T.C.
KASTAMONU ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ORMAN MÜHENDİSLİĞİ ANA BİLİM DALI

MAKİNE ÖĞRENMESİ VE AÇIKLANABİLİR YAPAY ZEKA YAKLAŞIMI İLE BARAJ DOLULUK ORANININ TAHMİNİ; KASTAMONU - GERMEÇTEPE BARAJI ÖRNEĞİ

UĞUR ÇELİK

YÜKSEK LİSANS TEZİ

DOÇ. DR. SENEM GÜNEŞ ŞEN

TEMMUZ - 2026
KASTAMONU

TEZ ONAYI

Uğur ÇELİK tarafından hazırlanan "**MAKİNE ÖĞRENMESİ VE AÇIKLANABİLİR YAPAY ZEKA YAKLAŞIMI İLE BARAJ DOLULUK ORANININ TAHMİNİ; KASTAMONU - GERMEÇTEPE BARAJI ÖRNEĞİ**" adlı tez çalışmasının savunma sınavı **17.07.2026** tarihinde yapılmış olup aşağıda verilen jüri tarafından oy birliği ile Kastamonu Üniversitesi Fen Bilimleri Enstitüsü **Orman Mühendisliği Ana Bilim Dalı Yüksek Lisans Tezi** olarak kabul edilmiştir.

Danışman	Doç. Dr. Senem GÜNEŞ ŞEN Kastamonu Üniversitesi	.....
Jüri Üyesi	Prof. Dr. Miraç AYDIN Kastamonu Üniversitesi	.....
Jüri Üyesi	Doç. Dr. Semih EDİŞ Çankırı Karatekin Üniversitesi	.....

Jüri üyeleri tarafından kabul edilmiş olan bu tez Kastamonu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulunca onanmıştır.

Enstitü Müdürü Prof. Dr. Selahattin KAYMAKCI .....

TAAHHÜTNAME

Bu tezin tasarımı, hazırlanması, yürütülmesi, araştırmalarının yapılması ve bulgularının analizlerinde bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu; ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını, bilimsel etiğe uygun olarak kaynak gösterildiğini bildirir ve taahhüt ederim.

Uğur ÇELİK

ÖZET

YÜKSEK LİSANS TEZİ

MAKİNE ÖĞRENMESİ VE AÇIKLANABİLİR YAPAY ZEKA YAKLAŞIMI İLE BARAJ DOLULUK ORANININ TAHMİNİ; KASTAMONU - GERMEÇTEPE BARAJI ÖRNEĞİ

UĞUR ÇELİK

KASTAMONU ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ
ORMAN MÜHENDİSLİĞİ ANA BİLİM DALI
DANIŞMAN:DOÇ. DR. SENEM GÜNEŞ ŞEN

Bu çalışmada, Kastamonu ilinde bulunan Germeçtepe Barajı'nın günlük aktif doluluk oranının farklı makine öğrenmesi algoritmaları kullanılarak tahmin edilmesi ve meteorolojik değişkenlerin model üzerindeki etkilerinin açıklanabilir yapay zekâ yöntemleri ile değerlendirilmesi amaçlanmıştır. Araştırmada 01 Ocak 2008–31 Aralık 2025 dönemine ait toplam 6574 günlük gözlem kullanılmıştır. Bağımlı değişken olarak günlük aktif doluluk oranı; bağımsız değişkenler olarak ise bir önceki güne ait aktif doluluk oranı (Lag-1), günlük ortalama sıcaklık, günlük toplam yağış, ortalama nispi nem ve ortalama aktüel basınç verileri modele dahil edilmiştir. Doğrusal Regresyon, Karar Ağacı, Random Forest ve XGBoost algoritmaları uygulanmış, ağaç tabanlı modellerde hiperparametre optimizasyonu gerçekleştirilmiştir. Model performansları MAE, RMSE ve R^2 ölçütleri kullanılarak değerlendirilmiş; en başarılı model üzerinde Feature Importance, Permutation Importance, SHAP ve Partial Dependence Plot analizleri uygulanmıştır.

Elde edilen sonuçlara göre optimize edilmiş Random Forest modeli en başarılı algoritma olmuş ve test veri setinde MAE=0,3994, RMSE=0,6718 ve R^2=0,9993 değerleri elde edilmiştir. Açıklanabilir yapay zekâ analizleri, model tahminlerinde en yüksek katkıya sahip değişkenin Lag-1 olduğunu göstermiştir. Lag-1 değişkeni model dışında bırakıldığında ise meteorolojik değişkenler arasında günlük ortalama sıcaklığın en etkili değişken olduğu belirlenmiştir. SHAP ve PDP analizleri, meteorolojik değişkenlerin aktif doluluk oranı üzerindeki etkilerinin doğrusal olmadığını ortaya koymuştur.

Sonuç olarak, optimize edilmiş Random Forest algoritmasının Germeçtepe Barajı'nın günlük aktif doluluk oranının tahmininde güvenilir sonuçlar verdiği belirlenmiştir. Ayrıca açıklanabilir yapay zekâ yöntemlerinin kullanılması sayesinde yalnızca model performansı değil, modelin karar mekanizması da ayrıntılı biçimde yorumlanmıştır. Çalışmanın, rezervuar yönetimi ve su kaynakları planlamasında karar destek sistemlerinin geliştirilmesine katkı sağlayacağı değerlendirilmektedir.

ANAHTAR KELİMELER: Makine Öğrenmesi, Random Forest, XGBoost, Açıklanabilir Yapay Zekâ, SHAP, Rezervuar Yönetimi, Aktif Doluluk Oranı

Temmuz 2026, 72 Sayfa

ABSTRACT

MSC THESIS

ESTIMATION OF RESERVOIR STORAGE RATIO USING MACHINE LEARNING AND EXPLAINABLE ARTIFICIAL INTELLIGENCE: A CASE STUDY OF GERMEÇTEPE DAM, KASTAMONU

UĞUR ÇELİK

KASTAMONU UNIVERSITY INSTITUTE OF SCIENCE
DEPARTMENT OF FOREST ENGINEERING
SUPERVISOR:ASSOC. PROF. DR. SENEM GÜNEŞ ŞEN

This study aimed to predict the daily active reservoir storage ratio of the Germeçtepe Dam, located in Kastamonu, Türkiye, using different machine learning algorithms and to evaluate the effects of meteorological variables on model predictions through Explainable Artificial Intelligence (XAI) methods. A total of 6,574 daily observations covering the period from 2 January 2008 to 31 December 2025 were used. The daily active reservoir storage ratio was considered as the dependent variable, while the previous day's active storage ratio (Lag-1), daily mean temperature, daily total precipitation, mean relative humidity, and mean actual atmospheric pressure were used as independent variables. Linear Regression, Decision Tree, Random Forest, and XGBoost algorithms were implemented, and hyperparameter optimization was performed for the tree-based models. Model performance was evaluated using MAE, RMSE, and R^2. Furthermore, Feature Importance, Permutation Importance, SHAP, and Partial Dependence Plot analyses were applied to the best-performing model.

The results indicated that the optimized Random Forest model achieved the highest predictive performance, with MAE = 0.3994, RMSE = 0.6718, and R^2 = 0.9993 on the test dataset. Explainable Artificial Intelligence analyses revealed that the Lag-1 variable made the greatest contribution to model predictions. When the Lag-1 variable was excluded, daily mean temperature emerged as the most influential meteorological variable. SHAP and PDP analyses further demonstrated that the effects of meteorological variables on the active reservoir storage ratio were predominantly non-linear.

In conclusion, the optimized Random Forest algorithm proved to be a reliable approach for predicting the daily active reservoir storage ratio of the Germeçtepe Dam. Moreover, the application of Explainable Artificial Intelligence methods enabled not only the evaluation of model performance but also a comprehensive interpretation of the model's decision-making process. The findings are expected to contribute to the development of decision support systems for reservoir operation and water resources management.

KEYWORDS: Machine Learning, Random Forest, XGBoost, Explainable Artificial Intelligence, SHAP, Reservoir Management, Active Reservoir Storage Ratio

July 2026, 72 Page

TEŞEKKÜR

Bu tez çalışmasının hazırlanması sürecinde bilgi, deneyim, destek ve katkılarıyla yanımda olan, akademik ve mesleki gelişimime değer katan tüm kişi ve kurumlara en içten teşekkürlerimi sunarım.

Öncelikle, tez danışmanım Doç. Dr. Senem Güneş Şen'e; çalışmanın tüm aşamalarında engin bilgi ve akademik birikimiyle bana yol gösterdiği, bilimsel yaklaşımı, yapıcı eleştirileri, değerli önerileri ve desteklerinden dolayı en içten saygı ve şükranlarımı sunarım. Araştırmacı kimliğimin gelişmesine ve bu çalışmanın nitelikli bir şekilde tamamlanmasına sağladığı katkılar benim için her zaman yol gösterici olacaktır.

Görev yapmaktan onur duyduğum Devlet Su İşleri 23. Bölge Müdürlüğü bünyesinde, tez çalışmam süresince göstermiş oldukları destek ve anlayışlarından dolayı Bölge Müdürümüz Sayın Yasin Devrim'e, Bölge Müdür Yardımcımız Sayın Mustafa Üstün'e, Havza Yönetimi, İzleme ve Tahsisler Şube Müdürümüz Sayın Ahmet Yalçın'a, İşletme ve Bakım Şube Müdürümüz Sayın Hüseyin Salih Akdaş'a ve çalışma arkadaşım Meteoroloji Mühendisi Sayın Ali Ertürk'e teşekkür ederim.

Hayatım boyunca desteklerini esirgemeyen anneme ve kardeşime, yoğun çalışma sürecinde sabrı, anlayışı, sevgisi ve fedakârlığıyla bana her zaman güç veren sevgili eşim Esma Çelik'e ve hayatıma anlam ve mutluluk katan canım çocuklarım Gökçe Hüma ile Alparslan Uğur Gazi'ye gönülden teşekkür ederim.

Dürüstlük, çalışkanlık ve vatan sevgisini öğreten, karakterimin şekillenmesinde büyük emeği olan ve bugün aramızda olmasa da hatırası her zaman benimle olan kıymetli babamı rahmet, özlem ve minnetle anıyorum. Bu çalışmayı onun aziz hatırasına ithaf ediyor, emeğini daima şükranla taşıyacağımı ifade ediyorum.

Son olarak, bu çalışmanın hazırlanmasında doğrudan veya dolaylı emeği geçen tüm akademisyenlere, mesai arkadaşlarıma ve katkı sağlayan herkese teşekkür eder, saygılarımı sunarım.

UĞUR ÇELİK

Kastamonu, 2026

İÇİNDEKİLER

Sayfa

TEZ ONAYI ..... **ii**
TAAHHÜTNAME ..... **iii**
ÖZET ..... **iv**
ABSTRACT ..... **v**
TEŞEKKÜR ..... **vi**
İÇİNDEKİLER ..... **vii**
ŞEKİLLER DİZİNİ ..... **ix**
TABLOLAR DİZİNİ ..... **x**
SİMGELER VE KISALTMALAR DİZİNİ ..... **xi**
1. GİRİŞ **1**
2. LİTERATÜR ÖZETİ **10**
2.1 Rezervuar Su Seviyesi Tahmininde Geleneksel Makine Öğrenmesi Yaklaşımları 10
2.2 Topluluk Öğrenmesi ve Derin Öğrenme Tabanlı Yaklaşımlar 12
2.3 Rezervuar Modellemesinde Güncel Eğilimler ve Araştırma Boşluğu 16
3. MATERYAL VE METOT **19**
3.1 Çalışma Alanı 19
3.2 Veri Seti ve Veri Kaynakları 20
3.3 Veri Ön İşleme Süreci 21
3.4 Kullanılan Makine Öğrenmesi Algoritmaları 24
3.4.1 Doğrusal Regresyon (Linear Regression, LR) 24
3.4.2 Karar Ağacı (Decision Tree, DT) 26
3.4.3 Random Forest (RF) 27
3.4.4 Extreme Gradient Boosting (XGBoost) 29
3.5 Model Optimizasyonu ve Hiperparametre Ayarları 30
3.6 Model Performansının Değerlendirilmesi 32
3.6.1 Ortalama Mutlak Hata (MAE) 33
3.6.2 Kök Ortalama Kare Hata (RMSE) 33
3.6.3 Belirlilik Katsayısı (R^2) 33
3.7 Yazılım Ortamı ve Kullanılan Kütüphaneler 34
3.8 Model Yorumlama ve Açıklanabilir Yapay Zekâ Analizleri 35
4. BULGULAR **37**
4.1 Çalışmada Kullanılan Değişkenlere Ait Tanımlayıcı İstatistikler 37
4.1.1 Değişkenlerin Normal Dağılım Özellikleri 38
4.1.2 Değişkenler Arasındaki Korelasyon İlişkileri 38
4.2 Baraj Doluluk Oranının Makine Öğrenmesi Modelleri ile Tahmini 40
4.2.1 Veri Setinin Hazırlanması 40
4.2.2 Linear Regresyon Sonuçları 41
4.2.3 Decision Tree Sonuçları 42
4.2.4 Random Forest Sonuçları 44
4.2.5 XGBoost Sonuçları 45
4.2.6 Modellerin Karşılaştırılması 47
4.2.7 En Başarılı Modelin Değerlendirilmesi 48
4.3 Meteorolojik Değişkenlerin Baraj Doluluğu Üzerindeki Etkileri 51

4.3.1 Korelasyon Analizi 51
4.3.2 Meteorolojik Değişkenlerin Önem Analizi (Feature Importance ve Permutation Importance) 53
4.3.3 SHAP Analizi 55
4.3.4 Partial Dependence Plot (PDP) Analizi 57
5. SONUÇLAR VE TARTIŞMA 59
5.1 Makine Öğrenmesi Modellerinin Performanslarının Değerlendirilmesi 59
5.2 Meteorolojik Değişkenlerin Baraj Doluluk Oranı Üzerindeki Etkilerinin Değerlendirilmesi 60
6. ÖNERİLER 64
KAYNAKLAR 66
ÖZGEÇMİŞ 72

ŞEKİLLER DİZİNİ

Sayfa

Şekil 3.1 Germeçtepe Barajının coğrafi konumu 20
Şekil 3.2 Makine öğrenmesi modelleme sürecine ait araştırma akış şeması 23
Şekil 4.1 Linear Regresyon modeli için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması 42
Şekil 4.2 Optimize edilmiş Decision Tree modeli için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması..... 43
Şekil 4.3 Optimize edilmiş Random Forest modeli için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması 45
Şekil 4.4 Optimize edilmiş XGBoost modeli için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması. 46
Şekil 4.5 Farklı makine öğrenmesi modelleri için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması 48
Şekil 4.6 Optimize edilmiş Random Forest modeli için gözlenen ve tahmin edilen aktif doluluk oranlarının saçılım grafiği..... 49
Şekil 4.7 Optimize edilmiş Random Forest modeli için artık değerlerin tahmin edilen aktif doluluk oranlarına göre dağılımı..... 50
Şekil 4.8 Optimize edilmiş Random Forest modeline ait değişken önem düzeyleri..... 50
Şekil 4.9 Aktif doluluk oranı ile meteorolojik değişkenler arasındaki Pearson korelasyon ısı haritası..... 52
Şekil 4.10 Optimize edilmiş Random Forest modeline ait Feature Importance sonuçları 54
Şekil 4.11 Optimize edilmiş Random Forest modeline ait Permutation Importance sonuçları..... 55
Şekil 4.12 Optimize edilmiş Random Forest modeline ait SHAP özet grafiği..... 56
Şekil 4.13 Optimize edilmiş Random Forest modeline ait ortalama mutlak SHAP değerleri 56
Şekil 4.14 Optimize edilmiş Random Forest modeline ait meteorolojik değişkenlerin kısmi bağımlılık (PDP) grafikleri..... 57

TABLOLAR DİZİNİ

Sayfa

Tablo 3.1 Çalışmada kullanılan değişkenler ve veri kaynakları 21
Tablo 4.1 Çalışmada kullanılan değişkenlere ait tanımlayıcı istatistikler..... 37
Tablo 4.2 Değişkenlere ait normal dağılım testi sonuçları 38
Tablo 4.3 Pearson korelasyon katsayıları..... 39
Tablo 4.4 Bağımsız değişkenlere ait Varyans Artış Faktörü (VIF) sonuçları..... 40
Tablo 4.5 Makine öğrenmesi modellerinde kullanılan veri setinin genel özellikleri..... 41
Tablo 4.6 Linear Regresyon modelinin performans sonuçları..... 42
Tablo 4.7 Decision Tree modelinin performans sonuçları..... 43
Tablo 4.8 Optimize edilmiş Decision Tree modeline ait hiperparametreler..... 43
Tablo 4.9 Random Forest modelinin performans sonuçları..... 44
Tablo 4.10 Optimize edilmiş Random Forest modeline ait en iyi hiperparametreler 44
Tablo 4.11 XGBoost modelinin performans sonuçları 45
Tablo 4.12 Optimize edilmiş XGBoost modeline ait en iyi hiperparametreler 46
Tablo 4.13 Makine öğrenmesi modellerinin performans karşılaştırması..... 47
Tablo 4.14 Aktif doluluk oranı ile meteorolojik değişkenler arasındaki Pearson korelasyon katsayıları..... 51
Tablo 4.15 Optimize edilmiş Random Forest modeline ait performans ölçütleri..... 53

SİMGELER VE KISALTMALAR DİZİNİ

Kısaltmalar

AI : Artificial Intelligence (Yapay Zekâ)
ANN : Artificial Neural Network (Yapay Sinir Ağı)
CV : Cross Validation (SDoğrulama)
DSİ : Devlet Su İşleri Genel Müdürlüğü
DT : Decision Tree (Karar Ağacı)
GRU : Gated Recurrent Unit
LSTM : Long Short-Term Memory
MAE : Mean Absolute Error (Ortalama Mutlak Hata)
ML : Machine Learning (Makine Öğrenmesi)
PDP : Partial Dependence Plot
R^2 : Coefficient of Determination (Belirleme Katsayısı)
RF : Random Forest
RMSE : Root Mean Square Error (Kök Ortalama Kare Hata)
SHAP : SHapley Additive exPlanations
SVR : Support Vector Regression (Destek Vektör Regresyonu)
VIF : Variance Inflation Factor (Varyans Artış Faktörü)
XAI : Explainable Artificial Intelligence (Açıklanabilir Yapay Zekâ)
XGBoost : Extreme Gradient Boosting

1. GİRİŞ

Su, yeryüzündeki yaşamın devamlılığını sağlayan en temel doğal kaynaklardan biri olup ekosistemlerin işleyişi, ekonomik kalkınma, gıda güvenliği ve toplumsal refah açısından vazgeçilmez bir öneme sahiptir. Dünya nüfusundaki hızlı artış, kentleşme, sanayileşme ve tarımsal üretimdeki genişleme, kullanılabilir tatlı su kaynakları üzerindeki baskıyı her geçen yıl artırmaktadır. UNESCO (2023)'ya göre kullanılabilir tatlı su kaynakları dünya su varlığının oldukça sınırlı bir bölümünü oluşturmakta ve artan nüfus ile iklim değişikliği bu kaynaklar üzerindeki baskıyı giderek artırmaktadır. FAO (2021) özellikle tarımsal üretimde artan su talebinin mevcut su kaynakları üzerindeki baskıyı giderek artırdığını, su kullanım verimliliğinin geliştirilmesinin sürdürülebilir kalkınma açısından kritik önem taşıdığını belirtmektedir.

