderilmesinde de kullanılmaktadır.

2.2.1.4. Renk

Şeffaf yapıya sahip olan su; bulanıklığın oluşmasına neden olan askıda katı maddelerin giderilmesi sonucu oluşan renk (hakiki renk), suda çöken veya kolloidal madde bulundurması sonucu oluşan suya filtre ve santrifüj işlemi yapılmaksızın oluşan renk (zahiri renk), metalik iyonları bulundurması sonucu oluşan renk, sanayi atıkları sonucu oluşan renk, humus, turba, yabani otlar ve algler sonucu suda çeşitli renklenme meydana gelmektedir.

Suda oluşan renklenme durumu tamamen tehlikeli ve korkulacak bir durum değilken incelenmesi ve sağlık açısından istenilen seviyede olması gerekmektedir. Çünkü suyun renkli olması durumunda bazı zararlı mikroorganizmaların oluşacağı öngörülmektedir bu durumun önüne geçmek için su kalitesi izleniminde takip edilmesi gereken bir fiziksel parametreyi oluşturmaktadır.

2005 yılında yayınlanan İnsani Tüketim Amaçlı Sular Hakkında Yönetmeliğe göre su kalitesini etkileyen renk parametresi herhangi bir olumsuz durum oluşmaması durumunda tüketicilerce kabul edilebilir ölçüde yer almaktadır.

2.2.1.5. Askıda Katı Madde (AKM)

Erozyonun sonucu oluşan madde taşınması, organik maddelerin bulunması ve su içerisinde yer alan planktonlardan dolayı suda askıda katı madde (AKM) oluşmaktadır. AKM'ler suda yüzebilir veya çökelmiş halde de bulunabilir.

AKM'ler suda geçici renk oluşumuna neden olarak suyun bulanıklaşmasına neden olmaktadır dolayısıyla suda ki ışık geçişini engelleyecektir.

2.2.1.6. Toplam Tuz Konsantrasyonu ve Elektriksel İletkenlik (Eİ)

Su içerisinde yer alan çözünmüş iyonların cinsine ve yoğunlaşma miktarına bağlı olarak suyun elektriksel iletkenlik (Eİ) durumu ölçülmektedir. Ayrıca Eİ miktarına bağlı olarak tuz konsantrasyon miktarı da artış göstermektedir.

Tuz oranında ki artış ozmatik basınç değerini de artıracığından dolayı bitkiler için kullanılan su hem bitkilerin hem de toprağın fiziksel ve kimyasal yapısını olumsuz yönde

etkileyecektir. Suların elektriksel iletkenlik değerin birimi mS/m veya mho/cm cinsinden ifade edilmektedir (1mho/m =10 mS/m) [21].

2.2.2. Kimyasal Parametreler

2.2.2.1. Çözünmüş Oksijen (ÇO)

ÇO, sucul ortamda çözünmüş halde bulunan oksijen yoğunluğunun bir ölçüsüdür. ÇO miktarının belirlenmesinde basıncın büyük bir etkisi yer almaktadır. Buna göre basıncın yükselmesi çözünmüş oksijen miktarını artırırken basıncın düşmesi durumunda ise düşüşe geçen basınç miktarı boyunca gaz çıkışı olacağından ötürü çözünmüş oksijen miktarı azalmış olacaktır.

Bunun için ÇO konsantrasyonun belirlenmesi sucul ortam ortamların verimliliğinin değerlendirilmesinde kalite derecesinin belirlenmesinde önemli bir parametredir. Su sıcaklığının artması oksijenin suda çözünme hızını azaltacağından dolayı balıkların ihtiyaç duyacakları oksijen miktarını artıracaktır. Balıkların yaşamlarını devam ettirebilecekleri çözünmüş oksijen miktarı en az 5 mg/L olmalıdır. Bu değer ise suyun sıcaklığına, suyun pH değerine ve çözünmüş tuz konsantrasyon oranına bağlıdır. Suda yeteri kadar oksijen miktarı bulunmaz ise su içerisinde çürümeler meydana gelmektedir. Oluşan çürümeler sonucunda ise amonyak, sülfür, metan gibi kötü kokulu gazların oluşmasına sebebiyet verirken su kalitesinin düşmesine de neden olacaktır. Bu durumu önlemek için arıtma tesisleri kurulabilir [22].

2.2.2.2. Biyolojik Oksijen İhtiyacı (BOİ)

Suda bulunan organik maddelerin mikroorganizmalar vasıtasıyla parçalanması için gerek duyulan oksijen miktarıdır. Suda BOİ değerinin yükselmesi suda kirlilik seviyesin yüksek olması şeklinde yorumlanmaktadır.

Beş gün sürede ve 20 °C sıcaklıkta organik maddelerin oksitlenmesi için kullanılan oksijen miktarının BOİ5 testi sonucunda BOİ değeri tespit edilmektedir. Ayrıca BOİ arıtma tesisinde arıtma yöntemlerinin saptanması için kullanılan bir parametredir.

2.2.2.3. Kimyasal Oksijen İhtiyacı (KOİ)

Suyun asitli yapısından etkilenen organik maddelerin oksitlenmesi sonucu ihtiyaç duyulan oksijen miktarıdır. Biyolojik yöntemler sonucu ayrışmayan maddeler ile beraber kimyasal oksijen ihtiyaç değeri atık suların kirlilik seviyesini belirlemede kullanılmaktadır. KOİ ve BOİ arasındaki ilişki ise biyolojik yollarla ayrışmayan bazı maddeleri KOİ'nin içermesi şeklinde de özetlenebilmektedir.

2.2.2.4. pH Değeri

Suyun asidik veya bazik değerde olduğunu anlamamıza yardımcı olan parametredir. Asidik su daha fazla hidrojen iyonundan oluşurken pH değeri 7'den küçüktür. Bazik su ise hidroksil iyonunun fazla olması durumunda oluşurken pH değerinin 7'den büyüktür. İçilebilir suyun ise pH değer aralığı 6,5-8,5 olması istenmektedir.

2.2.2.5. Nitrat

Nitrat, çözülmüş organik maddelerin yapısındaki azot oksidasyonu sonucu ortaya çıkmış su kirliliği için belirli seviyeye kadar zararsız kabul edilmiş parametredir. İçme sularında yer alan nitrat bağırsakta 4-12 saat içerisinde emildikten sonra böbrekler vasıtasıyla vücuttan atılmaktadır. Bunun yanı sıra ise içme sularında bulunan yüksek nitrat yoğunluğu bebeklerde methaemoglobinaemia hastalığına yakalanmasına sebep olmaktadır çünkü bebekler altı ayına kadar mide pH'ı 4,9 üzerindedir.

İnsani Tüketim Amaçlı Sular Hakkındaki Yönetmeliğe (Resmi Gazete 2005) göre suda ki nitrat oranı 50 mg/L değerini aşmaması belirtilmektedir.

2.2.2.6. Serbest Klor

Su içerisinde yer alan organik ve inorganik maddelerin emilerek okside olan veya klorüre olmak üzere kalan bunlara ek olarak suyun dezenfeksiyon işlemlerinde önemli rol oynayan klor miktarına serbest klor denilmektedir.

Sudaki serbest klor miktarının fazla olması suyun klor kokmasına ve içiminin zor olmasına neden olmaktadır. Dezenfeksiyon amaçlı kullanılması durumunda klor ölçümü

şebekenin en uç noktasından alınarak en fazla 0,5 mg/L olması istenilmektedir (Demirer 1995).

2.2.2.7. Sodyum Adsorbsiyon Oranı (SAO)

Su kalitesinin belirlenmesinde önemli bir parametre olan sodyum adsorbsiyon oranı magnezyum (Mg^{+2}), sodyum (Na^{+}) ve kalsiyum (Ca^{+2}) katyonları ile hesaplanmaktadır. Birimi mg/L'dir. Sodyum adsorbsiyon oranı toprakta ve sulama sularında bulunan Na^{+} katyonların değişim reaksiyonlarında görev almaktadır. Toprak yapısı için önemli olan SAO sulama suyu takibinde önemli rol almaktadır.

2.2.3. Biyolojik Parametreler

2.2.3.1. Fekalkoliform

İnsanların veya hayvanların idrar, kalıntı ve dışkı miktarını bakteriyolojik açıdan tanımlayan parametredir. İndikatör olarak koliform bakterileri kullanılmaktadır.

Su Kalitesi Kontrolü Yönetmeliği'ne göre 100 mL'deki fekalkoliform konsantrasyonu sıfır (0) olması gerektiği vurgulanmıştır.

2.2.3.2. Klorofil-a

Besin durumlarını ve plankton kütlelerini gözlemlemek için kullanılan ayrıca su kalitesinin belirlenmesinde yer alan bir parametredir. Klorofil-a'nın yoğunluğunun sularda yüksek seviyede olması istenmemektedir.

2.2.4. Ağır Metaller

2.2.4.1. Siyanür (CN^{-})

Son derece zehirli yapıya sahip olan siyanür hidrosiyanik asit ve bu asitten meydana gelebilen tuzların genel adını oluşturmaktadır. Bu tuzlardan en önemli olanları potasyum siyanür ve sodyum siyanürdür. Siyanürün içme sularında bulunması durumunda insanda başlıca olarak tiroid bezlerine ve sinir sistemine zarar vermektedir. Türk Standartlarına göre siyanür konsantrasyon oranı 50 µg/L olarak belirlenmiştir.

2.2.4.2.Civa (Hg)

Doğal su kaynaklarında az sayıda karşılaşılan civa elementine daha çok atık sularda karşılaşılmaktadır. İnsan sağlığına son derece zararlı olan civa elementi ağız ve diş etlerinde yaralanmaya yol açarken böbreklerin tahrip olmasına ayrıca vücuttan uzaklaştırılması zor olduğu için de çeşitli sağlık sorunlarının oluşmasına yol açmaktadır. En önemli kaynağı doğal gaz olan civa sularda bulunmasının nedeni ise en çok ilaç ve sanayi atıklarının suya karışması sonucu olduğu görülmüştür.

2.2.4.3.Arsenik (As)

Doğada arsenik sülfür ve oksit yapıda bulunmaktadır. Doğal su kaynaklarında arsenik konsantrasyon oranı çok azdır. Fakat suda bol miktarda bulunması insan sağlığını tehdit edecektir. Arsenik dolaşım sistemi hastalıklarına, kanser riskine, doku bozukluklarına yol açmaktadır. Endüstri ve ilaç sanayisi atıkları sonucu sularda görülmektedir. Türk Standartlarına göre arsenik konsantrasyon oranı 50 µg/L olarak belirlenmiştir.

2.3. KASKİ'den Alınan Veriler

Yazılan bu tez kapsamında KASKİ' den temin edilen 2015-2022 yıllarına ait merkez saha gözlem istasyonlarından alınan numuneler laboratuvar ortamında test edildikten sonra su kalitesi parametrelerinin 'İnsani Tüketim Amaçlı Sular Hakkında Yönetmelik' standartlarına uygunluğu incelenmiş elde edilen veriler ile 'Water Quality Index' (WQI) değerinin modellenmesi yapılarak su kalitesi tahmini yapılmıştır. WQI değerini hesaplamak için kullanılan parametreler 2015-2022 yıllarına ait aylık ölçümler sonucu elde edilen verilerdir. Bu veriler Kayseri iline ait olup merkez gözlemevi raporlarından alınan verilerdir. Gözlem evleri ise; Karpuzatan motor evi, Keykubat motor evi, Mahrumlar motor evi, Eğribucak motor evi, Gediris motor evi, Beştepeler motor evi, Germiraltı motor evi, Anneler Parkı motor evi, Keykubat Nuh Naci motor evi, Erkilet Hastek motor evi, Çaybağları motor evi, Anayurt motor evi, Erkilet Botaş motor evi ve Dokuzpınarlar dağıtım depoları şeklindedir. Toplamda 14 farklı gözlem evlerinden alınan veriler yılın farklı aylarında ölçülmüş olup her aydan alınan veriler modelleme verisi olarak kullanılmıştır. Modelleme verileri olarak 2022 yılına ait ocak, şubat ve mart ayları; 2021 yılına ait ocak, şubat, mart, nisan, mayıs, haziran, temmuz, ağustos, eylül, ekim,

kasım ve aralık ayları; 2020 yılına ait Ocak, Şubat, Nisan, Mayıs, Haziran, Temmuz, Ağustos, Eylül, Ekim, Kasım ve Aralık ayları; 2019 yılına ait Ocak, Şubat, Nisan, Mayıs, Haziran, Temmuz, Ekim, Kasım ve Aralık ayları; 2018 yılına ait Ocak, Şubat, Nisan, Mayıs, Haziran, Temmuz, Ağustos, Eylül, Ekim, Kasım ve Aralık ayları; 2017 yılına ait Ocak, Şubat, Mayıs, Haziran, Ağustos ve Kasım ayları; 2016 yılına ait Ocak, Şubat, Nisan, Mayıs, Haziran, Temmuz, Eylül, Ekim, Kasım ve Aralık ayları; 2015 yılına ait Ocak, Şubat, Nisan, Mayıs, Haziran, Temmuz ve Eylül aylarında ölçülen parametre değerleri kullanılmıştır. Modelleme giriş verileri olarak pH, iletkenlik, nitrat, sülfat, sodyum, serbest klor ve toplam siyanür kullanılırken çıkış verisi olarak ise Su Kalitesi İndeksi (WQI) değerleri kullanılmıştır. WQI değeri giriş parametreleri değerleri kullanılarak hesaplanmıştır.

Su kalitesi tahmin edilmesinde yapay zekâ tekniği olarak bilinen bir makine öğrenimi yöntemleri olan Yapay Sinir Ağı (ANN), Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS), Destek Vektör Makinesi (SVM), Mars Gezgin Model (MARS), Gauss Süreç Regresyon (GPR), Yakın Komşu Sınıflandırılması Modeli (KNN), Araştırılmış Regresyon Topluluğu Yöntemi (LSBooST), En Küçük Kareler Destek Vektör Makinesi(LSSVM) ve M5 Model Ağacı Yöntemi (M5TREE) teknikler kullanılmıştır. Kullanılan tekniklerden elde edilen sonuçlar birbirleri ile karşılaştırılarak hem yıllara ait eksik verilerin doldurulması hem de gelecekle ilgili su kalitesi parametre değerlerinin tahmin edilmesi amaçlanmaktadır. Böylelikle yazılan tez yakın gelecekte kullanım alanlarını artıran yapay zekâ çalışmalarına örnek olarak bundan sonraki araştırmalara yol göstermek amacıyla yazılmıştır. Diğer yandan zamandan ve maliyetlerden tasarruf edilmesine yardımcı olmaktadır. Yapılan çalışmanın çevreye bir zararı olmamasından dolayı sürdürülebilir bir çevre için de önemli bir yaklaşım oluşturmaktadır.

2.4. Yapay Sinir Ağları (ANN)

Büyüyen ve gelişen teknolojinin yanı sıra giderek hız kazanan yapay zekâ yöntemlerine verilen değer son zamanlarda giderek artış göstermektedir. İnsan beyninin öğrenebilme ve düşünebilme gibi fonksiyonel özellikleri göz önüne alınarak bilgisayar programlarına uyarlanması sonucu geliştirilmiştir. Bilgisayarların insan beynine özgü öğrenme ve düşünme faaliyetlerinin kazandırılması yapay zekâ çalışmasının temelini oluşturmaktadır. Yapay sinir ağları konusu ise yapay zekâyâ yönelik çalışmalarda önemli

yer tutmaktadır. İnsan beyninin nasıl çalıştığı üzere çalışmalar yürüten bilim insanları araştırma ve sorgulamaları sonucunda beynin çalışma prensibinden yararlanarak matematiksel modeller oluşturmuşlardır. Modeller oluşturulurken beynin sahip olduğu tüm bileşenlerin doğru olması son derece önemlidir. Bunun içinde birçok ağ modelleri oluşturulmuştur. Modelleme oluşturma çalışmaları sonucunda ise günümüz algoritma hesaplama yöntemlerinden farklı ve yeni bir bilim dalı olarak kabul edilen YSA ortaya çıkmıştır.

Bilgi işlem teknolojisi olan YSA insan beyninden yararlanarak bilgi işleme tekniğinden yararlanarak matematiksel modellemesi oluşturulmuştur. Yani, biyolojik nöron hücrelerinin birbirleri arasında sinaptik iletiminden yararlanarak dijital ortamda modellenmesi işlemini oluşturmaktadır. Nöronların farklı sayılarda ve çeşitli şekillerde bağlanarak ağları meydana getirmektedir. Oluşturulan ağlar öğrenme, veriler arasındaki ilişkiyi çözme, hafızasında tutma, düşünme ve tahmin ederek çözüm üretme gibi işlevlere sahiptir. Ağlar örnekler yardımı ile öğrenme işlemini yapmaktadır. Öğrenme işleminde bazı emir komutları tanımlayarak giriş ve çıkış bilgileri girilerek gerçekleştirilir. Öğrenme işleminin istenilen işlemi verimli yapabilmesi için verilerin doğru ve bol miktarda olması önemlidir.

2.4.1. Yapay Sinir Ağının Tarihçesi ve Kullanım Alanları

Geçmiş dönemlerde YSA alanında yapılan çalışmaların tarihçesine bakıldığında, Antik Yunan düşünürlerinden olan Aristoteles (İ.Ö. 384-320) ve Plato (İ.Ö. 427-327) dönemlerinde insan beyninin araştırılması ve incelenmesi için çeşitli teorilerin geliştirildiği ön görülmektedir. Zamanın ilerlemesiyle beraber teorilerdeki yaklaşımlar ve gelişmeler giderek hız kazanmıştır. Beynin nörofiziksel yapısından etkilenerek davranışların durumlarını modelleyebilmek için çeşitli yapay hücre ağı geliştirilmiştir. Geliştirilen hücre ağları üzerinde çalışmalar yürüten Pavlov, Poincare gibi bilim insanlarının teorileri mühendislik değeri taşımadığı için YSA'nın gelişmesinde öncü roller üstlenememişlerdir. İlk hesaplama yöntemlerini 1940 yılında fizyolog Warren McCulloch ile matematikçi Walter Pitts beraber çalışarak 1943 yılında yaptıkları çalışmalar makale olarak yayınlandığı için YSA ile ilgili ilk modelleme çalışması olmuştur. 1954 yılında W.A. Clark ve B.G.Farley tarafından oluşturulan ağ model çalışmasında sinir ağı içerisinde uyarılara karşı tepki veren ve tepkilere karşı adapte

olabilen model çalışmasında bulunmuşlardır. İlk neural bilgisayar ise 1960 yılında ortaya çıkmıştır. Sonraki yıllarda ise modeller üzerindeki çalışmalar incelenerek eksikliklerin giderilmesi hız kazanmıştır. 1985 yılında ise yapılan YSA çalışmaları derlenerek toplanmış yoğun çalışmalarda bulunularak tanınip bilinmesini sağlanmıştır. 1987 yılında ise ilk yapay sinir ağı sempozyumu gerçekleştirilmiştir. 1990'lı yıllarla beraber çalışmalar giderek hız kazanarak günümüze kadar gelmiştir. Kısacası YSA'nın tarihçesinde ki bazı önemli gelişmeler aşağıda belirtilmiştir [22].

- 1943 —► W. McCulloch ve W.Pitts elektrik devrelerini tanımlamak için beyindeki nöronların nasıl çalışacağını modelleyerek sinir ağı kavramını başlatmışlardır.
- 1956 —► Taylor tarafından ilişkisel bir bellek ağı tanıtılmıştır.
- 1958 —► W. McCulloch ve W.Pitts nöron modeli için Perceptron adlı öğrenimi yöntemini bulmuşlardır.
- 1960 —► B. Widrow ve Marci 'Adaline' ve 'Madaline' adlı modeller geliştirmişlerdir.
- 1961 —► Rosenblatt çok katmanlı ağlar için 'geri yayılım' planı önermiştir.
- 1969 —► Minsky ve Papert tarafından çok katmanlı algılayıcı (MLP) bulunmuştur.
- 1982 —► Hopfield 'enerji' yaklaşımı sayesinde YSA'da en büyük gelişme gerçekleşmiştir.
- 1988 —► Kosko ikili ilişkisel belleği (BAM)'ı geliştirerek çalışmalarında Bulanık Mantık kavramına yer vermiştir.

Günümüze kadar birçok çalışmaya konu olmuş YSA çeşitli alanlarda kullanılarak insan beyninin işlevlerinin öğretilmesinin etkileri kullanılarak modellemeler oluşturulmuştur. Kısa dönemde adından sıkça söz ettiren yapay zekâ çalışmaları içerisinde çok fazla sayıda yapay sinir ağı modellerinin oluşmasına yardımcı olmaktadır. Bilgisayar sistemi üzerinde etkin olarak kullanılan YSA çoğunlukla bilgilerin derlenip düzeltilmesinde, bilgileri modellemede, verilerin sınıflandırılmasında ve verilerin tahmin edilmesi gibi uygulamalarda kullanılmaktadır. Karmaşık ve uzun yapıda olan verilerin kopuk ya da

kesin olmayan bilgiler içereceğinden dolayı hata olasılığı yüksek modelleme sonuçları vermektedir.

Mühendislik alanları, tıp, finans, bankacılık, uzay, otomotiv sektörü, sigortacılık, savunma sanayi, elektronik, eğlence merkezleri, telekomünikasyon, petrokimya, güvenlik, askeri, robotik kodlama ve dil alanı olmak üzere çeşitli alanlarda YSA uygulamaları kullanılmaktadır. Bu sektörlerde yapılan çalışmalarda kullanılan YSA, günümüzde insan hayatını doğrudan etkileyen uygulama alanlarının gelişmesine yardımcı olurken bu alanlardan elde edilecek olan verilerin güçlü bir veri tabanı oluşturulmasına ve verilerin güvenli bir şekilde korunmasına yardımcı olmaktadır.

2.4.2. Yapay Sinir Ağlarının Avantaj ve Dezavantajları

Yapay sinir ağlarının avantajları aşağıda belirtilmiştir;

- Doğrusal olmaması nedeniyle her alanda uygulanabilmektedir. Sürekli fonksiyona ve türevlerine de yakınsama özelliğine sahip olduğu için evrensel fonksiyon yakınsayıcı olarak tanımlanabilmektedir. YSA'nın doğrusal olmayan zaman serilerinde modellenmesi verilerin tahmin aracı olarak kullanılmasını sağlamaktadır.
- Öğrenme yeteneği sayesinde eldeki verileri kullanarak daha öncesinde rastlanmamış durumlarda genelleme yaparak varmak istenilen durumlara çözüm üretebilmektedir. Yani kayıp ve eksik verileri kullanarak öğrenme yeteneği sayesinde eksik bilgilerin modellenmesi yapılabilmektedir.
- Sınırsız sayıda değişiklik gösteren parametreler ile rahatlıkla çalışabilmektedir. YSA, doğrusal olmayan, çok boyutlu, eksik veya hatalı gözlem içeren veriler için çözüm üretmek kesin matematiksel modellerin veya algoritmaların bulunmadığı işlemler de bir sonuca ulaşabilmektedir.
- Yapı olarak çok fazla hücrenin çeşitli şekillerde bağlanması sonucu olduğundan dolayı paralel dağılmış bir yapıya sahiptir. Bu yapısı sayesinde sahip olduğu bilgiler bütün hücrelere iletilmiş durumdadır. YSA'nın bazı bağlantılarında eksik veriler olması, bazı hücrelerin etkisiz hale gelmesi gibi aksi durumlar yaşanmasında ağ paralel dağıldığından dolayı veri akışında ve bilgi üretiminde önemli derece etkilenmemektedir. Bu yüzden hatayı tolerans etme ihtimalleri oldukça yüksektir.

- Veri setleri karmaşık yapıya sahipse ve verilerin dağılımı bilinen istatistiksel yöntem dağılımından farklıysa, istatistiksel sınıflama yöntemlerinden daha güvenilir sonuçlar vermektedir.
- Hata toleransı bulundurabilirler.

Yapay sinir ağlarının dezavantajları aşağıda belirtilmiştir;

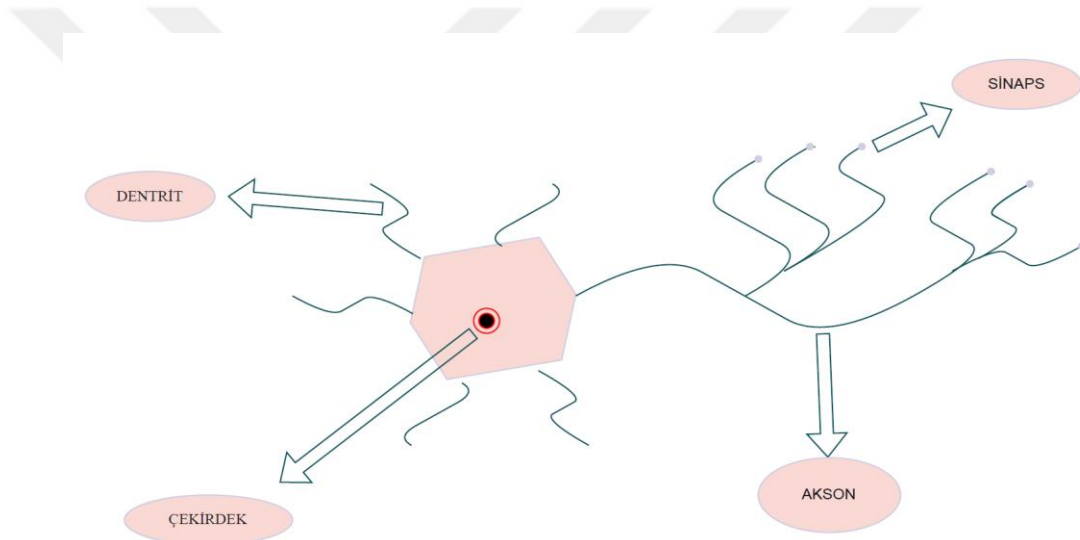
- Başlangıç olarak sinir ağlarının eğitim ve test edilmesi aşamasına geçilebilmesi için yeteri kadar genişlikte veri setlerine ihtiyaç duyulmaktadır.
- Modelleme yapılırken çok fazla sayıdaki ağ yapısının tespit edilmesi için deneme yanılma yöntemleri kullanılarak uygun ağ yapısı oluşturulmaktadır.
- Gizli katmandaki yani ara tabakadaki nöron sayısı eğitim hızını etkilemektedir. Gizli tabaka nöron sayısının çok büyük olması durumunda hesaplama yükü de fazla olacaktır.
- Bazen basit olarak görülebilecek modelleme yapılarının bile zaman zaman uygulanması zor ve karmaşık olabilmektedir.
- Hatalı öğrenmeye yatkındır. (overfitting)
- YSA'da herhangi bir probleme çözüm üretileceği zaman sonucun nereden ve nasıl bulunduğu dair açıklaması bulunmamaktadır.