Su kaynakları üzerindeki baskıyı artıran en önemli etkenlerden biri de iklim değişikliğidir. Küresel sıcaklıklardaki artış, yağış rejimlerinde meydana gelen değişimler, kar örtüsünün azalması, buharlaşma miktarındaki artış ve ekstrem hidrolojik olayların daha sık görülmesi doğal su döngüsünü önemli ölçüde değiştirmektedir. IPCC (2021) tarafından yayımlanan Altıncı Değerlendirme Raporu'nda, küresel sıcaklık artışının devam etmesi durumunda hidrolojik döngüdeki değişimlerin daha belirgin hâle geleceği, birçok bölgede kuraklık şiddeti ve süresinin artacağı, buna karşılık kısa süreli ancak yüksek şiddetli yağış olaylarının daha sık görüleceği belirtilmektedir. Bu değişimler, rezervuarların işletilmesinde hidrolojik belirsizliği artırmakta ve su yönetiminde daha güvenilir tahmin yöntemlerine olan gereksinimi güçlendirmektedir. Wang vd. (2025) ise rezervuar işletmesinde tahmin belirsizliğinin karar verme sürecini doğrudan etkilediğini ve tahmin bilgilerindeki belirsizliğin dikkate alınmasının hem taşkın riskinin azaltılması hem de su kaynaklarının daha etkin kullanılabilmesi açısından kritik önem taşıdığını göstermiştir. Liang vd. (2018) ise değişen hidrolojik koşulların rezervuar su seviyelerinde önemli dalgalanmalara neden olduğunu ve bu nedenle veri odaklı tahmin yaklaşımlarının su kaynakları yönetiminde giderek daha fazla önem kazandığını ifade etmektedir.

İklim değişikliğinin su döngüsü üzerindeki etkileri yalnızca toplam yağış miktarındaki değişimlerle sınırlı değildir. Artan hava sıcaklığına bağlı olarak buharlaşma ve evapotranspirasyon miktarlarının yükselmesi, kar erime zamanlarının değişmesi, yüzeysel akışın mevsimsel dağılımının farklılaşması ve yeraltı suyu beslenmesindeki değişiklikler havza hidrolojisini önemli ölçüde etkilemektedir. Ibañez vd. (2022), rezervuar su seviyelerinin güvenilir biçimde tahmin edilmesinin su kaynakları yönetimi ve rezervuar işletmesi açısından kritik öneme sahip olduğunu, su seviyelerinin yağış, giriş akımları, sulama talepleri ve diğer hidrolojik değişkenlerdeki belirsizliklerden etkilendiğini belirtmektedir. Bu nedenle, gelişmiş tahmin modellerinin kullanılması rezervuar işletme kararlarının doğruluğunu artırmada önemli bir araç olarak değerlendirilmektedir. Buna ek olarak, birçok havzada suyun zamansal ve mekânsal dağılımı daha değişken hâle gelmekte, su arzının güvenilirliği azalmaktadır (IPCC, 2021). Özellikle yarı kurak ve kurak iklim bölgelerinde bu değişimler, mevcut su depolama sistemlerinin işletilmesini daha karmaşık bir süreç hâline getirmektedir. Bu nedenle günümüzde yalnızca mevcut su kaynaklarının yönetimi değil, gelecekte oluşabilecek hidrolojik koşulların güvenilir biçimde tahmin edilmesi de sürdürülebilir su yönetiminin temel bileşenlerinden biri olarak kabul edilmektedir.

Bu süreçte barajlar, değişken hidrolojik koşullar altında suyun depolanmasını, düzenlenmesini ve farklı amaçlarla kullanılmasını sağlayan en önemli su yapılarıdır. Barajlar; içme ve kullanma suyu temini, tarımsal sulama, hidroelektrik enerji üretimi, taşkın kontrolü, rekreasyon ve ekolojik su ihtiyaçlarının karşılanması gibi çok sayıda işlevi aynı anda yerine getirmektedir. FAO (2021) verilerine göre dünya genelinde çekilen tatlı suyun yaklaşık %70'i tarımsal sulamada, %20'si sanayide ve %10'u evsel amaçlarla kullanılmaktadır. Bu nedenle özellikle sulama ve içme suyu amacıyla işletilen rezervuarlarda depolanan su miktarının doğru biçimde yönetilmesi hem ekonomik faaliyetlerin sürdürülebilirliği hem de toplumsal refah açısından kritik öneme sahiptir. Das vd. (2016) ve Hipni vd. (2013), rezervuar su seviyesinin güvenilir biçimde tahmin edilmesinin rezervuar işletmesi, su kaynakları planlaması ve su yönetimi açısından kritik öneme sahip olduğunu vurgulamaktadır. Güvenilir tahminler sayesinde su temini, sulama, taşkın kontrolü ve hidroelektrik üretimine ilişkin işletme kararlarının daha etkin biçimde planlanabilmesi mümkün olmaktadır.

Barajların etkin biçimde işletilebilmesi, rezervuarda depolanan su miktarının doğru tahmin edilmesine bağlıdır. Rezervuar aktif doluluk oranı, mevcut su potansiyelini doğrudan gösteren temel göstergelerden biri olup işletme kararlarının büyük bölümü bu bilgiye dayanmaktadır. Günlük aktif doluluk oranının doğru tahmin edilmesi; su tahsis planlarının hazırlanması, hidroelektrik enerji üretiminin optimize edilmesi, kuraklık dönemlerinde su kullanım önceliklerinin belirlenmesi ve taşkın risklerinin azaltılması açısından büyük önem taşımaktadır. Sapitang vd. (2020), makine öğrenmesi tabanlı rezervuar tahmin modellerinin işletme kararlarının doğruluğunu artırabileceğini ve su kaynaklarının daha etkin yönetilmesine katkı sağlayabileceğini göstermiştir. Dolayısıyla iklim değişikliğine bağlı hidrolojik belirsizliklerin arttığı günümüzde, rezervuar doluluk oranlarının güvenilir biçimde tahmin edilmesi yalnızca operasyonel bir gereklilik değil, aynı zamanda sürdürülebilir su kaynakları yönetiminin temel unsurlarından biri hâline gelmiştir.

ürkiye, sahip olduğu iklim özellikleri, topoğrafik yapısı ve düzensiz yağış rejimi nedeniyle su kaynaklarının planlı ve sürdürülebilir biçimde yönetilmesinin büyük önem taşıdığı ülkeler arasında yer almaktadır. Ülkenin yıllık kullanılabilir su potansiyeli yaklaşık 112 milyar m^3 olup, kişi başına düşen yıllık kullanılabilir su miktarı yaklaşık 1.346 m^3'tür (DSİ, 2023). Bu değer, uluslararası sınıflandırmalara göre Türkiye'nin su stresi yaşayan ülkeler arasında yer aldığını göstermektedir. Nüfus artışı, kentleşme, sanayileşme ve tarımsal sulama faaliyetlerinin genişlemesiyle birlikte suya olan talebin önümüzdeki yıllarda daha da artacağı öngörülmektedir (FAO, 2021). Buna ek olarak iklim değişikliğine bağlı yağış düzensizlikleri, sıcaklık artışları ve artan buharlaşma, mevcut su kaynaklarının sürdürülebilir yönetimini daha da zorlaştırmaktadır (IPCC, 2021). Bu nedenle mevcut su potansiyelinin etkin kullanılması, su arz güvenliğinin sağlanması ve entegre su kaynakları yönetimi uygulamalarının güçlendirilmesi, Türkiye'nin ulusal su politikalarının öncelikli hedefleri arasında yer almaktadır.

Türkiye'de su kaynaklarının yönetiminde en önemli görevlerden biri barajlara düşmektedir. İçme ve kullanma suyu temini, tarımsal sulama, hidroelektrik enerji üretimi ve taşkın kontrolü amacıyla işletilen yüzlerce baraj, ülkenin su yönetim sisteminin temel bileşenlerini oluşturmaktadır. Özellikle son yıllarda büyükşehirlerin

içme suyu ihtiyacını karşılayan barajlarda gözlenen doluluk oranı değişimleri kamuoyunun da yakından takip ettiği önemli göstergeler hâline gelmiştir (Yılmaz vd., 2020). Kurak dönemlerde rezervuar doluluk oranlarında meydana gelen hızlı düşüşler, su kullanım kısıtlamaları ve enerji üretiminde yaşanabilecek olası sorunlar açısından önemli riskler oluşturmaktadır. Hipni vd. (2013), güvenilir su seviyesi tahminlerinin rezervuar işletme kararlarının başarısını doğrudan etkilediğini belirtirken, Das vd. (2016) günlük rezervuar su seviyesi tahminlerinin operasyonel rezervuar yönetiminde karar verme süreçlerinin iyileştirilmesine önemli katkılar sağladığını ifade etmektedir. Dolayısıyla rezervuarlardaki su seviyesinin ve aktif doluluk oranının güvenilir biçimde tahmin edilmesi, yalnızca mühendislik açısından değil, ekonomik, çevresel ve sosyal açıdan da stratejik öneme sahiptir.

Baraj doluluk oranının doğru tahmin edilmesi, rezervuar işletme süreçlerinde alınan kararların etkinliğini doğrudan etkilemektedir. Günlük veya kısa dönemli doluluk tahminleri sayesinde rezervuardan bırakılacak su miktarı planlanabilmekte, sulama programları optimize edilebilmekte, hidroelektrik enerji üretimi daha verimli gerçekleştirilebilmekte ve kuraklık dönemlerinde uygulanacak su yönetim stratejileri önceden belirlenebilmektedir. Bunun yanında taşkın dönemlerinde rezervuarda yeterli depolama hacminin oluşturulabilmesi için de gelecekteki su seviyelerinin doğru tahmin edilmesi büyük önem taşımaktadır. Sapitang vd. (2020), makine öğrenmesi tabanlı rezervuar tahmin modellerinin işletme kararlarının doğruluğunu önemli ölçüde artırabileceğini göstermiştir. Benzer şekilde García-Feal vd. (2022), veri odaklı tahmin modellerinin rezervuar işletmeciliğinde etkili karar destek sistemleri oluşturulmasına önemli katkılar sağladığını belirtmektedir. Bu nedenle rezervuar doluluk oranının tahmini, günümüzde birçok ülkede karar destek sistemlerinin temel bileşenlerinden biri olarak değerlendirilmektedir.

Rezervuar su seviyesi ve aktif doluluk oranının tahmini amacıyla literatürde uzun yıllar boyunca çoklu doğrusal regresyon, zaman serisi analizleri ve çeşitli istatistiksel modelleme yaklaşımları kullanılmıştır. Stojanovic vd. (2013), klasik regresyon modellerinin belirli koşullarda başarılı sonuçlar üretebildiğini ancak veri yapısı karmaşıklaştıkça performanslarının azaldığını belirtmiştir. Wong vd. (2019) ise rezervuar su seviyesi tahmininde kullanılan geleneksel yöntemlerin doğrusal olmayan

ilişkileri açıklamada yetersiz kalabildiğini ifade etmektedir. Çünkü rezervuar davranışı yalnızca yağış ve sıcaklık gibi meteorolojik değişkenlerden değil; havza özellikleri, yüzeysel akış, yeraltı suyu katkısı, rezervuar işletme politikaları, su çekimleri, buharlaşma ve önceki dönem su depolama koşulları gibi çok sayıda faktörün birlikte etkisi altında şekillenmektedir. Bu değişkenler arasındaki ilişkiler çoğu zaman doğrusal değildir ve zamana bağlı karmaşık etkileşimler içermektedir. Kisi vd. (2012), bu nedenle yapay zekâ tabanlı yöntemlerin klasik zaman serisi modellerine göre daha başarılı sonuçlar verdiğini göstermiştir. Wee vd. (2021) ise son yıllarda rezervuar modellemelerinde makine öğrenmesi yaklaşımlarının giderek yaygınlaştığını ve geleneksel yöntemlerin yerini almaya başladığını vurgulamaktadır.

Son yıllarda yapay zekâ ve makine öğrenmesi yöntemlerinde yaşanan hızlı gelişmeler, rezervuar modelleme çalışmalarında yeni bir dönemin başlamasını sağlamıştır. Özellikle makine öğrenmesi algoritmaları, doğrusal olmayan ilişkileri öğrenebilme, yüksek boyutlu veri setlerini işleyebilme ve farklı değişkenler arasındaki karmaşık etkileşimleri modelleyebilme yetenekleri sayesinde hidrolojik tahmin çalışmalarında yaygın olarak kullanılmaya başlanmıştır. Günümüzde Yapay Sinir Ağları (Artificial Neural Networks, ANN), Destek Vektör Regresyonu (Support Vector Regression, SVR), Karar Ağaçları (Decision Tree), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Uzun Kısa Süreli Bellek Ağları (Long Short-Term Memory, LSTM), Geçitli Tekrarlayan Birimler (Gated Recurrent Unit, GRU) ve Transformer tabanlı modeller rezervuar su seviyesi tahmini başta olmak üzere birçok hidrolojik modelleme çalışmasında başarıyla uygulanmaktadır (Azad vd., 2025; Wee vd., 2021). Ouma vd. (2022), Random Forest algoritmasının iklim değişkenleri ile rezervuar su seviyeleri arasındaki karmaşık ilişkileri yüksek doğrulukla modelleyebildiğini göstermiştir. Benzer şekilde Rehamnia ve Mahdavi-Meymand (2025), Random Forest ve derin öğrenme modellerinin farklı barajlarda istikrarlı ve yüksek tahmin başarısı sergilediğini bildirmiştir.

Makine öğrenmesi yöntemlerinin hidrolojik modelleme çalışmalarındaki başarısının temel nedenlerinden biri, değişkenler arasındaki karmaşık ilişkileri önceden herhangi bir matematiksel varsayım gerektirmeksizin öğrenebilmesidir. Özellikle Random Forest ve XGBoost gibi topluluk öğrenmesi algoritmaları, farklı karar ağaçlarının

birlikte çalışmasına dayanan yapıları sayesinde yüksek doğrulukta tahminler üretebilmekte ve aşırı uyum problemini önemli ölçüde azaltabilmektedir (Breiman, 2001; Chen ve Guestrin, 2016). Liang vd. (2018), LSTM modelinin su seviyesi tahmininde geleneksel yöntemlerden daha başarılı olduğunu göstermiştir. Benzer şekilde Özdoğan-Sarıkoç vd. (2023), rezervuar hacmi tahmininde LSTM modelinin klasik makine öğrenmesi algoritmalarından daha yüksek doğruluk sağladığını belirtmiştir. Yi ve Yi (2024) ise rezervuar verilerine dayalı taşkın tahmininde GRU modelinin derin öğrenme yöntemleri arasında en başarılı performansı gösterdiğini rapor etmiştir. Bu gelişmeler, makine öğrenmesi yöntemlerini günümüzde su kaynakları yönetimi ve rezervuar işletmeciliği alanında en hızlı gelişen araştırma konularından biri hâline getirmiştir.

Makine öğrenmesi modellerinin hidrolojik tahmin çalışmalarında yaygın biçimde kullanılmasına rağmen, bu modellerin önemli bir bölümü "kara kutu (black-box)" modeller olarak değerlendirilmektedir. Özellikle topluluk öğrenmesi ve derin öğrenme algoritmaları yüksek tahmin doğruluğu sağlayabilmelerine karşın, modelin hangi değişkenlerden nasıl etkilendiğinin ve belirli bir tahmin sonucuna hangi mekanizmalarla ulaştığının açıklanması çoğu zaman güç olmaktadır. Wee vd. (2021), rezervuar modelleme çalışmalarında son yıllarda tahmin doğruluğunun önemli ölçüde arttığını ancak geliştirilen modellerin yorumlanabilirliğinin hâlen önemli bir araştırma konusu olduğunu belirtmektedir. Bu durum, özellikle su kaynakları yönetimi gibi karar verme süreçlerinin yüksek güvenilirlik gerektirdiği alanlarda model sonuçlarının yorumlanabilirliğini sınırlandırabilmektedir. Son yıllarda yalnızca yüksek doğruluk sağlayan modeller geliştirilmesi yeterli görülmemekte; aynı zamanda bu modellerin karar mekanizmalarının anlaşılabilir ve yorumlanabilir olması da beklenmektedir (Azad vd., 2025).

Bu gereksinim doğrultusunda geliştirilen Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence, XAI) yaklaşımları, makine öğrenmesi modellerinin karar verme süreçlerinin daha şeffaf biçimde yorumlanmasına olanak sağlamaktadır. Özellikle SHAP (SHapley Additive exPlanations), Feature Importance, Permutation Importance ve Partial Dependence Plot (PDP) analizleri, model tahminlerinde hangi değişkenlerin ne düzeyde etkili olduğunu ve bu etkinin yönünü ortaya koyabilmektedir. Böylece

yalnızca en başarılı modelin belirlenmesi değil, aynı zamanda modelin neden bu sonucu ürettiğinin de bilimsel olarak açıklanması mümkün olmaktadır. Lundberg ve Lee (2017) tarafından geliştirilen SHAP yöntemi, her bir değişkenin tahmin sonucuna yaptığı katkıyı ayrı ayrı hesaplayarak karmaşık makine öğrenmesi modellerinin yorumlanabilirliğini önemli ölçüde artırmıştır. Molnar (2022) ise açıklanabilir makine öğrenmesi yaklaşımlarının yalnızca model performansının değerlendirilmesi açısından değil, geliştirilen modellerin güvenilirliğinin artırılması ve karar destek sistemlerinde kullanılabilirliğinin sağlanması açısından da büyük önem taşıdığını vurgulamaktadır.

Literatürde rezervuar su seviyesi ve doluluk oranının tahminine yönelik çok sayıda araştırma bulunmasına rağmen, çalışmaların büyük bölümünde temel amaç farklı makine öğrenmesi veya derin öğrenme algoritmalarının tahmin performanslarının karşılaştırılması olmuştur. Wee vd. (2021) tarafından gerçekleştirilen kapsamlı derleme çalışması, son yirmi yılda rezervuar modellemelerinde ANN, ANFIS, SVR ve çeşitli hibrit yaklaşımların yoğun biçimde kullanıldığını göstermektedir. Azad vd. (2025) ise son yıllarda özellikle Random Forest, XGBoost, LSTM ve diğer derin öğrenme modellerinin rezervuar modelleme çalışmalarında yaygın olarak kullanılmaya başlandığını ortaya koymuştur. Bununla birlikte geliştirilen modellerin karar mekanizmalarının açıklanmasına yönelik çalışmaların oldukça sınırlı olduğu görülmektedir. Özellikle SHAP, Feature Importance, Permutation Importance ve Partial Dependence Plot analizlerinin birlikte kullanıldığı rezervuar doluluk oranı tahmin çalışmalarına literatürde oldukça az rastlanmaktadır. Ayrıca Türkiye'de gerçekleştirilen araştırmaların büyük bölümünün belirli algoritmaların performanslarının değerlendirilmesine odaklandığı, açıklanabilir yapay zekâ yaklaşımlarının hidrolojik modelleme çalışmalarına yeterince entegre edilmediği dikkati çekmektedir.

Bu tez çalışması, söz konusu literatür boşluğunun giderilmesine katkı sağlamayı amaçlamaktadır. Araştırma, Kastamonu ilinde bulunan Germeçtepe Barajı üzerinde gerçekleştirilmiş olup, 01 Ocak 2008–31 Aralık 2025 tarihleri arasındaki günlük hidrolojik ve meteorolojik gözlemler kullanılarak aktif doluluk oranının tahmini hedeflenmiştir. Modelleme sürecinde bağımlı değişken olarak günlük aktif doluluk oranı kullanılırken; bağımsız değişkenler olarak bir önceki güne ait aktif doluluk oranı

(Lag-1), günlük ortalama sıcaklık, günlük toplam yağış, ortalama nispi nem ve ortalama aktüel basınç verileri modele dahil edilmiştir. Aynı veri seti üzerinde Doğrusal Regresyon (Linear Regression), Karar Ağacı (Decision Tree), Random Forest ve Extreme Gradient Boosting (XGBoost) algoritmaları geliştirilmiş; ağaç tabanlı modellerde hiperparametre optimizasyonu uygulanarak model performansları karşılaştırılmıştır.

Bu çalışmanın temel amacı, Germeçtepe Barajı'nın günlük aktif doluluk oranını farklı makine öğrenmesi algoritmaları kullanarak yüksek doğrulukla tahmin etmek ve geliştirilen modellerin performanslarını karşılaştırmaktır. Bunun yanında çalışma yalnızca en başarılı tahmin modelinin belirlenmesini değil, aynı zamanda açıklanabilir yapay zekâ yöntemlerinden Feature Importance, Permutation Importance, SHAP ve Partial Dependence Plot analizleri kullanılarak meteorolojik değişkenlerin aktif doluluk oranı üzerindeki etkilerinin ayrıntılı biçimde ortaya konulmasını amaçlamaktadır. Böylece geliştirilen modeller yalnızca tahmin doğruluğu bakımından değil, aynı zamanda yorumlanabilirlik açısından da değerlendirilmiştir.