2.4.3. Yapay Sinir Ağlarının Yapısı

Bir sinir sisteminin fonksiyonel ve uyarlanabilir görevlerine paralel olarak işlenmesi, bazı problemleri geliştirerek tahmin yapılması, insan beyinde bulunan sinir hücrelerinden yararlanarak bileşenlerin çoğaltılması ve simüle edilmesi amacıyla YSA kullanılmaktadır. Çok fazla sayıda sinir hücresinin insan beyinde birleşmesiyle sinir ağlarını meydana gelmektedir. Sinir ağları insan düşüncelerini, hareketlerini ve işlevlerini meydana getirmektedir. Kompleks bir yapıya sahip olan biyolojik sinir ağları sinir hücrelerinden gelen uyarıları dentritler vasıtasıyla hücre gövdesine taşımaktadır. Gelen uyarılarla birlikte kararlı yapısı bozulan hücre içerisinde kimyasal süreç başlamış sayılmaktadır. Uyarıların diğer hücrelere taşınması sırasında akson uçları ve dentritler

arasındaki sinaptik boşluklar görev almaktadır. Sinaptik boşluklar içerisinde ‘sinaptik kesecikler’ yer almaktadır. Sinaptik kesecikler ise, gelen uyarıların diğer hücrelere dentritler aracılığıyla taşınmasını sağlamaktadır. Hücrelere gelen uyarılar diğer hücrelere aktarılacak kendisiyle uyumlu diğer hücreleri etkilemektedir böylece hücreler arasında ki öğrenme işlemine de karşılık gelmektedir.

Nöronlar arasındaki iletişim elektriksel yolla iletilmektedir. Nöronların çevresiyle etkileşime girdikleri ara yüzü birbirine bağlı birkaç dentritten oluşmaktadır. Diğer nöronlara ise uyarı iletimi akson boyunca sinapslara iletilmektedir. Giriş sinyallerinin toplamı eşik durumu aşması durumunda ise diğer nöronlara iletimi sağlanmaktadır.



Şekil 1. Biyolojik nöron yapısı ve elemanları [23]

Şekil-1’ de biyolojik nöron yapısı ve elemanları gösterilmiştir. Her bir nöronun birbirine bağlanması sonucunda biyolojik sinir ağı oluşmaktadır. Buna göre biyolojik yapısını oluşturan elemanlar aşağıda açıklanmıştır.

2.4.3.1. Hücre Gövdesi

Biyolojik nöronun baş kısmı olarak nitelendirilen hücre gövdesi soma olarak da adlandırılmaktadır. Hücre gövdesi içerisinde hücrenin çekirdeğini barındırırken etrafında dışarıdan gelen uyarıları ileten dendritler ile çevrelenmiştir. Yani dendritlerden gelen sinyallerin birleştirildiği ve iletildiği yerdir. Hücre gövdesi uyarıların iletilmesinde aktif bir rol oynamazken hücrenin yapısının korunmasında ve gelen uyarıların fonksiyonel

olarak değerlendirilmesinde görev almaktadır. Ayrıca hücre içi enerjinin tutulduğu biyolojik nöron elemanıdır. İçerisinde bulunan çekirdek ise, dendrit aracılığıyla gelen uyarıları bir araya toplayıp değerlendirerek aksona iletmekle görevlidir.

2.4.3.2. Dendrit

Dendritler bir nöronun dallı çıkıntılarını oluşturmaktadır. Bağlı oldukları nöron hücrelerine gelen elektrokimyasal uyarıları kendisinin bağlı olduğu diğer nöronun çekirdeğine iletmektedir. Elektriksel uyarıların iletilmesinde görev alan dendritler çeşitli noktalarda bulunan sinapslar vasıtasıyla ilgili nöronlar uyarılmaktadır.

2.4.3.3. Akson

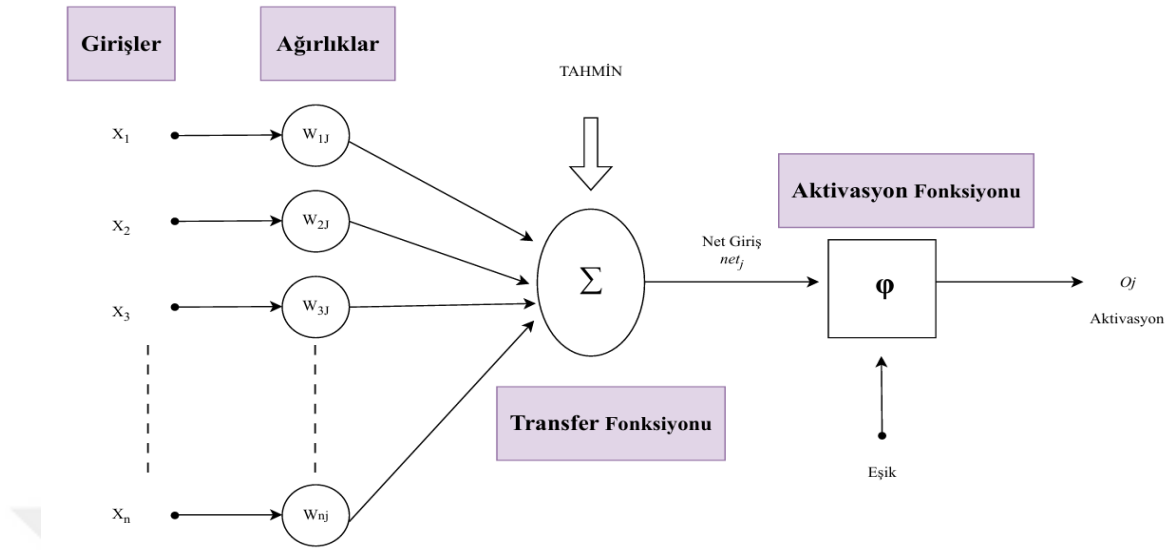
Sinir lifi olarak da bilinen akson, bir sinir hücresinin uzun, ince bir çıkıntısıdır. Hücrede toplanan uyarılar akson aracılığıyla işlenerek nöronun diğer ucunda yer alan sinapslara iletilmektedir.

Hiçbir nöronun birden fazla aksonu yoktur; ancak omurgasız hayvanlarda, böceklerde veya sülüklerde aksonun bazen birden fazla işlev gören birkaç bölgeden oluştuğu gözlemlenmiştir.

2.4.3.4. Sinaps

Sinir hücresinde sinaps, bir nörona gelen uyarıların değerlendirilmiş iletisini diğer nöron hücresine iletmekte görev alan yapıdır. Nöronal işlev için gerekli olan sinapslar, hücreden gelen sinyalleri tek tek hedef hücrelere iletmek için görev alan araçlardır.

Biyolojik sinir sistemlerinden esinlenerek oluşturulan yapay sinir ağı hücreleri yapı ve işlev olarak birbirlerine bezemektedir. Yapay sinir hücreleri de tıpkı biyolojik sinir hücreleri gibi nöronlara gelen bilgileri toplayarak kendi içerisinde değerlendirip işledikten sonra çıkıntılarını ilgili hücrelere transfer etmektedir. Şekil-2’ de bir yapay sinir ağı hücresinin yapısı ve elemanları gösterilirken Tablo- 2’de ise yapay sinir ağı sisteminin biyolojik sinir sisteminde ki karşılığı gösterilmiştir.



Şekil 2. Yapay sinir hücresinin yapısı ve elemanları [24]

Tablo 2. Yapay sinir ağı sisteminin biyolojik sinir sistemindeki karşılığı. [25]

Biyolojik Sinir Sistemi	Yapay Sinir Ağı Sistemi
Hücre Gövdesi	Transfer Fonksiyonu
Nöron	İşlemci Eleman
Akson	Yapay Nöron Çıkışı
Dentrit	Toplama Fonksiyonu
Sinaps	Ağırlıklar

Yapay sinir ağı sistemi toplamda 5 kısımdan oluşmaktadır. Bu kısımlar ise girdiler, ağırlıklar, toplama (birleştirme) fonksiyonu, aktivasyon fonksiyonu ve çıktı elemanlarıdır.

- **Girdiler:** Bir yapay sinir hücresinin dışardan gelen bilgileridir. Girişler ($x_1, x_2, x_3, \dots, x_n$) dışardan gelen bilgileri toplayarak bir sonraki adım olan çekirdeğe iletmekle görevli olan elemanlardır. Bir nörona sınırsız sayıda veri girişi olabilirken sadece bir tane veri çıkışı olabilmektedir.
- **Ağırlıklar:** Yapay hücreye gelen iletinin hücre üzerindeki etkisini gösteren ağırlıklar ($w_1, w_2, w_3, \dots, w_n$) sinir ağı tarafından öğrenilen bilgilerin saklanması ve depolanmasında görev alan elemanlardır. Hücreye gelen bilgiler girdiler üzerindeki bağlantıların ağırlıklarıyla çarpılması sonucunda çekirdeğe iletilmektedir. Bu süreç

sonucunda ise çıktı verilerinin üzerindeki etki ise giriş ağırlıklarıyla belirlenmektedir. Ağırlıkların büyük veya küçük değerde olması önemli veya önemsiz değerde olduğu anlamını taşımamaktadır. Sadece ağırlık değeri sıfır olanların çıktı üzerinde bir etkisi olmamaktadır.

- **Toplama (Birleştirme) Fonksiyonu:** Hücre içerisine gelen girdi değerlerini ağırlıklarıyla beraber çarparak toplanması sonucunda hücreye gelen net girdi değerini hesaplayan fonksiyondur. Bunun için çeşitli fonksiyonlar kullanılmaktadır fakat en yaygın olarak kullanılan ağırlıklı toplam fonksiyonudur. Hücreye gelen bilgilerin bazen sayısı önemli iken bazen de bilgilerin değerleri önemli olmaktadır. Bir sinir ağı uygulamasında giriş verilerinin toplanması için geliştirilmiş bir yöntem olmadığı için deneme ve yanılma yöntemleri kullanılarak giriş fonksiyonları seçilmektedir.

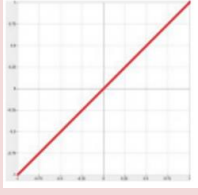
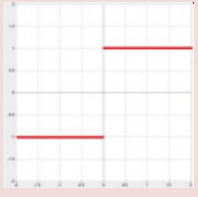
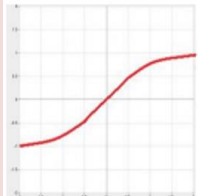
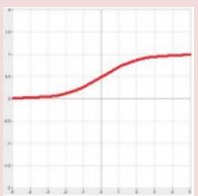
Buna göre aşağıda yer alan Tablo-3’de bazı toplama fonksiyonlarına yer verilmiştir.

Tablo 3. Toplama Fonksiyonları. [26]

Toplam $Net = \sum_{i=1}^N X_i * W_i$	Toplam fonksiyonunda girdi değerleri ile ağırlık değerleri çarpılarak çıkan sonuçlar toplanır. Böylece Net girdi bilgisi bulunmasıdır.
Çarpım $Net = \prod_{i=1}^N X_i * W_i$	Çarpım fonksiyonunda ağırlık değerleri ile girdi değerleri çarpılarak sonradan oluşan sonuçlar ile yeniden çarpılarak Net girdi bulunmasıdır.
Minimum $Net = \min(X_i * W_i)$	n adet girdi verileri ağırlıkları ile çarpılarak n adedin içinden en küçüğü Net girdi olarak seçilmesidir.
Maksimum $Net = \max(X_i * W_i)$	n adet girdi verileri ağırlıkları ile çarpılarak n adedin içinden en büyüğü Net girdi olarak seçilmesidir.
Kümülatif $Net = Net(eski) + \sum_{i=1}^N X_i * W_i$	Hücreye gelen bilgiler ağırlıklarıyla beraber toplanması sonucunda eski net girdi ile yeni hesaplanan girdi değerleri eklenerek hücrenin net girdi değerinin bulunmasıdır.
Çoğunluk $Net = \sum_{i=1}^N Sgn(X_i * W_i)$	n adet girdi değeri ağırlıkları ile çarpılarak pozitif ve negatif olanların sayısı bulunur. Büyük olan sayı hücrenin net girdisi olarak kabul edilmektedir.

- **Aktivasyon Fonksiyonu:** Hücre içerisine gelen bilgiler toplama fonksiyonunda ağırlıklarıyla beraber toplanarak aktivasyon fonksiyonuna aktarılmaktadır. Burada hücreye giren net girdiler hesaplanarak çıkış değerleri oluşturulmaktadır. Oluşturulan çıkış değeri üretilirken genellikle doğrusal olmayan fonksiyon tercih edilmektedir. Bu aktivasyon fonksiyonu tercihi de yapay sinir ağlarının ‘doğrusal olmama’ özelliğinin oluşmasından kaynaklanmaktadır. Ayrıca aktivasyon fonksiyonu belirlenirken türevi kolay hesaplanabilen fonksiyonlar tercih edilmektedir. Daha sonrasında ise üretilen çıkış değerleri diğer katmana iletilmektedir. Günümüzde ise en yaygın olarak kullanılan aktivasyon fonksiyonları tanjant, sigmoid gibi fonksiyonlardır. Aşağıda Tablo-4’de yapay sinir ağlarında kullanılan aktivasyon fonksiyonları verilmiştir.

Tablo 4. Aktivasyon fonksiyonları. [27]

Doğrusal (Lineer) Aktivasyon Fonksiyonu		$F(Net) = A * Net$ (A sabit bir sayı)	Doğrusal problemler için kullanılan bir fonksiyondur. Toplama fonksiyonundan çıkan değer, belirlenen katsayı ile çarpılarak hücrenin çıktısı olarak hesaplanmaktadır.
Adım (Step) Aktivasyon Fonksiyonu		$F(Net) = \begin{cases} 1 & \text{if } Net > \text{Eşik Değer} \\ 0 & \text{if } Net \leq \text{Eşik Değer} \end{cases}$	Net girdi değeri, saptanan bir eşik değere eşit veya altında kalmasına ya da üstünde olmasına göre hücrenin çıktısı 0 veya 1 değerini almaktadır.
Tanjant Hiperbolik Aktivasyon Fonksiyonu		$F(Net) = \frac{e^{Net} + e^{-Net}}{e^{Net} - e^{-Net}}$	Özellik olarak Sigmoid fonksiyonuna benzemektedir. Bu fonksiyonun çıkış değerleri -1 ile 1 arasında değişiklik göstermektedir.
Sigmoid Aktivasyon Fonksiyonu		$F(Net) = \frac{1}{1 + e^{-Net}}$	Türevi alınabilen sürekli olarak devam eden ve doğrusal olmayan bir fonksiyondur. Bu fonksiyon her girdi değerleri için 0 ile 1 arasında bir değer üretmektedir.
Sinüs Aktivasyon Fonksiyonu		$F(Net) = \sin(Net)$	Öğrenilmesi istenilen durumlarda sinüs fonksiyonuna ait dağılım görüldüğü yerlerde kullanılır.
Eşik Değer Aktivasyon Fonksiyonu		$F(Net) = \begin{cases} 0 & \text{if } Net \leq 0 \\ Net & \text{if } 0 < Net < 1 \\ 1 & \text{if } Net \geq 1 \end{cases}$	Gelen bilgiler 0 veya 0'dan küçük olması durumunda 0 değerini, 1 veya 1'den küçük olması durumunda 1 değerini, 0 ile 1 arasında olması durumunda ise kendi değerini alan çıktılar oluşturur.

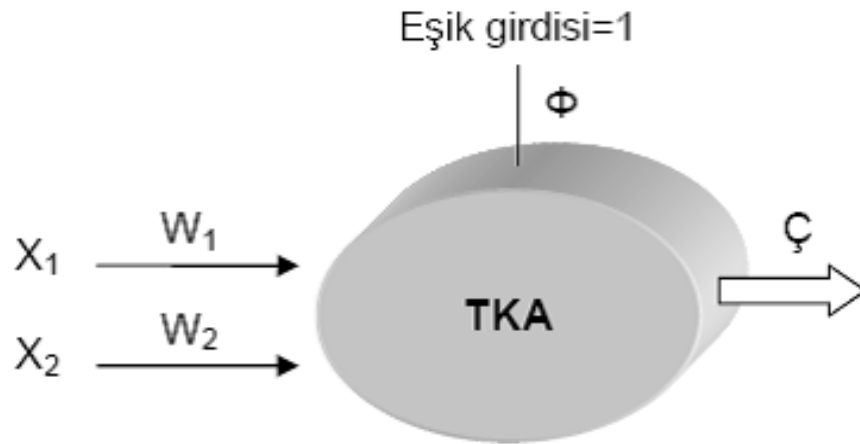
- **Çıktılar:** Yapay sinir hücresine gelen girdilerin aktivasyon fonksiyonuyla işlenerek tek bir çıkış hücresi üreten çıktı elemanlarıdır. Oluşturulan çıktı değerleri başka bir hücreye gönderilir ya da dış dünyada kullanılabilir hale gelir.

2.4.4. Yapay Sinir Ağı Mimarileri

- I. Tek katmanlı ağlar
- II. Çok katmanlı ağlar

2.4.4.1. Tek Katmanlı Ağlar

Tek katmanlı ağlar, giriş ve çıkış katmanı olmak üzere iki kısımdan oluşmaktadır. Giriş katmanı nöronların giriş sinyallerini alırken çıkış katmanı nöronların çıkış sinyallerini vermektedir. İki katmana sahip olmasına rağmen ağın tek katmanlı olarak adlandırılmasının sebebi sadece çıktı katmanının hesaplama yapmasından kaynaklanmaktadır. Çıktı verileri bütün girdi birimlerine bağlanmaktadır. Buna göre her bir bağlantının bir ağırlığı bulunmaktadır. Şekil-3’de iki girdi ve çıktıdan oluşan tek katmanlı yapay ağı modeli aşağıda verilmektedir.



Şekil 3. Tek katmanlı iki girdi ve bir çıktıdan oluşan yapay sinir ağı modeli [28]

Tek katmanlı ağlarda çıktı değerinin sıfır olmasını engelleyen eşik değeri (Φ) bulunmakta olup daima 1’e eşittir. Yapay sinir ağının çıktı değeri, her bir girdi değerinin ağırlıklarıyla çarpılması sonucunda değerlerinin eşik değeri ile toplanması sonucunda

hesaplanmaktadır. Tek katmanlı algılayıcı modellerde çıktı fonksiyonu doğrusal fonksiyondur. Doğrusal fonksiyonun değeri 1 veya -1 değer aralığında yer almaktadır.

2.4.4.2. Çok Katmanlı Ağlar

Çok katmanlı ağlar, sırasıyla giriş katmanı, gizli katman ve çıktı katmanı olmak üzere çok sayıda katmandan oluşmaktadır. Yapı olarak doğrusal olmayan aktivasyon fonksiyonlarının birçok nöronla bağlandığı ağlardır. Çok katmanlı ağların ortaya çıkışında kullanılan bazı yöntemlerin eksik ve yetersiz kalmasından kaynaklanmaktadır.

Giriş katmanı nöronları, gizli katman nöronlarına ve bu katmandaki girişlerin ağırlıklarına bağlıdır ve buna da giriş gizli katman ağırlıkları denilmektedir. Gizli katman nöronları, çıktı katman nöronları ve bu katman üzerindeki ağırlıklarıyla beraber birleştirilmesi sonucunda oluşan katmana ise gizli çıktı katmanı olarak isimlendirilmektedir.

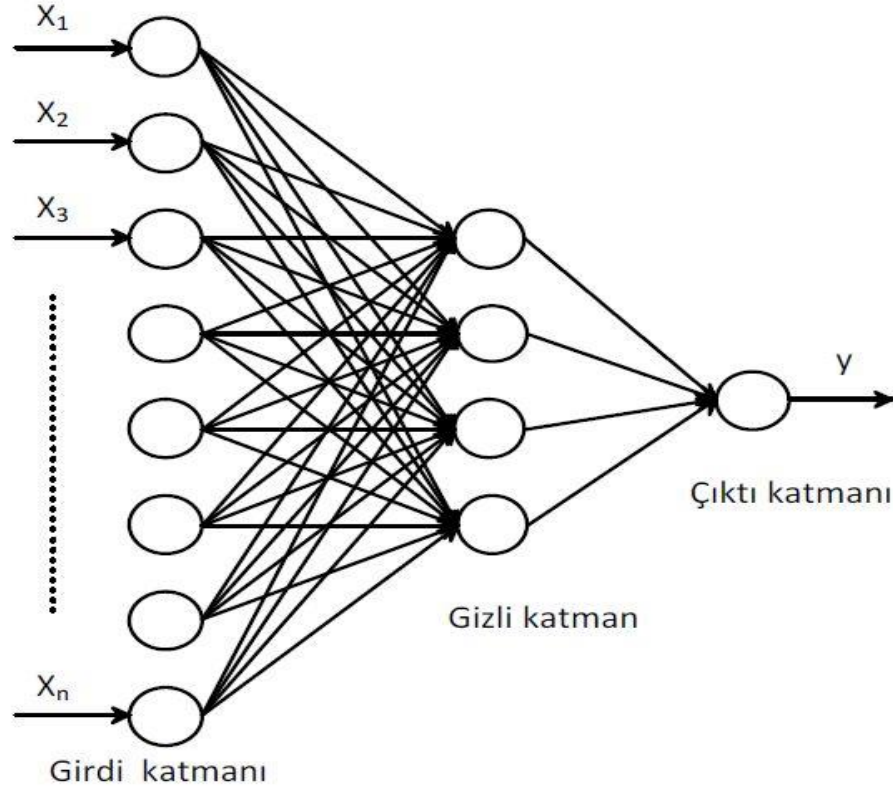
Giriş Katmanı: Nöronların dışarıdan uyarıldığı yani ilk iletilerin geldiği katmandır. Kendisine gelen iletileri gizli tabakaya aktarmakla görevlidir. Bu katman herhangi bir performans göstermez iken giriş katmanında ki nöron sayısını verilerin giriş sayısı belirlemektedir. Her girdi verileri bir nörona ait olup bu katmandaki nöronlara ise giriş nöronları denilmektedir.

Gizli (Ara) Katman: Giriş katmanı ve çıkış katmanı arasında kalan ara katmanlara gizli (ara) katman denilmektedir. Giriş katmanından gelen iletileri değerlendirerek ilgili katmana iletmektedir. Girdi katmanından gelen bilgileri çıktı katmanına iletmekten önce kendi içerisinde ara hesaplamalar yapmaktadır.

Bazı yapay sinir ağlarında ara katman bulunurken bazı yapay sinir ağlarında ara katman bulunmamaktadır. Bilgilerin işlendiği yer olan gizli katman, giriş katmanı ve çıkış katmanına göre nöron sayısı oldukça fazladır. Gizli katman sayısının fazla olması karmaşık yapıdaki yapay sinir ağının kontrol altına almasında da önemli bir göreve sahiptir.

Çıkış Katmanı: Gizli katmandan gelen sinyalleri aldıktan sonra giren veriye uygun olarak hesaplamalar yaparak çıkış verileri oluşturmaktadır. Çıkış verileri her nörona tek bir çıkış verileri gelecek şekilde dışarıya aktarılacak için üretilmektedir.

Şekil-4’ de çok katmanlı yapay sinir ağı modeli katmanlarıyla beraber gösterilmektedir.



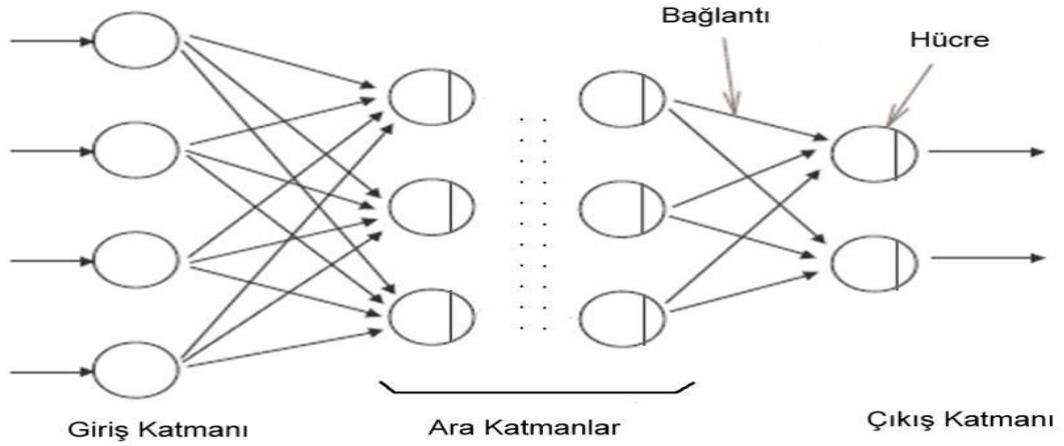
Şekil 4. Tek çıkışlı çok katmanlı yapay sinir ağı modeli [29]

2.4.5. Yapılarına Göre Yapay Sinir Ağları

- I. İleri Beslemeli Yapay Sinir Ağları
- II. Geri Beslemeli Yapay Sinir Ağları

2.4.5.1. İleri Beslemeli Yapay Sinir Ağları

İleri beslemeli yapay sinir ağlarında nöronlar giriş katmanından çıkış katmanına doğru düzenli bir şekilde konumlanmışlardır. Nöronlar arasında bir döngü oluşturmazlar. Bir katman sadece kendinden sonraki katmanla bağ oluşturmaktadır. Giriş katmanına gelen bilgiler bir değişime uğramadan gizli katmana iletilmektedir. Çıkış katmanında işlenerek dış ortama gönderilmek üzere sistemi çalışmaktadır. İleri beslemeli yapay sinir ağları tek yönlü sinyal iletiminin akışı için gelen sinyallere izin vermektedir. Buna göre Şekil-5’ de ileri beslemeli yapay sinir ağı modeline örnek verilmiştir.

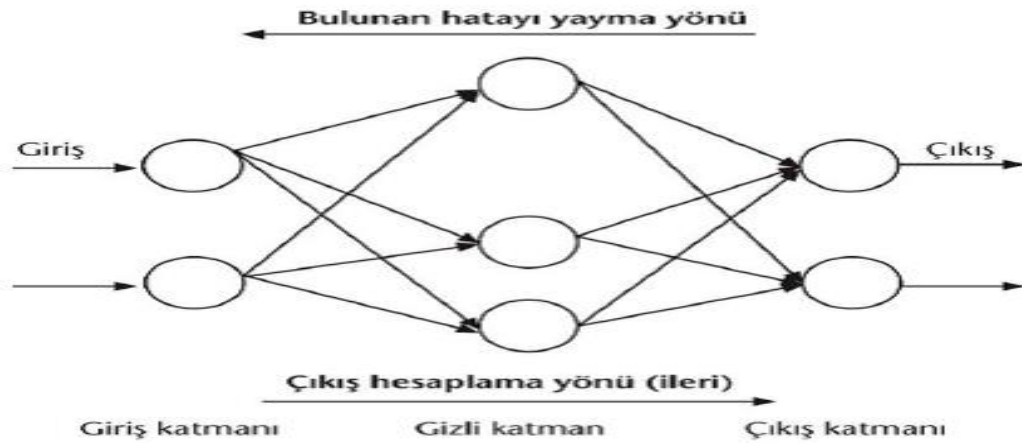


Şekil 5. İleri Beslemeli Yapay Sinir Ağı Model Gösterimi [30]

En yaygın olarak kullanılan algoritma olan geriye yayılım öğrenme algoritması ileri beslemeli yapay sinir ağı eğitiminde sıkça kullanılmaktadır.

2.4.5.2. Geri Beslemeli Yapay Sinir Ağları

Çalışma prensibi ileri beslemeli yapay sinir ağlarından farklı olarak bir nöronun çıktı verilerini sadece kendisinden sonra gelen nörona girdi verisi olarak iletmez, kendinden önceki katmana veya kendi katmanlarında bulunan herhangi bir nörona da girdi verilerini iletebilmektedir. Yapısı bakımından doğrusal olmayan dinamik bir davranış göstermektedir. Geri beslemenin uygulama şekline göre farklı yapı ve davranışta yeni ağlar elde edilebilmektedir. Geri beslemeli yapay sinir ağları kendi içerisinde de döngüye girdiğinden dolayı eğitim süreci yavaş gerçekleşmektedir.



Şekil 6. Geri beslemeli yapay sinir ağı modeli [31]

2.4.6. Öğrenme Algoritmalarına Göre Yapay Sinir Ağları

Yapay sinir ağlarının çıktı verilerini üretebilmesi için girdi verilerini öğrenme algoritması ile yapısında tanımlaması gerekmektedir. Ayrıca öğrenme işlemi ağırlıkları güncellemek için de kullanılmaktadır. Yapay sinir ağları öğrenme algoritmalarına göre denetimli, denetimsiz ve takviyeli öğrenme algoritmalarına göre sınıflandırılmaktadır.

2.4.6.1. Denetimli Öğrenme

Yapay sinir ağlarında analiz işlemlerine başlamadan önce sisteme girdi verileri tanımlanırken çıktı verilerinin de ne olduğu tanıtılmaktadır. Analiz ya da tahmin işleminin başlatılması durumunda sisteme tanımlı olan girdi verilerine göre istenen çıktı verilerini oluşturmak için kendi ağırlıklarını oluşturmaktadır. Sistemin problemlere çözüm üretebilmesi için verilen örneklerden yararlanarak yeni çözüm önerileri üretmektedir. Geliştirdiği çözüm yollarında ise çıktı verileri ile ulaşılması istenen veriler kıyaslanarak aralarındaki hata payı yeni ağırlık oranına göre hesaplanmaktadır. Hata payının hesaplanması ise, ağın çıktı verilerinin bulunması istenen veriler arasındaki farkın hesaplanması ile bulunmaktadır. Böylece her nörona düşen hata payı bulunmaktadır.

2.4.6.2. Denetimsiz Öğrenme

Girdi verileri sisteme girdiğinde sistem kendi içerisinde özellik ve yapı olarak birbirine en çok benzeyen verileri gruplar halinde bölerek her bir grup verilerinin kendi kurallarını yani öğrenimini gerçekleştirmektedir.

Gruplandırma işlemini, aynı özellikte olan girdi verilerini ağırlıklarını oluşturarak düzenlemektedir.

2.4.6.3. Takviyeli Öğrenme

Denetimli öğrenme modeli hesaplamasına göre benzer bir metot uygulanmaktadır. Yapay sinir ağı analiz sırasında her bir döngüye denk gelecek çıktı verilerinin uygunluğunu değerlendiren bir kriter kullanmaktadır. Sinir ağı kritere göre oluşturan bilgileri kendi hafızasında sürekli olarak güncellemektedir.

Bu sayede ağ hem eğitilmiş bilgileri öğrenerek hem de işlenmiş verilerden çıkardığı sonuçlarla sürekli olarak döngüye devam ettirmektedir.

2.5. Uyarlamalı Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS)

Uyarlamalı ağ tabanlı bulanık mantık çıkarım sisteminin yapısı öncül ve sonuç bölümleri olarak adlandırılan iki bölümden oluşmaktadır. ANFIS'in eğitimi ise, bu bölüm parçalarına ait verilerinde optimizasyon algoritması geliştirmek için kullanılmaktadır. ANFIS, eğitim esnasında mevcut giriş ve çıkış veri setlerini özümseyerek bu verilerin birbirine bağlı olmasına göre 'If –Then' bulanık kuralı geliştirmiştir (Jang 1993).

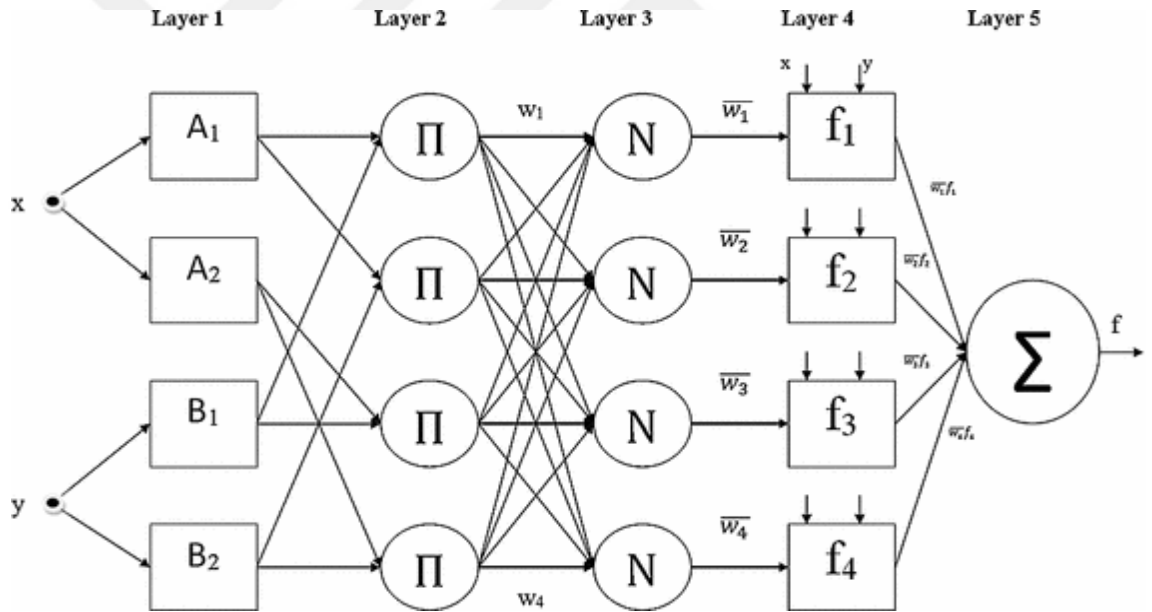
ANFIS yöntemi en çok doğrusal olmayan problemlerin çözümlerinde kullanılmaktadır. Problemlerde yaşanan belirsizlik durumunun çözülmesi için ele alınan ANFIS yönteminde iki farklı yaklaşım yöntemi uygulanmaktadır. Bu yaklaşımlar ise, Mamdani ve Takagi Sugeno yöntemleridir [32].

Mamdani yöntemi, tecrübeli insan deneyimlerinden elde edilen dilsel kontrol kuralından ortaya çıkarılan bir sistemdir. Her bir kuralın ayrı ayrı çıktı değeri olup bu çıktı değerlerinin bir de çıkış üyelik fonksiyonu bulunmaktadır. Bu çıktı verileri bulanık çıkarım sisteminden türetilen bulanık bir kümedir. Oluşan bulanık çıkarım kümeleri tek

bir yerde birleşmektedir. Birleşen çıktı kümelerinin istenen çıktı değerine ulaşması için berraklaştırma (durulaştırma) işlemi yapılmaktadır. Mamdani yöntemi sezgisel olup anlaşılması ve yorumlanması kolay olan bir yöntemdir.

Takagi-Sugeno yöntemi ise, doğrusal ve sabit fonksiyonlu girdi değerleri, tekil çıkış üyelik fonksiyonlarına sahiptir. Çıkış üyelik fonksiyonları ile bulanıklaştırma işleminden gelen fonksiyon üyelik seviyeleri polinom ile ilişkilendirilmektedir. Berraklaştırma işlemi ise veri noktalarının ağırlıklı ortalaması alınarak yapılmaktadır. Takagi-Sugeno yöntemi matematiksel analizler için uygundur buna ek olarak optimizasyon ve uygulanabilirlik açısından iyi çalışan bir yöntemdir.

ANFIS'in yapısı Şekil-7'de görüldüğü 2 giriş ve 1 çıkıştan oluşan 5 katmanlı olarak çözümlenen yapısı verilmiştir.



Şekil 7. İki girdi ve bir çıktıya sahip anfis yapısı[33]

Doğrusal kombinasyonlar ile en küçük kareler yöntemi katsayısı belirlenmekte ve üyelik fonksiyonları ile geri yayımlı öğrenme algoritmasının öğrenmesi sağlanmaktadır. Bu fonksiyonların kullanılması ANFIS modelinde girdi ve çıktı parametrelerinin girilmesi sonucunda bulanık bir çıkarım yapılmaktadır. İki adet girdisi (x ve y), bir adet çıktısı (f) olan ANFIS denklemi eşitlik (2.1)'de gösterilmektedir.

$$f = p_1x + q_1y + r_1 \quad (2.1)$$

Katman 1: Bulanıklaştırma katmanı olarak adlandırılmaktadır. Girdi değerlerinden bulanık kümeler elde etmek için üyelik fonksiyonları kullanılmaktadır. Her düğüm için çıkış değeri eşitlik (2.2)'de verilmektedir.

$$0_{1i} = \mu_{A_i}(x), i = 1, b$$

$$0_{1i} = \mu_{B_{i-2}}(y), i = 3,4 \quad 0_{2i} = \mu_{A_i}(x) * \mu_{B_{i-2}}(y) \quad (2.2)$$

Katman 2: Bu katmana kural katmanı denilmektedir. Kurallar için Takagi Sugeno kuralı bulanıklaştırma katmanında hesaplanan belirti değerleri ile birleştirilerek her bir durum için üyelik değerleri kullanılarak sonuca varıldığı katmandır.

Katman 3: Bu katmana normalizasyon katmanı denilmektedir. Normalizasyon katmanında bulanıklaştırılmış girdi değerleri birleştirilip ağırlık fonksiyonlarıyla eklenmesi sonucunda oluşan değerlerin standardize edildiği katmandır. Buna göre bu katmanda hesaplanan formül değeri eşitlik (2.3)'de verilmektedir.

$$0_{3i} = \overline{w_i} = \frac{\omega_i}{\sum_{(i=1)}^2 \omega_i} \quad (2.3)$$

Katman 4: Bu tabakaya defuzzifikasyon tabakası denilmektedir. Burada berraklaştırma işlemi yapılmaktadır. Bulanıklaşmış girdi verilerinin tekrardan net hale gelmesi için ve her bir noktanın model çıktısı üzerindeki etkisini belirlemek için bu katman oluşturulmuştur. Buna göre dördüncü katmanda hesaplanan formül değeri eşitlik (2.4)'de gösterilmektedir.

$$0_{4,i} = \overline{w_i} f_i = \overline{w_i} (p_i x + q_i y + r_i) \quad (2.4)$$

$\overline{w_i}$ normalleştirme katmanının çıktısıdır. (p_i, q_i, r_i) parametre kümesini oluşturmaktadır. Bu değerlere ise sonuç parametreleri adı verilmektedir. Her kuralın sonuç parametrelerinin sayısı, girdi verilerinden bir fazladır.

Katman 5: Katmanlardan sonuncusu olan beşinci katman toplama katmanı olarak adlandırılmaktadır. ANFIS'in gerçek çıktı verisi olarak belirlediği değeri, defuzzifikasyon katmanındaki her bir kural için elde edilen çıktı değerlerinin toplanması sonucunda ulaşıldığı değerdir. Aşağıda verilen eşitlik (2.5)'de ki gibi hesaplanmaktadır.

$$0_5 = \sum_i \overline{w_i} f_i \quad (2.5)$$

ANFIS eğitim sürecinin amacı yapısındaki parametrelerin bir optimizasyon algoritmaları ile birleştirilmesi ilkesine dayanmaktadır. ANFIS ile etkili sonuçlara ulaşabilmek için başarılı bir eğitim süreci şarttır. Daha iyi performanslar elde etmek için farklı eğitim yaklaşımları öne sürülmüştür. Bu yaklaşımlardan en çok kullanılanı sezgisel tabanlı, türev tabanlı ve hibrit yaklaşımlarıdır.

2.6. GAUSS Süreç Regresyonu Modeli (GPR)

Rasmussen ve Williams (2006) tarafından ortaya atılan Gauss Süreç Regresyonu (Gaussian Process Regression-GPR), Bayes öğrenme yaklaşımından yararlanarak doğrusal olmayan regresyon problemlerin çözümünde kullanılan başarılı bir makine öğrenmesi yöntemidir. Buna ek olarak çekirdek (kernel) fonksiyonlarını temel alarak olasılık dağılımlarını tahmin eden öğrenme yaklaşımlarına sahiptir. Bu özelliği sayesinde daha az veriye ihtiyaç duyarak istenilen tahmin işlemini veri koruma özelliğini kullanarak daha esnek denetimli öğrenme aracı olarak kullanılmaktadır.

Öğrenme algoritmaları, elde edilen çıktı verileri sürekli bir yapıya sahip ise regresyon, kesikli bir yapıya sahip ise sınıflandırma problemlerinde analiz için kullanılmaktadır. Birden çok sayıdaki değişken arasındaki ilişkileri tahmin etmek için kullanılan denetimli öğrenme yöntemlerine regresyon denilmektedir. Gauss sürecinde ise regresyon analizi bu tanımlamaya dayalı olarak girmiştir. GPR yöntemi ise, tahmin belirsizlik problemlerinin çözüm çalışmalarında yaygın olarak kullanılmaktadır.

Gauss süreç regresyon modeli ile ilgili olarak standart regresyon uygulamasına ait Bayes analizi ve Gauss dağılımı eşitlik (2.6)'da bahsedilmektedir [34].

$$f(x) = x^T w \quad y = f(x) + \varepsilon \quad (2.6)$$

Buna göre eşitlik (2.6)'da verilen f fonksiyon değerini, w ağırlık vektörünü, x giriş vektörünü, ε sapma veya hata değerini ve y gözlenen hedef vektörünü göstermektedir.

Gözlenen hedef vektörü y , $f(x)$ fonksiyonunda ki hata (sapma) değeri kadar farklılık göstermesine modelin gürültüsü denilmektedir. Gürültü değerinin varyans değeri, sıfır

ortalama ve bağımsız değeri Gauss dağılımına uygun olduğu kabul edilmiştir. Bu kabulün gösterilimi eşitlik (2.7)' de gösterilmiştir.

$$\varepsilon \sim N(0, \sigma^2) \quad y = f(x) + N(0, \sigma^2) \quad (2.7)$$

Eşitlik (2.7)'de verilen gürültü hakkındaki varsayım, eğitme veri setindeki parametrelerin olasılık yoğunluklarına ulaşmasını sağlamaktadır. Gürültü varsayımından yola çıkılarak eşitlik (2.8)'de ki olasılık yoğunluğu elde edilmiştir.

$$p(y|X, w) = \prod_{i=1}^n p(y_i|x_i, w) = N(X^T w, \sigma^2 I) \quad (2.8)$$

Eşitlik (2.8)'den Bayes formülüne geçiş yapabilmek için gözlemlerden önce çarpan parametrelerini ilgilendiren çeşitli varsayımlarda bulunarak ağırlık vektörlerine kovaryans matrisi ve sıfır ortalamalı bir Gauss dağılımı eklenmiştir.

$$w \sim N(0, \Sigma_p) \quad (2.9)$$

Eşitlik (2.9)'a göre, gözlem verilerinin analiz yapılmaya başlamadan önce ağırlık vektörleri üzerinde öncül bir dağılım uygulanmıştır. Bu işlemten sonra Bayes formülüne göre gözlem kümesi üzerinde soncul olasılık dağılımı belirlenmiştir. Bayes teoremi, rastgele bir değişken için olasılık dağılımı içerisindeki marjinal olasılıklar ve koşullu olasılıklar arasındaki ilişkiyi göstermektedir.

$$p(y|X, w) = \frac{p(y|X, w)p(w)}{p(y|X)} \quad (2.10)$$

Eşitlik (2.10)'da ise Bayes teoreminin soncul dağılımı ifade edilmiştir. Soncul olasılık dağılımı, olasılık değeri ile öncül verilerin çarpımının marjinal olasılık dağılımına oranıyla hesaplanmaktadır.

Gauss süreci ise, Gauss dağılımına ait sonlu sayıdaki değişkenlerin toplanması sonucunda bulunmaktadır. Yukarda bahsedildiği gibi hesaplanan ağırlık vektörlerine sıfır ortalamalı Gauss ve kovaryans matrisi eklenmektedir. Böylece Gauss süreci başlamaktadır. Eşitlik (2.11) Gauss sürecini temsil etmektedir.

$$f(x) \sim GP(m(x_i), k(x_i, x_j)) \quad (2.11)$$

Eşitlik (2.11) ile ifade edilen $m(x_i)$ ortalama fonksiyon değerini, $k(x_i, x_j)$ kovaryans fonksiyonunu göstermektedir. Kovaryans matrisi değişkenler arasındaki benzerlikleri ifade etmektedir. Değişkenler arasındaki benzerlikler ise üstel fonksiyon aracılığıyla eşitlik (2.12)'de gösterilmiştir.

$$k(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \quad (2.12)$$

Eşitlik (2.12)'de σ parametresi çekirdek genişliğini ifade etmektedir. Bu parametrelerinin olasılıksal öğrenme yöntemi ile çekirdek fonksiyonlarının optimize edilerek birleştirilmesi Gauss sürecinde Bayes teoreminin kullanılmasıdaki en önemli nedenlerinden birisidir. Gauss sürecinde gözlem kümesinden hariç (x_*) değişken verisi ve bu değişken verisine bağlı olarak ağırlık vektörü üzerindeki (y_*) çıktı değerinin tahmin dağılımını göstermek için eşitlik (2.13)'de verilen gömüleme fonksiyonu kullanılmıştır. Buna göre (μ_*) tahmini ortalama ile $\sigma_*^2(x_*)$ kovaryans ifadelerinin değerleri aşağıda yer verilen eşitlik (2.14), (2.15)'de gösterilmiştir.

$$p(y_* | (x, y), x_*) = N(\mu_*(x_*), \sigma_*^2(x_*)) \quad (2.13)$$

$$\mu_*(x_*) = k_*^T (K_\theta + \sigma^2 I)^{-1} y \quad (2.14)$$

$$\sigma_*^2(x_*) = k_{**} - k_*^T (K_\theta + \sigma^2 I)^{-1} k_* \quad (2.15)$$

Kovaryans değeri, çekirdek fonksiyonlarını kullanarak eğitim ve test çıktıları arasındaki bağlantıyı mesafelerine göre belirler. Bu açıdan kovaryans değerinin seçimi model performansının belirlenmesinde oldukça önemlidir. Gauss sürecinde radyal tabanlı fonksiyonlar, kovaryans fonksiyonları ve çekirdek fonksiyonları kullanılmaktadır. Kullanılan bazı radyal tabanlı temel fonksiyonlar; doğrusal, saf ikinci dereceden, sabit ve sıfır fonksiyonlarıdır. Çekirdek fonksiyonları; üstel, karesel üstel, ard üstel, rasyonel ikinci dereceden, matern3/2, ard matern3/2, matern5/2 ve ard matern5/2 ve ard rasyonel ikinci dereceden fonksiyonlarıdır.

GPR, öğrenme aşamasında daha az veriyi kullanarak belirsizlik problemleri üzerinde etkili olan doğrusal olmayan fonksiyonların özellikleri öğrenme işleminde başarılıdır.

2.7. Mars Gezgin Model (MARS)

Jerome Friedman (1991) tarafından geliştirilen çok değişkenli uyarlanabilir regresyon eğrileri (MARS) modeli, bazı veri setlerinin doğrusal olmayan eğilim göstermesinden kaynaklı doğrusal eğilim gösteren modellerde başarılı sonuç vermemektedir. Model kurulurken bağımlı ve bağımsız değişkenler arasındaki algoritmik bağlantının şekline dair varsayımda bulunmaz iken model, bağlantı katsayılarından ve taban fonksiyonlarından oluşturulmaktadır. MARS modeli, bağımsız değişkenlerin değerlerini alanlara bölerek her oluşan alanı regresyon eşitliğine dayandırır. MARS modeli iki adımdan oluşmaktadır [35].

1. Adım: Model üzerinde temel fonksiyonlar kurularak modelin karmaşık seviyesi en üst seviyeye gelinceye kadar model geliştirilmeye devam edecektir. Temel fonksiyonlar, çok değişkenli veya tek değişkenli verilerin birbiri ile etkileşim içerisinde olan terimlerinden meydana gelmektedir. Temel fonksiyonlar, bağımsız değişkenli verilerin birbiriyle etkileşime girmesiyle ya da doğrusal olmayan dönüşümlerinin tahmin edilmesi sürecinde kullanılan matematiksel bir yöntemdir. Buna göre eşitlik (2.16)'da MARS modeline ait formül hesabı verilmiştir.

$$Y = \beta_0 + \sum_{k=1}^K \alpha_k \beta_k(X) \quad (2.16)$$

Eşitliğe göre, (k) düğüm sayısını, (X) bağımsız değişkeni, (β_0) model üzerindeki sabit terimi, (α_k) k temel fonksiyon katsayısını ve (K) temel fonksiyon sayısını vermektedir. Eşitlik (2.17)'de ise temel fonksiyonun formül gösterimi verilmiştir.

$$\beta_k(X) = \prod [S_{k,j}(Xw_{(k,j)} - t_{(k,j)})] \quad (2.17)$$

2. Adım: Temel fonksiyonların tahmin hatasının istenilen seviyeye gelene kadar modelde tutulmadan çıkarılmasıdır. Bu işlemin gerçekleştiği basamağa 'budama (puring)'denilir. Bu başmağın çalışma prensibi MARS modelinin geriye doğru basamak algoritmasını çalıştırmasıdır. Geriye doğru basamak algoritmasının çalışması aşağıdaki işlem basamaklarının gerçekleştirilmesi ile oluşturulmaktadır. Modele başlamadan önce, temel fonksiyon sabitinin olduğu en küçük model ile başlanır.

- Uygun hesaplamalar yapılarak tahmin hatasının en düşük olduğu değişkenler ile temel fonksiyon uzayı araştırılmalıdır.