Bu araştırmanın literatüre sağladığı temel katkılar şu şekilde özetlenebilir:

- Aynı veri seti üzerinde doğrusal istatistiksel yöntem ile farklı makine öğrenmesi algoritmalarının sistematik olarak karşılaştırılması,

- Random Forest ve XGBoost modellerinde hiperparametre optimizasyonunun etkisinin değerlendirilmesi,

- Feature Importance, Permutation Importance, SHAP ve Partial Dependence Plot analizlerinin birlikte kullanılarak model sonuçlarının açıklanabilir yapay zekâ yaklaşımı kapsamında yorumlanması ve

- Germeçtepe Barajı için uzun dönemli günlük gözlemler kullanılarak geliştirilen modelleme yaklaşımının su kaynakları yönetimine uygulanabilirliğinin ortaya konulmasıdır.

Bu yönleriyle çalışma, yalnızca yüksek doğrulukta tahmin modeli geliştirmeyi değil, aynı zamanda geliştirilen modelin karar mekanizmasını bilimsel olarak açıklamayı hedeflemesi bakımından da mevcut literatürdeki benzer çalışmalardan ayrılmaktadır.

Bu araştırma kapsamında aşağıdaki hipotezler test edilmiştir:

H1. Makine öğrenmesi algoritmaları, Germeçtepe Barajı'nın günlük aktif doluluk oranını yüksek doğrulukla tahmin edebilmektedir.

H2. Topluluk öğrenmesi algoritmaları (Random Forest ve XGBoost), doğrusal regresyon ve tek karar ağacı modellerine göre daha yüksek tahmin performansı göstermektedir.

H3. Açıklanabilir yapay zekâ yöntemleri (Feature Importance, Permutation Importance, SHAP ve Partial Dependence Plot), meteorolojik değişkenlerin aktif doluluk oranı üzerindeki etkilerinin yorumlanmasına önemli katkılar sağlamaktadır.

H4. Günlük ortalama sıcaklık, meteorolojik değişkenler içerisinde aktif doluluk oranını etkileyen en önemli meteorolojik değişkendir.

Bu hipotezler doğrultusunda elde edilen bulguların, makine öğrenmesi ve açıklanabilir yapay zekâ yaklaşımlarının rezervuar doluluk oranı tahminindeki etkinliğini ortaya koyması ve su kaynakları yönetimine yönelik karar destek süreçlerine bilimsel katkı sağlaması beklenmektedir.

2. LİTERATÜR ÖZETİ

2.1 Rezervuar Su Seviyesi Tahmininde Geleneksel Makine Öğrenmesi Yaklaşımları

Baraj ve rezervuarlarda su seviyesi tahmini üzerine gerçekleştirilen ilk çalışmaların büyük bölümü doğrusal regresyon, zaman serisi modelleri, yapay sinir ağları (Artificial Neural Networks, ANN), destek vektör regresyonu (Support Vector Regression, SVR), adaptif bulanık çıkarım sistemleri (ANFIS) ve genetik programlama (Genetic Programming, GP) gibi geleneksel makine öğrenmesi ve istatistiksel yöntemlere odaklanmıştır. Bu çalışmaların temel amacı, geçmiş hidrolojik ve meteorolojik verilerden yararlanarak rezervuar su seviyesindeki değişimleri doğru biçimde tahmin etmek ve su kaynakları yönetimine karar desteği sağlamaktır. Özellikle doğrusal olmayan hidrolojik süreçlerin modellenmesinde klasik istatistiksel yöntemlerin yetersiz kalması, araştırmacıları veri odaklı makine öğrenmesi algoritmalarına yöneltmiştir.

Egawa vd. (2011), Japonya'daki Tadami–Agano Havzası'nda bulunan on bir farklı baraj için su akışının tahmini amacıyla yapay sinir ağları ve regresyon modellerinin performanslarını karşılaştırmışlardır. Çalışmada normal ve taşkın dönemlerine ait hidrolojik veriler kullanılmış, veri miktarının yeterli olduğu barajlarda yapay sinir ağları, veri miktarının sınırlı olduğu durumlarda ise regresyon modelleri uygulanmıştır. Sonuçlar, yapay sinir ağlarının doğrusal olmayan hidrolojik süreçleri başarılı şekilde modellediğini ve özellikle geniş veri setlerinde yüksek tahmin doğruluğu sağladığını göstermiştir. Araştırmacılar geliştirilen sistemin gerçek zamanlı tahminlerde de kullanılabildiğini belirtmişlerdir.

Kisi vd. (2012), günlük göl su seviyesi tahmini amacıyla Genetik Programlama (GP), Yapay Sinir Ağları (ANN), ANFIS ve ARMA modellerini karşılaştırmışlardır. Çalışmada geçmiş su seviyesi kayıtlarından yararlanılarak farklı tahmin senaryoları oluşturulmuş ve yapay zekâ tabanlı yöntemlerin geleneksel zaman serisi modellerine göre daha başarılı sonuçlar verdiği belirlenmiştir. Ayrıca çoklu giriş değişkenlerinin kullanıldığı modellerin tahmin doğruluğunu önemli ölçüde artırdığı ifade edilmiştir.

Stojanovic vd. (2013), Bosna-Hersek'teki Bocac Barajı'nın davranışlarının modellenmesi amacıyla çoklu doğrusal regresyon ve genetik algoritmaları birleştiren adaptif bir model geliştirmişlerdir. Sensör verileri kullanılarak oluşturulan modelde genetik algoritmalar yardımıyla model parametreleri optimize edilmiş ve geliştirilen yaklaşımın geleneksel regresyon yöntemlerine göre daha başarılı sonuçlar verdiği belirlenmiştir. Araştırmacılar özellikle veri eksikliklerinin bulunduğu durumlarda hibrit modellerin önemli avantajlar sağlayabileceğini vurgulamışlardır.

Büyükyıldız vd. (2014), Beyşehir Gölü'nde aylık su seviyesi değişimlerinin tahmin edilmesi amacıyla ANN, PSO-ANN, SVR, RBNN ve ANFIS modellerini karşılaştırmışlardır. Araştırmada aylık giriş-çıkış akımları, yağış ve buharlaşma verileri modele giriş değişkeni olarak kullanılmış; model performansları RMSE, MSE, MAE ve R^2 ölçütleri ile değerlendirilmiştir. Sonuçlar, ε-SVR modelinin diğer yöntemlere göre daha başarılı performans gösterdiğini ve yapay zekâ yöntemlerinin göl su seviyesi tahmininde yüksek doğruluk sağlayabildiğini ortaya koymuştur.

Nwobi-Okoye ve Igboanugo (2013), günlük su seviyesi tahmini için geliştirdikleri yapay sinir ağı modellerinde farklı giriş değişkeni kombinasyonlarını değerlendirmişlerdir. Araştırma sonucunda giriş değişkeni sayısının belirli bir seviyeye kadar model performansını artırdığı, ancak gereğinden fazla değişken kullanımının tahmin doğruluğunu olumsuz etkileyebileceği belirlenmiştir. Bu bulgu, model başarısında yalnızca algoritmanın değil, uygun değişken seçiminin de kritik öneme sahip olduğunu göstermektedir.

Damla vd. (2020), Yalova Gökçe Barajı'nda su seviyesi tahmini amacıyla Yapay Sinir Ağları yöntemini kullanmışlardır. Çalışmada akım, yağış, buharlaşma, sızıntı miktarı, su tahliyeleri ve rezervuar su seviyesi gibi çok sayıda hidrolojik değişken modele dahil edilmiş ve geliştirilen model için %94,14 determinasyon katsayısı elde edilmiştir. Araştırmacılar, yapay sinir ağlarının rezervuar su seviyesinin tahmininde güvenilir sonuçlar üretebildiğini belirtmişlerdir.

Wong vd. (2019), Malezya'daki Klang Gates Barajı'nda zaman serisi regresyonu ile Destek Vektör Makineleri yöntemlerini karşılaştırmışlardır. Sonuçlar, bazı

senaryolarda zaman serisi regresyon modelinin daha başarılı performans gösterdiğini ortaya koymuş, ayrıca veri ön işleme ve uygun değişken seçiminin model doğruluğu üzerinde önemli etkileri olduğu vurgulanmıştır.

Anwar (2023), Doğancı Barajı'nda meteorolojik değişkenler ile rezervuar su seviyesi arasındaki ilişkileri Yapay Sinir Ağları kullanarak modellemiştir. Hava sıcaklığı, güneş radyasyonu, buharlaşma ve evapotranspirasyon gibi değişkenlerin modele dahil edildiği çalışmada korelasyon katsayısı 0,99 düzeyine ulaşılmış ve meteorolojik değişkenlerin rezervuar davranışının tahmininde önemli rol oynadığı belirlenmiştir.

Arslan vd. (2020), Keban Baraj Gölü'nde ANFIS ve Destek Vektör Makineleri yöntemlerini Çoklu Doğrusal Regresyon modeli ile karşılaştırmışlardır. Araştırmada enerji üretimi amacıyla gerçekleştirilen su salımları, günlük su tüketimi ve buharlaşma verileri kullanılmış; yapay zekâ tabanlı yöntemlerin klasik regresyon modellerine göre daha başarılı sonuçlar verdiği bildirilmiştir.

Bu çalışmalar genel olarak değerlendirildiğinde, yapay sinir ağları, destek vektör regresyonu, ANFIS ve benzeri geleneksel makine öğrenmesi yöntemlerinin rezervuar ve göl su seviyesi tahmininde başarılı sonuçlar verdiği görülmektedir. Bununla birlikte, bu modellerin hiperparametre seçimine duyarlı olmaları, büyük veri setlerinde hesaplama maliyetlerinin artabilmesi ve bazı durumlarda genelleme performanslarının sınırlı kalabilmesi nedeniyle son yıllarda araştırmaların odağı Random Forest ve XGBoost gibi topluluk öğrenmesi algoritmalarına yönelmiştir.

2.2 Topluluk Öğrenmesi ve Derin Öğrenme Tabanlı Yaklaşımlar

Son yıllarda rezervuar ve baraj su seviyesi tahminlerinde geleneksel makine öğrenmesi yöntemlerinin yanında topluluk öğrenmesi ve derin öğrenme tabanlı yaklaşımlar daha yaygın biçimde kullanılmaya başlanmıştır. Topluluk öğrenmesi algoritmaları, birden fazla modelin tahminlerini birleştirerek tekil modellere göre daha kararlı ve yüksek doğruluklu sonuçlar üretebilmektedir. Özellikle Random Forest, XGBoost, Bagged Trees ve Boosted Decision Tree gibi yöntemler; doğrusal olmayan ilişkilerin modellenmesi, değişkenler arası etkileşimlerin öğrenilmesi ve yüksek genelleme kapasitesi nedeniyle hidrolojik tahmin çalışmalarında ön plana çıkmaktadır. Derin

öğrenme tabanlı LSTM, GRU, NARX ve benzeri modeller ise zaman serilerindeki uzun dönemli bağımlılıkları öğrenebilme özellikleri nedeniyle rezervuar su seviyesi ve giriş-çıkış akımı tahminlerinde önemli avantajlar sağlamaktadır.

Yang vd. (2017), Amerika Birleşik Devletleri ve Çin'de bulunan iki farklı rezervuarda aylık giriş akımlarının tahmin edilmesi amacıyla Random Forest, Yapay Sinir Ağları ve Destek Vektör Regresyonu modellerini karşılaştırmışlardır. Çalışmada geçmiş hidrolojik verilerle birlikte 17 farklı iklim göstergesi modele dahil edilmiştir. Sonuçlar, tüm modellerin rezervuar giriş akımlarını başarılı şekilde tahmin edebildiğini, ancak Random Forest algoritmasının diğer yöntemlere göre daha yüksek performans gösterdiğini ortaya koymuştur.

Sapitang vd. (2020), Malezya'daki Kenyir Rezervuarı'nda su seviyesi tahmini amacıyla Boosted Decision Tree Regression, Decision Forest Regression, Bayesian Linear Regression ve Neural Network Regression modellerini karşılaştırmışlardır. Araştırmada yağış, rezervuar su seviyesi ve rezervuardan bırakılan su miktarı değişkenleri kullanılmıştır. Sonuçlar, farklı tahmin senaryolarında Bayesian Linear Regression ve Boosted Decision Tree Regression modellerinin yüksek doğruluk sağladığını göstermiştir.

Ouma vd. (2022), Botsvana'da bulunan Gaborone ve Bokaa barajlarında su seviyesi değişimlerinin tahmin edilmesi amacıyla Çoklu Doğrusal Regresyon, Vektör Otoregresyon, Random Forest Regresyonu ve Çok Katmanlı Algılayıcı Yapay Sinir Ağı modellerini karşılaştırmışlardır. Çalışmada yağış, sıcaklık, iklim indisleri ve arazi kullanım/arazi örtüsü verileri giriş değişkenleri olarak kullanılmıştır. Sonuçlar, doğrusal regresyon modelinin karmaşık doğrusal olmayan ilişkileri açıklamada sınırlı kaldığını, Random Forest ve MLP-ANN modellerinin ise daha yüksek tahmin doğruluğu sağladığını göstermiştir.

Almubaidin vd. (2023), günlük baraj su seviyesi tahmini amacıyla yedi farklı makine öğrenmesi algoritmasını karşılaştırmışlardır. Çalışmada yalnızca geçmiş su seviyesi verileri kullanılarak farklı gecikmeli giriş kombinasyonları test edilmiştir. Sonuçlar, yedi günlük gecikmeli giriş verileriyle eğitilen Bagged Trees modelinin en yüksek

doğruluğu sağladığını göstermiştir. Bu çalışma, gecikmeli değişkenlerin günlük su seviyesi tahmininde model başarısını önemli ölçüde artırabileceğini ortaya koymaktadır.

Pathan vd. (2024), Malezya'daki Batu Barajı'nda yalnızca yağış verilerini kullanarak su seviyesi tahmini amacıyla MLP, XGBoost, SVR ve Random Forest modellerini karşılaştırmışlardır. Çalışmada farklı gecikmeli giriş senaryoları denenmiş ve modeller MAE, RMSE ve R^2 ölçütleriyle değerlendirilmiştir. Sonuçlar, üç günlük gecikmeli giriş senaryosunda MLP modelinin en başarılı performansı gösterdiğini ortaya koymuştur.

Soria-Lopez vd. (2023), İspanya'nın Miño-Sil Havzası'nda bulunan sekiz farklı rezervuarda bir gün sonraki çıkış debisinin tahmin edilmesi amacıyla Yapay Sinir Ağları (ANN), Destek Vektör Makineleri (SVM) ve Random Forest (RF) algoritmalarını karşılaştırmışlardır. Çalışmada rezervuar hacmi, giriş debisi ve çıkış debisi modele giriş değişkenleri olarak kullanılmıştır. Sonuçlar, ANN modelinin sekiz rezervuarın beşinde en başarılı performansı gösterdiğini ortaya koymuştur. Bununla birlikte Random Forest modelinin de yüksek tahmin doğruluğu sergilediği ve makine öğrenmesi yöntemlerinin rezervuar işletmesi ile karar destek sistemlerinde etkin biçimde kullanılabileceği vurgulanmıştır.

Güneş Şen (2025), Karaçomak Barajı'nda günlük su seviyesi tahmini amacıyla doğrusal regresyon, karar ağacı, Random Forest ve XGBoost algoritmalarını karşılaştırmıştır. Çalışmada 2008–2024 yılları arasındaki günlük hidro-meteorolojik veriler kullanılmıştır. Sonuçlar, Random Forest ve XGBoost algoritmalarının doğrusal regresyon ve karar ağacı modellerine göre daha yüksek tahmin doğruluğu sağladığını göstermiştir.

Rehamnia ve Mahdavi-Meymand (2025), Cezayir'de bulunan üç farklı barajda günlük su seviyesi dalgalanmalarının tahmini amacıyla çoklu doğrusal regresyon, sığ sinir ağı, derin sinir ağı, destek vektör regresyonu, sınıflandırma ve regresyon ağacı ile Random Forest modellerini karşılaştırmışlardır. Sonuçlar, özellikle Random Forest, derin sinir ağı ve hibrit destek vektör regresyonu modellerinin yüksek doğruluk sağladığını

göstermiştir. Araştırmacılar Random Forest algoritmasının farklı barajlarda istikrarlı sonuçlar üretebildiğini belirtmişlerdir.

Liang vd. (2018), Çin'deki Dongting Gölü su seviyelerinin tahmin edilmesi ve Üç Boğaz Barajı'nın göl üzerindeki etkilerinin belirlenmesi amacıyla LSTM ağlarını kullanmışlardır. Çalışmada yedi farklı hidrolojik ve meteorolojik değişken giriş verisi olarak kullanılmıştır. Sonuçlar, LSTM modelinin SVM modelinden daha başarılı tahminler ürettiğini ve uzun dönemli zaman bağımlılıklarının modellenmesinde güçlü bir yaklaşım sunduğunu göstermiştir.

Özdoğan-Sarıkoç vd. (2023), Türkiye'deki Ladik ve Yedikır rezervuarlarında rezervuar hacimlerinin tahmin edilmesi amacıyla ANN, SVR ve LSTM modellerini karşılaştırmışlardır. Çalışmada LSTM modelinin rezervuar hacmindeki zamansal değişimleri öğrenmede daha başarılı olduğu ve diğer modellere göre daha yüksek tahmin doğruluğu sağladığı belirlenmiştir.

Yi ve Yi (2024), Güney Kore Han Nehri Havzası'nda rezervuar verilerine dayalı taşkın tahmini amacıyla Random Forest, SVR, LSTM ve GRU modellerini karşılaştırmışlardır. Çalışmada üç rezervuarın birlikte değerlendirildiği senaryoda GRU modelinin en başarılı derin öğrenme yöntemi olduğu, Random Forest algoritmasının ise geleneksel makine öğrenmesi yöntemleri arasında en yüksek performansı gösterdiği belirlenmiştir.

Zhou ve Li (2021), rezervuarlara gelen akım miktarının tahmini amacıyla Recurrent Neural Ordinary Differential Equations tabanlı FlowODE modelini geliştirmişlerdir. İki büyük hidroelektrik barajından elde edilen verilerle uygulanan modelin farklı tahmin ufuklarında yüksek doğruluk sağladığı ve rezervuar işletmesi, taşkın kontrolü ve enerji üretimi planlamasında kullanılabileceği belirtilmiştir.

García-Feal vd. (2022), İspanya'nın Miño Nehri üzerindeki sekiz farklı rezervuarda çıkış debisinin tahmini amacıyla MLP, NARX, LSTM ve çoklu doğrusal regresyon modellerini karşılaştırmışlardır. Sonuçlar, NARX modelinin genel olarak en başarılı tahmin performansını gösterdiğini ve veri odaklı modellerin rezervuar işletmeciliğinde karar destek sistemleri açısından önemli katkılar sağlayabileceğini ortaya koymuştur.

Genel olarak değerlendirildiğinde, topluluk öğrenmesi ve derin öğrenme tabanlı yaklaşımların rezervuar su seviyesi, hacim ve akım tahminlerinde yüksek doğruluk sağladığı görülmektedir. Random Forest, XGBoost, Bagged Trees ve Boosted Decision Tree gibi topluluk öğrenmesi algoritmaları, doğrusal olmayan ilişkileri modelleyebilme ve yüksek genelleme kapasitesi nedeniyle birçok çalışmada başarılı sonuçlar üretmiştir. Derin öğrenme tabanlı LSTM, GRU, NARX ve FlowODE gibi modeller ise özellikle uzun dönemli zaman bağımlılıklarının bulunduğu veri setlerinde ön plana çıkmaktadır. Bununla birlikte bu modellerin performansı kullanılan veri setinin büyüklüğüne, giriş değişkenlerinin niteliğine, gecikmeli değişkenlerin seçimine ve model optimizasyonuna bağlı olarak değişmektedir. Bu nedenle rezervuar tahmin çalışmalarında yalnızca model performansının değil, modelin hangi değişkenlere dayanarak tahmin ürettiğinin de değerlendirilmesi önemli hale gelmiştir.