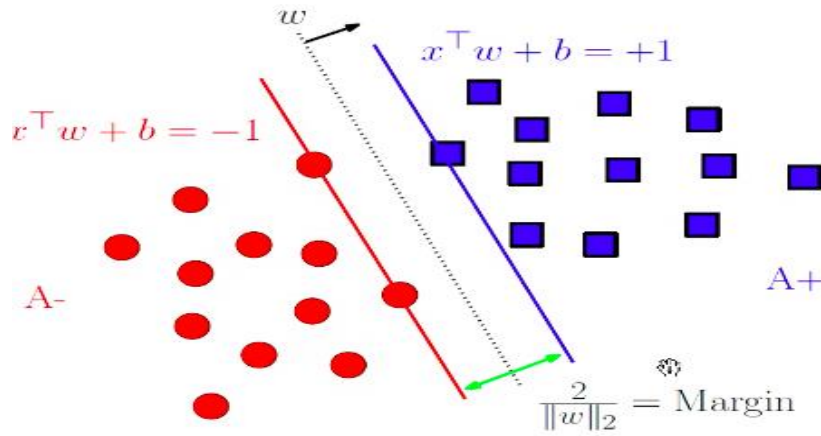
- En karmaşık modele sahip değere ulaşıncaya kadar değişkenlerin ve bütün düğüm noktalarına bakma işlemi tekrarlanır.
- Son adımda ise budama (puring) aşamasına geçilir.

Sonuç olarak istenilen seviyedeki modele ulaşılır. Daha sonra en uygun model için hata kareler toplamının en düşük olduğu seviye belirlenir. Budama modeli, genelleştirilmiş çapraz tekniği kullanılarak yapılmaktadır. Bu teknik, model karmaşıklığını ile artıkların hatasında göz önünde bulundurur.

MARS algoritması amacı ise, bağımlı ve bağımsız değişkenler arasındaki dağılımlarla ilgili herhangi bir varsayım olmaması üzerine bağımlı ve bağımsız değişkenler arasındaki ilişkiyi inceleyerekten tahmin denklemlerinin oluşturulması prensibine dayanmaktadır.

2.8. Destek Vektör Makinesi Modeli (SVM)

Cortes ve Vapnik (1995) tarafından öne sürülen Destek Vektör Makineleri (Support Vector Machine – SVM), verileri sınıflandırmak için istatistiksel öğrenme teorisi kullanılarak yapısal risk sınıflandırmasına dayalı makine öğrenme modelidir. Küçük sayıdaki ve büyük boyuttaki veri setlerinde başarılı bir şekilde öğrenme gerçekleştirmektedir. Regresyon ve sınıflandırma problemlerinin çözülmesinde yaygın olarak kullanılan denetimli öğrenme yaklaşımıdır. Sınıflandırmadaki amaç incelenen veri setlerinin doğru bir şekilde gruplanmasını sağlayarak bir hiper düzlem oluşturmaktır. Hiper düzlemi iki farklı sınıfın oluşması için doğrusal bir sınır çizgisi çizilerek marjin olarak adlandırılan bir aralık oluşturulur. Marjin, sınıflandırma yapmak için en iyi değeri ve en geniş sınır aralığını bulmak için kullanılmaktadır. Buna göre SVM modelini kullanarak sapma değerlerinin belirli bir eşik değerde kalmasını sağlayarak hataları minimize ederek risk minimize yönteminden faydalanmaktadır. Şekil-8’de SVM modeline ait sınıflandırma durumu gösterilmiştir [5][32].



Şekil 8. SVM model gösterimi [36]

Şekil-8’de gösterilen SVM model gösterimine göre, n elemandan oluşan örnek uzayın eğitme verilerinin $A+$ ve $A-$ sınıflarına ayrılması eşitlik (2.18)’de ki fonksiyon ile yapılmaktadır.

$$f(x) = w \cdot x + b = 0 \quad (2.18)$$

Eşitlik (2.15)’de gösterilen $f(x)$ fonksiyonu optimum hiperdüzlemi, (x) düzlem üzerinde yer alan veri noktalarını, (w) ağırlık vektörünü, (b) hata (sapma ya da bias) değerini temsil etmektedir. Optimum hiper düzlemin oluşturulması için sınıra en yakın noktaları içeren veriler için $|w \cdot x_i + b| = 1$ eşitliğinin çözümüne göre sınıflandırılması gerekmektedir. Fonksiyonlar sonucu hesaplanan hiper düzlemin sınıflandırılması ile oluşan marjın bölgesi $2/\|w\|$ formülü ile toplam uzunluğu belirlenmektedir. Optimum hiper düzleme üzerinde bulunan (en yakın) konumda bulunan veriler destek vektörleridir. Ayrıca ağırlık vektörü (w) farklı değerler almasına göre farklı ayırıcı fonksiyonlar oluşturulmaktadır. Bu doğrultuda marjın sınırının genişliği ayarlanmalıdır. Ağırlık vektörünün küçük olması durumunda marjın genişliği maksimum seviyededir. Marjın genişliğinin yüksek sınırdan olmasının olumlu yanı verilerin daha büyük bir marjında daha iyi genellemeler yapabilecek olmasıdır.

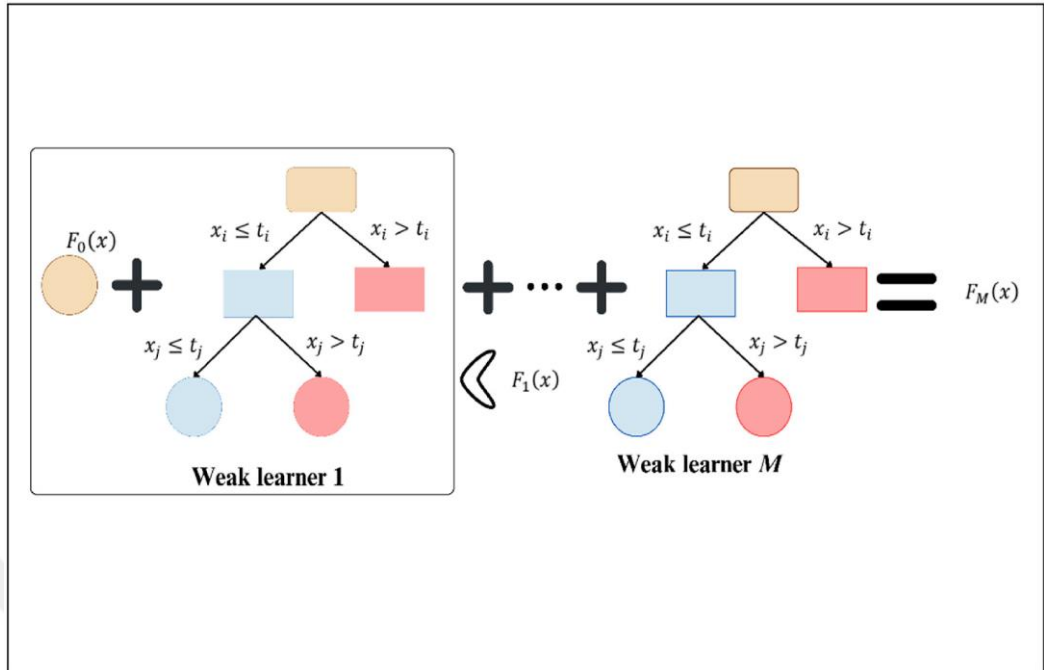
Destek vektör makineleri başlangıçta iki farklı sınıflandırma yaparken sonradan ihtiyaç durumuna göre çoklu sınıflandırma yöntemi olarak da kullanılabilir. SVM, eğitme hatalarının bulunmadığı doğrusal sınıflandırma problemlerinin çözülmesinde ve eğitme

hatalarının bulunduğu doğrusal olmayan sınıflandırma problemlerin çözülmesinde sıkça kullanılmaktadır.

SVM kullanımı birçok alanı kapsamaktadır. Bu alanlar ise şu şekilde örneklendirilebilir; bu tez kapsamını da ilgilendiren hidroloji ve su kalitesi alanında, yağış-akış model seçiminde uygun modelin seçilmesinde, coğrafi amaçlı veri analizlerinde, uydu görüntülerinde, mimari tasarımlarda, harita çalışmalarında, gen çalışmalarında biyolojik sınıflandırmada, yüz tanımlama çalışmalarında, ses ve nesnelerin sınıflandırılması işlemlerinde gibi birçok alan içerisinde SVM model çalışmalarında kullanılmaktadır.

2.9. En Küçük Kareler Artırma Modeli (LSBoost)

En Küçük Kareler Artırma (LSBoost) algoritması, Friedman tarafından (2001) en küçük kareler (LS) stratejisinin uygulandığı genel gradyan Artırma algoritmasının bir çeşididir. LSBoost, karmaşık veri kümelerini işleyerek doğrusal olmayan ilişkiler arasındaki bağlantıyı yakalamak için uygun bir modeldir. Regresyon problemlerinde veya sınıflandırma işlemlerinde kullanmak için uygun bir makine öğrenimi algoritmasıdır. Modele tekrarlama olacak şekilde regresyon ağaçları ekleyerek çalışmaktadır. LSBoost algoritması, gerçek değerler ile tahmin edilen değerler arasındaki karesi alınıp hata değerlerinin toplamının en aza indirilmesini sağlamaktadır. Tahmin modelleri yüksek doğruluk oranına sahip olduğundan dolayı aşırı makine öğrenimini uyumunu kontrol etmek için parametrelerinin dikkatli bir şekilde ayarlanması önem arz etmektedir. Şekil-9’ da LSBoost modeline ait şematik gösterilimi verilmiştir.



Şekil 9. LSBoost modeli şematik gösterimi [37]

Gradient boosting yöntemlerinden biri olan LSBoost modeli, özellikle düşük ortalama karesel hatalar üzerine odaklı çalışmaktadır. Gradient boosting metoduna dayalı olması, zayıf tahmin modellerinin bir araya getirilmesi ile güçlü bir tahmin oluşturmayı amaçlamaktadır.

Ağaç tabanlı bir yöntem olarak çalışmakta olup her iterasyon döngüsünde bir karar ağacı ekleyerek önceki ağaç tahminlerinin hatasını en aza indirmeye çalışmaktadır. Ayrıca LSBoost modelinin performansı, ağaç derinliğine, ağaç sayısına, öğrenme oranına bağlı olmak üzere aynı zamanda parametrelerin iyi ayarlanması ile de doğrudan ilişkilidir.

2.10. K-En Yakın Komşu Modeli (KNN)

Evelyn Fix ve Joseph Hodges (1951) tarafından geliştirilen K-en yakın komşu (K-Nearest Neighbours-KNN) algoritması, örüntü tanımlama alanında kullanılan yöntemdir. KNN, veri setlerinin parametrik olmayan veri setleri hakkında bilginin olmadığı veya veri setleri hakkında çok az bilginin olduğu zaman tercih edilen bir yöntem olarak kullanılmaktadır. Birden fazla gruplara bölünerek sınıflandırılan veri setlerinin ağırlıklarına göre ayrıştırılmasından yararlanarak veri tabanı üzerinde yeni bir örnek nokta oluşturulmasını tahmin etmek için k-en yakın komşu algoritması kullanılmaktadır. Veri

madenciliği tekniğine dayalı olan k-en yakın komşu modeli, tembel öğrenci algoritmasına da benzetilerek verilerin sınıflandırılmasında oldukça başarılı bir yöntem olarak kullanılmaktadır.

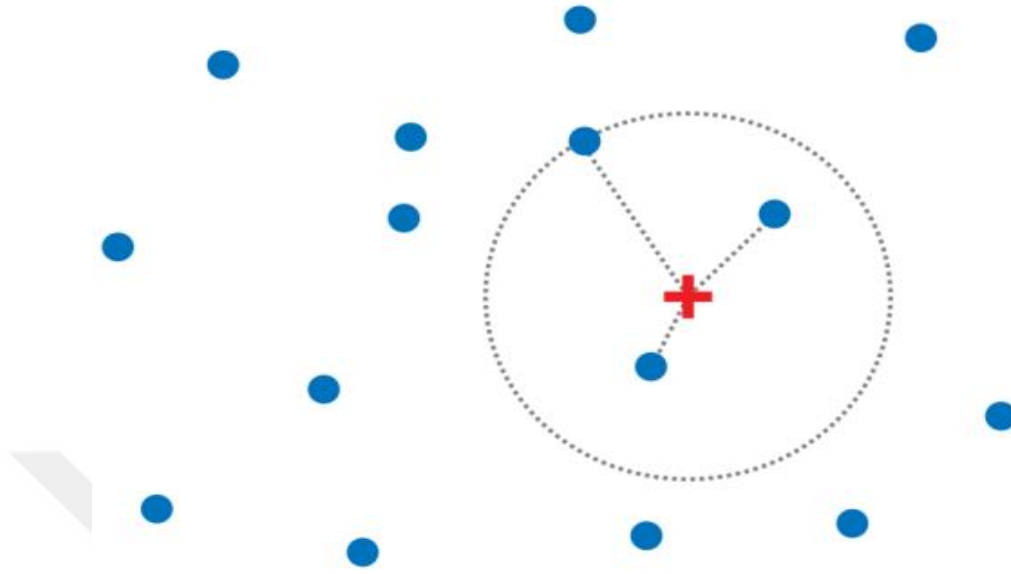
Bir örneklem üzerinde ki veri seti üzerinden nesneleri ön hazırlık süreci gerektirmeden kategorize ederek bir eğitim seti hazırlar. Eksik veriyi, k-en yakın komşular modeli kullanarak sınıflara ayırır. Modelin tahmin gücü, veri çıktısının kolayca yorumlanabilmesi ve çıktıların kısa sürede hesaplanabilmesi özelliğinden dolayı bu algoritmanın tercih edilmesindeki önemli hususlarından bazılarıdır.

Veri madenciliğinde sınıflandırma için kullanılan k-en yakın komşular yöntemi, sınıflandırma işleminde başarı elde etmesi k katsayısının iyi bir değer seçilmesine bağlıdır. Problemlere göre seçiminde farklılık gösteren k katsayısının seçilmesinde evrensel olarak iyi bir değerin seçilmesi belirsiz olup duruma göre deneme-yanılma yapılarak seçilmektedir. Sınıflandırma aşamasında etkisiz bir vektör olan k, genellikle pozitif tamsayı olarak seçilir. Eğer k değeri bir seçilirse (k=1) çıktı verisi en yakın komşusuyla beraber tahmin edilir. 'k' sayısının değeri azaldıkça tahmin hata değerleri de artmaktadır.

Veri setlerindeki eğitim ve test datalar arasındaki ilişkilendirmeyi yapmak için Öklid mesafesi kullanılır ve hesaplamalar bu mesafeye göre yapılır. Eşitlik (2.19)'da Öklid uzaklık formülü verilmiştir. Formüle göre n boyutlu, iki nokta arasındaki uzaklık hesabı yapılmaktadır.

$$d(x, y) = \sqrt{\sum_i^n (X_i - Y_i)^2} \quad (2.19)$$

Minimum ve maksimum mesafesi Öklid mesafesine göre hesaplandıktan sonra örnek veriler artan sırasına göre düzenlenmektedir. Buna göre Şekil-9'da k-en yakın komşu modeli örneği verilmiştir.



Şekil 10. K-en yakın komşu modeli örneği [38]

Şekil-10'da verilen k-en yakın komşular modelinde $k=3$ olarak belirlenmiştir. Öklid fonksiyonları kullanılarak en yakın üç nokta uzaklığı bulunarak ağırlıklarına göre sınıflandırılması bu noktalara göre yapılmıştır.

2.11. Radyal Tabanlı M5 Model Ağacı (M5TREE)

Quinlan (1992) tarafından tanıtılan M5 model ağacı (M5Tree), karar ağacının (DT) geliştirilmiş algoritmasıdır. Önerildiğinden beri, M5Tree'nin son birkaç yılda tanıtılan etkili ve sağlam makine öğrenimi algoritmalarından biri olarak kullanılmaktadır. M5Tree model hesaplama yönteminin temelinde bağımlı değişkenleri kullanarak bağımsız değişkenlerin tahmin edilmesi esasına dayanmaktadır. Bu model, klasik karar ağacı modelinden farklı olarak hem kategorize edilmiş veriler için hem de nicel veriler için kullanılmaktadır. Geleneksel regresyon modellerinde tüm verileri tanımlamak için kapsayan tek denklem kullanılırken M5Tree model ağacı yönteminde ise veriler alt bölgelere bölünerek yaprak etiketleri atanmaktadır.

M5Tree model ağacı yönteminde, veriler içerisindeki sayısal değerleri tahmin etmesi ya da düğümler arasındaki ilişkiyi gösterebilmek için doğrusal regresyon denklemi kullanılmaktadır. M5Tree model ağacının temel yapısı incelenecek olduğunda ise, bir

ağacın yapısına benzer olacak şekilde kökler, dallar, düğümler ve yapraklardan oluşmaktadır. Kök kısmı, ilk düğüm yeri olarak en yukarı seviyede yer almaktadır. Yapraklara kadar uzanan dalları ve düğümlerden oluşan zincirlerden oluşmaktadır. Model üzerinde bulunan düğümler, bir tahmin değişkenini temsil etmekte olup her tahmin değişkeninde yeni bir dallanma oluşmaktadır. Dallar arasındaki ilişki incelenecek olduğunda ise, bir dal düğüme ulaşabilmek için üstündeki düğümden kopan sayısal bir aralık içermekte olup her bir üst düğüm yeri alt kısmından ikiye doğru bölünmektedir.

Bir M5Tree modeli, yapraklarda karar ağacını ve doğrusal regresyon fonksiyonlarını birleştirmektedir. Bir M5Tree modeli geliştirmek iki aşama gereklidir. Birincisi verileri birkaç alt kümeye bölerek ve standart sapma azaltımını sağlayarak bir bölme kriteri olarak karar ağacı modeli oluşturmak ikincisi ise birkaç alt ağaç budanarak doğrusal regresyon fonksiyonları ile değiştirilerek oluşturulur. Bölme kriteri standart sapma azaltılmasını en üst seviyeye çıkarmaktadır. Standart sapma daha fazla düşürülemeyecek hale geldiğinde ise ana düğüm bölünmesi son bulur ve son düğüm yani yaprak elde edilir. Buna göre standart sapma azalması aşağıdaki eşitlik (2.20) ve (2.21)'de ki gibi hesaplanmaktadır [39].

$$\sigma_r = \sigma(t) - \sum_{i=1}^n \left(\frac{t_i}{t}\right) \sigma(t_i) \quad (2.20)$$

$$\sigma = \sqrt{\left(\frac{1}{n}\right) \left[\sum_{i=1}^n Y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n Y_i \right)^2 \right]} \quad (2.21)$$

Eşitlik (2.17)' ye göre (σ_r) alt düğümdeki standart sapma azalmasını, (t_i) üst düğüme giriş verilerin alt kümesini, (t) ana düğüme giriş verilerini ifade etmektedir. Eşitlik (2.18)'de ise σ standart sapma ifadesinin hesaplama formülünü göstermektedir. Alt düğümdeki standart sapma değeri bölme işleminden kaynaklı ana düğümün standart sapmasından daha az ve daha fazla homojen verilerin dağıldığını görmekteyiz.

2.12. En Küçük Kareler Destek Vektör Makinesi (LSSVM)

Vladimir Vapnik ve arkadaşları (1995) tarafından öner sürülen En Küçük Kareler Destek Vektör Makinesi (LSSVM), giriş ve çıkış verileri arasındaki doğrusal olmayan ilişkiyi minimum hata payı ile hesaplamak için tasarlanan makine öğrenimi yöntemidir. SVM modelinden türetilen LSSVM, sınıflandırma, tahmin ve regresyon problemleri için

karşılaşılan durumlarda çözüm üretmek için kullanılan bir yöntemdir. Suykens ve Vandewalle (1999) tarafından karmaşık bir zaman serisi kullanarak LSSVM modeli projelendirilmiştir.

SVM ile LSSVM arasındaki ilişki kıyaslandığında, LSSVM giriş verilerinin eğitimi aşamasında doğrusal denklemler kullanırken, SVM ikinci dereceden optimize edilmiş denklemler kullanmaktadır. Diğer makine öğrenimi yöntemlerinden [(Kısmi En Küçük Kare Regresyon (PLS), Geri Yayılımlı Sinir Ağları (BPNN), Çok Değişkenli Doğrusal Regresyon Modeli (MLR)] gibi geleneksel modellere göre matematiksel hesaplama açısından daha kapsamlı olması daha kolay uygulanabilir olduğunu ortaya koymaktadır. LSSVM modeli, mevcut zaman serisindeki parametreleri önceki zaman serisindeki parametreler ile kıyaslayarak model için en uygun giriş ve çıkış verilerini bulmak için kullanılmaktadır. Doğrusal olmayan ve karmaşık yapıdaki problem çözümünde kullanılması uygun bir modeldir. Buna göre eşitlik (2.22)'de regresyon fonksiyonu verilmiştir [40].

$$q(p) = \text{sign} [w^T p + b] \quad (2.22)$$

Eşitlik (2.22)'de verilen regresyon fonksiyonu iki sınıfın verileri ayrılabilir olduğu zaman ise eşitlik (2.23)'de gösterilmiştir.

$$\begin{cases} w^T Pk + b \geq +1, \text{ eğer } qk = +1 \\ w^T Pk + b \leq -1, \text{ eğer } qk = -1 \end{cases} \quad (2.23)$$

Eşitlik (2.23)'de verilen hesaplama formülünde yer alan (Pk) çıktı verilerini ifade ederken R sınıf eğitim kümesinin elamanı olup $(-1, +1)$ aralığında yer alan doğrusal bir sınıflandırıcıdır.

Bir öğrenme algoritması olarak kullanılan LSSVM'nin temel amacı, regresyon yönteminde karşılaşılan veri setlerini kullanarak belirli bir fonksiyon ile kullanarak en iyi modellemesini sağlamaktır. SVM ile kıyaslandığı zaman ise temelinde regresyon problemlerinde uygulanabilir olduğu görülmektedir. LSSVM'nin temel ilkesinde veri noktalarının hiper düzlem üzerinde veya bir fonksiyon ile en küçük kareler ilkesine göre optimizasyonunu sağlamaktır. LSSVM'nin optimizasyon ilkesi, regresyon problemlerinde hem hatayı minimize etmek hem de ayrık olmayan çözümler üretmek için en küçük kareler ilkesini kullanmaktadır. Ayrıca yüksek boyutlu verilerde çalışabilme

yeteneği sayesinde karmaşık sınıflandırma problemlerinde de tercih edilen bir öğrenme modelidir.

2.13. Performans Kriterleri

Sınıflandırma ve regresyon gibi farklı çalışmaların değerlendirilmesi için farklı değerlendirme bulunmaktadır. Bunun içinde makine öğrenim algoritmalarının uygun olanının seçilmesi için uygun metriklerin kullanılması gerekmektedir. Bu tez kapsamında su kalitesi tahmini için kullanılan modellerin performans değerlendirilmesi işleminde en başarılı model performansını seçmek için; Determinasyon Katsayısı (R^2), Ortalama Mutlak Yüzde Hata (MAPE), Bağıl Kök Ortalama Karesel Hata (RRMSE), Ortalama Mutlak Hata (MAE), Nash-Sutcliffe Etkinliği (NSE), Kling-Gupta Verimliliği (KGE) , model performansının genel indeksi (Overall Index (OI)) ve Wilmott'un rafine indeksi kullanarak sekiz metrikten oluşan performans değerlendirme yöntemleri kullanılmıştır. Bu metrikler günümüzde yaygın olarak kullanılan performans değerlendirme ölçütleri arasında yer almaktadır.

2.13.1. Ortalama Mutlak Yüzde Hata (MAPE)

İstatistiksel bir ölçü aracı olan Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error- MAPE), tahmin modelinin gerçek değeri ile ne kadar uyumlu olduğunu ölçen hata metriği arasında yer almaktadır. Özellikle öngörü problemlerinde ve tahmin etme problemlerinin hesaplanmasında sıklıkla kullanılan MAPE değerinin formülü eşitlik (2.24)'de gösterilmektedir.

$$MAPE (\%) = \frac{1}{n} \cdot \sum_{i=1}^n \left| \left(\frac{y_{model,i} - y_{gercek,i}}{\bar{y}_{model}} \right) \right| \quad (2.24)$$

Eşitlik (2.17)'ye göre, veri noktalarının toplam sayısını (n), model tahmin değerini ($y_{model,i}$), gözlenen gerçek değeri ($y_{gercek,i}$), model tahmin değerlerinin ortalamasını ise ($\bar{y}_{model}$) ifade etmektedir. MAPE değeri sıfıra yaklaştıkça tahmin değerlerinin gerçeğe o derece yaklaştığı söylenebilir. Tablo-5'de MAPE performans kriterlerinin sınıflandırılması verilmiştir.

Tablo 5. MAPE hata kriterlerinin sınıflandırılması. [41]

MAPE Derecelendirmesi (%)	Performans
% 50 < MAPE	Hatalı Öngörü
% 20 < MAPE ≤ % 50	Mantıklı Öngörü
% 10 < MAPE ≤ % 20	İyi Öngörü
MAPE ≤ % 10	Yüksek Başarılı Öngörü

2.13.2. Bağlı Kök Ortalama Karesel Hata (RRMSE)

Tahmin değerlerinin gerçek değerleri ile ne kadar uyumlu olduğunu ölçmek için Bağlı Kök Ortalama Karesel Hata (Rrelative Root Mean Squared Error-RRMSE) performans ölçütü kullanılmaktadır. RRMSE, yüzde cinsinden ifade edilerek yüzde cinsinden hata payını temsil etmektedir.

Eşitlik (2.25)'de yer verilen RRMSE hesaplama formülü aşağıda verilmiştir.

$$RRMSE = \sqrt{\frac{\frac{1}{n} \cdot \sum_{i=1}^n (y_{model,i} - y_{gercek,i})^2}{\frac{1}{n} \cdot \sum_{i=1}^n y_{gercek,i}}} \quad (2.25)$$

RRMSE değeri sıfıra yaklaştıkça tahmin değerlerinin gerçek değerine yaklaşması artmaktadır. RRMSE değerinin küçük olması, tahmin modelinin daha iyi performans gösterdiğini ve daha düşük hata paylarının olduğu sonucuna varılmaktadır. RRMSE değerinin büyük olması ise tahmin modelinin gerçek değerinden uzaklaştığı ve hata paylarının yüksek sonuç verdiği sonucuna ulaşılmıştır. Buna göre Tablo-6'da yer alan RRMSE hata kriteri sınıflandırılması aşağıdaki tabloda verilmiştir.

Tablo 6. RRMSE hata kriterleri sınıflandırılması. [42]

RRMSE Derecelendirmesi	Performans
%30 < RRMSE	Hatalı Tahmin
%20 < RRMSE ≤ %30	Mantıklı Tahmin
%10 < RRMSE ≤ %20	İyi Tahmin
RRMSE ≤ %10	Yüksek Başarılı Tahmin

2.13.3. Ortalama Mutlak Hata (MAE)

Tahmin hatalarının mutlak büyüklüğünü değerlendirmek için kullanılan istatistiksel değerlendirme ölçütü olan Ortalama Mutlak Hata (Mean Absolute Error-MAE), gerçek veri setlerinin değeri ile modellenerek tahmini yapılan veri setleri arasındaki mutlak hata cinsinden birbirine olan uzaklık miktarını göstermektedir. MAE değeri sıfıra yaklaştıkça modelin başarısı da o derece yüksek olarak değerlendirilmektedir. Eşitlik (2.26)'de MAE değeri için hesap formülü gösterilmiştir.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{model,i} - y_{gercek,i}| \quad (2.26)$$

2.13.4. Determinasyon Katsayısı (R²)

Oluşturulan regresyon modelinin doğrusal bir eğri üzerinde ne kadar iyi uyum sağladığının bir ölçüsü olan determinasyon katsayısı, bağımsız girdi verilerinin bağımlı çıktı verilerine göre oluşturulan yüzde ifadesi olarak açıklanır. Determinasyon katsayısı ifadesi değeri 0 ile 1 arasında değer almaktadır. Çıktı verileri 1'e yaklaştıkça modelin iyi uyum sağladığını gösterirken, çıktı verilerinin 0'a doğru yaklaştıkça da modelin kötü uyum sağladığını göstermektedir. Determinasyon katsayısı açıklanan varyasyonların toplam varyasyonlarına oranı olarak da bilinmektedir. Buna göre aşağıda yer alan eşitlik (2.27)'de determinasyon katsayısının hesaplama formülü verilmiştir.

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \left[\frac{\sum_{i=1}^n (y_{model,i} - y_{gercek,i})^2}{\sum_{i=1}^n (y_{model,i} - \bar{y}_{gercek,i})^2} \right] \quad (2.27)$$

Tablo 7'de R² hata kriterlerinin performans değerlerine göre sınıflandırılması verilmiştir.

Tablo 7. R² hata kriterlerinin sınıflandırılması. [42]

R ² Derecelendirmesi	Performans
0.80 < R ² ≤ 1.00	Çok İyi
0.60 < R ² ≤ 0.80	İyi
0.40 < R ² ≤ 0.60	Normal
R ² ≤ 0.40	Yetersiz

2.13.5. Nash-Sutcliffe Etkinliği (NSE)

Hidrolojik modellerin performans tahmini için kullanılan bir metrik olan Nash-Sutcliffe Etkinliği (Nash-Sutcliffe Efficiency –NSE) katsayısı, gerçek ve model değerlerinin hangi oranda birbirine benzediğinin bir ölçüsüdür. NSE katsayısı, $-\infty$ ile 1 arasında problemin durumuna göre farklı değerler almaktadır. Eşitlik (2.28)’de Nash-Sutcliffe etkinliğinin katsayı hesaplama formülü verilmiştir.

$$NSE = 1 - \frac{\sum_{i=1}^n (y_{model,i} - y_{gercek,i})^2}{\sum_{i=1}^n (y_{gercek,i} - \bar{y}_{gercek})^2} \quad (2.28)$$

Tablo 8’de ise Nash-Sutcliffe etkinliği katsayısına ait hata kriter performanslarına göre sınıflandırılması verilmiştir.

Tablo 8. NSE Hata Kriterlerinin Sınıflandırılması .[42]

NSE Derecelendirmesi	Performans
$0.75 < NSE \leq 1.00$	Çok İyi
$0.65 < NSE \leq 0.75$	İyi
$0.50 < NSE \leq 0.65$	Tatmin Edici
$0.40 < NSE \leq 0.50$	Kabul Edilebilir
$NSE \leq 0.40$	Yetersiz

2.13.6. Kling-Gupta Verimliliği (KGE)

Kling-Gupta Verimliliği (Kling Gupta Efficiency) hidrolojik modellerin tahmin performansı için kullanılan bir metriktir. Kling-Gupta Verimliliği gerçek değerler ve model değerler arasında ki benzerlikleri ve farklılıkları değerlendiren bir kriteridir. Kling-Gupta Verimliliği $-\infty$ ile 1 arasında değer almaktadır. KGE değeri 1’e doğru yaklaştıkça modelin gösterdiği performans değeri de yükselmektedir. Buna göre aşağıda yer alan (2.29), (2.30), (2.31), (2.32) eşitlikler KGE hesaplama formülünü göstermektedir.

$$KGE = 1 - (\sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}) \quad (2.29)$$

$$r = \frac{\sum_{i=1}^n [(y_{model,i} - \bar{y}_{model}) \cdot (y_{gercek,i} - \bar{y}_{gercek})]}{\sum_{i=1}^n (y_{model,i} - \bar{y}_{model})^2 \cdot \sum_{i=1}^n (y_{gercek,i} - \bar{y}_{gercek})^2} \quad (2.30)$$

$$\beta = \frac{\bar{y}_{model}}{\bar{y}_{gerçek}} \quad (2.31)$$

$$\gamma = \frac{\frac{\sigma_{gerçek}}{\bar{y}_{gerçek}}}{\frac{\sigma_{model}}{\bar{y}_{model}}} \quad (2.32)$$

Tablo 9. KGE hata kriterlerinin sınıflandırılması. [42]

Range of KGE	Performance
$KGE \leq 0.40$	Yetersiz
$0.40 < KGE \leq 0.50$	Kabul Edilebilir
$0.50 < KGE \leq 0.60$	Tatmin Edici
$0.60 < KGE \leq 0.70$	İyi
$0.70 < KGE \leq 1.00$	Çok İyi

2.13.7. Overall Index (OI)

Overall Index (OI), gerçek değer ile tahmin modeli arasındaki uyumu istatistiksel bir ölçüt olarak değerlendiren metriktir. Overall Index, ∞ ile 1 değer aralığında yer almaktadır. Overall Index değeri 1'e doğru yaklaştıkça tahmini yapılan modelin gerçek ile uygunluğunun çok yüksek oranda benzemektedir. Overall Index formülü eşitlik (2.33), (2.34)'de verilmiştir.

$$RMSE = \frac{1}{n} \cdot \sum_{i=1}^n \sqrt{(y_{model,i} - y_{gerçek,i})^2} \quad (2.33)$$

$$OI = \frac{1}{2} \left(2 - \frac{RMSE}{y_{gerçek,max} - y_{gerçek,min}} - \frac{\sum_{n=1}^N (y_{gerçek,i} - y_{model,i})^2}{\sum_{n=1}^N (y_{gerçek,i} - \bar{y}_{gerçek})^2} \right) \quad (2.34)$$

2.13.8. Willmott Rafine İndeksi (WI)

1980 yılında yılın da Willmott ve arkadaşları tarafından geliştirilen Willmott Rafine İndeksi, son derece karmaşık ve geniş kapsamlı model performanslarında oldukça rahat uygulanabilir istatistiksel yaklaşımlarda kullanılmaktadır. Model tahminlerinin birbirleri ile karşılaştırılması metoduna dayanmaktadır. Willmott ve çalışma arkadaşları, model-

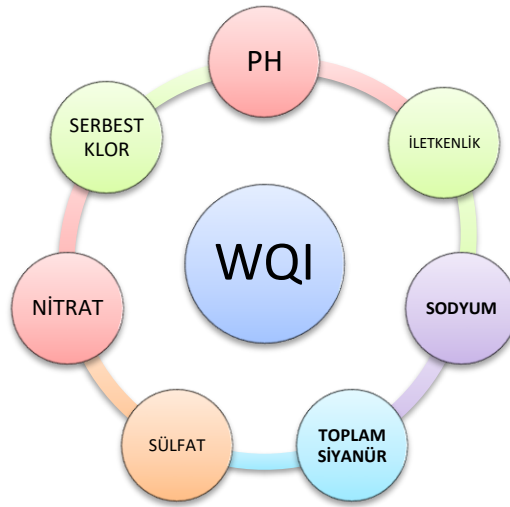
performans problemlerinde çok sayıda uygulama yapıldığında hataları toplamadan önce karelerinin alınmasının bir sorun teşkil edeceğini düşünmeleri üzerine daha büyük hata değerlerinin karesi alındığında, bu hataların karesi alınan hataların toplamı üzerindeki etkisini araştıran istatistiksel hesaplama formülü ortaya koymuşlardır. Buna göre eşitlik (2.35)'de verilen Willmott Rafine İndeksi hesaplama formülü verilmiştir [43].

$$d = 1 - \frac{\sum_{i=1}^n |(Pi-\bar{O})-(Oi-\bar{O})|^2}{\sum_{i=1}^n |(Pi-\bar{O})+(Oi-\bar{O})|^2} \quad (2.35)$$

2.14. Su Kalite İndeksi (WQI)

Çevre için çok önemli bileşen olan su, yüzey ve yer altı sularından oluşmaktadır. Fakat hem insan hem de doğal faaliyetlerden dolayı su kalitesi bozulmaktadır. Su kalitesini etkileyen doğal faaliyetlere hidrolojik, iklimsel, atmosferik ve topografik faaliyetler örnek gösterilirken; su kalitesini olumsuz etkileyen faaliyetlere hayvancılık, madencilik, atık üretimi ve bertarafı ve ağır metal kirliliği gibi faaliyetler örnek gösterilebilir. Son zamanlarda nüfus artışıyla beraber su hizmetleri ve su kalitesinde oluşan sorunları iyileştirmek için su kalitesi indeks çalışmalarına olan yönelim artmıştır.

Su Kalite İndeksi (WQI), zamansal ve mekânsal olarak değişen büyük su kalitesi veri setlerinin analize izin verilmesini sağlayan tek bir değerdir. Yani su kütlesinin kalitesini gösteren, su kalitesi indeksi üretmesine izin veren bir toplama fonksiyonudur. WQI tipik olarak dört bileşenden oluşmaktadır. Bunlardan birincisi, su kalitesi parametrelerinin seçilmesidir. İkincisi, su kalitesi verileri okunarak her bir su kalitesi parametresi için yoğunlaşma değerlerinin tek değerli boyutsuz bir alt indekse dönüştürülmesidir. Üçüncüsü, her bir su kalitesi parametresi için ağırlıklandırma faktörünün belirlenmesidir. Dördüncüsü ise, tüm su kalitesi parametrelerini ağırlıklandırma faktörlerini kullanarak bir toplama fonksiyonu ile birleştirerek tek değerli bir su kalitesi indeksi oluşturmaktır. Buna göre bu tez çalışması kapsamında WQI değeri için kullanılan su kalitesi parametreleri Şekil-11'de verilmiştir.



Şekil 11. WQI değeri için kullanılan su kalitesi parametreleri

WQI modelinde toplama işlemi, su kalitesi parametreleri alt endekslerinin tek bir değerde toplanmasında uygulanır. Buna göre eşitlik (2.36)'da verilen hesaplama formülü su kalitesi indeksini belirlemek için kullanılan toplama fonksiyonunu temsil etmektedir.

$$WQI = \sum_{i=1}^n S_i W_i \quad (2.36)$$

Eşitlik (2.36)'da verilen (W_i) değeri, fonksiyonun ağırlık değeri olup 0 ile 1 arasında değişen değer almaktadır.

WQI parametre değerlerini alt indekse dönüştürmek için su kalitesinin yoğunluk konsantrasyon seviyesine veya kirlilik durumuna göre sınıflandırıldığı doğrusal bir ölçeklendirme işlevi vardır. Alt indeks değerleri 0-100 arasında değişmekte olup en kötü kalitedeki koşullar için 0 değeri atanırken mükemmel sonuç veren su kalitesinde ise 100 değeri atanır. Buna göre su kalitesi sınıf değerlerine göre gruplandırılmış değerleri Tablo-10'da verilmiştir.

Tablo 10. WQI modeli su kalitesi sınıfları. [43]

WQI Modeli Su Kalitesi Sınıfı	
WQI= 90-100	Mükemmel
WQI= 70-89	İyi
WQI= 50-69	Orta
WQI= 25-49	Kötü
WQI= 0-24	Çok Kötü Kalite

3. BÖLÜM

BULGULAR

3.1. Model Gelişimi

Son dönemlerde hızla gelişen yapay zekâ uygulamaları birçok alanda kullanılmaktadır. Bu tez kapsamında yapılan çalışmalar ise, su kalitesi tahmini için kullanılan geleneksel yöntemler yerine daha hızlı ve pratik olarak kullanılan yapay zekâ kullanım alanlarından birisi olan subilim dalına ait çalışmalarına örnek gösterilebilir. Yapılan tahmin modelleme çalışmaları için kullanılan veriler ise; Kayseri ilinde yer alan Su ve Kanalizasyon İdaresi Genel Müdürlüğüne bağlı 2015-2022 yılları arasındaki içme suyu şebeke verileridir. Bu veriler, merkez ilçelere ait olmakla beraber toplamda 14 gözlem evi istasyonundan toplanan su kalitesi parametre verileridir. Alınan su kalitesi parametreleri modelleme çalışmaları için girdi verisi olarak kullanılmakta olup toplamda 7 adet girdi parametresi kullanılmıştır. Bu parametreler ise şunlardır: pH, iletkenlik, sodyum, serbest klor, nitrat, sülfat ve nitrat olup bu parametreler yılın ayları şeklinde düzenli olarak ölçülmüştür. Modellemelerde girdi parametrelerini kullanarak çok sayıdaki değişkenden tek değerli sonuç elde edilen Su Kalitesi İndeksi (WQI) değeri 1 adet çıkış verisi olarak kullanılmıştır. Bu verileri MATLAB 2023a programı yardımıyla toplamda 9 adet makine öğrenmesi modeli kullanarak içme suyu kalitesi tahmin modelleri elde edilmiştir. Sonuçlar ve modeller birbiri arasında karşılaştırılmıştır. Kullanılan makine öğrenme modelleri ise şunlardır: ANN, ANFIS, GPR, MARS, SVM, LSBoost, KNN, M5TREE, LSSVM'dir. Modellemelerde kullanılan veri sayısı toplamda 74 adet olup bu verilerin 59 (%80) tanesi eğitim verisi, 15 (%20) tanesi test veri olarak kullanılmıştır. Eğitim verileri model elde etmek için kullanılırken, test verileri ise oluşturulan modelin doğruluğunu test etmek için kullanılmıştır. Tahmin modelleri ise test verilerini kullanarak elde edilmiştir.

3.2. Makine Öğrenmesi Modellerinin Elde Edilmesi

Su kalitesi parametreleri kullanarak toplamda 9 farklı makine öğrenmesi yöntemleriyle modellemeler yapılarak WQI tahmini ve modellerin tahmin performansları karşılaştırılmıştır. Kullanılan yöntemler, ANN, ANFİS, GPR, MARS, SVM, LSBoost, M5TREE, LSSVM olup aşağıda yapılan model çalışmaları anlatılmıştır. Model çalışmaları MATLAB 2023a yazılımı kullanarak oluşturulmuştur.

3.2.1. ANFIS Model Gelişimi

ANFIS tahmin modelinde, bulanık çıkarım sistemi olarak Takagi-Sugeno yaklaşımı kullanılmıştır. Optimizasyon ve uyarlanabilir teknikleriyle beraber ele alınan Takagi-Sugeno yaklaşımı, Mamdani yaklaşımı ile kıyaslandığında hesap adımları daha pratik olup işlem adımlarında daha başarılı sonuçlar veren bir yaklaşımdır. Sinir ağı modeli olarak ızgara kümeleme (grid partition) yöntemi kullanılmıştır. Izgara kümeleme yöntemi kullanılmasının sebebi ise, daha az bağlantı sayısı ile katmanların birbirine daha hızlı bağlanmasından kaynaklanmaktadır. ANFIS’de öğrenme algoritması olarak hem sinir ağı kurallarını kolaylıkla öğrenebilen hem de bulanık mantık öğrenmesi olan hibrit algoritması kullanılmıştır. Buna ek olarak bu algoritmanın tercih edilmesinin bir diğer sebebi ise, diğer öğrenme algoritmalarına göre daha hızlı olmasıdır. ANFIS model mimarisinde giriş ile ara tabaka arası üyelik fonksiyonları olarak Gauss (gaussmf) ve üçgen (triangular;trimf), ara tabaka ve çıkış üyelik fonksiyonları arasında ise doğrusal (linear) ve sabit (constant) fonksiyonlarını deneme-yanılma sonucunda kullanılmıştır. Ayrıca deneme-yanılma ile tahmin modelleri, 2 ve 3 üyelik fonksiyonu ve iterasyon sayısı 1 ile 5 arasında denenmiştir. Bu deneme-yanılmalar sonucunda en düşük MAPE ve en yüksek R^2 değerini veren 2 üyelik fonksiyon sayısı, 5 iterasyon sayısı, Gauss (gaussmf) giriş üyelik fonksiyonu ve Sabit (constant) çıkış üyelik fonksiyonu ANFIS mimari yapısında tespit edilmiştir.

ANFIS modelinde, pH, iletkenlik, sodyum, serbest klor, nitrat, sülfat ve nitrat olarak toplamda 7 adet giriş verisi girilirken, 1 adet çıkış verisi olarak ise WQI değeri girilmiştir. Modelleme sonucunda ise determinasyon katsayısı (R^2) 0.985 değerini vermiş olup sonucun 1’e yaklaşma durumuna göre tahmin modelinin mükemmelliği de artmaktadır.

Bu çalışmada kullanılan ANFIS yönteminde kullanılan model parametreleri Tablo-11’de verilmiştir.

Tablo 11. ANFIS tahmin modeli parametreleri.

Parametre adı	Parametre özelliği veya değeri
Bulanık çıkarım tipi	Takagi-Sugeno
Sinir ağı modeli	Izgara kümeleme (Grid partition)
Girdi sayısı	7 adet (pH, iletkenlik, sodyum, serbest klor, nitrat, sülfat ve nitrat)
Çıktı sayısı	1 adet (WQI)
İterasyon sayısı	5
Üyelik fonksiyon sayısı	2
Giriş üyelik fonksiyonları	Gauss (gaussmf)
Çıkış üyelik fonksiyonları	Sabit (constant)
Eğitme optimizasyonu	Hibrit algoritması

3.2.2. ANN Model Gelişimi

WQI değerini tahmin etmek için kullanılan YSA modelinde, deneme –yanılma işlemleri sonucunda en iyi YSA mimari yapısı belirlenmeye çalışılmıştır. YSA mimari yapısında çok katmanlı YSA modellerinden yararlanılmıştır. YSA, tahmin problemlerindeki başarısı ve güçlü öğrenme algoritmaları sebebiyle WQI tahmini için bu çalışmada tercih edilmiştir. YSA tahmin modellerine girecek olan eğitme ve test verileri, tüm veri setinin %80’i eğitme ve %20’si test olacak şekilde kabul edilmiştir. Giriş verileri, diğer öğrenme yöntemlerine göre daha doğru tahminler veren ve daha hızlı sonuçlar veren sayısal bir optimizasyon tekniği olan Levenberg-Marquardt öğrenme algoritması (trainlm) ile eğitilmiştir. YSA modelinde, bilgi akışının iki yönlü olarak sağlanması istenildiği için geri beslemeli ağ yapısı tercih edilmiştir. Ayrıca bu çalışmada bir ara (gizli) katmana sahip çok katmanlı bir sinir ağı modeli kullanılmıştır. Katmanlar arası bilgi aktarımını sağlayan transfer fonksiyonu olarak logaritmik sigmoid fonksiyonu (logsig) ve hiperbolik tanjant sigmoid (tansig) deneme-yanılma ile belirlenmiştir. Girdilere dayalı olarak

çıktıların özelliklerini belirleyen ve çıktıların elde edilmesini sağlayan aktivasyon fonksiyonu olarak ise doğrusal transfer fonksiyonu (purelin), logsig ve tansig fonksiyonları denenmiştir. Çalışmada tercih edilen transfer ve aktivasyon fonksiyonları ile birlikte 2 ile 14 arası nöron sayısı ve 2 ile 100 arası iterasyon sayısı denenerek tahmin modelleri oluşturulmuştur. Tablo-12’de YSA tahmin modellerine ait parametreler özet olarak verilmiştir.

Tablo 12. YSA tahmin modeli parametreleri.

Parametre adı	Parametre özelliği veya değeri
Sinir ağı modeli	Çok katmanlı
Ağın bağlantı yönü	Geri beslemeli
Öğrenme algoritması	Levenberg-Marquardt (trainlm)
Girdi sayısı	7 adet
Çıktı sayısı	1 adet
İterasyon sayısı	16
Nöron sayısı	5
Transfer fonksiyonları	Logaritmik sigmoid (logsig)
Aktivasyon fonksiyonu	Doğrusal transfer fonksiyonu (purelin)

Bu denemeler sonucunda en düşük MAPE ve en yüksek R^2 değerini veren bir gizli katmanda logsig ve çıkış katmanında purelin aktivasyon fonksiyonları, 1 ara katmandaki 5 gizli nöron sayısı ve 16 adet iterasyon sayısı bulunarak YSA mimari yapısı belirlenmiştir. Modelleme sonucunda ise determinasyon katsayısı (R^2) 0,9901 değerini vermiş olup sonucun 1’e yaklaşma durumunu göre tahmin modelinin mükemmelliği de artmaktadır. R^2 değerine göre yapılan modellemeler arasında en iyi sonucu YSA modeli vermiştir.

3.2.3. GPR Model Gelişimi

GPR yöntemi bu çalışmada ele alınan bir diğer makine öğrenmesi yöntemidir. GPR yönteminin tercih edilmesinin başlıca nedenlerinden birisi de tahmin problemleri kullanılırken daha az veriye ihtiyaç duyması ve parametrik olmayan öğrenme algoritması olmasıdır. Çekirdek (kernel) fonksiyonları ile oldukça iyi uyum gösteren GPR, tahmin problemlerindeki başarısından kaynaklı olarak bu tez çalışmasında tercih edilmiştir. GPR modellerindeki en kritik parametre, optimum kovaryans matrisi ve buna bağlı olarak oluşturulan kovaryans fonksiyonunun belirlenmesidir. Bu tez çalışmasına göre; GPR'nin tahmin aralığını kapsamlı bir şekilde belirleyebilmek ve model performanslarını artırabilmek için birçok farklı fonksiyonlar kullanılmıştır. GPR modellerinde yaygın olarak kullanılan fonksiyonlar olarak radyal tabanlı temel fonksiyonlar ve diğer çekirdek fonksiyonları olarak sınıflandırılmaktadır. Buna göre radyal tabanlı temel fonksiyonlara doğrusal, sabit, sıfır, saf ikinci dereceden fonksiyonlar örnek verilirken çekirdek fonksiyonlarına ise üstel, karesel üstel, rasyonel ikinci dereceden, matern 3/2, matern 5/2, ard üstel, ard karesel üstel, ard rasyonel ikinci dereceden, ard matern 3/2 ve ard matern 5/2 fonksiyonları örnek verilmektedir. Buna göre su kalitesi tahmin modellemesi çalışmasında GPR modelinde toplamda 10 adet kernel fonksiyonu kullanılırken 3 adet de basic fonksiyonları kullanılmıştır. Kernel fonksiyonu olarak, ardexponential, ardmatern32, ardmatern52, matern52, matern32, ardrationalquadratic, ardsquaredexponential, exponential, rationalquadratic ve squaredexponential fonksiyonları kullanılmıştır. Basic fonksiyonları olarak ise constant, linear ve pureQuadratic fonksiyonları kullanılmıştır. Ayrıca bu kadar çok fonksiyonun kullanılmasının nedeni ise, WQI tahmini için en uygun olan modellemeyi seçmek için 9 farklı yöntemi MATLAB 2023a yazılım programı aracılığıyla rahatlıkla deneyerek yapay zekâ çalışmalarına örnek teşkil etmek içindir. GPR modeli ise bu yöntemlere örnek olurken tahmin işlemlerinin hızlı ve pratik bir şekilde yapılabilmesi de çalışmamıza kolaylık sağlamıştır.

GPR tahmin modellerinde, önceden herhangi bir ön parametre bilgisine ihtiyaç duymadan öğrenme algoritması ve olasılık tabanlı olarak kullanılan Bayes optimizasyon tekniğine yer verilmiştir. GPR modellerinde gerçek değerleri bulunmaya çalışılan tahmini ortalama değeri (β) bazı kaynaklarda (μ) olarak gösterilmektedir. Ayrıca çekirdek genişlik parametresi (σ), tamamen bağımsız koşullu yaklaşım (fully independent conditional

approximation) ile kullanılarak varyans problemlerinin önüne geçilerek çekirdek fonksiyonlarını daha da işlevsel hale getirmektedir. Kernel ölçeği GPR tahmin modellerinde de 1 alınmıştır. 1–30 aralığında değişim gösteren iterasyonlarla ve doğru standardizasyon yöntemi ile analizler yapılarak model çıktıları elde edilmiştir. GPR yöntemi kullanılarak oluşturulan tahmin modelleri, MATLAB 2023a yazılımı ile analiz edilmiştir. GPR tahmin modeli parametreleri Tablo-13’de özetlenmiştir

Tablo 13. GPR tahmin modeli parametreleri.