2.3 Rezervuar Modellemesinde Güncel Eğilimler ve Araştırma Boşluğu

Rezervuar su seviyesi ve aktif doluluk oranının tahmini üzerine gerçekleştirilen çalışmalar incelendiğinde, son yıllarda kullanılan modelleme yaklaşımlarında önemli değişimler yaşandığı görülmektedir. İlk dönem çalışmalarda çoklu doğrusal regresyon, ARMA ve benzeri geleneksel istatistiksel yöntemler yaygın olarak kullanılırken, hidrolojik süreçlerin doğrusal olmayan yapısının daha iyi temsil edilebilmesi amacıyla zaman içerisinde makine öğrenmesi ve derin öğrenme tabanlı yöntemlere yönelim artmıştır. Özellikle son yıllarda yayımlanan derleme çalışmaları, rezervuar modelleme alanında veri odaklı yaklaşımların hızla yaygınlaştığını ve klasik istatistiksel yöntemlerin yerini giderek daha gelişmiş yapay zekâ tabanlı modellere bıraktığını göstermektedir (Azad vd., 2025; Wee vd., 2021).

Makine öğrenmesi alanındaki bu gelişime paralel olarak, rezervuar modelleme çalışmalarında kullanılan algoritmaların çeşitliliği de önemli ölçüde artmıştır. Geleneksel yapay sinir ağları ve destek vektör regresyonu gibi yöntemlerin yanı sıra Random Forest, XGBoost, LightGBM ve CatBoost gibi topluluk öğrenmesi algoritmaları ile LSTM, GRU ve NARX gibi derin öğrenme modelleri yaygın biçimde kullanılmaya başlanmıştır. Özellikle zaman serisi verilerinin modellenmesinde LSTM ve GRU gibi tekrarlayan sinir ağı tabanlı yöntemlerin başarılı sonuçlar verdiği

bildirilmektedir (Liang vd., 2018; Özdoğan-Sarıkoç vd., 2023; Yi ve Yi, 2024). Bununla birlikte birçok araştırmada Random Forest ve XGBoost algoritmalarının da yüksek tahmin doğruluğu ve güçlü genelleme yetenekleri sayesinde rezervuar modellemesinde başarılı sonuçlar ürettiği belirtilmektedir (Ouma vd., 2022; Rehamnia ve Mahdavi-Meymand, 2025; Yang vd., 2017).

Bununla birlikte son yıllarda makine öğrenmesi çalışmalarında yalnızca yüksek tahmin doğruluğu elde etmek yeterli görülmemektedir. Geliştirilen modellerin karar mekanizmalarının anlaşılabilmesi ve model çıktılarının yorumlanabilmesi amacıyla Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence, XAI) yaklaşımları giderek daha fazla önem kazanmaktadır. Özellikle SHAP (SHapley Additive exPlanations) (Lundberg ve Lee, 2017), Feature Importance (Breiman, 2001), Permutation Importance (Fisher vd., 2019) ve Partial Dependence Plot (PDP) analizleri (Friedman, 2001), makine öğrenmesi modellerinin hangi değişkenlerden nasıl etkilendiğinin ortaya konulmasına olanak sağlamaktadır. Lundberg ve Lee (2017) tarafından geliştirilen SHAP yöntemi, karmaşık makine öğrenmesi modellerinin yorumlanabilirliğini artıran en önemli yöntemlerden biri olarak kabul edilmektedir. Benzer şekilde Molnar (2022), açıklanabilir makine öğrenmesi yaklaşımlarının yalnızca model performansının değerlendirilmesine değil, aynı zamanda model güvenilirliğinin artırılması, karar süreçlerinin şeffaflaştırılmasına ve model çıktılarının yorumlanabilirliğinin geliştirilmesine önemli katkılar sağladığını vurgulamaktadır.

Literatür genel olarak değerlendirildiğinde, rezervuar su seviyesi ve aktif doluluk oranının tahminine yönelik çok sayıda çalışma bulunmasına rağmen, bu araştırmaların büyük bölümünde modellerin yalnızca tahmin performanslarının karşılaştırıldığı görülmektedir. Buna karşın geliştirilen modellerin karar mekanizmalarının açıklanmasına yönelik analizler oldukça sınırlı sayıdadır. Özellikle Feature Importance, Permutation Importance, SHAP ve Partial Dependence Plot (PDP) analizlerinin birlikte kullanıldığı çalışmaların sayısı oldukça azdır. Türkiye'de gerçekleştirilen araştırmaların önemli bir kısmı ise yapay sinir ağları veya klasik makine öğrenmesi algoritmalarının performans karşılaştırmalarına odaklanmış olup,

açıklanabilir yapay zekâ yaklaşımlarını bütüncül biçimde kullanan çalışmalar sınırlı düzeydedir.

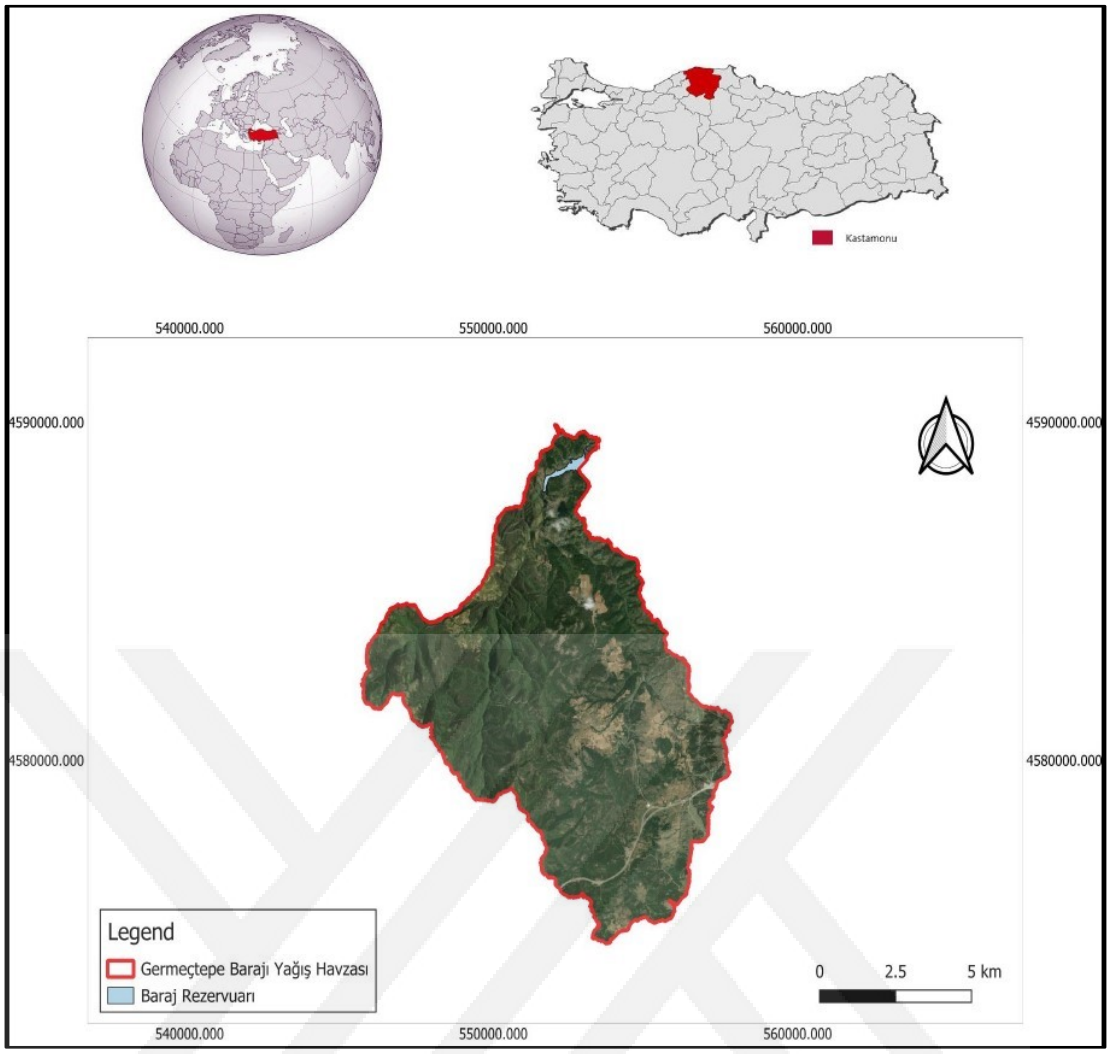
Bu tez çalışması, literatürdeki söz konusu boşluğu doldurmayı amaçlamaktadır. Çalışmada doğrusal regresyon, Decision Tree, Random Forest ve XGBoost algoritmalarının tahmin performansları karşılaştırılmış; en başarılı model belirlendikten sonra Feature Importance, Permutation Importance, SHAP ve Partial Dependence Plot (PDP) analizleri kullanılarak meteorolojik değişkenlerin Germeçtepe Barajı'nın aktif doluluk oranı üzerindeki etkileri ayrıntılı biçimde değerlendirilmiştir. Böylece çalışma yalnızca yüksek doğrulukta bir tahmin modeli geliştirmeyi değil, aynı zamanda model sonuçlarını açıklanabilir yapay zekâ yöntemleriyle yorumlayarak rezervuar doluluk oranını etkileyen temel faktörleri ortaya koymayı hedeflemesi bakımından literatüre özgün katkı sunmaktadır.

3. MATERYAL VE METOT

3.1 Çalışma Alanı

Çalışma, Türkiye'nin Batı Karadeniz Bölgesi'nde yer alan Kastamonu ili sınırları içerisindeki Germeçtepe Barajı'nda yürütülmüştür. Germeçtepe Barajı, Kuzeybatı Anadolu'da Kastamonu ili sınırları içerisinde, Kızılırmak Havzası'nda, Daday Çayı'nın yan kollarından biri olan Şadibey Deresi üzerinde yer almaktadır. Baraj, Kastamonu il merkezinin yaklaşık 10 km kuzeybatısında bulunmaktadır.

Germeçtepe Barajı, 33°38′09″–33°46′58″ doğu boylamları ile 41°25′17″–41°26′47″ kuzey enlemleri arasında yer almaktadır (Şekil 3.1). Sulama amacıyla inşa edilen barajın yapımına 1977 yılında başlanmış, 1986 yılında ise işletmeye alınmıştır. Kil çekirdekli kaya dolgu tipinde inşa edilen barajın gövde hacmi 364.000 m^3, akarsu yatağından yüksekliği 41,70 m, normal su kotundaki göl hacmi 9,36 hm^3 ve göl alanı 0,601 km^2'dir. Germeçtepe Barajı yaklaşık 3.226 hektarlık tarım alanına sulama suyu sağlamaktadır (DSİ, 2025).



Şekil 3.1 Germeçtepe Barajının coğrafi konumu

3.2 Veri Seti ve Veri Kaynakları

Bu çalışmada Germeçtepe Barajı'nın günlük aktif doluluk oranının makine öğrenmesi yöntemleri kullanılarak tahmin edilmesi amacıyla hidrolojik ve meteorolojik verilerden oluşan bir veri seti oluşturulmuştur. Araştırmada kullanılan hidrolojik veriler Devlet Su İşleri Genel Müdürlüğü'nden (DSİ), meteorolojik veriler ise Meteoroloji Genel Müdürlüğü'nden (MGM) temin edilmiştir.

Çalışmada kullanılan veri seti 01 Ocak 2008–31 Aralık 2025 tarihleri arasındaki dönemi kapsamakta olup toplam 6575 günlük gözlemden oluşmaktadır. Tüm veriler günlük zaman adımında değerlendirilmiş ve tarih bilgileri esas alınarak tek bir veri tabanında birleştirilmiştir. Oluşturulan veri setinde eksik gözlem bulunmamaktadır.

Araştırmada bağımlı değişken olarak günlük aktif doluluk oranı (%) kullanılmıştır. Bağımsız değişkenler ise günlük ortalama sıcaklık (°C), günlük toplam yağış (mm), ortalama nispi nem (%) ve ortalama aktüel basınç (hPa) olarak belirlenmiştir. Meteorolojik veriler Kastamonu Meteoroloji İstasyonu'ndan elde edilmiş olup istasyon çalışma alanına yaklaşık 15 km uzaklıkta yer almaktadır. Aktif doluluk oranına ait günlük veriler ise DSİ 23. Bölge Müdürlüğü kayıtlarından temin edilmiştir.

Çalışmada kullanılan veri seti; bir bağımlı değişken (aktif doluluk oranı) ve dört meteorolojik bağımsız değişkenden (günlük ortalama sıcaklık, günlük toplam yağış, ortalama nispi nem ve ortalama aktüel basınç) oluşmaktadır. Modelleme sürecinde kullanılan değişkenler ve veri kaynakları Tablo 3.1'de verilmiştir.

Tablo 3.1 Çalışmada kullanılan değişkenler ve veri kaynakları

Değişken	**Birim**	**Veri Kaynağı**	**Modeldeki Rolü**
Aktif Doluluk	%	DSİ	Bağımlı değişken
Günlük Ortalama Sıcaklık	°C	MGM	Bağımsız değişken
Günlük Toplam Yağış	mm	MGM	Bağımsız değişken
Ortalama Nispi Nem	%	MGM	Bağımsız değişken
Ortalama Aktüel Basınç	hPa	MGM	Bağımsız değişken

3.3 Veri Ön İşleme Süreci

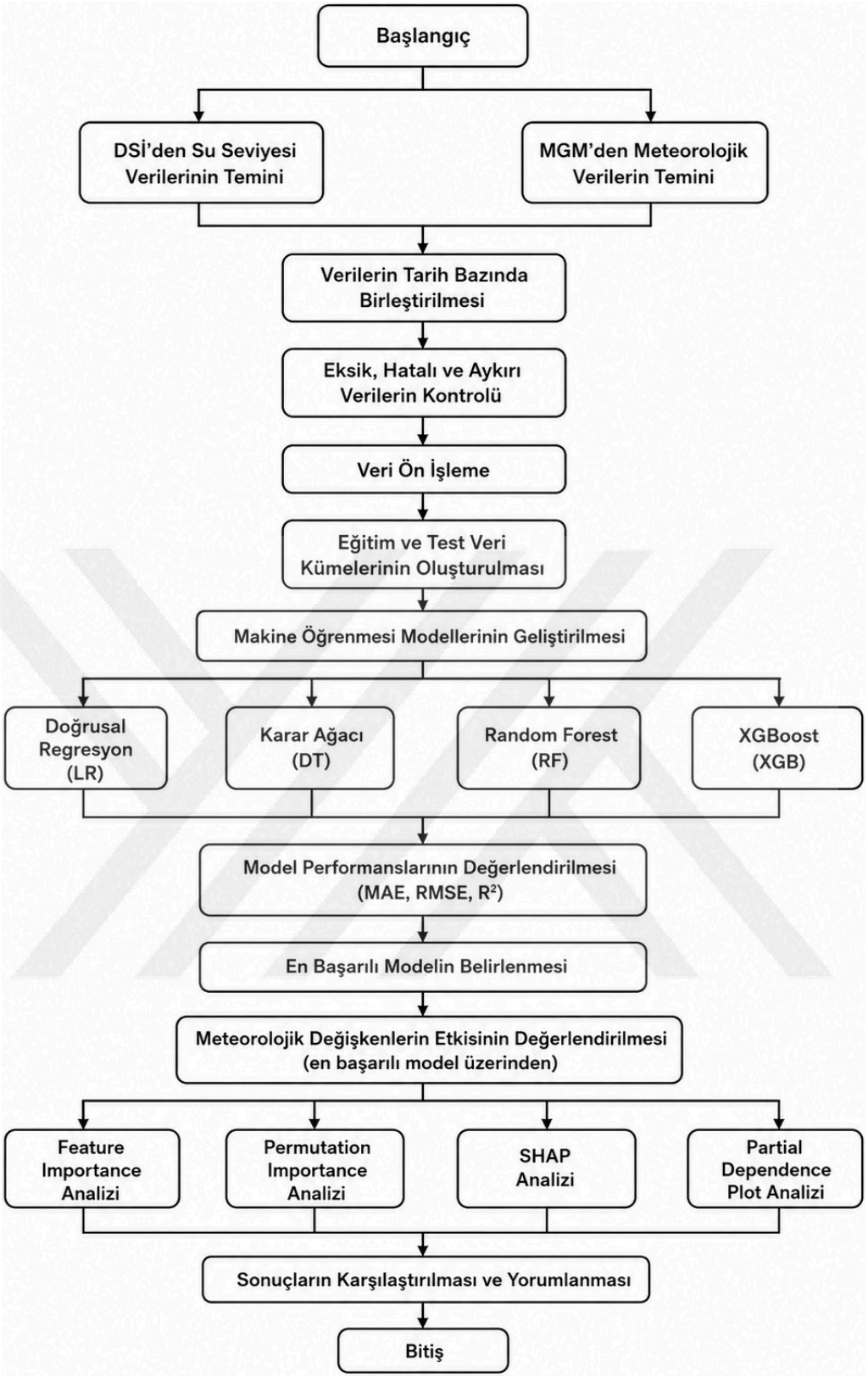
Makine öğrenmesi modellerinin güvenilir ve yüksek doğrulukta tahminler üretebilmesi için veri setinin modelleme sürecine uygun şekilde hazırlanması gerekmektedir. Bu nedenle model geliştirme aşamasına geçilmeden önce veri seti üzerinde çeşitli veri ön işleme işlemleri gerçekleştirilmiştir.

İlk olarak, zaman serilerinin sürekliliğini modele yansıtabilmek amacıyla bağımlı değişken olan aktif doluluk oranının bir önceki güne ait değerini temsil eden Lag-1 (aktif_doluluk_lag1) değişkeni oluşturulmuştur. Lag-1 değişkeninin oluşturulması nedeniyle veri setinin ilk gözleminde eksik değer oluşmuş, bu kayıt veri setinden çıkarılmış ve modelleme analizleri toplam 6574 günlük gözlem üzerinden gerçekleştirilmiştir.

Daha sonra bağımsız değişkenler arasındaki doğrusal ilişkiler Pearson korelasyon analizi ile incelenmiş, çoklu doğrusal bağlantının değerlendirilmesi amacıyla Varyans Artış Faktörü (Variance Inflation Factor, VIF) analizi uygulanmıştır. VIF analizi sonucunda günlük maksimum sıcaklık ve günlük minimum sıcaklık değişkenlerinin günlük ortalama sıcaklık değişkeni ile yüksek düzeyde çoklu doğrusal bağlantı gösterdiği belirlenmiştir. Bu nedenle modelleme sürecinde yalnızca günlük ortalama sıcaklık değişkeni kullanılmış, günlük maksimum ve minimum sıcaklık değişkenleri analizlerden çıkarılmıştır.

Model performansının tarafsız olarak değerlendirilebilmesi amacıyla veri seti kronolojik sıra korunarak eğitim ve test veri kümelerine ayrılmıştır. Veri setinin %80'i (5259 gözlem) model eğitimi, %20'si (1315 gözlem) ise model performansının değerlendirilmesi amacıyla test verisi olarak kullanılmıştır. Böylece modeller yalnızca geçmiş dönem verileri kullanılarak eğitilmiş ve sonraki döneme ait gözlemler üzerinde test edilmiştir.

Veri ön işleme aşamalarının tamamlanmasının ardından elde edilen nihai veri seti, makine öğrenmesi modellerinin geliştirilmesi, hiperparametre optimizasyonu ve performans değerlendirme analizlerinde kullanılmıştır. Araştırmada izlenen veri hazırlama ve modelleme süreci Şekil 3.2'de sunulmuştur.



Şekil 3.2 Makine öğrenmesi modelleme sürecine ait araştırma akış şeması

3.4 Kullanılan Makine Öğrenmesi Algoritmaları

Baraj aktif doluluk oranının tahmini, meteorolojik ve hidrolojik değişkenler arasındaki karmaşık ve çoğu zaman doğrusal olmayan ilişkilerin modellenmesini gerektirmektedir. Son yıllarda makine öğrenmesi algoritmaları, bu tür ilişkileri başarılı bir şekilde modelleyebilmeleri, büyük veri setlerini işleyebilmeleri ve yüksek tahmin doğruluğu sağlayabilmeleri nedeniyle hidroloji ve su kaynakları yönetimi alanında yaygın olarak kullanılmaktadır (Mosavi vd., 2018; Shen, 2018).

Makine öğrenmesi yöntemleri, önceden tanımlanmış matematiksel denklemlere ihtiyaç duymadan geçmiş gözlemlerden öğrenerek geleceğe yönelik tahminler gerçekleştiren veri odaklı modelleme yaklaşımlarıdır. Bu yöntemler, özellikle değişkenler arasındaki doğrusal olmayan ve karmaşık ilişkileri öğrenebilme yetenekleri sayesinde akarsu akımı, taşkın, kuraklık, yeraltı suyu seviyesi ve rezervuar doluluk oranı tahmini gibi birçok hidrolojik çalışmada başarılı bir şekilde uygulanmaktadır (Jordan ve Mitchell, 2015; Sit vd., 2020).

Bu çalışmada, Germeçtepe Barajı'nın günlük aktif doluluk oranının tahmini amacıyla bir doğrusal istatistiksel yöntem ve üç farklı makine öğrenmesi algoritması kullanılmıştır. Karşılaştırma modeli olarak Doğrusal Regresyon (Linear Regression, LR), ağaç tabanlı algoritmalar olarak ise Karar Ağacı (Decision Tree, DT), Random Forest (RF) ve Extreme Gradient Boosting (XGBoost) modelleri uygulanmıştır. Geliştirilen modellerin performansları karşılaştırılarak çalışma alanı için en uygun tahmin modelinin belirlenmesi amaçlanmıştır.

3.4.1 Doğrusal Regresyon (Linear Regression, LR)

Doğrusal Regresyon (Linear Regression, LR), bağımlı bir değişken ile bir veya daha fazla bağımsız değişken arasındaki doğrusal ilişkinin modellenmesinde kullanılan en temel istatistiksel yöntemlerden biridir. Modelin temel amacı, bağımlı değişkendeki değişimi bağımsız değişkenler yardımıyla açıklayan doğrusal ilişkiyi belirlemek ve bu ilişkiyi kullanarak tahminler üretmektir. Basit yapısı, düşük hesaplama maliyeti ve sonuçlarının kolay yorumlanabilmesi nedeniyle mühendislik, çevre bilimleri ve hidroloji alanlarında yaygın olarak kullanılmaktadır (Montgomery vd., 2021).

Doğrusal regresyon modelinde bağımlı değişken, bağımsız değişkenlerin doğrusal bir fonksiyonu olarak ifade edilmektedir. Çoklu doğrusal regresyon modelinin genel matematiksel gösterimi aşağıdaki gibidir:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon \quad (3.1)$$

Burada Y bağımlı değişkeni, β_0 sabit terimi, $\beta_1 \ldots . \beta_n$ regresyon katsayılarını, $X_1 \ldots .$ X_2 bağımsız değişkenleri ve ε ise hata terimini ifade etmektedir (Kutner vd., 2005).

Doğrusal regresyon modelinde regresyon katsayıları genellikle En Küçük Kareler Yöntemi (Ordinary Least Squares, OLS) kullanılarak tahmin edilmektedir. Bu yöntemde amaç, gözlenen değerler ile model tarafından tahmin edilen değerler arasındaki hata kareleri toplamını en aza indiren katsayıları belirlemektir (Hastie vd., 2009).

Doğrusal regresyon modelinin güvenilir sonuçlar verebilmesi için değişkenler arasında doğrusal ilişkinin bulunması, hata terimlerinin bağımsız olması, sabit varyans (homoskedastisite) varsayımının sağlanması ve bağımsız değişkenler arasında ciddi düzeyde çoklu doğrusal bağlantının bulunmaması beklenmektedir (Kutner vd., 2005). Özellikle çoklu doğrusal bağlantı, regresyon katsayılarının güvenilirliğini azaltabileceğinden modelleme öncesinde değerlendirilmesi gereken önemli bir varsayımdır.

Her ne kadar hidrolojik süreçler çoğu zaman doğrusal olmayan ilişkiler içerse de doğrusal regresyon modeli, basit yapısı ve yorumlanabilirliği nedeniyle makine öğrenmesi çalışmalarında sıklıkla referans model olarak kullanılmaktadır. Bu sayede daha karmaşık algoritmaların sağladığı performans artışı nesnel olarak değerlendirilebilmektedir (James vd., 2021).

Bu çalışmada Doğrusal Regresyon modeli, günlük aktif doluluk oranının tahmininde temel karşılaştırma modeli olarak kullanılmıştır. Modelin bağımsız değişkenlerini günlük ortalama sıcaklık, günlük toplam yağış, ortalama nispi nem, ortalama aktüel basınç ve bir önceki güne ait aktif doluluk oranını temsil eden Lag-1 değişkeni oluşturmuştur. Model, eğitim veri seti kullanılarak oluşturulmuş ve tahmin

performansı test veri seti üzerinde değerlendirilmiştir. Elde edilen sonuçlar, ağaç tabanlı makine öğrenmesi algoritmalarından elde edilen performans değerleri ile karşılaştırılarak değerlendirilmiştir.

3.4.2 Karar Ağacı (Decision Tree, DT)

Karar Ağacı (Decision Tree, DT), hem sınıflandırma hem de regresyon problemlerinde yaygın olarak kullanılan denetimli öğrenme algoritmalarından biridir. Algoritma, veri setini belirli karar kurallarına göre ardışık olarak alt gruplara ayırarak tahminler üretmektedir. Basit yapısı, kolay yorumlanabilir olması ve doğrusal olmayan ilişkileri modelleyebilmesi nedeniyle makine öğrenmesi uygulamalarında yaygın olarak kullanılmaktadır (Breiman vd., 1984; Quinlan, 1986).

Regresyon problemlerinde karar ağacı algoritmasının temel amacı, bağımlı değişkendeki varyansı en fazla azaltacak bölünme noktalarını belirleyerek veri setini giderek daha homojen alt gruplara ayırmaktır. Her düğümde farklı bölünme seçenekleri değerlendirilmekte ve hata varyansını en fazla azaltan değişken ile eşik değeri seçilmektedir. Bu işlem belirlenen durdurma kriterlerine ulaşılıncaya kadar yinelemeli olarak devam etmekte, son aşamada ise yaprak düğümlerde bulunan gözlemlerin ortalama değeri tahmin sonucu olarak kullanılmaktadır (Hastie vd., 2009; Loh, 2011).

Karar ağacı algoritmalarının en önemli avantajlarından biri, değişkenler arasındaki doğrusal olmayan ilişkileri herhangi bir matematiksel fonksiyon tanımlamaya ihtiyaç duymadan öğrenebilmesidir. Ayrıca algoritma, değişkenler arasındaki etkileşimleri otomatik olarak belirleyebilmekte ve sonuçlarının görsel olarak yorumlanmasına olanak sağlamaktadır. Buna karşılık tek bir karar ağacı veri setine aşırı uyum (overfitting) gösterebilmekte ve farklı eğitim veri kümelerinde farklı ağaç yapıları oluşturabilmektedir. Bu nedenle model performansının artırılabilmesi için maksimum ağaç derinliği (max_depth), düğümün bölünmesi için gerekli minimum örnek sayısı (min_samples_split) ve yaprak düğümde bulunması gereken minimum örnek sayısı (min_samples_leaf) gibi hiperparametrelerin uygun şekilde belirlenmesi gerekmektedir (James vd., 2021).

Karar ağacı algoritmaları, özellikle doğrusal olmayan hidrolojik süreçlerin modellenmesinde başarılı sonuçlar vermektedir. Bununla birlikte tek bir karar ağacının yüksek varyans göstermesi ve aşırı öğrenmeye yatkın olması nedeniyle günümüzde çoğu çalışmada Random Forest ve XGBoost gibi topluluk öğrenmesi (ensemble learning) algoritmaları tercih edilmektedir. Bununla birlikte, Decision Tree algoritması bu yöntemlerin temel yapı taşı olması ve yorumlanabilirliği nedeniyle önemini korumaktadır (Breiman vd., 1984; Mosavi vd., 2018).

Bu çalışmada Decision Tree algoritması, günlük aktif doluluk oranının tahmin edilmesi amacıyla karşılaştırma modellerinden biri olarak kullanılmıştır. Model girdileri olarak günlük ortalama sıcaklık, günlük toplam yağış, ortalama nispi nem, ortalama aktüel basınç ve Lag-1 aktif doluluk oranı kullanılmıştır. Modelin hiperparametreleri GridSearchCV yöntemi ve Scikit-learn kütüphanesinin varsayılan yöntemi olan beş katlı K-Fold çapraz doğrulama (5-fold K-Fold Cross Validation) kullanılarak optimize edilmiş, daha sonra model performansı test veri seti üzerinde değerlendirilmiş ve diğer makine öğrenmesi algoritmaları ile karşılaştırılmıştır.

3.4.3 Random Forest (RF)

Random Forest (RF), Breiman (2001) tarafından geliştirilen ve karar ağaçlarının topluluk öğrenmesi (ensemble learning) yaklaşımıyla bir araya getirilmesine dayanan denetimli bir makine öğrenmesi algoritmasıdır. Algoritma, çok sayıda karar ağacının oluşturulması ve bu ağaçlardan elde edilen tahminlerin birleştirilmesi esasına dayanmaktadır. Bu yaklaşım, tek bir karar ağacında görülebilen yüksek varyans ve aşırı öğrenme (overfitting) problemlerini azaltarak daha kararlı ve yüksek doğrulukta tahminler üretilmesini sağlamaktadır (Breiman, 2001).

Random Forest algoritmasının temelini bootstrap örnekleme (bootstrap sampling) ve rastgele değişken seçimi (random feature selection) oluşturmaktadır. Bootstrap yönteminde eğitim veri setinden tekrar seçmeye izin verilerek çok sayıda alt veri seti oluşturulmakta ve her karar ağacı bu alt veri setlerinden biri kullanılarak eğitilmektedir. Ayrıca her düğümde tüm bağımsız değişkenler yerine yalnızca rastgele seçilen bir değişken alt kümesi değerlendirilmektedir. Böylece karar ağaçları

arasındaki benzerlik azaltılmakta, model çeşitliliği artırılmakta ve genelleme performansı iyileştirilmektedir (James vd., 2021).

Regresyon problemlerinde Random Forest algoritmasının nihai tahmini, oluşturulan tüm karar ağaçlarının tahminlerinin ortalaması alınarak elde edilmektedir. Bu yaklaşım aşağıdaki eşitlik ile ifade edilmektedir:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \tag{3.2}$$

Burada $\hat{y}$ modelin nihai tahminini, N oluşturulan karar ağacı sayısını ve $T_i(x)$ ise her bir karar ağacının ürettiği tahmin değerini ifade etmektedir (Breiman, 2001).

Random Forest algoritması, doğrusal olmayan ilişkileri modelleyebilmesi, değişkenler arasındaki karmaşık etkileşimleri öğrenebilmesi ve yüksek boyutlu veri setlerinde başarılı sonuçlar üretebilmesi nedeniyle hidrolojik modelleme çalışmalarında yaygın olarak kullanılmaktadır. Bunun yanında veri setindeki gürültüye ve aykırı gözlemlere karşı görece dayanıklı olması, algoritmanın önemli avantajları arasında yer almaktadır (Mosavi vd., 2018; Tyralis vd., 2019).

Random Forest algoritmasının önemli özelliklerinden biri de bağımsız değişkenlerin modele katkı düzeylerini belirleyebilmesidir. Model oluşturulduktan sonra her bir değişken için önem değerleri hesaplanabilmekte, böylece tahmin performansına en fazla katkı sağlayan değişkenler belirlenebilmektedir (Cutler vd., 2007). Bu özellik, yalnızca model performansının değerlendirilmesine değil, aynı zamanda modelin yorumlanmasına da önemli katkı sağlamaktadır.

Bununla birlikte Random Forest algoritmasının performansı, oluşturulan karar ağacı sayısı (n_estimators), maksimum ağaç derinliği (max_depth) ve düğüm bölünme kriterleri gibi hiperparametrelerin uygun şekilde belirlenmesine bağlıdır. Bu nedenle modelin yüksek tahmin başarısı elde edebilmesi için hiperparametre optimizasyonu büyük önem taşımaktadır (Liaw ve Wiener, 2002).

Bu çalışmada Random Forest algoritması, günlük aktif doluluk oranının tahmin edilmesi amacıyla uygulanmıştır. Model girdileri olarak günlük ortalama sıcaklık,

günlük toplam yağış, ortalama nispi nem, ortalama aktüel basınç ve Lag-1 aktif doluluk oranı kullanılmıştır. Modelin hiperparametreleri GridSearchCV yöntemi ve Scikit-learn kütüphanesinin varsayılan yöntemi olan beş katlı K-Fold çapraz doğrulama (5-fold K-Fold Cross Validation) kullanılarak optimize edilmiş, daha sonra optimize edilen modelin performansı test veri seti üzerinde değerlendirilerek diğer makine öğrenmesi algoritmaları ile karşılaştırılmıştır.

3.4.4 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost), Chen ve Guestrin (2016) tarafından geliştirilen ve gradyan artırma (gradient boosting) yaklaşımına dayanan güçlü bir topluluk öğrenmesi (ensemble learning) algoritmasıdır. XGBoost, ardışık olarak oluşturulan karar ağaçları sayesinde önceki modellerin tahmin hatalarını azaltmayı hedeflemekte ve özellikle büyük veri setlerinde yüksek tahmin doğruluğu, hızlı hesaplama kapasitesi ve aşırı öğrenmeyi (overfitting) azaltan yapısı nedeniyle makine öğrenmesi çalışmalarında yaygın olarak kullanılmaktadır (Chen ve Guestrin, 2016).

Topluluk öğrenmesi yöntemleri genel olarak bagging ve boosting olmak üzere iki temel yaklaşıma ayrılmaktadır. Random Forest algoritması bagging yaklaşımını benimserken, XGBoost boosting yaklaşımını kullanmaktadır. Bagging yönteminde çok sayıda model birbirinden bağımsız olarak eğitilirken, boosting yaklaşımında modeller ardışık olarak oluşturulmakta ve her yeni model bir önceki modelin tahmin hatalarını azaltmaya odaklanmaktadır. Bu sayede modelin tahmin doğruluğu olarak artırılmaktadır (Hastie vd., 2009).

XGBoost algoritmasında amaç fonksiyonu, tahmin hatasını temsil eden kayıp fonksiyonu ile model karmaşıklığını kontrol eden düzenlileştirme (regularization) teriminden oluşmaktadır. Amaç fonksiyonu aşağıdaki şekilde ifade edilmektedir.

$$L_{xgb} = \sum_{i=1}^{N} L(y_i, F(x_i)) + \sum_{m=1}^{M} \Omega(h_m) \tag{3.3}$$

$$\Omega(h) = \gamma T + \frac{1}{2}\lambda \|w\| 2 \tag{3.4}$$

Burada $L\ (y_i, F(xi))$ gözlenen ve tahmin edilen değerler arasındaki kayıp fonksiyonunu, $\Omega(h)$ ise model karmaşıklığını kontrol eden düzenleme terimini ifade etmektedir. Düzenleme terimi içerisinde yer alan T ağaçtaki yaprak düğüm sayısını, w yaprak ağırlıklarını, γ yaprak sayısına bağlı ceza katsayısını ve λ ise L2 düzenlileştirme katsayısını ifade etmektedir (Chen ve Guestrin, 2016). Düzenlileştirme mekanizması sayesinde model karmaşıklığı kontrol edilmekte ve aşırı öğrenme riski azaltılmaktadır.

XGBoost algoritması, doğrusal olmayan ilişkileri modelleyebilmesi, değişkenler arasındaki karmaşık etkileşimleri öğrenebilmesi ve yüksek boyutlu veri setlerinde başarılı sonuçlar üretebilmesi nedeniyle hidrolojik modelleme çalışmalarında yaygın olarak kullanılmaktadır. Ayrıca eksik verilerle çalışabilmesi, paralel işlem desteği sunması ve değişken önem analizleri gerçekleştirebilmesi algoritmanın önemli avantajları arasında yer almaktadır (Mosavi vd., 2018). Bununla birlikte model performansı; maksimum ağaç derinliği (max_depth), öğrenme oranı (learning_rate), karar ağacı sayısı (n_estimators), alt örnekleme oranı (subsample) ve değişken örnekleme oranı (colsample_bytree) gibi hiperparametrelerin uygun şekilde belirlenmesine bağlıdır. Bu nedenle hiperparametre optimizasyonu, model başarısını doğrudan etkileyen önemli bir aşamadır (Prokhorenkova vd., 2018).

Bu çalışmada XGBoost algoritması, günlük aktif doluluk oranının tahmin edilmesi amacıyla uygulanmıştır. Model girdileri olarak günlük ortalama sıcaklık, günlük toplam yağış, ortalama nispi nem, ortalama aktüel basınç ve Lag-1 aktif doluluk oranı kullanılmıştır. Model hiperparametreleri GridSearchCV yöntemi ve Scikit-learn kütüphanesinin varsayılan yöntemi olan beş katlı K-Fold çapraz doğrulama (5-fold K-Fold Cross Validation) kullanılarak optimize edilmiş, daha sonra optimize edilen modelin performansı test veri seti üzerinde değerlendirilerek doğrusal regresyon, Decision Tree ve Random Forest modelleri ile karşılaştırılmıştır.

3.5 Model Optimizasyonu ve Hiperparametre Ayarları

Makine öğrenmesi algoritmalarının tahmin performansı yalnızca kullanılan veri setine değil, aynı zamanda model hiperparametrelerinin uygun şekilde belirlenmesine de bağlıdır. Hiperparametreler, modelin öğrenme sürecini kontrol eden ve eğitim

öncesinde kullanıcı tarafından belirlenen ayarlardır. Bu parametrelerin uygun seçilmemesi durumunda model veri setine aşırı uyum gösterebilmekte (overfitting) veya verideki örüntüleri yeterince öğrenemeyerek düşük performans sergileyebilmektedir. Bu nedenle çalışmada model performansının artırılması amacıyla hiperparametre optimizasyonu gerçekleştirilmiştir.

Çalışmada kullanılan Doğrusal Regresyon (Linear Regression, LR), Karar Ağacı (Decision Tree, DT), Random Forest (RF) ve Extreme Gradient Boosting (XGBoost) algoritmalarına ait hiperparametreler GridSearchCV yöntemi kullanılarak optimize edilmiştir. GridSearchCV, önceden belirlenen hiperparametre kombinasyonlarını sistematik olarak değerlendirerek en yüksek performansı sağlayan parametre setini belirleyen yaygın bir optimizasyon yöntemidir (Pedregosa vd., 2011). Bu yaklaşım, deneme-yanılma esasına dayalı parametre seçiminden kaynaklanabilecek öznelliği ortadan kaldırmakta ve model performansının nesnel olarak değerlendirilmesine olanak sağlamaktadır.

Hiperparametre optimizasyonunda her parametre kombinasyonu Scikit-learn kütüphanesinin varsayılan yöntemi olan beş katlı K-Fold çapraz doğrulama (5-fold K-Fold Cross Validation) yöntemi ile değerlendirilmiştir. Bu yöntemde eğitim veri seti beş alt kümeye ayrılmış, her iterasyonda dört alt küme model eğitimi için kullanılırken kalan bir alt küme doğrulama amacıyla değerlendirilmiştir. Bu işlem her alt küme doğrulama veri seti olacak şekilde beş kez tekrarlanmış ve elde edilen performans değerlerinin ortalaması alınarak en uygun hiperparametre kombinasyonları belirlenmiştir. Böylece model performansının yalnızca tek bir eğitim-doğrulama ayrımına bağlı olarak değerlendirilmesi önlenmiş ve daha güvenilir sonuçlar elde edilmiştir (Kohavi, 1995).