Parametre adı	Parametre özelliği veya değeri
Radyal tabanlı temel fonksiyonlar	constant, linear ve pureQuadratic
Çekirdek fonksiyonları	ardexponential, ardmatern32, ardmatern52, matern52, matern32, ardrationalquadratic, ardsquaredexponential, exponential, rationalquadratic ve squaredexponential
Parametre hesap yöntemi	Tamamen bağımsız koşullu yaklaşım
Optimizasyon algoritması	Bayes Optimizasyon Tekniği
Kernel ölçeği	1
İterasyon sayısı	1-30
Girdi sayısı	7 adet
Çıktı sayısı	1 adet

Deneme- yanılma sonucunda modele ait en düşük MAPE ve en yüksek R^2 bakımından en uygun GPR mimari yapısı olarak Kernel işlevinde ardsquaredexponential fonksiyonu belirlenmiştir. Modelleme sonucunda ise determinasyon katsayısı (R^2) 0,9791 değerini vermiş olup sonucun 1’e yaklaşma durumunu göre tahmin modelinin mükemmelliği de artmaktadır.

3.2.4. SVM Model Gelişimi

Çalışmada tercih edilen bir diğer makine öğrenmesi yöntemi olarak SVM modeli kullanılmıştır. Yapısal risk minimizasyonu ve istatistiksel öğrenme teorisine dayalı bir yöntem olan SVM'nin tahmin edilmesinde yani regresyon problemlerine yönelik çözüm geliştiren bir alt grubu olarak kullanılan SVMR (Destek Vektör Makinesi Regresyonu) geliştirilmiştir. Doğrusal olmayan problemlerde çekirdek (kernel) fonksiyonları vasıtasıyla kolaylıkla çözüm üretebilmesi ve iyi bir genelleme yapabilmesinden dolayı tercih edilmiştir. SVM tahmin modelleri, MATLAB 2023a yazılım programında geliştirilerek açık kodlar kullanımıyla beraber tamamlanmıştır. SVM tahmin modellerinde; Gaussian, polinom ve doğrusal (lineer) olmak üzere üç farklı çekirdek fonksiyonu kullanılmıştır.

Ayrıca, Sıralı Minimum Optimizasyon (SMO) yöntemi SVM'nin öğrenme aşaması esnasında ortaya çıkan ikinci dereceden programlama problemlerini çözen modellerde optimizasyon algoritması olarak kullanılmıştır. SVM modellerinde kutu sınırlama seviyesini belirlemek için optimum seviyede hiper düzlem aralığı için gerekli işlemler ve parametre hesaplamaları bu aralıkta yapılmaktadır. Bu model çalışmasında, kutu sınırlama seviyesi 1 olarak seçilmiştir. Tahmin modellerindeki girdi ve çıktılar, 1–30 aralığında değişen iterasyonlarla ve doğru standardizasyon yöntemi ile analiz edilerek elde edilmiştir. Buna göre SVM tahmin modeli parametreleri Tablo-14'de özetlenmiştir.

Tablo 14. SVM tahmin modeli parametreleri.

Parametre adı	Parametre özelliği veya değeri
Standardize yöntemi	Doğru
Optimizasyon algoritması	Sıralı Minimum Optimizasyon (SMO)
Çekirdek fonksiyonları	Gauss, Polinom, ve Doğrusal
İterasyon sayısı	1-30
Girdi sayısı	7 adet
Çıktı sayısı	1 adet (WQI)
Kernel ölçeği	1
Kutu sınırı	1-1000
C parametresi	Otomatik
ϵ parametresi	0.09

Deneme-yanılma sonucunda modele ait en düşük MAPE ve en yüksek R^2 bakımından en uygun SVM mimari yapısı olarak Kernel işlevinde lineer fonksiyonu belirlenmiştir.

Lineer fonksiyonu en yüksek determinasyon katsayısı değerini vermiş olup R^2 değeri 0,8681 değerini vermiştir. R^2 değerinin 1'e yaklaşmasına göre oluşturulan tahmin modellerinin de gerçeğe uygunluğu yaklaşımaktadır.

3.2.5. M5TREE Model Gelişimi

Yönetimsel (managerial), WQI parametresinin tahmin edilmesi için kullanılan bir diğer yöntem ise M5Tree modelidir. M5Tree yöntemi birkaç basit doğrusal bağlantının kombinasyonu ile tahmin yapmak için kullanılmaktadır. Bu modelde kullanılan su kalitesi parametre değerleri giriş verisi olarak seçilmiş olup düğümün temsil edebileceği minimum eğitim gözlemi sayısı 2 ve yumuşatma katsayısı (smoothing coefficient) olarak 15 alınmıştır. Wang and Witten (1997) yaptığı çalışmada, bölünmüş eşik sayısının azalması ile model sayının artacağını belirtmişlerdir. Model yapısının uzamaması için ise bölünmüş eşik (split threshold) değeri 0.05 alınmıştır (Wang & Witten, 1997, Heddam & Kisi, 2018). Düğümdeki çıkış değişkeni değerlerinin SD'si, tüm eğitim verilerine ait çıkış

değişkeninin SD'sinin bölünme eşiğinden küçük ise, düğümü bölmeye gerek kalmadan model sayısı durdurulmaktadır.

3.2.6. KNN Model Gelişimi

KNN algoritması, bu çalışmada kullanılan diğer tüm yöntemlere göre uygulaması en kolay yöntemdir. Eğitim setine bağlı olarak temsil yeteneği yüksek “k” sayısının belirlenmesinde deneme–yanılma yapılarak tespit edilmiştir. Deneme–yanılma yöntemlerinin yanı sıra her bir “k” katsayısı exhaustive arayıcısı yardımıyla eğitme verilerinde optimize edilmiştir. Ayrıca literatürde yer alan euclidean, cityblock, minkowski, chebychev, spearman, jaccard, mahalanob, correlation, hamming, cosine ve seuclidean olmak üzere toplamda 11 farklı mesafe fonksiyonu denenmiştir. Deneme ve yanılmaların sonucunda bütün kombinasyonlarda en düşük MAPE ve en yüksek R^2 değerine göre k sayısı 1 ve cosine mesafe fonksiyonu ($R^2 = 0,8347$) en iyi tahmin WQI değerini verdiği tespit edilmiştir.

3.2.7. MARS Model Gelişimi

Bu çalışmada, kesişme terimini ifade eden maksimum temel fonksiyon sayısı 1 ile 200 arasında; düğüm başına genelleştirilmiş doğrulama sayısı 1 ile 3 arasında; giriş değişkenleri arasındaki maksimum etkileşim derecesini ise 1 ile 3 arasında denenmiştir. Deneme–yanılmaların sonucunda bütün kombinasyonlarda en düşük MAPE ve en yüksek R^2 değerine göre maksimum temel fonksiyon sayısını (maxFuncs) 101, düğüm başına genelleştirilmiş doğrulama sayısı (c) 1, giriş değişkenleri arasındaki maksimum etkileşim derecesini (maxInteractions) ise 2 belirlenerek en iyi MARS mimari yapısı oluşturulmuştur.

3.2.8. LSSVM Model Gelişimi

LSSVM modeli, giriş ve çıkış verileri arasındaki doğrusal olmayan ilişkiyi minimum hata payı ile hesaplamak için tasarlanan bir makine öğrenimi yöntemidir. SVM modelinden türetilen LSSVM modeli, sınıflandırma, tahmin ve regresyon problemleri için karşılaşılan durumlarda çözüm üretmek için kullanılmaktadır. SVM ile LSSVM arasındaki ilişkiyi kıyasladığımızda ise, LSSVM modelinde giriş verilerinin eğitimi sırasında doğrusal denklemler kullanırken, SVM modelinde ise ikinci dereceden optimize edilmiş

denklemler kullanmaktadır. LSSVM modeli, mevcut zaman serisindeki parametreleri kullanarak önceki zaman serisindeki parametreler ile kıyaslayıp model için en uygun giriş ve çıkış verilerini bulmak için kullanılmaktadır. LSSVM'nin temel amacı, regresyon yönteminde karşılaşılan veri setlerini kullanarak belirli bir fonksiyon ile birleştirilmesi sonucunda en iyi modellemeyi oluşturmaktır. SVM ile kıyaslandığı zaman ise temelinde regresyon problemlerinde uygulanabilir olduğu görülmektedir.

Sigma ve gamma değerleri için 1'er fark artışlarla 1'den 50'ye kadar değişen sayılar denenmiş, en düşük RMSE değerini veren parametreler en iyi model parametresi olarak kabul edilmiş ve LSSVM modelinde Radial Basis fonksiyonu kullanılmıştır. Deneme-yanılmaların sonucunda bütün kombinasyonlarda en düşük MAPE ve en yüksek R^2 değerine göre sigma değerini 50 ve gamma değerini 4 veren en iyi LSSVM mimari yapısı tespit edilmiştir.

3.2.9. LSBoost Model Gelişimi

Bu çalışmada, optimum hiper parametrelerinin değerlerini ayarlamak ve elde etmek için Bayesian optimizasyon arama tabanlı sürecini kullandık. Ayarlama için odaklanılan üç ana hiper parametre, öğrenme hızı, eğitim döngüsü sayısı (ağaç maksimum derinliği) ve maksimum bölünmelerden oluşmaktadır. LSBoost algoritmasında öğrenme hızı (LearnRate) parametresi 0 ile 1 arasında değişmekte olup 1 alındığında topluluk maksimum hızda öğrenme göstermektedir. Bu çalışmada, LearnRate parametresi 0.01 alınmıştır. Ayrıca LSBoost yönteminin genel performansını iyileştirmek amacıyla öğrenme döngüsü sayısını (NumLearningCycles) 5000 ve ağaç başına maksimum karar bölümü sayısını (MaxNumSplits) ise 100 olarak alınmıştır.

3.3. Analiz Sonuçları

Su kalitesi parametrelerini kullanarak elde edilen WQI değerinin tahmini için ANFIS, ANN, GPR, MARS, SVM, LSBoost, KNN, M5TREE ve LSSVM modellerini kullanarak tahmin edilmiş ve test sonuçlarına ait performans kriterleri karşılaştırılması yapılmıştır. 9 makine öğrenmesi modelinin eğitime ve test verilerine ait performans kriterleri Tablo-15'de verilmiştir. Tablo-5 göre performans kriterlerinden MAPE değerinin birimi yüzde (%) cinsinden hesaplanmakta olup %10 ve %20 arasında olduğu zaman iyi tahmin;

%10'un altında kaldığı zaman yüksek başarılı tahmin sonucu vermiş olarak değerlendirilmektedir. Buna göre test verileri Tablo-15'e göre incelendiğinde kullanılan 9 farklı modelleme yönteminin %10 değerinin altında kaldığını ve yüksek başarılı tahmin sonuçları verdiği görülmüştür. Bu modeller ise kendi aralarında kıyaslandığında en yüksek başarılı tahmin sonucunu LSSVM modeli vermiştir. Kısacası, tahmin değerinin en yüksek başarılı sonuç veren modeli ise 0,1547 MAPE değeri ile LSSVM modeli olarak bulunmuştur. Yani bu çalışmada MAPE değeri %10'dan küçük olduğundan dolayı tahmin değerini hesaplamak için kullanılan modellerde çoğunlukla LSSVM modeli yüksek başarılı tahmin sonucu vermiştir.

RRMSE değeri de tıpkı MAPE değeri gibi yüzde cinsinden hesaplanmaktadır. Tablo-6'ya göre RRMSE değeri %10 ve %20 değeri arasında kalırsa iyi tahmin verirken %10 değerinin altında kalması durumunda çok daha iyi tahmin sonucu vermektedir. Buna göre Tablo-15'de görüleceği üzere, en düşük RRMSE değerine sahip olan LSSVM modelinin RRMSE değeri 0,2694 olup en başarılı tahmin modelini vermiştir. LSSVM modeli MAPE kriterinde olduğu gibi RRMSE kriterinde de yüksek başarılı tahmin modeli oluşturmuştur.

MAE performans kriteri ise, sıfır değerine yaklaştıkça modelin tahmin başarısı da o kadar yüksek olarak yorumlanmaktadır. Bu nedenle Tablo-15'e göre test verileri karşılaştırıldığında en küçük değeri LSSVM modeli 0,0708 değerini vererek en başarılı tahmin modeli olarak değerlendirilmiştir.

Tablo 7'ye göre $0.80 < R^2 \leq 1.00$ arasında kalan tahmin sonuçları çok iyi tahmin, değer 1'e yaklaşmasına göre de tahmin modellerinin mükemmelliği artmaktadır. $0.40 < R^2 \leq 0.60$ arasında kalan tahmin sonuçları ise normal derecede tahmin sonucu olarak değerlendirilmektedir. Tablo-15'e göre modellerin R^2 değeri karşılaştırıldığında tüm modeller çok iyi sonuç vermesine rağmen içlerinde LSSVM modeli 1 değerine daha yakın sonuç vererek en mükemmel tahmin modelini vermiştir. (LSSVM modeli R^2 için 0,9985 değerini vermiştir.) Böyle LSSVM modeli en başarılı tahmin modeli olarak kabul edilmiştir.

Tablo-8'e göre $NSE \leq 0.40$ ' dan düşük tahmin sonucu veriyor ise yetersiz $0.75 < NSE \leq 1.00$ arasında kalıyorsa çok iyi tahmin olarak değerlendirilmektedir. Bu kriterlere göre Tablo-15'de NSE kriteri karşılaştırıldığında 0.75 ve 1 değeri arasında kaldığı için çok iyi

tahmin modelleri olduğu görülmüştür. Modeller içerisinde NSE kriterine göre en iyi tahmin modelini 0,9983 değeri ile LSSVM modeli vermiştir.

Tablo-9'a göre $0.60 < KGE \leq 0.70$ aralığında tahmin sonuçları veren modeller iyi performans sergilerken $0.70 < KGE \leq 1.00$ aralığında çok iyi derecede tahmin sonuçları vermektedir. Buna göre tahmin sonuçlarının 1'e yaklaşma durumuna göre tahmin sonuçlarının mükemmelliği de artacaktır. Yapılan bu çalışmada ise modellerin KGE değerine göre karşılaştırılması yapıldığında 1'e yakınlık derecesine göre 0,9886 tahmin değerini vererek ANFIS modeli olmuştur. KGE kriterinde en iyi tahmin modeli ANFIS modeli iken R^2 ve NSE kriterlerinde en iyi model LSSVM modeli olmuştur.

OI değeri ise $-\infty$ ile 1 arasında değişkenlik göstermektedir. OI değerinin 1'e yakın olmasına göre tahmin modelinin mükemmellik derecesi de artmaktadır. Buna göre tüm modellerin karşılaştırılması yapıldığında 0,9932 değeri ile LSSVM modeli 1'e en yakın ve en yüksek değeri alarak en iyi tahmin modeli olarak belirlenmiştir.

WI değerinde ise tahmin sonuçlarının 1'e yaklaşmasına göre sonuçların iyileşme derecesi artış göstermekte olup Tablo-15 bakılıp modeller birbiri arasında kıyaslandığında en iyi tahmin sonucunu veren model 0,9859 değeri ile LSSVM modeli olmuştur.

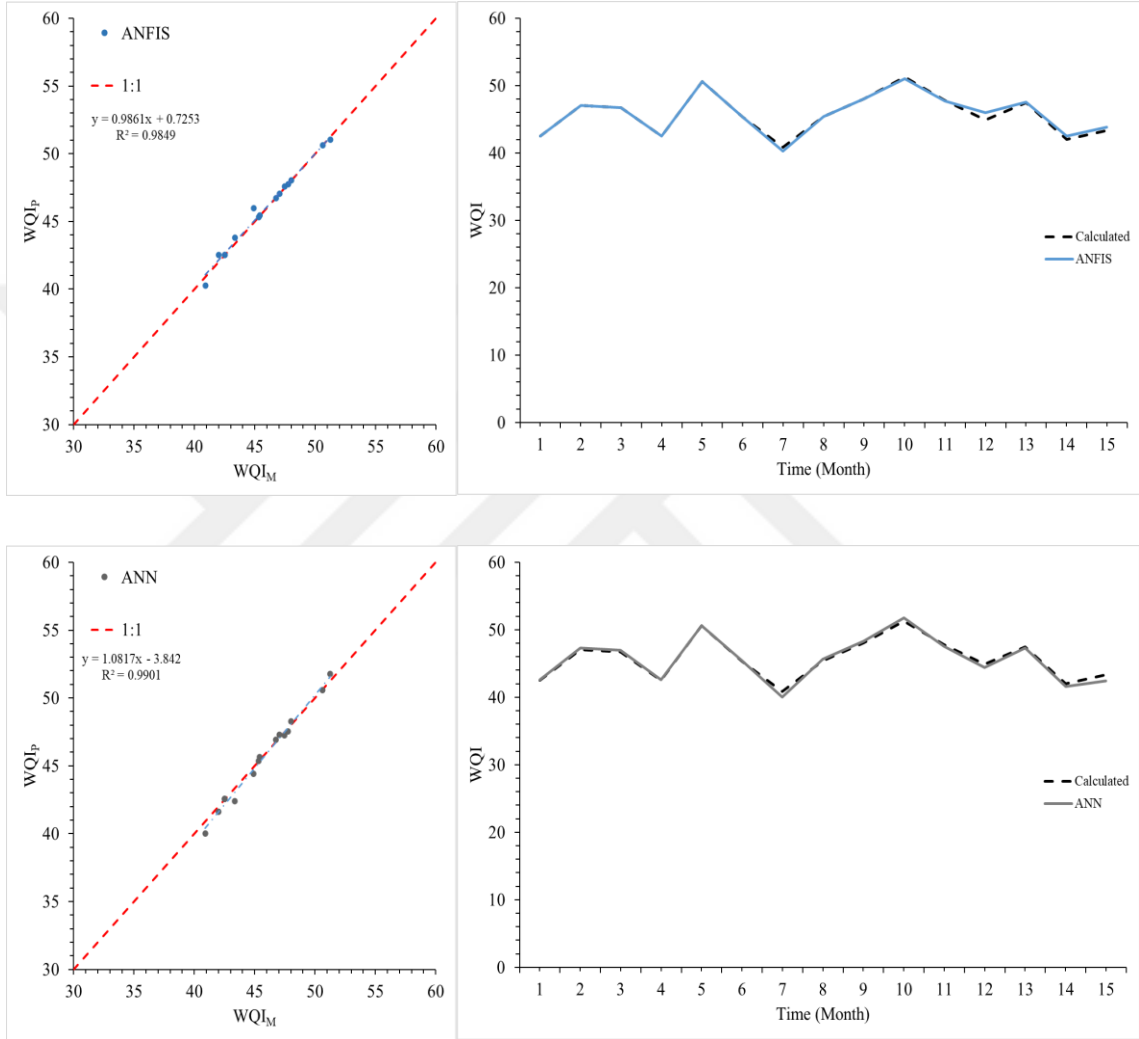
Tablo 15. WQI tahmini model performanslarının karşılaştırılması

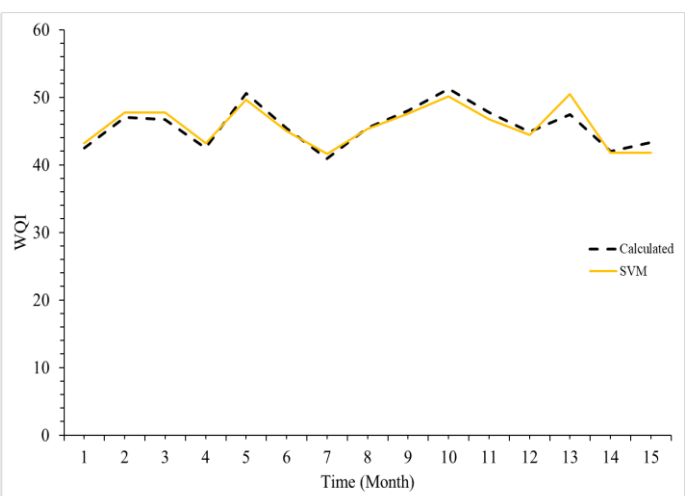
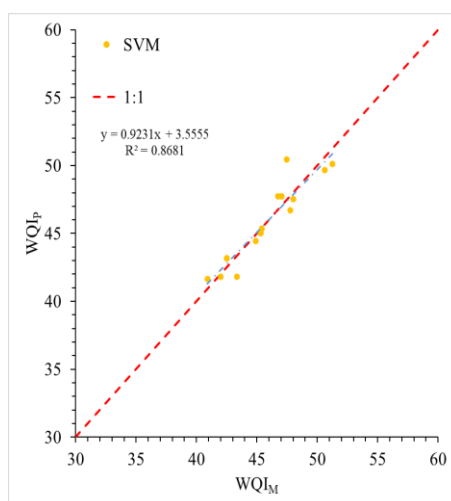
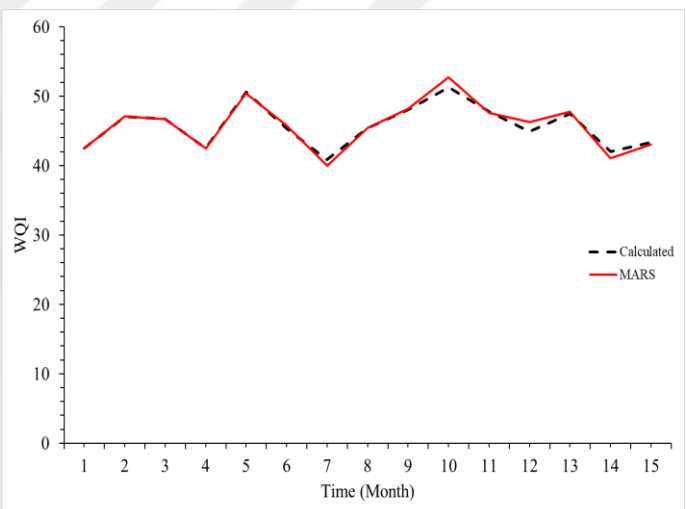
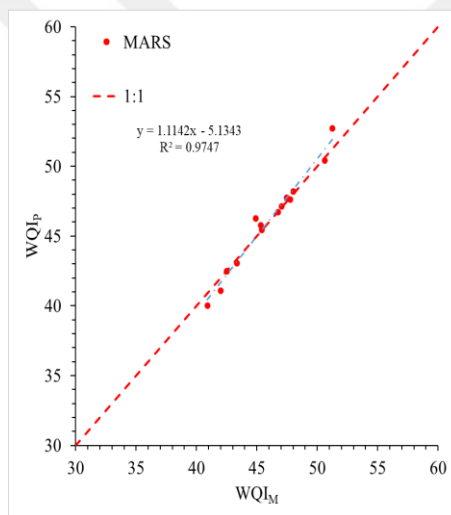
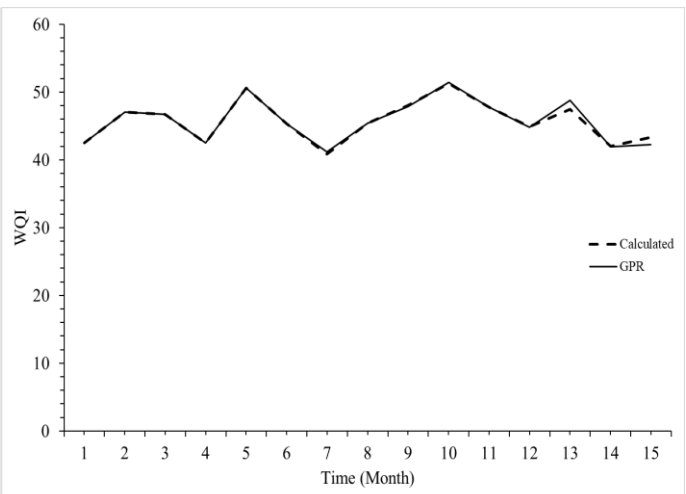
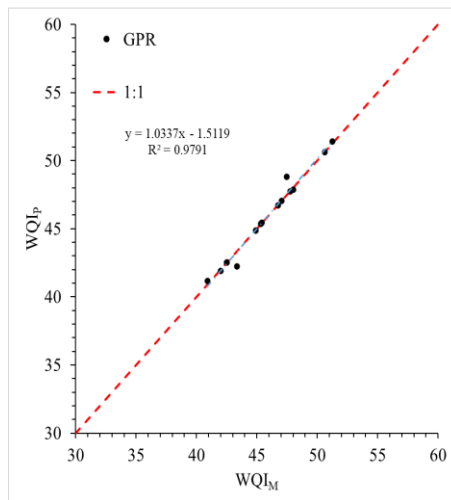
Eğitme								
	MAPE	RRMSE	MAE	RKARE	NSE	KGE	OI	WI
ANFIS	0.0171	0.028	0.008	1.0000	1.0000	0.9999	0.9996	0.9986
ANN	0.3289	0.4177	0.1511	0.9970	0.9970	0.9973	0.9923	0.9729
GPR	0.009	0.0127	0.0041	1.0000	1.0000	0.9999	0.9998	0.9993
MARS	0.2073	0.2873	0.0952	0.9986	0.9986	0.9990	0.9950	0.9829
SVM	1.912	2.8086	0.8769	0.8702	0.8664	0.9300	0.8916	0.8427
LSBoost	0.000	0.000	0.000	1.0000	1.0000	1.0000	1.0000	1.0000
KNN	0.000	0.000	0.000	1.0000	1.0000	1.0000	1.0000	1.0000
M5TREE	1.041	1.830	0.487	0.9433	0.9433	0.9593	0.9445	0.9127
LSSVM	0.114	0.130	0.052	0.9998	0.9997	0.9916	0.9979	0.9907
Test								
	MAPE	RRMSE	MAE	RKARE	NSE	KGE	OI	WI
ANFIS	0.4756	0.8264	0.2084	0.9849	0.9840	0.9886	0.9738	0.9585
ANN	0.6997	0.9168	0.3124	0.9901	0.9804	0.9102	0.9700	0.9379
GPR	0.4768	1.0129	0.2159	0.9791	0.9760	0.9548	0.9657	0.9571
MARS	0.9353	1.4057	0.4253	0.9747	0.9538	0.8729	0.9459	0.9154
SVM	1.8762	2.4087	0.8665	0.8681	0.8644	0.9310	0.8791	0.8276
LSBoost	1.3182	2.6163	0.6215	0.8530	0.8401	0.7980	0.8623	0.8764
KNN	1.4572	2.8524	0.6391	0.8347	0.8099	0.8888	0.8420	0.8729
M5TREE	1.9798	2.8133	0.9112	0.8312	0.8151	0.9034	0.8455	0.8187
LSSVM	0.1547	0.2694	0.0708	0.9985	0.9983	0.9869	0.9932	0.9859