Çalışmada doğrusal regresyon modeli için sabit terimin (fit_intercept) modele dahil edilip edilmeyeceği değerlendirilmiştir. Karar Ağacı algoritmasında maksimum ağaç derinliği (max_depth), düğümün bölünmesi için gerekli minimum örnek sayısı (min_samples_split) ve yaprak düğümde bulunması gereken minimum örnek sayısı (min_samples_leaf) optimize edilmiştir. Random Forest algoritmasında bu parametrelere ek olarak karar ağacı sayısı (n_estimators) değerlendirilmiştir. XGBoost

algoritmasında ise maksimum ağaç derinliği (max_depth), öğrenme oranı (learning_rate), karar ağacı sayısı (n_estimators), gözlem alt örnekleme oranı (subsample) ve değişken alt örnekleme oranı (colsample_bytree) için hiperparametre optimizasyonu gerçekleştirilmiştir.

Optimizasyon sonucunda elde edilen en uygun hiperparametreler kullanılarak modeller yeniden eğitilmiş ve test veri seti üzerinde performans değerlendirmeleri gerçekleştirilmiştir. Her algoritma için belirlenen optimum hiperparametre değerleri ve bu parametrelerle elde edilen model performansları Bulgular bölümünde sunulmuştur.

Hiperparametre optimizasyonu tamamlandıktan sonra modellerin tahmin performansları çeşitli performans ölçütleri kullanılarak değerlendirilmiştir.

3.6 Model Performansının Değerlendirilmesi

Makine öğrenmesi modellerinin başarısının belirlenebilmesi için model tarafından üretilen tahmin değerlerinin gözlenen değerlerle karşılaştırılması gerekmektedir. Bu amaçla çeşitli performans ölçütleri kullanılmaktadır. Performans ölçütleri, modelin gözlenen verileri ne ölçüde doğru temsil ettiğini ve daha önce görmediği veriler üzerindeki tahmin başarısını ortaya koymaktadır (Willmott ve Matsuura, 2005).

Bu çalışmada geliştirilen Doğrusal Regresyon (Linear Regression, LR), Karar Ağacı (Decision Tree, DT), Random Forest (RF) ve Extreme Gradient Boosting (XGBoost) modellerinin performanslarının değerlendirilmesinde Ortalama Mutlak Hata (Mean Absolute Error, MAE), Kök Ortalama Kare Hata (Root Mean Squared Error, RMSE) ve Belirlilik Katsayısı (Coefficient of Determination, R^2) kullanılmıştır. Bu performans ölçütleri hidroloji, su kaynakları yönetimi ve makine öğrenmesi alanlarında gerçekleştirilen çalışmalarda yaygın olarak kullanılmaktadır (Chai ve Draxler, 2014; Moriasi vd., 2007).

3.6.1 Ortalama Mutlak Hata (MAE)

Ortalama Mutlak Hata (MAE), gözlenen ve tahmin edilen değerler arasındaki mutlak farkların ortalamasını ifade etmektedir. MAE değeri model hatalarının büyüklüğünü doğrudan göstermekte olup yorumlanması oldukça kolaydır. MAE değerinin düşük olması model tahminlerinin gözlenen değerlere daha yakın olduğunu göstermektedir.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i| \tag{3.5}$$

Burada y_i gözlenen değeri, $\hat{y}_i$ model tarafından tahmin edilen değeri ve n toplam gözlem sayısını ifade etmektedir. MAE değerinin sıfıra yaklaşması model tahminlerinin gözlenen değerlere daha yakın olduğunu göstermektedir.

3.6.2 Kök Ortalama Kare Hata (RMSE)

RMSE, gözlenen ve tahmin edilen değerler arasındaki farkların karelerinin ortalamasının karekökü alınarak hesaplanmaktadır. Büyük tahmin hatalarına daha duyarlı olması nedeniyle hidrolojik modelleme çalışmalarında yaygın olarak tercih edilen performans ölçütlerinden biridir (Chai ve Draxler, 2014).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{3.6}$$

Burada y_i gözlenen değeri, $\hat{y}_i$ model tarafından tahmin edilen değeri ve n toplam gözlem sayısını ifade etmektedir. RMSE değerinin düşük olması modelin tahmin performansının yüksek olduğunu göstermektedir.

3.6.3 Belirlilik Katsayısı (R^2)

Belirlilik katsayısı (R^2), bağımlı değişkendeki toplam değişimin model tarafından ne ölçüde açıklanabildiğini göstermektedir. R^2 değeri 0 ile 1 arasında değişmekte olup değerin 1'e yaklaşması modelin açıklayıcılık düzeyinin arttığını ifade etmektedir (Montgomery vd., 2021).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3.7)$$

Burada y_i gözlenen değeri, $\hat{y}_i$ tahmin edilen değeri, $\bar{y}$ gözlenen değerlerin ortalamasını ve n toplam gözlem sayısını ifade etmektedir.

Bu çalışmada geliştirilen tüm modeller için MAE, RMSE ve R^2 değerleri test veri seti üzerinde hesaplanmış ve modeller bu ölçütler esas alınarak karşılaştırılmıştır. Model performanslarının değerlendirilmesinde MAE ve RMSE değerlerinin düşük, R^2 değerinin ise yüksek olması tercih edilmiştir. En düşük hata değerlerine ve en yüksek R^2 değerine sahip model çalışma alanı için en başarılı tahmin modeli olarak değerlendirilmiştir.

3.7 Yazılım Ortamı ve Kullanılan Kütüphaneler

Bu çalışmada veri hazırlama, makine öğrenmesi modelleme süreçleri, hiperparametre optimizasyonu, performans değerlendirmeleri ve görselleştirme işlemleri Python programlama dili kullanılarak gerçekleştirilmiştir. Python, açık kaynak kodlu yapısı, geniş kütüphane desteği ve veri bilimi alanındaki yaygın kullanımı nedeniyle makine öğrenmesi çalışmalarında en sık tercih edilen programlama dillerinden biridir (VanderPlas, 2016).

Veri setinin düzenlenmesi, birleştirilmesi ve veri ön işleme işlemlerinde Pandas, sayısal hesaplamalarda ise NumPy kütüphanelerinden yararlanılmıştır (Harris vd., 2020; McKinney, 2010).

Makine öğrenmesi modellerinin geliştirilmesi ve uygulanmasında Scikit-learn kütüphanesi kullanılmıştır. Doğrusal Regresyon (Linear Regression), Karar Ağacı (Decision Tree) ve Random Forest algoritmaları bu kütüphane aracılığıyla oluşturulmuş; hiperparametre optimizasyonu, 5 katlı K-Fold çapraz doğrulama ve model performans değerlendirmeleri de Scikit-learn kütüphanesinin ilgili modülleri kullanılarak gerçekleştirilmiştir (Pedregosa vd., 2011).

Extreme Gradient Boosting (XGBoost) modelinin geliştirilmesinde ise Python ortamında çalışan XGBoost kütüphanesinden yararlanılmıştır (Chen ve Guestrin, 2016).

Model sonuçlarının görselleştirilmesinde Matplotlib ve Seaborn kütüphaneleri kullanılmıştır. Ayrıca modelin açıklanabilirliğinin değerlendirilmesi amacıyla gerçekleştirilen SHAP analizlerinde SHAP kütüphanesinden yararlanılmış; değişkenlerin modele olan etkilerinin yorumlanmasında bu kütüphanenin sunduğu görselleştirme araçları kullanılmıştır (Lundberg ve Lee, 2017).

Çalışmada gerçekleştirilen tüm veri işleme, model geliştirme, optimizasyon, görselleştirme ve performans değerlendirme süreçleri aynı Python ortamında yürütülmüş olup analizlerin tekrarlanabilirliği sağlanmıştır.

3.8 Model Yorumlama ve Açıklanabilir Yapay Zekâ Analizleri

Makine öğrenmesi modelleri yüksek tahmin doğruluğu sağlayabilmelerine rağmen, çoğu zaman karar mekanizmalarının doğrudan yorumlanamaması nedeniyle "kara kutu (black-box)" modeller olarak tanımlanmaktadır. Bu nedenle son yıllarda yalnızca model performansının değerlendirilmesi değil, aynı zamanda modelin tahminlerini hangi değişkenlere dayanarak oluşturduğunun ortaya konulması da büyük önem kazanmıştır (Molnar, 2022). Bu amaçla geliştirilen Açıklanabilir Yapay Zekâ (Explainable Artificial Intelligence, XAI) yöntemleri, makine öğrenmesi modellerinin yorumlanabilirliğini artırmakta ve değişkenlerin model üzerindeki etkilerinin daha ayrıntılı biçimde değerlendirilmesine olanak sağlamaktadır.

Bu çalışmada, model karşılaştırmaları sonucunda en başarılı performansı gösteren optimize edilmiş Random Forest modeli esas alınarak meteorolojik değişkenlerin baraj aktif doluluk oranı üzerindeki etkileri ayrıntılı olarak incelenmiştir. Bu kapsamda değişken önem düzeylerinin belirlenmesi amacıyla Feature Importance, değişkenlerin modele katkılarının doğrulanması amacıyla Permutation Importance, değişkenlerin tahminler üzerindeki yönü ve büyüklüğünün değerlendirilmesi amacıyla SHapley Additive exPlanations (SHAP) analizi ve değişkenlerin model çıktısı üzerindeki

marjinal etkilerinin belirlenmesi amacıyla Partial Dependence Plot (PDP) analizleri uygulanmıştır.

Feature Importance analizi ile Random Forest modelinde yer alan her bir bağımsız değişkenin tahmin performansına yaptığı göreceli katkı hesaplanmıştır. Permutation Importance analizinde ise her bir değişkenin değerleri rastgele karıştırılarak model performansındaki değişim değerlendirilmiş ve değişkenlerin gerçek önem düzeyleri belirlenmiştir.

SHAP analizi, oyun teorisine dayanan Shapley değerleri yaklaşımı kullanılarak her bir değişkenin model tahminlerine yaptığı katkının yönünü ve büyüklüğünü hesaplamaktadır (Lundberg ve Lee, 2017). Böylece yalnızca değişkenlerin önem sıralaması değil, aynı zamanda değişken değerlerindeki artış veya azalışın model tahminlerini hangi yönde etkilediği de ortaya konulabilmektedir.

Partial Dependence Plot (PDP) analizi ise diğer değişkenlerin ortalama etkisi sabit tutulurken tek bir bağımsız değişkendeki değişimin model çıktısı üzerindeki etkisini göstermektedir. Bu analiz sayesinde meteorolojik değişkenler ile baraj aktif doluluk oranı arasındaki doğrusal olmayan ilişkiler görsel olarak değerlendirilmiş ve model tarafından öğrenilen örüntüler ayrıntılı biçimde yorumlanmıştır.

Bu analizlerden elde edilen sonuçlar, makine öğrenmesi modellerinin yalnızca tahmin performanslarının değil, aynı zamanda tahmin mekanizmalarının da değerlendirilmesine olanak sağlamış ve meteorolojik değişkenlerin baraj aktif doluluk oranı üzerindeki göreceli etkilerinin kapsamlı biçimde ortaya konulmasını sağlamıştır.

4. BULGULAR

4.1 Çalışmada Kullanılan Değişkenlere Ait Tanımlayıcı İstatistikler

Çalışmada kullanılan aktif doluluk oranı ile meteorolojik değişkenlere ilişkin tanımlayıcı istatistikler Tablo 4.1'de sunulmuştur.

Tablo 4.1 Çalışmada kullanılan değişkenlere ait tanımlayıcı istatistikler

Değişken	N	Minimum	Maksimum	Ortalama	Standart Sapma
Aktif Doluluk (%)	6575	0,09	100,92	58,51	27,03
Günlük Maksimum Sıcaklık (°C)	6575	-9,20	41,60	17,91	9,80
Günlük Minimum Sıcaklık (°C)	6575	-20,20	22,90	4,98	6,75
Günlük Ortalama Sıcaklık (°C)	6575	-13,90	30,10	10,66	7,99
Günlük Toplam Yağış (mm)	6575	0,00	82,60	1,50	4,48
Ortalama Nispi Nem (%)	6575	27,10	99,30	76,07	13,42
Ortalama Aktüel Basınç (hPa)	6575	901,30	941,50	923,85	4,98

Tablo 4.1 incelendiğinde, çalışma kapsamında değerlendirilen tüm değişkenler için toplam 6575 günlük gözlem bulunduğu ve veri setinde eksik gözlem yer almadığı görülmektedir. Bağımlı değişken olarak kullanılan aktif doluluk oranının ortalama değeri %58,51 olup minimum %0,09 ve maksimum %100,92 arasında değişmektedir.

Meteorolojik değişkenler incelendiğinde, günlük maksimum sıcaklığın -9,20 ile 41,60°C, günlük minimum sıcaklığın -20,20 ile 22,90°C ve günlük ortalama sıcaklığın ise -13,90 ile 30,10°C arasında değiştiği belirlenmiştir. Günlük toplam yağış miktarının ortalama 1.50 mm olduğu, maksimum günlük yağışın ise 82,60 mm'ye ulaştığı görülmektedir. Ortalama nispi nem %76,07, ortalama aktüel basınç ise 923,85 hPa olarak hesaplanmıştır. Meteorolojik değişkenlere ait minimum ve maksimum değerler, çalışma döneminde gözlenen değişim aralıklarını ortaya koymaktadır.

4.1.1 Değişkenlerin Normal Dağılım Özellikleri

Parametrik analizlerin uygulanabilirliğinin değerlendirilmesi amacıyla değişkenlerin normal dağılıma uygunluğu Kolmogorov–Smirnov testi ile incelenmiştir. Test sonuçları Tablo 4.2'de verilmiştir.

Tablo 4.2 Değişkenlere ait normal dağılım testi sonuçları

Değişken	**KS**	**p**	**Skewness**	**Kurtosis**
Aktif Doluluk	0,103	<0,001	0,035	-1,333
Günlük Maksimum Sıcaklık	0,066	<0,001	-0,167	-0,976
Günlük Minimum Sıcaklık	0,063	<0,001	-0,211	-0,620
Günlük Ortalama Sıcaklık	0,069	<0,001	-0,123	-0,958
Günlük Toplam Yağış	0,369	<0,001	6,133	58,884
Ortalama Nispi Nem	0,043	<0,001	-0,290	-0,579
Ortalama Aktüel Basınç	0,024	<0,001	0,015	0,319

Tablo 4.2 incelendiğinde, tüm değişkenler için Kolmogorov–Smirnov test sonuçlarının istatistiksel olarak anlamlı olduğu görülmektedir ($p<0,001$). Bununla birlikte, örneklem büyüklüğünün yüksek olması ($N=6575$) nedeniyle normal dağılım değerlendirmesinde yalnızca Kolmogorov–Smirnov testi esas alınmamış, değişkenlere ait çarpıklık (Skewness) ve basıklık (Kurtosis) değerleri de birlikte değerlendirilmiştir.

Çarpıklık ve basıklık değerleri incelendiğinde, günlük toplam yağış değişkeni dışında diğer değişkenlerin normal dağılıma yakın özellik gösterdiği belirlenmiştir. Günlük toplam yağış değişkeninin ise yüksek çarpıklık (Skewness = 6,133) ve basıklık (Kurtosis = 58.884) değerlerine sahip olduğu görülmektedir.

4.1.2 Değişkenler Arasındaki Korelasyon İlişkileri

Çalışmada kullanılan değişkenler arasındaki doğrusal ilişkiler Pearson korelasyon analizi ile değerlendirilmiş olup sonuçlar Tablo 4.3'te sunulmuştur.

Tablo 4.3 Pearson korelasyon katsayıları

Değişken	Aktif Doluluk	T_{max}	T_{min}	T_{ort}	Yağış	Nem	Basınç
Aktif Doluluk	1						
T_{max}	0,295**	1					
T_{min}	0,286**	0,846**	1				
T_{ort}	0,315**	0,960**	0,950**	1			
Yağış	0,103**	-0,038**	0,129**	0,032**	1		
Nem	-0,293**	-0,637**	-0,348**	-0,549**	0,236**	1	
Basınç	-0,240**	-0,231**	-0,348**	-0,300**	-0,179**	0,060**	1

Not: ** p < 0,01 düzeyinde anlamlıdır

Korelasyon analizi sonuçlarına göre aktif doluluk oranı ile günlük ortalama sıcaklık arasında pozitif yönde zayıf düzeyde ilişki belirlenmiştir (r=0,315; p<0,001). Günlük maksimum sıcaklık (r=0,295) ve günlük minimum sıcaklık (r=0,286) ile aktif doluluk arasında da benzer şekilde zayıf pozitif ilişkiler tespit edilmiştir. Günlük toplam yağış ile aktif doluluk arasındaki ilişkinin oldukça zayıf olduğu görülmüştür (r=0,103; p<0,001). Buna karşılık aktif doluluk ile ortalama nispi nem (r=-0.293) ve ortalama aktüel basınç (r=-0,240) arasında zayıf düzeyde negatif ilişkiler belirlenmiştir.

Bağımsız değişkenler arasındaki ilişkiler incelendiğinde günlük maksimum sıcaklık ile günlük ortalama sıcaklık (r=0,960) ve günlük minimum sıcaklık ile günlük ortalama sıcaklık (r=0,950) arasında çok güçlü pozitif korelasyonlar bulunduğu görülmektedir. Benzer şekilde günlük maksimum ve minimum sıcaklık arasında da güçlü düzeyde pozitif ilişki tespit edilmiştir (r=0,846). Korelasyon katsayıları, sıcaklık değişkenleri arasında yüksek düzeyde doğrusal ilişki bulunduğunu ortaya koymaktadır.

Sıcaklık değişkenleri arasındaki çoklu doğrusal bağlantının daha ayrıntılı olarak değerlendirilmesi amacıyla Varyans Artış Faktörü (Variance Inflation Factor, VIF) analizi gerçekleştirilmiş ve sonuçlar Tablo 4.4'te sunulmuştur.

Tablo 4.4 Bağımsız değişkenlere ait Varyans Artış Faktörü (VIF) sonuçları

Değişken	Tolerance	VIF
Günlük Maksimum Sıcaklık (°C)	0,035	28,882
Günlük Minimum Sıcaklık (°C)	0,031	32,503
Günlük Ortalama Sıcaklık (°C)	0,010	99,094
Günlük Toplam Yağış (mm)	0,871	1,148
Ortalama Nispi Nem (%)	0,385	2,598
Ortalama Aktüel Basınç (hPa)	0,851	1,175

Tablo 4.4 incelendiğinde, günlük maksimum sıcaklık (VIF=28,882), günlük minimum sıcaklık (VIF=32,503) ve günlük ortalama sıcaklık (VIF=99,094) değişkenlerinin yüksek VIF değerlerine sahip olduğu görülmektedir. Buna karşılık günlük toplam yağış (VIF=1,148), ortalama nispi nem (VIF=2,598) ve ortalama aktüel basınç (VIF=1,175) değişkenlerinin VIF değerleri kabul edilebilir sınırlar içerisinde yer almaktadır. Elde edilen VIF sonuçları dikkate alınarak sonraki modelleme analizlerinde yalnızca günlük ortalama sıcaklık değişkeni kullanılmış, günlük maksimum ve minimum sıcaklık değişkenleri modele dahil edilmemiştir.

4.2 Baraj Doluluk Oranının Makine Öğrenmesi Modelleri ile Tahmini

4.2.1 Veri Setinin Hazırlanması

Baraj doluluk oranının tahmin edilmesine yönelik makine öğrenmesi modellerinin geliştirilmesinde 01 Ocak 2008–31 Aralık 2025 tarihleri arasındaki günlük gözlem verileri kullanılmıştır. Modelleme sürecinde kullanılan veri setine ilişkin genel özellikler Tablo 4.5'te sunulmuştur. Veri seti toplam 6574 günlük gözlemden oluşmakta olup bağımlı değişken olarak günlük aktif baraj doluluk oranı (%), bağımsız değişkenler olarak ise günlük ortalama sıcaklık (°C), günlük toplam yağış (mm), ortalama nispi nem (%) ve ortalama aktüel basınç (hPa) kullanılmıştır. Ayrıca zaman serisinin sürekliliğini temsil etmek amacıyla bir önceki güne ait aktif doluluk oranını ifade eden Lag-1 (aktif_doluluk_lag1) değişkeni oluşturularak modele dahil edilmiştir.