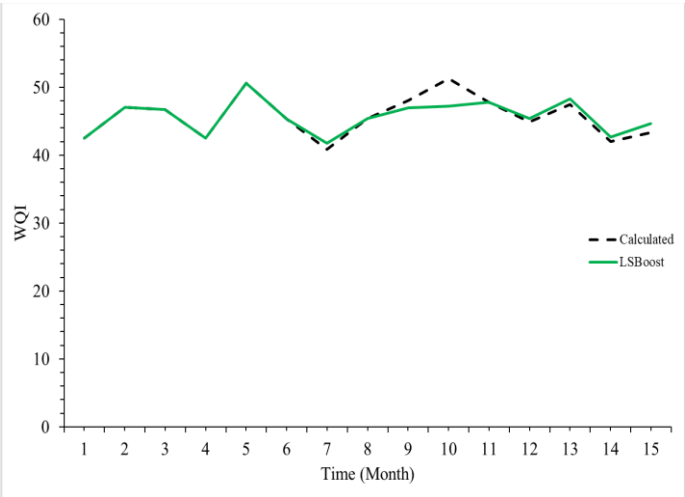
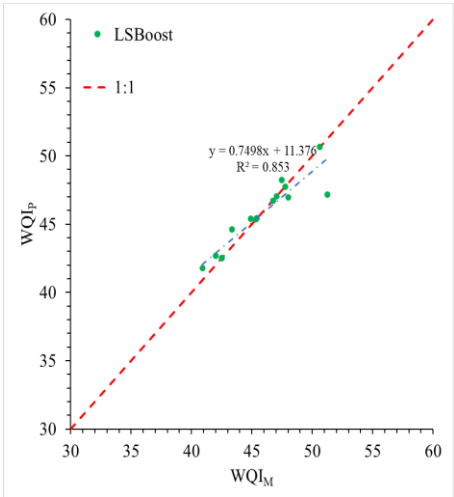
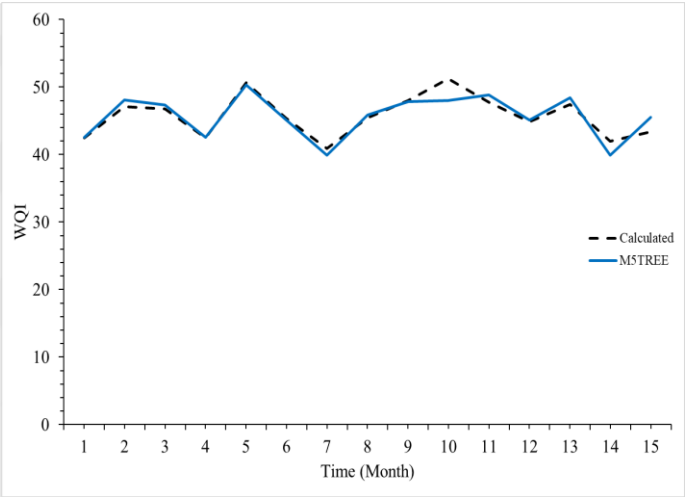
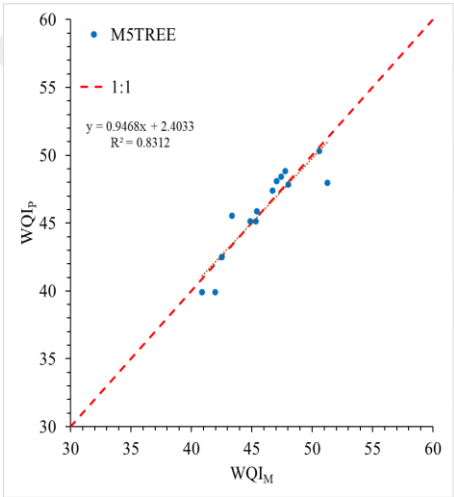
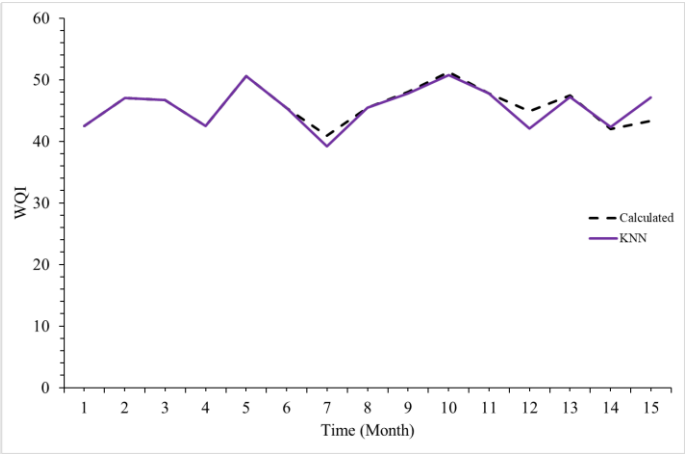
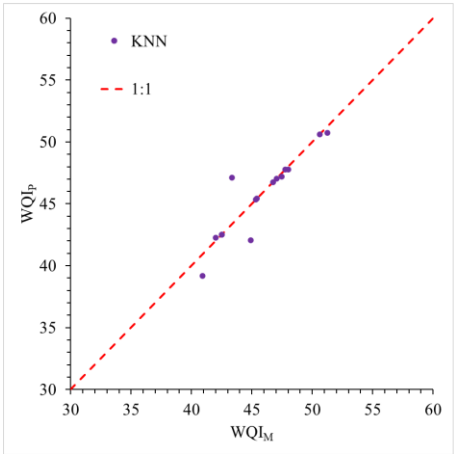
MAPE, RRMSE MAE, performans kriterlerinde en düşük değer; R^2 , NSE, ANFIS, OI ve WI performans kriterlerinde 1' e en yakın model LSSVM modeli vermiş olup dokuz modelden en başarılısıdır. LSSVM modeline ait performans değerleri; MAPE değeri 0,1547, RRMSE değeri 0.2694, MAE değeri 0.0708, R^2 değeri 0,9985 NSE değeri 0.9983, KGE değeri 0,9869, OI değeri 0,9932 ve WI değeri 0.9859 olarak bulunmuştur.

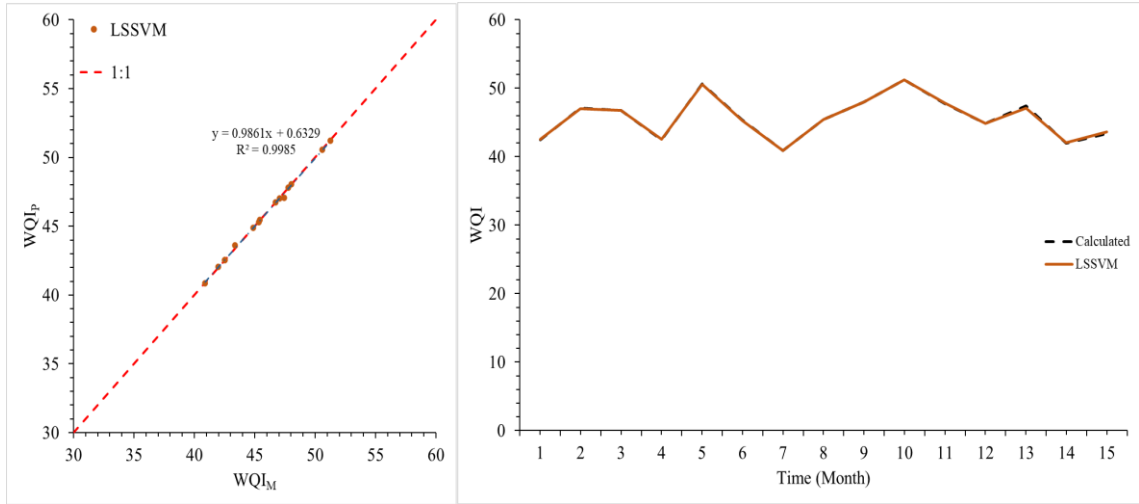
Şekil 12'de WQI tahmin modellerine ait saçılma diyagramları ve zaman serisi grafikleri verilmiştir. Saçılma diyagramları ölçülen veri modelleri ile tahmin edilen veri modelleri arasında 1:1'lik bir doğru oluşturarak bu doğru üzerindeki tahmin modellerinin saçılmasını göstermektedir. Buna göre noktaların 1:1'lik doğru üzerinde yer almasına göre tahmin modellerinin doğruluğu da artmaktadır. Şekil 12'de yer alan zaman serisi grafiklerinde ise WQI değerinin aylara göre tahmin modellerinin ölçülen değerlerine göre gösterimi verilmiştir. Ölçülen değerler grafikte kesikli çizgilerle gösterilirken modellere ait tahmin değerleri düze çizgilerle gösterilmektedir. Grafiklere bakıldığında ise kesikli

çizgiyle örtüşen veya en yakın noktasında geçen zaman aralığında en iyi sonuçları verdiği görülmüştür. Sonuç olarak performans kriterlerinde de olduğu gibi LSSVM modeline ait saçılma ve zaman serisi grafikleri diğer modellerle karşılaştırıldığında gerçek değere en yakın tahmin modelleri vererek oldukça iyi sonuç vermektedir.



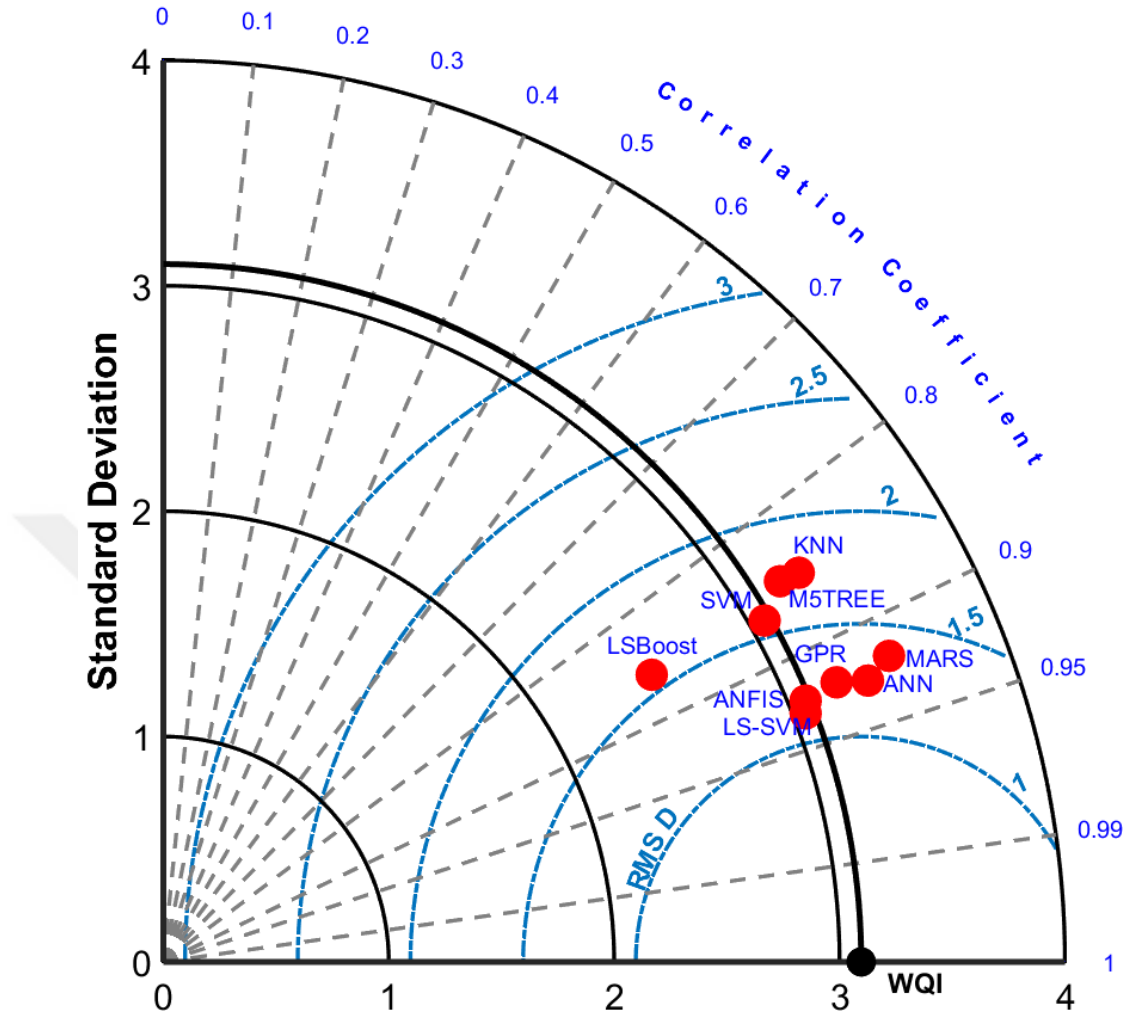






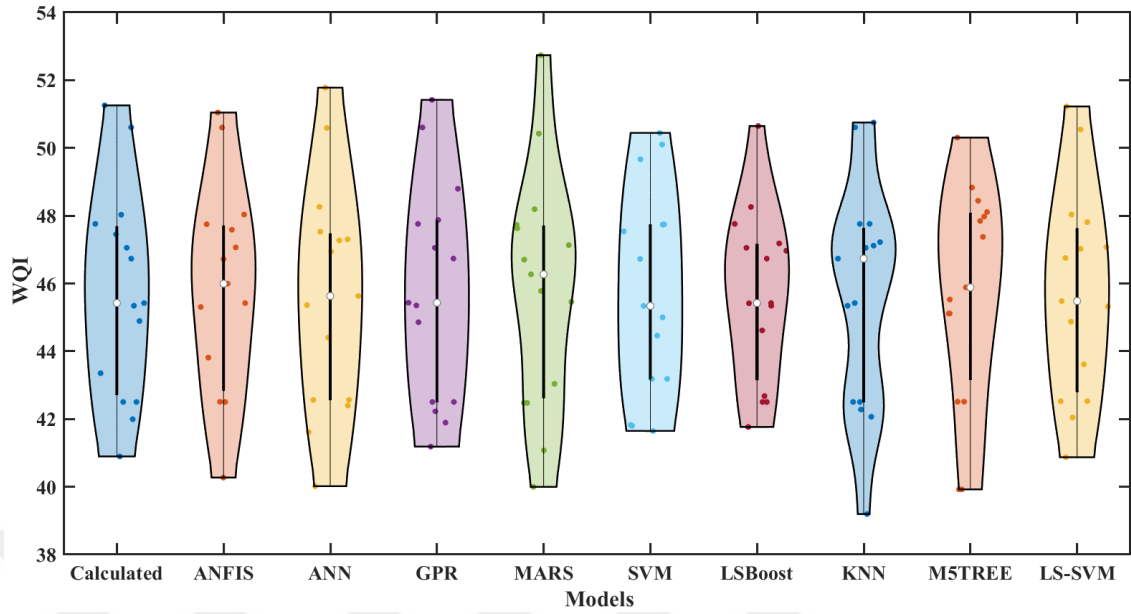
Şekil 12. WQI modellerine ait saçılma diyagramları ve zaman serisi grafikleri

Bu tez çalışmasında kullanılan modellere ait test sonuçlarının karşılaştırılması yapıldığında performans kriterleri, zaman serisi ve saçılma diyagramlarına ek olarak Taylor diyagramından da faydalanılmıştır. Taylor diyagramının kullanılmasında amaç modellere ait test sonuçlarını bir arada görebilmek ve daha iyi karşılaştırmasını yapabilmek için kullanılmıştır. WQI değerinin gerçek değeri ile modellerden elde edilen WQI tahmin değerlerine göre standart sapma, korelasyon katsayısı ve merkezlenmiş kök ortalama karesel farkı istatistiksel kriterlerine göre Taylor diyagramı üzerinden karşılaştırması yapılmıştır. Şekil 13’ de görüldüğü üzere WQI tahmini için elde edilen Taylor diyagramında, modellere göre tahmini test sonuçlarının standart sapma, korelasyon katsayısı ve merkezlenmiş kök ortalama karesel farkı istatistiksel kriterlere göre en yakın ve en uyumlu olan LSSVM tahmin modelidir. SVM, KNN, M5TREE ve LSBoost tahmin modellerinde ise standart sapma, korelasyon katsayısı ve merkezlenmiş kök ortalama karesel farkı istatistiksel kriterlere göre en uzak olan ve WQI tahmin modellerinden diğer modellere göre daha başarısız modelleri veren yöntem olarak görülmektedir.



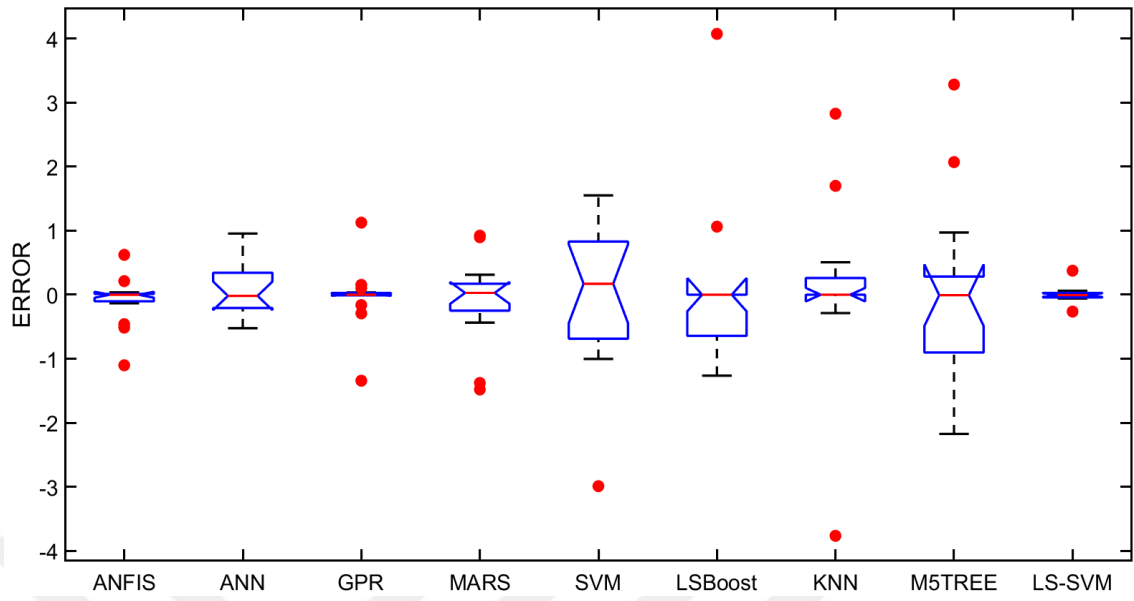
Şekil 13. WQI Tahmini İçin Kullanılan Modellere Ait Taylor Diyagramı

Ayrıca çalışılan tüm modeller için Violin ve Hata-Boxplot diyagramları da çizilmiş ve 9 makine öğrenmesi modeli karşılaştırılmıştır. Her bir modele ait Violin diyagramları Şekil 14’ de; Hata-Box-Plot diyagramları ise Şekil 15’ de yer verilmiştir. Şekil 14’ de yer verilen WQI tahmini için oluşturulan modellere ait Violin diyagramları yer almaktadır. Buna göre şekil olarak ölçülen modele benzerlik açısından en iyi sonucu veren model LSSVM ve ANFIS modelleri olmuştur. Fakat LSSVM modelinin ortalama ve medyan değerleri WQI parametresinin ortalama ve medyan değerlerine yakındır. Ayrıca LSSVM modelinin kutu grafiği WQI parametresinin kutu grafiğine daha yakındır. ANN modelinde ise WQI değerinin en yüksek ve en düşük değerleri tahminin de bir miktar sapmalar söz konusudur. KNN, LSBoost, M5TREE ve MARS modelleri ise şekilsel olarak WQI parametresinin şeklinden oldukça uzaktır.



Şekil 14. WQI tahmini için kullanılan modellere ait violin diyagramları

Şekil 15’ de yer alan WQI Hata-Boxplot diyagramında ise tahmin modelleri kıyaslandığında LSBoost, M5TREE ve SVM modellerinin kutu grafikleri oldukça yüksektir. SVM modelinin aykırı değeri -3; LSBoost modelinin aykırı değeri 1 ile 4 aralığında; M5TREE modelinin aykırı değeri 1.5 ile 3.5 aralığındadır. Bu sonuçlara göre aykırı değerlerde SVM modeli yüksek tahmin ederken LSBoost ve M5TREE modeli düşük tahmin yaptığı gözlenmiştir. Violin grafiğinde olduğu ANFIS ve LSSVM modellerinin kutu grafiği benzer olup kutu grafikleri dar modellerdir. Fakat ANFIS modelinin aykırı değerleri -1 ile 1 arasında değişirken LSSVM modelinin aykırı değerleri -0.5 ile 0.5 aralığında değişmektedir. GPR modelinin kutu grafiği en dar model olmasına karşın aykırı değerleri -1.5 ile 1.5 arasında değişmektedir.



Şekil 15. WQI tahmini için kullanılan modellere ait hata violin diyagramları

Çalışmanın son kısmında ise istatistiksel anlamlılık testi olan Kruskal Wallis (KW) testi ile kontrolü yapılmıştır. Modellere ait sonuçların doğruluğunu test etmek için KW testinden yararlanılmıştır. KW testi yardımıyla, tahmin modelinin ortalaması ile ölçülen WQI değerinin ortalamasının aynı olup olmadığı belirlenmektedir. Tablo 16’de KW testinin sonuçları verilmiştir. Bu test sonuçlarına göre tüm modeller için tahmin sonuçları KW testinde de gerçek ölçüm değerlerini yansıtmaktadır. KW testine göre H_0 değeri ölçülen ve tahmin edilen modellerin birbirine denk olmadığını temsil ederken H_1 değeri ise ölçüm ve tahmin değerinin denk olduğunu temsil etmektedir. Buna göre tüm modeller için H_0 değeri reddedilerek tahmin değerlerinin WQI parametresi temsil ettiği belirlenmiştir.

Tablo 16. KW testinin sonuçları.

Kruskal–Wallis test		
Methods	p_value	H ₀
ANFIS	0.8195	Reject
ANN	0.9835	Reject
GPR	0.9504	Reject
MARS	0.9504	Reject
SVM	0.8519	Reject
LSBoost	0.9989	Reject
KNN	0.9834	Reject
M5TREE	0.6934	Reject
LSSVM	0.6934	Reject

H₀: $\mu_{ölçüm} \neq \mu_{tahmin}$, H₁: $\mu_{ölçüm} = \mu_{tahmin}$

4. BÖLÜM

TARTIŞMA-SONUÇ VE ÖNERİLER

4.1. Tartışma

İnsan hayatı için son derece önemli olan su kaynakları günümüzde çeşitli kullanım alanları için tüketilmektedir. Bu kullanım alanlarına kentsel atık, tarımsal faaliyetler, taşımacılık, sanayileşme ve insan aktiviteleri gibi faaliyet alanları su tüketimine örnek gösterilmektedir. Tüketim oranı artan su kaynaklarının bilinçsizce kullanılmasının yanı sıra su kaynakları üzerinde kirlilik oluşumunu da artırmakta ve su kalitesini etkilemektedir. Bu faktörler su kalitesi üzerinde olumsuz etkiler oluşturmaktadır. Bu gibi problemlerden dolayı su kalitesi hakkında daha detaylı ve daha güvenilir bilgiyi bulmak için su kalitesinde önem arz eden çok kapsamlı veri kümelerinin yapay zekâ tekniklerini kullanarak modellemeleri ve tahminlerini yapmak mümkün hale gelmiştir. Bu çalışma kapsamında da içme suyuna ait kalite parametreleri kullanılarak makine öğrenmesi yöntemleri vasıtasıyla WQI tahmin modellemeleri oluşturulmuştur. Yapılan çalışmada, Kayseri Merkez ilçelerine ait 7 adet (pH, iletkenlik, toplam siyanür, sülfat, nitrat, sodyum ve serbest klor) su kalitesi parametreleri girdi verisi olarak WQI değeri ise çıktı verisi olarak kullanılmıştır. WQI tahmini için LSSVM, KNN, MARS, M5TREE, LSBoost, GPR, SVM, ANFİS ve ANN olmak üzere 9 farklı makine öğrenmesi yöntemi kullanılmış olup bu modellerin performansları karşılaştırılmıştır. Yapılan karşılaştırmalar sonucunda LSSVM modeli diğer yöntemlere göre daha iyi sonuçlar vermiştir. Ayrıca karşılaştırılan modellerin etkinliği ise KW testi ile desteklenmiştir.

Literatür incelendiğinde ise su kalitesi parametreleri kullanılarak yapay zekâ yöntemleri ile tahmin modellemeleriyle ilgili pek çok çalışma mevcuttur. Ayrıca literatür çalışmaları incelendiğinde su kalitesine ait parametrelerin tahmin modellemeleri sadece içme sularında değil aynı zamanda nehirler üzerinde, atık sularda, deniz ve tatlı sularda, yeraltı

sularında ve akarsularda olmak üzere çeşitli su kaynaklarına etkiyen su kalitesi parametrelerinin makine öğrenmesi modelleriyle tahmin sonuçlarını ortaya koymuşlardır.

Buna göre Najah Ahmed A. et al. (2019) yaptığı çalışmalarında çeşitli kalkınma ve insan faaliyetlerinden dolayı kaynaklanan Johor nehrinin kirlilik seviyesini kontrol altına almak için su kalitesi parametrelerinin modellemesini yapmışlardır. Makine öğrenmesi yöntemleri olarak ANFIS, RBF-ANN, MLP-ANN yaklaşımlarını kullanmıştır. Fakat izleme istasyonlarından elde edilen veriler gürültüden kaynaklı olarak sistematik hatalar vermiştir. Bu durumu engellemek için WDT-ANFIS kullanımını artırarak gürültüden oluşan sistematik hataların önüne geçileceği görülmüştür. Modelleme verileri olarak pH, AN ve AKM kullanmışlardır. Yapılan bu çalışmalar sonucunda aynı parametrelerin her istasyondaki değerlerine ait tahmin modelleri geliştirmişlerdir. Sonuç olarak ise tahmin modellerinin tatmin edici sonuçlardan verdiği ortaya koymuşlardır. Yapılan çalışma incelendiğinde ise su kalitesi parametrelerinin sadece içme suyunda değil aynı zaman da nehir sularında kalitesinin belirlenmesinde de önemli rol oynadığı ve tahmin modellemelerinin makine öğrenimi yöntemleri vasıtasıyla kolaylıkla yapılabileceği ortaya konmuştur. Ayrıca içme suyu kalitesinden alınan verilerde gürültü faktörü olmazken verilerin tahmini için ANFIS modelinin her iki çalışmada da kullanıldığı ve tahmin modellerinin tatmin edici sonuçlar verdiği görülmüştür.

Zhu, M. et al. (2022) yapmış oldukları çalışmada içme suyu, atık su, yeraltı suyu, yüzey suyu, deniz ve tatlı su kaynaklarının farklı su kalitesi parametrelerine sahip oldukları için verilerin toplanmasında zorluklar yaşanmaktadır. Saha çalışmalarının güç olduğu durumları engellemek amacıyla makine öğrenmesi metodlarını kullanarak farklı yöntemlerin karşılaştırılmasını yapmışlardır. SVM, YSA, PCA, XGBoost, DT, DO, LSTM, WQI gibi çeşitli algoritmalarından yararlanarak su kirliliğinin kontrol altına alınması, su arıtımının izlenmesi ve su ortamlarının gelecekteki durumları hakkında tahminlerin yapılması amaçlanmaktadır. Buradaki literatür ile bizim çalışmamız kıyaslandığı zaman ise benzer amaç için çalışmaya başlanıldığına ve SVM, YSA, WQI gibi ortak algoritmaların içme suları için kullanımına uygun olduğuna ve çeşitli su kaynakları için de makine öğrenmesi modellerinin tahmin çalışmasına elverişli olduğu sonucuna varılmıştır.

Zamili H., (2023) Yeşilırmak nehrindeki su kalitesi parametrelerini kullanarak WQI değerini tahmin etmek için yaptığı çalışmada pH, BOI, ÇO, Eİ, Ca^{+2} , Na^{+1} , Mg^{+2} , Cl^{-1} , NO_2^{-1} ve SO_4^{-2} olmak üzere toplamda 11 adet su kalitesi parametresi kullanmıştır. CPSOCSGA, PSO, MPA algoritmaları YSA ile birleştirilerek içme suyuna uygun olmayan noktaların optimizasyonu yapılmıştır. Yapılan modellemelerde MAE=0,01627, RMSE=0,0187 ve $R^2 = 0,965$ değerini vermiş olup CPSOCSGA-ANN performansının MPA-ANN ve PSO-ANN performanslarına göre daha üstün olduğunu ortaya koymuştur. İncelenen literatür ile yaptığımız çalışma karşılaştırıldığında ise toplamda 7 farklı su kalitesi parametresi kullanılarak 9 farklı makine öğrenmesi yöntemi kullanılmıştır. Kullanılan yöntemler arasında ise en iyi sonucu veren yöntem LSSVM modeli olup MAE=0.0708, RMSE=0.2694 ve $R^2 = 0.9985$ istatistiksel doğrulama değerini vermiştir.

Literatürde yer alan Shamsuddin, I. et al., 2022, Haghiabi, A. et al., 2018, Özeren Y., 2022, Altuntaş M., 2018, Najah Ahmed A, et al. 2019, Acar R. vd., 2022, Şemsüddin, I. et al., 2022, Haghiabi A., et al., 2018, çalışmalarda çeşitli bölgelerdeki nehir suyu havzalarına ait su kalitesi parametreleri kullanılarak su kalitesi tahmini için çeşitli modeller oluşturmak amacıyla makine öğrenmesi yöntemleri kullanmışlardır. Su kalitesi tahmin modellerinin yanı sıra Saplıoğlu K, 2022, Wilcoxon testi, R^2 ve ağırlıklı karesel hata değerlerini kullanarak su sertliğinin tahmin edilmesinde elde edilen modellemeleri de göstermiştir.

Ay M., 2014, çalışmasında ise farklı akarsu kaynaklarından ölçülen karmaşık ve doğrusal olmayan su kalitesi parametreleri ve debi parametrelerini kullanarak klasik denklemlerden ziyade yapay zeka tekniklerini kullanarak oluşan modellerin karşılaştırılmasını yapmıştır. Karşılaştırma sonucuna göre klasik yöntemlere göre su kalitesi parametrelerinin daha başarılı sonuç verdiğini ve tahmin performansının daha güçlü olduğunu belirlemiştir. Buna göre bu çalışmadan çıkarılacağı üzere sahadan alınan ve klasik uzun hesaplama yöntemleri gerektiren su kalitesi parametreleri ve buna ek olarak debi parametreleri makine öğrenimi yönteminde tahmin modelleme çalışmaları için uygun olduğu görülmüştür.

Dede Ö. ve Sezer M., 2017 yaptığı çalışmada 1 yıllık izlenen 36 adet su kalitesi parametre analizi sırasında çözünmüş oksijen, T.coli, E.coli ve fenol değerlerinin su kalitesi standartlarının uygunluğu dışında kalıp ana kirleticiler olarak belirlemiştir. Buna ek

olarak renk, bulanıklık, mangan, demir, alüminyum, cıva ve amonyak azotunun da sınır değerlerini aştığı görülmüştür. Bu amaçla verilerin anlaşılabilirliğinin sağlanması ve yorumlanması, su kalitesini etkileyen parametrelerin belirlenmesi için Kanada Su Kalitesi İndeks Modeli (CWQI) uygulamıştır. İstasyon değerleri için CWQI değerleri zayıf kategorisinde değerlendirilmiş hesaplanan indeks değerinin mantıklı sonuçlar verdiğini saptamışlardır. Buna göre yapılan çalışma ile literatürde ki çalışma kıyaslandığında 7 yıllık sahada izlenen ve laboratuvarda klasik hesaplama yöntemleri ile analiz edilen su kalitesi parametreleri kullanılarak karmaşık ve doğrusal olmayan veri setlerinin anlaşılabilirliğinin sağlanması ve yorumlanması için WQI tahmin modeli uygulanmıştır.

Özen S., 2018, KASKİ merkez ileri biyolojik atık su arıtma tesisine ait 2011-2014 yılları arasındaki KOİ ve BOİ değişkenleri kullanılarak atık su arıtma verilerinin tahmin edilmesi çalışmasını yürütmüştür. KOİ değişkeni tahmininde ölçülen pH, Q, BOİ, AKM değerleri kullanılırken BOİ değişkeni tahmininde pH, P, Q, KOİ, AKM değerleri bağımsız değişken olarak kullanmıştır. KOİ ve BOİ tahmininde YSA, ANFIS yöntemlerini kullanmıştır. Yapay zekâ metodları ile analiz edilmeden önce elde olan verileri eğitim ve test datası şeklinde iki ayrı küme şeklinde gruplandırmıştır. Toplamda 724 verinin %75'i yani 543 adet eğitim verisi olarak kullanılırken kalan %25'i yani 181 tanesini de test verisi şeklinde kullanmıştır. KOKH, OMH istatistiksel doğrulama yöntemleri kullanılarak BOİ değişkeninin tahminde bu verileri karşılaştırmıştır. Tahmin sonucuna göre YSA ve ANFIS modellerinin sonuçlarının birbirine çok yakın çıktığını gözlemlemiştir.

Citakoglu H. vd, 2022, hidroloji mühendisliği ve geoteknik mühendisliğine ait inşaat mühendisliği alanlarında karşılaşılan problemleri analiz etmek için makine öğrenmesi yöntemlerinden yararlanarak nasıl performans gösterdiğini araştırmışlardır. Yapılan çalışmada girdi parametrelerinin çıktı parametreleri üzerindeki etkinliği incelenmiştir. DSİ tarafından temin edilen veriler ANN, GPR, SVMR, LSTM, MGPP ve METree makine öğrenimi yöntemlerini kullanılarak modellenmeleri oluşturulmuştur. Oluşturulan modellerin performanslarını değerlendirmek için genel Performans İndeksi hesaplanarak makine öğrenimi modelinin göreceli performanslarını ise Taylor diyagramı, Keman diyagramı ve Tek kuyruklu Wilcoxon İşaretli Sıra testleri kullanılarak değerlendirilmesi yapılmıştır. Sonuç olarak doğal ortamda MGPP modeli en iyi tahminleri verirken YSA modelinde ise yöntemsel ve deneysel veri kümeleri kullanıldığında en iyi tahminleri

ürettiğini ortaya koymuşlardır. Taylor ve Keman diyagramlarına göre daha geniş açıdan bakıldığında ANN, MGGP ve GPR modelleri daha başarılı sonuçlar vermiştir. Ayrıca tüm modeller Kruskal - Wallis testini sorunlar açısından geçmiş bulunmaktadır.

4.2. Sonuç ve Öneriler

Bu tez çalışmasında; Kayseri Su ve Kanalizasyon İdaresinden 2015-2022 yıllarına ait su kalitesi parametre indeksleri temin edilerek WQI tahmin modelleri oluşturulmuştur. Makine öğrenmesi yöntemleri olarak; LSSVM, KNN, MARS, M5TREE, LSBoost, GPR, SVM, ANFİS ve ANN ile tahmin modellemesi yapılmıştır. WQI verileri MATLAB 2023a yazılım programı üzerinden işlenerek WQI verileri 9 farklı makine öğrenmesi modeli ile tahmin edilmeye çalışılmıştır. Modellemelerde toplamda 7 adet giriş verisi kullanılırken 1 adette çıkış verisi kullanılmıştır. Giriş verisi olarak; pH, iletkenlik, toplam siyanür, sülfat, nitrat, sodyum ve serbest klor kullanılırken çıkış verisi olarak WQI kullanılmıştır. Modellemelerde kullanılan veri sayısı toplamda 74 adet olup bu verilerin 59 (%80) tanesi eğitime verisi, 15 (%20) tanesi test veri olarak kullanılmıştır. Her bir makine öğrenmesi sırasıyla çalışılmış olup, oluşturulan tahmin modelleri performans kriterleriyle değerlendirilmiştir. R^2 , MAPE, RRMSE, MAE, NSE, KGE, OI ve WI ölçütleri kullanarak toplamda sekiz farklı metriktan oluşan performans değerlendirme yöntemleri kullanılmıştır. Performans kriterlerinin karşılaştırılması sonucunda LSSVM modeli yüksek başarılı tahmin modeli olarak seçilmiştir. Çalışılan diğer tahmin modelleri ise güzel sonuçlar vermesine rağmen en mükemmel sonucu LSSVM modeli vermiştir. Çalışma sonucunda ise performans kriterleri dışındaki değerlendirmeleri ise Taylor, Violin ve Hata Boxplot diyagramlarına göre en iyi tahmin sonucu veren modellerin doğruluğunu desteklemek için modellerin görsel olarak karşılaştırılması yapılmıştır. Diyagramların yorumlanması sonucuna göre ise yine yüksek başarılı tahmin modelini LSSVM modeli oluşturmuştur. Karşılaştırmada önerilen 9 makine öğrenmesi yönteminin etkinliği KW testi ile kanıtlanmıştır.

İçme suyu su kalitesi parametrelerinin değişimlerinin izlenmesi ve geleceğe yönelik öngörülerin yapılabilmesi için yapay zekâ ve alt dalları olan yapay sinir ağları ve makine öğrenme yöntemlerinden yararlanmakta fayda vardır. Günümüzde küresel ısınmanın etkisiyle beraber giderek azalmaya yüz tutan doğal su kaynaklarının içme ve kullanma suyu bakımından sağlığa ve çevreye olan etkisini incelemek son derece önemlidir. Bu

nedenle gelişen teknoloji ile su kalitesi parametrelerinin gelecekteki tahmini ve kullanımı son derece önem arz etmektedir. Kullanılan su kalitesi parametreleri ve WQI değerleri hem gelecekte üretilen WQI değerinin tahmininin yapılması hem de modellerin karşılaştırılması için yapılmıştır. Çalışmamız kapsamında kullanılan içme suyu şebeke sistemlerine ait veriler göz önüne alındığında aynı modelleme çalışmaları yapılarak nehir, akarsu ve atık su artıma parametrelerinin tahmin modelleme çalışmaları için, yüzey sularındaki bakteri formunun tahmin edilmesi gibi birçok çalışmada bizlere örnek teşkil eden bir çalışma oluşturmaktadır. Ayrıca kullanılan bu yöntemler sadece su kalitesi parametresi tahmini için değil çeşitli konularda da zaman serisi içeren veri setlerinin tahmin modellemelerinde de kullanılması tavsiye edilebilir. Bunların haricinde bu çalışma için farklı ön işleme teknikleri veya farklı makine öğrenmeleri ile de performans değerlendirmeleri incelenmesi önerilmektedir.

KAYNAKÇA

- [1] Çorba, S., 2021. Gelecekteki tehlike: su kıtlığı. (Web sayfası: <https://bilimgenc.tubitak.gov.tr/makale/gelecekteki-tehlike-su-kitligi>), (Erişim tarihi: Ocak 2024).
- [2] Aytüre, S., 2016. Su kanununun AB su çerçeve direktifi kapsamında değerlendirilmesi. **Türk Bilimsel Derlemeler Dergisi**, (1): 65-71.
- [3] Türkman, A., Tokgöz, S., Sarptaş, A. G. H., 1999. İçme Suyu Standartları ve Güvenilir İçme Suyu. 3. *Ulusal Çevre Mühendisliği Kongresi*, 2023, İzmir, Dokuz Eylül Üniversitesi Çevre Mühendisliği Bölümü.
- [4] Ahmed, A. N., Othman, F. B., Afan, H. A., Ibrahim, R. K., Fai, C. M., Hossain, M. S., Elshafie, A., 2019. Machine learning methods for better water quality prediction. **Journal of Hydrology**, 578: 124084.
- [5] Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Ye, L., 2022. A review of the application of machine learning in water quality evaluation. **Eco-Environment & Health**, 1 (2): 107-116.
- [6] Zamili, H., Bakan, G., Zubaidi, S. L., Alawsi, M. A., 2023. Water quality index forecast using artificial neural network techniques optimized with different metaheuristic algorithms. **Modeling Earth Systems and Environment**, 9 (4): 4323-4333.
- [7] Shamsuddin, I. I. S., Othman, Z., Sani, N. S., 2022. Water Quality Index Classification Based on Machine Learning: A Case from the Langat River Basin Model. *Water*, 14, 2939.
- [8] Ay, M., 2014. Su Kalitesi Parametrelerinin Yapay Zekâ Yöntemleri İle Değerlendirilmesi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, Kayseri, s. 1-136.

- [9] Haghiabi, A. H., Nasrolahi, A. H., Parsaie, A., 2018. Water quality prediction using machine learning methods. **Water Quality Research Journal**, 53 (1): 3-13.
- [10] Özeren, Y., 2022. Sakarya Havzasına Ait Su Kalitesi Parametrelerinin Yapay Zekâ Yöntemleri İle Modellenmesi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kayseri, s. 1-205.
- [11] Altuntaş, M., 2018. Yeşilırmak Havzası Su Kalitesi Parametrelerinin Yapay Zekâ Teknikleriyle Modellenmesi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kayseri, s. 1-80.
- [12] Dede, Ö. T., Sezer, M. 2017. The application of Canadian water quality index (CWQI) model for the assessment of water quality of Aksu creek. **Journal of the Faculty of Engineering and Architecture of Gazi University**, 32 (3): 909-917.
- [13] Özen, S., 2018. KASKİ Atık Su Arıtma Verilerinin Yapay Sinir Ağları Ve Bulanık Mantık Yöntemleri İle Tahmin Edilmesi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kayseri, s. 1-98.
- [14] Ahmed, A. N., Othman, F. B., Afan, H. A., Ibrahim, R. K., Fai, C. M., Hossain, M. S., Elshafie, A., 2019. Machine learning methods for better water quality prediction. **Journal of Hydrology**, 578: 124084.
- [15] Acar, R., Saplıoğlu, K., 2022. Etkili girdi parametrelerinin çoklu regresyon ile belirlendiği su sertliğinin anfis yöntemi ile tahmin edilmesi. **Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi**, 22(6): 1413-1424.
- [16] Shamsuddin, I. I. S., Othman, Z., Sani, N. S., 2022. Water quality index classification based on machine learning: A case from the langat river basin model. **Water**, 14: 2939.
- [17] Haghiabi, A. H., Nasrolahi, A. H., Parsaie, A., 2018. Water quality prediction using machine learning methods. **Water Quality Research Journal**, 53(1): 3-13.
- [18] Uncuoglu, E., Citakoglu, H., Latifoglu, L., Bayram, S., Laman, M., Ilkentapar, M., Oner, A. A., 2022. Comparison of neural network, Gaussian regression,

support vector machine, long short-term memory, multi-gene genetic programming, and M5 Trees methods for solving civil engineering problems. **Applied Soft Computing**, 129: 109623.

- [19] KASKİ., 2024. Kayseri Büyükşehir Belediyesi. (Web sayfası: <https://www.kaski.gov.tr/>), (Erişim tarihi: Mart 2024).
- [20] İçme Suyu Kabul Edilebilir Değerler, Marmara Üniversitesi. 2022. (Web sayfası: [https://www. İçme Suyu Kabul Edilebilir Değerler | Doğa Bitkileri ve Su Ürünleri Araştırma ve Uygulama Merkezi | Marmara Üniversitesi.gov.tr/](https://www.İçme Suyu Kabul Edilebilir Değerler | Doğa Bitkileri ve Su Ürünleri Araştırma ve Uygulama Merkezi | Marmara Üniversitesi.gov.tr/)), (Erişim tarihi: Kasım 2022)
- [21] T. C. Cumhurbaşkanlığı Mevzuat Bilgi Sistemi, 2005. İnsani tüketim amaçlı sular hakkında yönetmelik. (Web sayfası: <https://www.mevzuat.gov.tr/-mevzuat?MevzuatNo=7510&MevzuatTur=7&MevzuatTertip=5>), (Erişim tarihi: Şubat 2022).
- [22] Durmuş, M. T., 2021. Ankara Çayı su kalitesinin NSF-WQI indeksi kullanılarak belirlenmesi ve çözüm önerilerinin değerlendirilmesi.
- [23] Keskenler, M. F., Keskenler, E. F., 2017. Geçmişten günümüze yapay sinir ağları ve tarihçesi. **Takvim-i Vekayi**, 5 (2): 8-18.
- [24] Yazıcı, A. C., Ögüş, E., Ankaralı, S., Canan, S., Ankaralı, H., Akkuş, Z., 2007. Yapay Sinir Ağlarına Genel Bakış. **Türkiye Klinikleri Journal Of Medical Sciences**, 27 (1): 65-71.
- [25] Kabalci, E., 2017. Yapay Sinir Ağları. Artificial Neural Networks.
- [26] Kotan, M., 2022. Eko-469-Veri madenciliği - Sınıflandırma. (Web sayfası: [https://mkotan.sakarya.edu.tr/sites/mkotan.sakarya.edu.tr/file/EKO469_V M_H12_Siniflandirma3.pdf](https://mkotan.sakarya.edu.tr/sites/mkotan.sakarya.edu.tr/file/EKO469_V_M_H12_Siniflandirma3.pdf)), (Erişim Tarihi: Ocak 2024)
- [27] Ögücü, M. O., 2006. Yapay Sinir Ağları İle Sistem Tanıma, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi, İstanbul, s. 1-85.

- [28] Öztürk, K., Şahin, M.E., 2018. Yapay sinir ağları ve yapay zekâ' ya genel bir bakış. **Takvim-i Vekayi**, 6 (2): 25-36
- [29] Güzel, K., Güzel, K., 2019. Geri yayımlı çok katmanlı yapay sinir ağları-1. (Web sayfası: <https://medium.com/@billmuhh/geri-yayimli-cokkatmanli-yapay-sinir-aglari-1-47daa3856247>), (Erişim tarihi: Kasım 2018).
- [30] Şimsek, N. Y. Derin Öğrenme (Deep Learning) Nedir ve Nasıl Çalışır?.
- [31] Sengöz, N., 2017. Derin öğrenme/ Deep learning. (Web sayfası: www.derinogrenme.com), (Erişim Tarihi: Mart 2024)
- [32] Çoşkun, Ö., 2023. Sakarya Havzasındaki Kısa Dönem Meteorolojik Kuraklığın Hibrit Modeller İle Tahmin Edilmesi, Erciyes Üniversitesi, Doktora Tezi, Kayseri, s. 1-235.
- [33] Jang, J. S., 1993. ANFIS: adaptive-network-based fuzzy inference system. **IEEE Transactions on Systems, Man, and Cybernetics**, 23(3): 665-685.
- [34] Aran, V., 2019. Flexible And Robust Control Of Heavy Duty Diesel Engine Airpath Using Data Driven Disturbance Observers And GPR Models, Sabancı Üniversitesi, Doktora Tezi, İstanbul, s. 1-117.
- [35] Demirgöl, T., Demir, V., Sevimli, M. F., 2024. Farklı makine öğrenmesi yaklaşımları ile Türkiye'nin solar radyasyon tahmini. **Geomatik**, 9 (1): 106-122.
- [36] Vives, J., 2022. Incorporating machine learning into vibration detection for wind turbines. **Modelling and Simulation in Engineering**, 2022 (1): 6572298.
- [37] Alawi, O. A., Kamar, H. M., Abdelrazek, A. H., Mallah, A. R., Mohammed, H. A., Homod, R. Z., Yaseen, Z. M., 2024. Design optimization of solar collectors with hybrid nanofluids: An integrated ansys and machine learning study. **Solar Energy Materials and Solar Cells**, 271: 112822.

- [38] Gültekin, S. U., 2017. Veri Madenciliği: Yapay Sinir Ağı ve Doğrusal Regresyon Yöntemleri ile Fiyat Tahmini, Pamukkale Üniversitesi, Yüksek Lisans Tezi, Denizli, s. 1-161.
- [39] Başakın, E. E., Ekmekcioğlu, Ö., Çıtakoğlu, H., Özger, M., 2022. A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. **Neural Computing and Applications**, **34**(1), 783-812.
- [40] Sungur, H., 2023. Sivas İli ve Çevresi Rüzgar Hızı Verilerinin Farklı Makine Öğrenme Yöntemleri ile Tahmin Edilmesi ve Karşılaştırılması, Erciyes Üniversitesi, Yüksek Lisans Tezi, Kayseri, s. 1-97.
- [41] Bayram, S., Çıtakoğlu, H., 2023. Modeling monthly reference evapotranspiration process in Turkey: application of machine learning methods. **Environ Monit Assess**, **195** (67): 1-23.
- [42] Willmott, C. J., Robeson, S. M., Matsuura, K., 2012. A refined index of model performance. **International Journal of climatology**, **32** (13): 2088-2094.
- [43] Willmott, C. J., Robeson, S. M., Matsuura, K., 2012. A refined index of model performance. **International Journal of climatology**, **32** (13): 2088-2094.
- [44] 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.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı: Nur ORHAN

Uyruğu: T

Doğum Tarihi ve Yeri:

Medeni Durum:

e-mail:

Telefon:

EĞİTİM

Derece	Kurum	Mezuniyet Tarihi
Yüksek Lisans	Erciyes Üniversitesi, İnşaat Mühendisliği	2024
Lisans	Erciyes Üniversitesi, İnşaat Mühendisliği	2022
Lise	2000 Evler Anadolu Lisesi, Nevşehir	2017

YABANCI DİL

İngilizce

YAYINLAR

- Orhan, N., Çıtakoğlu, H., Öner, A., "Least Square Support Vector Regression Models for Water Quality Index Prediction," 7th International Azerbaijan Congress on Life, Engineering, Mathematical, and Applied Sciences, vol.1, no.1, Baku, Azerbaijan, pp.163, 2024