Tablo 4.5 Makine öğrenmesi modellerinde kullanılan veri setinin genel özellikleri

Parametre	**Açıklama**
Veri periyodu	02.01.2008–31.12.2025
Toplam gözlem sayısı	6574
Eğitim seti	5259 gözlem (%80)
Test seti	1315 gözlem (%20)
Bağımlı değişken	Aktif baraj doluluk oranı (%)
Bağımsız değişkenler	Lag-1 doluluk oranı, günlük ortalama sıcaklık, günlük toplam yağış, ortalama nispi nem, ortalama aktüel basınç

Veriler kronolojik sıraya göre düzenlenmiş ve zaman serilerinde bilgi eksikliğini önlemek amacıyla rastgele karıştırma yapılmaksızın eğitim ve test kümelerine ayrılmıştır. Veri setinin %80'i model eğitimi, kalan %20'si ise model performansının bağımsız olarak değerlendirilmesi amacıyla test verisi olarak kullanılmıştır. Bu kapsamda eğitim kümesi 5259 gözlemden, test kümesi ise 1315 gözlemden oluşmaktadır. Eğitim dönemi 02.01.2008–26.05.2022 tarihlerini, test dönemi ise 27.05.2022–31.12.2025 tarihleri arasındaki gözlemleri kapsamaktadır.

Model performanslarının değerlendirilmesinde Ortalama Mutlak Hata (MAE), Kök Ortalama Kare Hata (RMSE) ve Belirleme Katsayısı (R^2) performans ölçütleri kullanılmıştır. Tüm makine öğrenmesi modelleri aynı eğitim ve test veri kümeleri kullanılarak eğitilmiş ve elde edilen performans ölçütleri karşılaştırılmıştır.

4.2.2 Linear Regresyon Sonuçları

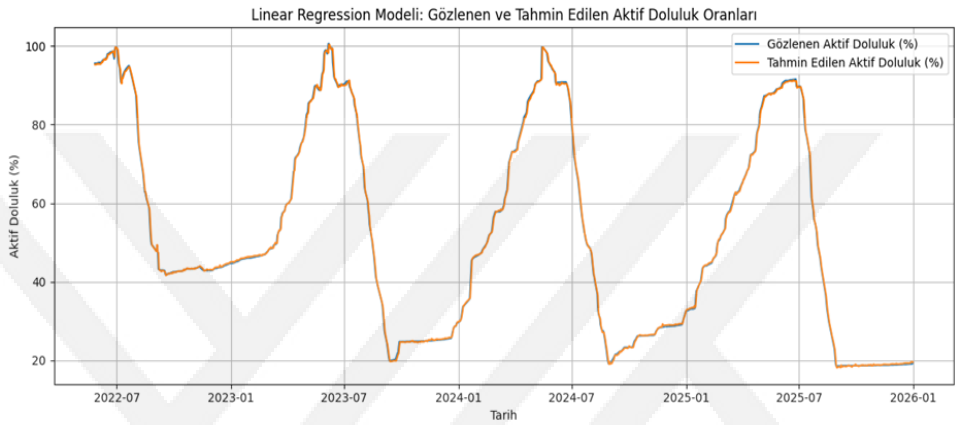
Linear Regresyon modeli, baraj aktif doluluk oranının tahmin edilmesinde temel doğrusal regresyon yaklaşımı olarak uygulanmıştır. Modelin eğitiminde eğitim veri seti kullanılmış, elde edilen model test veri seti üzerinde değerlendirilmiştir. Model performansı Ortalama Mutlak Hata (MAE), Kök Ortalama Kare Hata (RMSE) ve Belirlilik Katsayısı (R^2) performans ölçütleri kullanılarak belirlenmiştir. Linear Regresyon modeline ait performans sonuçları Tablo 4.6'te verilmiştir.

Tablo 4.6 Linear Regresyon modelinin performans sonuçları

Model	MAE	RMSE	R^2
Linear Regresyon	0,4119	0,6978	0,9993

Tablo 4.5 incelendiğinde, Linear Regresyon modeli için test veri setinde MAE değeri 0,4119; RMSE değeri 0,6978 ve R^2 değeri 0,9993 olarak hesaplanmıştır.

Linear Regresyon modeline ait test dönemindeki gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması Şekil 4.1'de verilmiştir.



Şekil 4.1 Linear Regresyon modeli için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması

Şekil 4.1 incelendiğinde, model tarafından tahmin edilen aktif doluluk oranlarının gözlenen değerlerle genel olarak benzer bir eğilim gösterdiği görülmektedir. Ayrıca modelin tahmin ettiği değerlerin zaman içerisindeki değişimleri büyük ölçüde takip ettiği görülmektedir.

Linear Regresyon modeli, çalışmada kullanılan makine öğrenmesi modelleri arasında temel karşılaştırma modeli olarak değerlendirilmiştir.

4.2.3 Decision Tree Sonuçları

Decision Tree algoritması, baraj aktif doluluk oranının tahmin edilmesinde kullanılan makine öğrenmesi modellerinden biri olarak uygulanmıştır. İlk aşamada model varsayılan parametrelerle oluşturulmuş, daha sonra hiperparametre optimizasyonu gerçekleştirilmiştir. Hiperparametre optimizasyonunda GridSearchCV yöntemi

kullanılmış ve en uygun model parametreleri Scikit-learn kütüphanesinin varsayılan yöntemi olan beş katlı K-Fold çapraz doğrulama (5-fold K-Fold Cross Validation) ile belirlenmiştir. Decision Tree modelinin varsayılan ve optimize edilmiş performans sonuçları Tablo 4.7'da verilmiştir.

Tablo 4.7 Decision Tree modelinin performans sonuçları

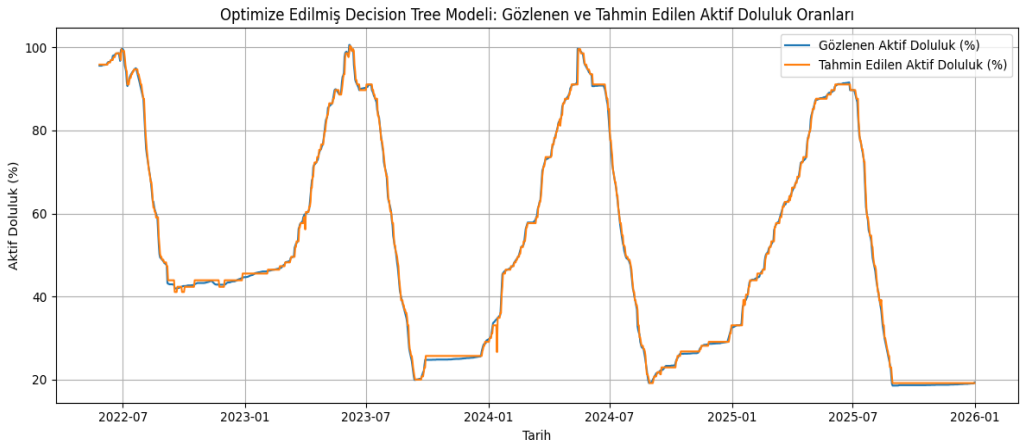
Model	**MAE**	**RMSE**	**R^2**
Decision Tree	0,5649	0,8834	0,9989
Optimize Edilmiş Decision Tree	0,5682	0,8470	0,9990

Tablo 4.6 incelendiğinde, varsayılan Decision Tree modeli için MAE, RMSE ve R^2 değerleri sırasıyla 0,5649; 0,8834 ve 0,9989 olarak hesaplanmıştır. Optimize edilmiş Decision Tree modeli için ise aynı performans ölçütleri sırasıyla 0,5682; 0,8470 ve 0,9990 olarak belirlenmiştir. GridSearchCV analizi sonucunda optimize edilmiş Decision Tree modeline ait en uygun hiperparametreler Tablo 4.8'de sunulmuştur.

Tablo 4.8 Optimize edilmiş Decision Tree modeline ait hiperparametreler

Hiperparametre	**En iyi değer**
Maksimum ağaç derinliği (*max_depth*)	7
Düğümün bölünmesi için gerekli minimum örnek sayısı (*min_samples_split*)	10
Yaprak düğümde bulunması gereken minimum örnek sayısı (*min_samples_leaf*)	2

Optimize edilmiş Decision Tree modeline ait gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması Şekil 4.2'de verilmiştir.



Şekil 4.2 Optimize edilmiş Decision Tree modeli için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması

Şekil 4.2 incelendiğinde, optimize edilmiş Decision Tree modeli ile tahmin edilen aktif doluluk oranlarının gözlenen değerlerle genel olarak benzer bir değişim eğilimi gösterdiği görülmektedir. Ayrıca model tarafından tahmin edilen değerlerin zaman içerisindeki değişimleri büyük ölçüde takip ettiği görülmektedir.

4.2.4 Random Forest Sonuçları

Random Forest algoritması, baraj aktif doluluk oranının tahmin edilmesinde kullanılan makine öğrenmesi modellerinden biri olarak uygulanmıştır. İlk aşamada model varsayılan parametrelerle oluşturulmuş, daha sonra GridSearchCV yöntemi kullanılarak hiperparametre optimizasyonu gerçekleştirilmiştir. Random Forest modelinin varsayılan ve optimize edilmiş performans sonuçları Tablo 4.9'de verilmiştir.

Tablo 4.9 Random Forest modelinin performans sonuçları

Model	**MAE**	**RMSE**	**R^2**
Random Forest	0,4405	0,7356	0,9992
Optimize Edilmiş Random Forest	0,3994	0,6718	0,9993

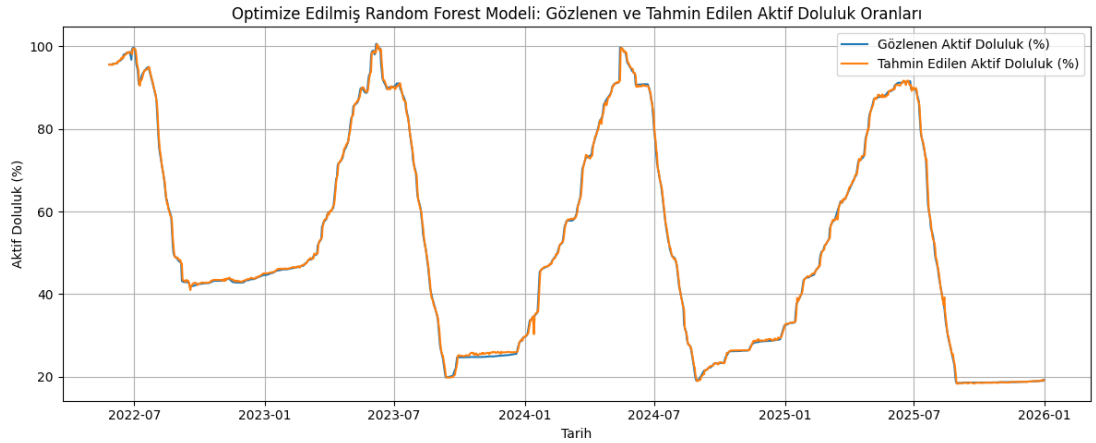
Tablo 4.9 incelendiğinde, varsayılan Random Forest modeli için MAE, RMSE ve R^2 değerleri sırasıyla 0,4405; 0,7356 ve 0,9992 olarak hesaplanmıştır. Optimize edilmiş Random Forest modeli için ise aynı performans ölçütleri sırasıyla 0,3994; 0,6718 ve 0,9993 olarak belirlenmiştir.

GridSearchCV analizi sonucunda optimize edilmiş Random Forest modeline ait en uygun hiperparametreler Tablo 4.10'da sunulmuştur.

Tablo 4.10 Optimize edilmiş Random Forest modeline ait en iyi hiperparametreler

Hiperparametre	**En iyi değer**
Ağaç sayısı (*n_estimators*)	200
Maksimum ağaç derinliği (*max_depth*)	None
Düğümün bölünmesi için gerekli minimum örnek sayısı (*min_samples_split*)	10
Yaprak düğümde bulunması gereken minimum örnek sayısı (*min_samples_leaf*)	4

Optimize edilmiş Random Forest modeline ait gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması Şekil 4.3'te verilmiştir.



Şekil 4.3 Optimize edilmiş Random Forest modeli için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması

Şekil 4.3 incelendiğinde, optimize edilmiş Random Forest modeli ile tahmin edilen aktif doluluk oranlarının gözlenen değerlerle genel olarak benzer bir değişim eğilimi gösterdiği görülmektedir. Ayrıca model tarafından tahmin edilen değerlerin zaman içerisindeki değişimleri büyük ölçüde takip etmektedir.

4.2.5 XGBoost Sonuçları

XGBoost (Extreme Gradient Boosting) algoritması, baraj aktif doluluk oranının tahmin edilmesinde kullanılan makine öğrenmesi modellerinden biri olarak uygulanmıştır. İlk aşamada model varsayılan parametrelerle oluşturulmuş, daha sonra GridSearchCV yöntemi kullanılarak hiperparametre optimizasyonu gerçekleştirilmiştir. XGBoost modelinin varsayılan ve optimize edilmiş performans sonuçları Tablo 4.11'de verilmiştir.

Tablo 4.11 XGBoost modelinin performans sonuçları

Model	**MAE**	**RMSE**	**R^2**
XGBoost	0,4774	0,7495	0,9992
Optimize Edilmiş XGBoost	0,4047	0,7016	0,9993

Tablo 4.11 incelendiğinde, varsayılan XGBoost modeli için MAE, RMSE ve R^2 değerleri sırasıyla 0,4774; 0,7495 ve 0,9992 olarak hesaplanmıştır. Optimize edilmiş

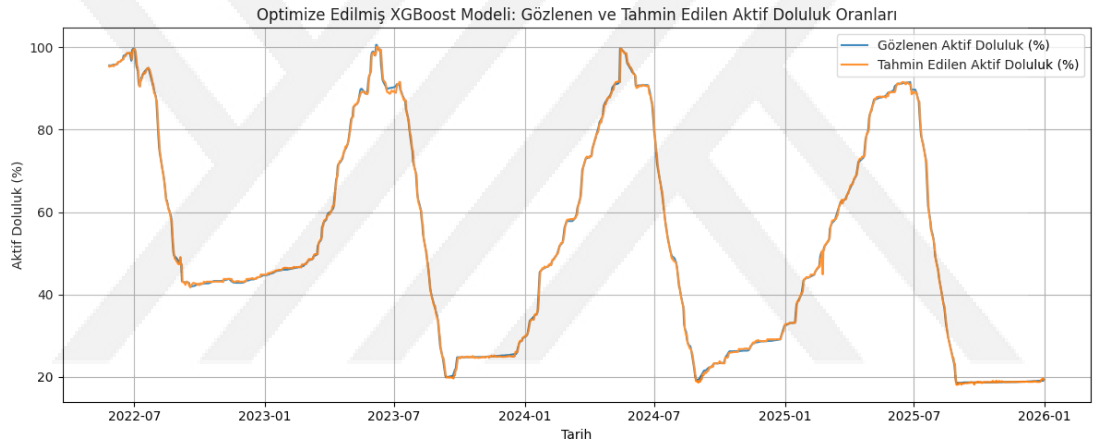
XGBoost modeli için ise aynı performans ölçütleri sırasıyla 0,4047; 0,7016 ve 0,9993 olarak belirlenmiştir.

GridSearchCV analizi sonucunda optimize edilmiş XGBoost modeline ait en uygun hiperparametreler Tablo 4.12'de sunulmuştur.

Tablo 4.12 Optimize edilmiş XGBoost modeline ait en iyi hiperparametreler

Hiperparametre	**En iyi değer**
Ağaç sayısı (*n_estimators*)	200
Maksimum ağaç derinliği (*max_depth*)	3
Öğrenme katsayısı (*learning_rate*)	0,05
Alt örnekleme oranı (*subsample*)	0,8
Değişken örnekleme oranı (*colsample_bytree*)	1,0

Optimize edilmiş XGBoost modeline ait gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması Şekil 4.4'te verilmiştir.



Şekil 4.4 Optimize edilmiş XGBoost modeli için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması.

Şekil 4.4 incelendiğinde, optimize edilmiş XGBoost modeli ile tahmin edilen aktif doluluk oranlarının gözlenen değerlerle genel olarak benzer bir değişim eğilimi gösterdiği görülmektedir. Ayrıca model tarafından tahmin edilen değerlerin zaman içerisindeki değişimleri büyük ölçüde takip etmektedir.

4.2.6 Modellerin Karşılaştırılması

Bu çalışmada baraj aktif doluluk oranının tahmin edilmesi amacıyla Linear Regresyon, Decision Tree, Random Forest ve XGBoost algoritmaları kullanılmıştır. Ayrıca Decision Tree, Random Forest ve XGBoost modelleri için hiperparametre optimizasyonu uygulanmıştır. Tüm modeller aynı eğitim ve test veri kümeleri kullanılarak eğitilmiş ve performansları MAE, RMSE ve R^2 ölçütleri esas alınarak karşılaştırılmıştır. Makine öğrenmesi modellerine ait performans sonuçlarının karşılaştırılması Tablo 4.13'de verilmiştir.

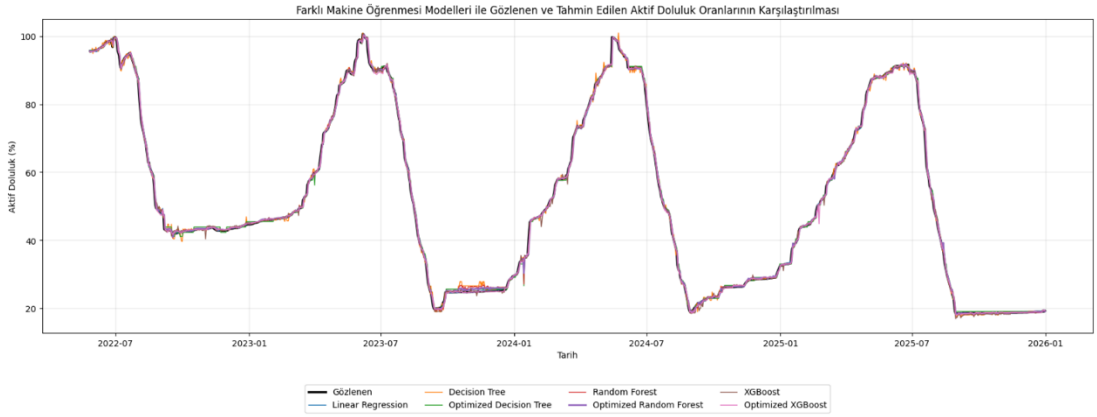
Tablo 4.13 Makine öğrenmesi modellerinin performans karşılaştırması

Model	**MAE**	**RMSE**	**R^2**
Linear Regresyon	0,4119	0,6978	0,9993
Decision Tree	0,5649	0,8834	0,9989
Optimize Edilmiş Decision Tree	0,5682	0,8470	0,9990
Random Forest	0,4405	0,7356	0,9992
Optimize Edilmiş Random Forest	**0,3994**	**0,6718**	**0,9993**
XGBoost	0,4774	0,7495	0,9992
Optimize Edilmiş XGBoost	0,4047	0,7016	0,9993

Tablo 4.13 incelendiğinde, optimize edilmiş Random Forest modelinin MAE, RMSE ve R^2 performans ölçütleri bakımından en yüksek performansı gösterdiği görülmektedir. Optimize edilmiş Random Forest modeli için MAE, RMSE ve R^2 değerleri sırasıyla 0,3994; 0,6718 ve 0,9993 olarak hesaplanmıştır. Optimize edilmiş XGBoost modeli benzer performans değerleri ile ikinci sırada yer alırken, Linear Regresyon modeli de yüksek performans göstermiştir. Decision Tree modeli ise diğer modellere kıyasla daha yüksek hata değerlerine sahip olmuştur.

Hiperparametre optimizasyonu sonrasında Random Forest ve XGBoost modellerinde MAE ve RMSE değerlerinde azalma, R^2 değerinde ise artış olduğu görülmektedir. Decision Tree modelinde ise optimizasyon sonrasında RMSE ve R^2 değerlerinde iyileşme gözlenirken, MAE değerinde sınırlı düzeyde artış meydana gelmiştir.

Modellerin test dönemindeki tahmin eğrilerinin gözlenen aktif doluluk oranları ile karşılaştırılması Şekil 4.5'te verilmiştir.



Şekil 4.5 Farklı makine öğrenmesi modelleri için test döneminde gözlenen ve tahmin edilen aktif doluluk oranlarının karşılaştırılması

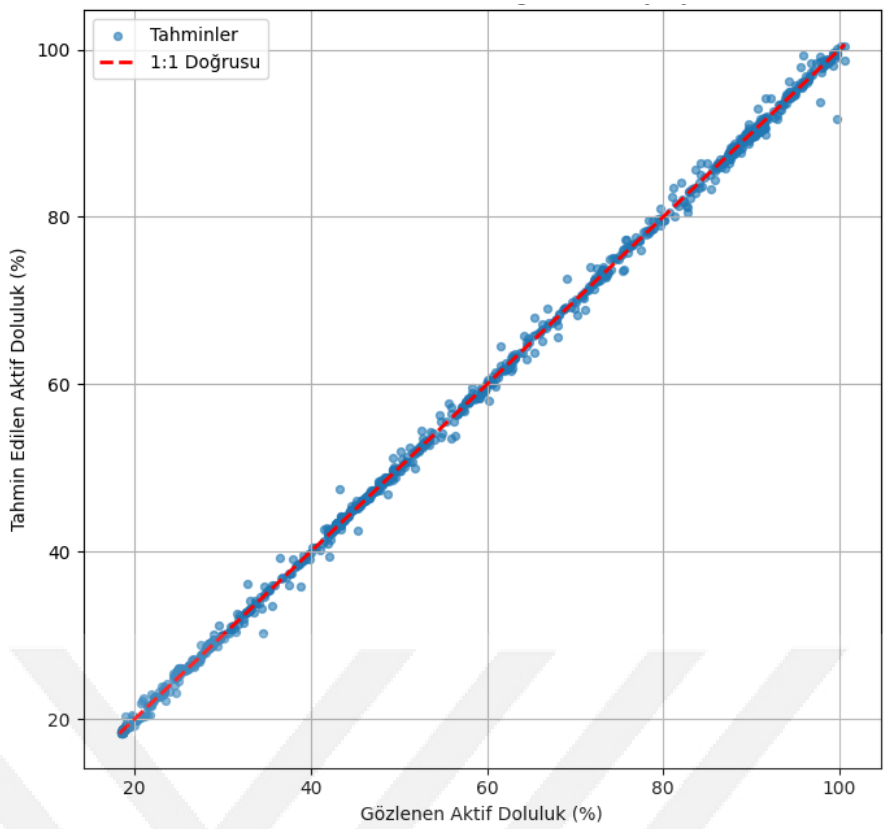
Şekil 4.5 incelendiğinde, tüm modellerin test dönemindeki aktif doluluk oranındaki zamansal değişimi genel olarak takip ettiği görülmektedir. Modeller arasında özellikle bazı dönemlerde tahmin eğrilerinin gözlenen değerlere yakınlığı bakımından farklılıklar bulunduğu dikkati çekmektedir. Optimize edilmiş Random Forest modeline ait tahmin eğrisinin test döneminin büyük bölümünde gözlenen değerlere daha yakın seyrettiği görülmektedir.

Tablo 4.13 ve Şekil 4.5 birlikte değerlendirildiğinde, optimize edilmiş Random Forest modelinin çalışmada değerlendirilen makine öğrenmesi modelleri arasında en yüksek performansı gösterdiği belirlenmiştir.

4.2.7 En Başarılı Modelin Değerlendirilmesi

Makine öğrenmesi modellerinin performans karşılaştırmaları sonucunda optimize edilmiş Random Forest modelinin en düşük MAE ve RMSE değerlerine, en yüksek R^2 değerine sahip olduğu belirlenmiştir. Bu nedenle optimize edilmiş Random Forest modeli üzerinden ayrıntılı grafiksel değerlendirmeler gerçekleştirilmiştir.

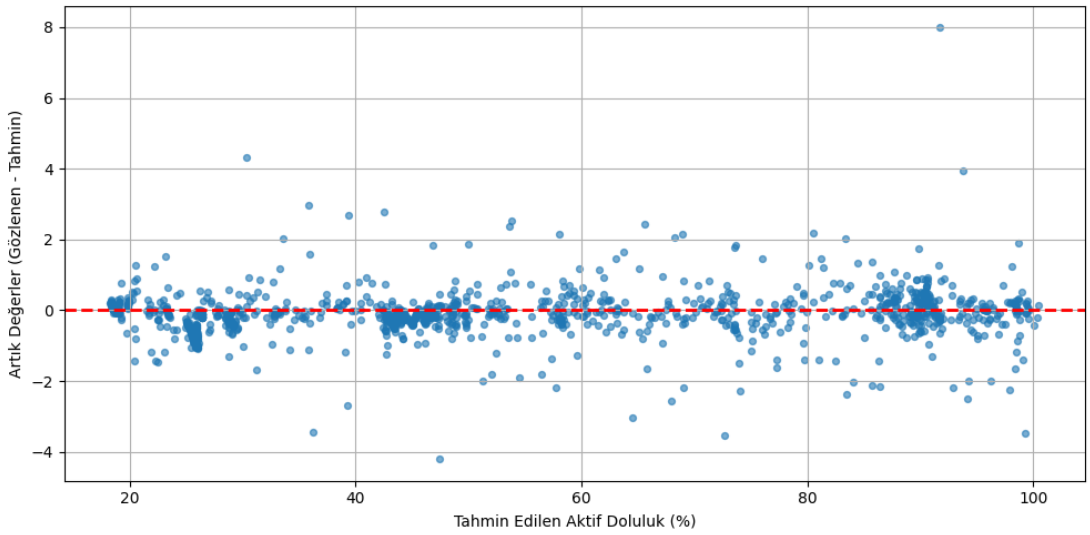
Optimize edilmiş Random Forest modelinin tahmin sonuçlarının görsel olarak değerlendirilmesi amacıyla gözlenen ve tahmin edilen aktif doluluk oranlarının saçılım grafiği oluşturulmuştur. Elde edilen sonuçlar Şekil 4.6'da verilmiştir.



Şekil 4.6 Optimize edilmiş Random Forest modeli için gözlenen ve tahmin edilen aktif doluluk oranlarının saçılım grafiği

Şekil 4.6 incelendiğinde, tahmin edilen değerlerin büyük bölümünün 1:1 doğrusu etrafında toplandığı görülmektedir. Ayrıca gözlenen ve tahmin edilen aktif doluluk oranlarının genel olarak birbirine yakın dağılım gösterdiği dikkati çekmektedir.

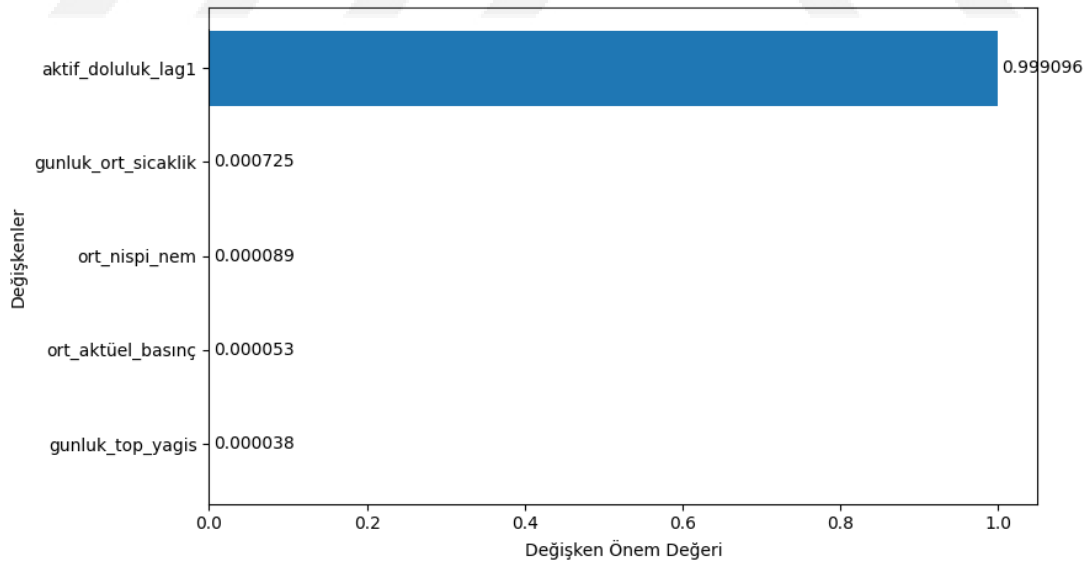
Model artıklarının dağılımını değerlendirmek amacıyla optimize edilmiş Random Forest modeline ait artık değerler hesaplanmış ve tahmin edilen aktif doluluk oranlarına göre dağılımları incelenmiştir. Elde edilen sonuçlar Şekil 4.7'de verilmiştir.



Şekil 4.7 Optimize edilmiş Random Forest modeli için artık değerlerin tahmin edilen aktif doluluk oranlarına göre dağılımı

Şekil 4.7 incelendiğinde, artık değerlerin sıfır çizgisi etrafında belirgin bir desen oluşturmadan dağıldığı görülmektedir. Ayrıca artık değerlerin test veri seti boyunca farklı tahmin düzeylerinde benzer bir dağılım gösterdiği dikkati çekmektedir.

Optimize edilmiş Random Forest modelinde bağımsız değişkenlerin modele katkı düzeylerini belirlemek amacıyla değişken önem (Feature Importance) analizi gerçekleştirilmiştir. Elde edilen değişken önem değerleri Şekil 4.8'de verilmiştir.



Şekil 4.8 Optimize edilmiş Random Forest modeline ait değişken önem düzeyleri

Şekil 4.8 incelendiğinde, değişken önem analizine göre en yüksek önem değerine sahip değişkenin Lag-1 olduğu görülmektedir. Lag-1 değişkenini sırasıyla günlük ortalama sıcaklık, ortalama nispi nem, ortalama aktüel basınç ve günlük toplam yağış değişkenleri takip etmektedir.

Feature Importance analizinden elde edilen bulgular doğrultusunda meteorolojik değişkenlerin baraj aktif doluluk oranı üzerindeki etkilerinin daha ayrıntılı olarak değerlendirilmesi amacıyla bir sonraki bölümde Lag-1 değişkeni modele dahil edilmeksizin ek analizler gerçekleştirilmiştir.

4.3 Meteorolojik Değişkenlerin Baraj Doluluğu Üzerindeki Etkileri

4.3.1 Korelasyon Analizi

Baraj aktif doluluk oranı ile meteorolojik değişkenler arasındaki doğrusal ilişkilerin belirlenmesi amacıyla Pearson korelasyon analizi gerçekleştirilmiştir. Korelasyon analizi sonucunda elde edilen katsayılar Tablo 4.14'de verilmiştir.

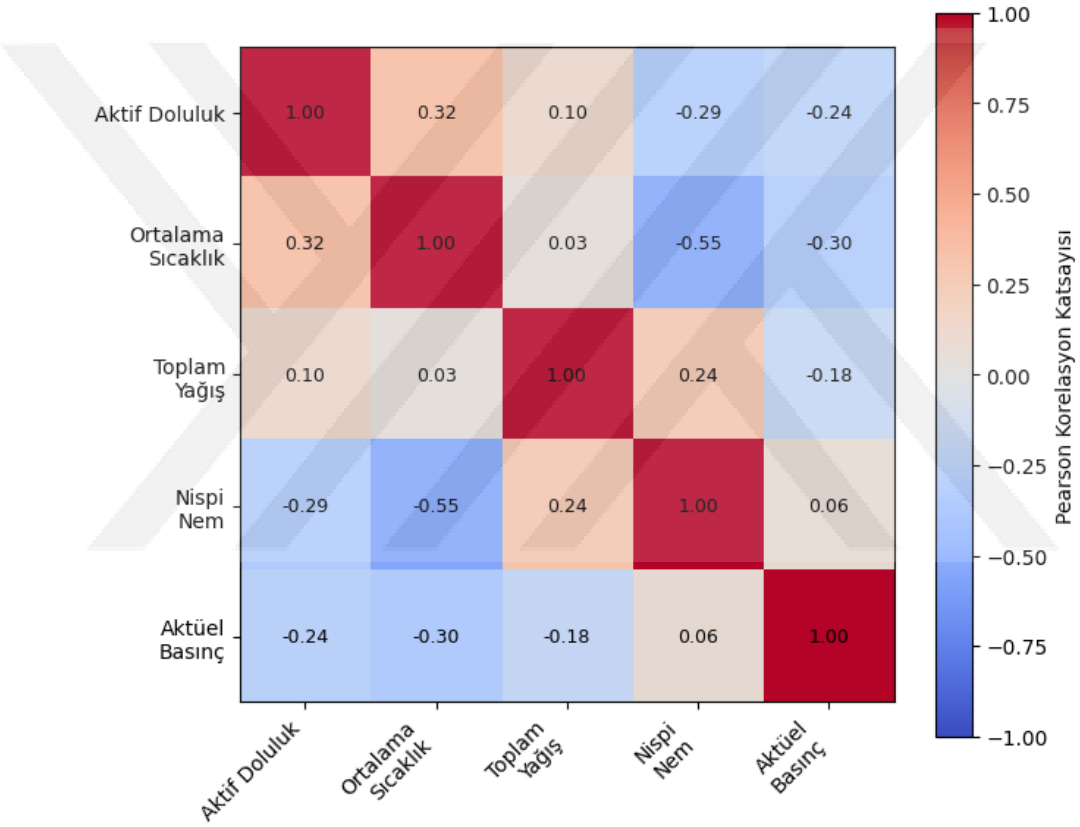
Tablo 4.14 Aktif doluluk oranı ile meteorolojik değişkenler arasındaki Pearson korelasyon katsayıları

Değişken	Aktif Doluluk	Ortalama Sıcaklık	Toplam Yağış	Nispi Nem	Aktüel Basınç
Aktif Doluluk	1,000				
Ortalama Sıcaklık	0,316	1,000			
Toplam Yağış	0,102	0,032	1,000		
Nispi Nem	-0,293	-0,549	0,236	1,000	
Aktüel Basınç	-0,240	-0,299	-0,179	0,060	1,000

Tablo 4.14 incelendiğinde, aktif doluluk oranı ile günlük ortalama sıcaklık arasında pozitif yönlü zayıf bir ilişki (r = 0,316) belirlenmiştir. Buna karşılık aktif doluluk oranı ile ortalama nispi nem (r = -0,293) ve ortalama aktüel basınç (r = -0,240) arasında negatif yönlü zayıf ilişkiler tespit edilmiştir. Günlük toplam yağış ile aktif doluluk oranı arasındaki korelasyon katsayısı ise 0,102 olarak hesaplanmıştır.

Meteorolojik değişkenlerin kendi aralarındaki ilişkileri incelendiğinde, günlük ortalama sıcaklık ile ortalama nispi nem arasında negatif yönlü orta düzeyde bir ilişki (r = -0,549) belirlenmiştir. Günlük toplam yağış ile ortalama nispi nem arasında pozitif yönlü (r = 0,236), günlük ortalama sıcaklık ile ortalama aktüel basınç arasında ise negatif yönlü (r = -0,299) ilişkiler belirlenmiştir. Diğer değişken çiftleri arasındaki korelasyon katsayılarının ise düşük düzeyde olduğu görülmüştür.

Korelasyon katsayılarının daha kolay değerlendirilebilmesi amacıyla değişkenler arasındaki ilişkiler ısı haritası şeklinde görselleştirilmiş olup sonuçlar Şekil 4.9'da sunulmuştur.



Şekil 4.9 Aktif doluluk oranı ile meteorolojik değişkenler arasındaki Pearson korelasyon ısı haritası

Şekil 4.9 incelendiğinde, aktif doluluk oranı ile günlük ortalama sıcaklık arasındaki pozitif ilişkinin, ortalama nispi nem ve ortalama aktüel basınç ile olan negatif ilişkilerin renk yoğunluğu ile görsel olarak desteklendiği görülmektedir. Ayrıca günlük ortalama sıcaklık ile ortalama nispi nem arasındaki negatif ilişkinin diğer değişken çiftlerine göre daha belirgin olduğu dikkat çekmektedir. Günlük toplam yağış

ile aktif doluluk oranı arasındaki ilişkinin ise düşük düzeyde olduğu ısı haritasında da görülmektedir.

Korelasyon analizinden elde edilen bulgular doğrultusunda meteorolojik değişkenlerin aktif doluluk oranına olan katkılarının daha ayrıntılı olarak değerlendirilmesi amacıyla bir sonraki bölümde optimize edilmiş Random Forest modeli kullanılarak değişken önem analizleri gerçekleştirilmiştir.

4.3.2 Meteorolojik Değişkenlerin Önem Analizi (Feature Importance ve Permutation Importance)

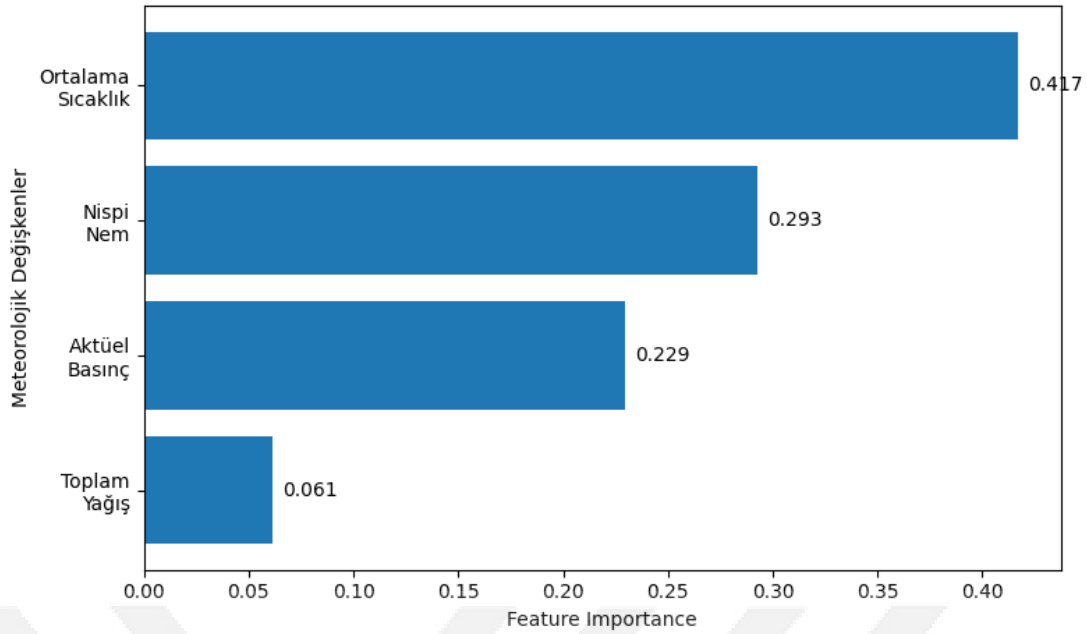
Meteorolojik değişkenlerin baraj aktif doluluk oranının tahminine olan katkılarının belirlenmesi amacıyla, yalnızca meteorolojik değişkenlerden oluşan optimize edilmiş Random Forest modeli kullanılarak değişken önem analizleri gerçekleştirilmiştir. Modelin hiperparametreleri GridSearchCV yöntemi ile optimize edilmiş ve elde edilen en iyi model kullanılarak Feature Importance ve Permutation Importance analizleri uygulanmıştır. Optimize edilmiş Random Forest modeline ait performans ölçütleri Tablo 4.15'da verilmiştir.

Tablo 4.15 Optimize edilmiş Random Forest modeline ait performans ölçütleri

Performans Ölçütü	Değer
MAE	19,837
RMSE	23,969
R^2	0,166

Tablo 4.15 incelendiğinde, yalnızca meteorolojik değişkenlerin kullanıldığı optimize edilmiş Random Forest modelinin test veri seti için ortalama mutlak hata (MAE), kök ortalama kare hata (RMSE) ve determinasyon katsayısı (R^2) değerleri sırasıyla 19,8366; 23,9688 ve 0,1655 olarak hesaplanmıştır.

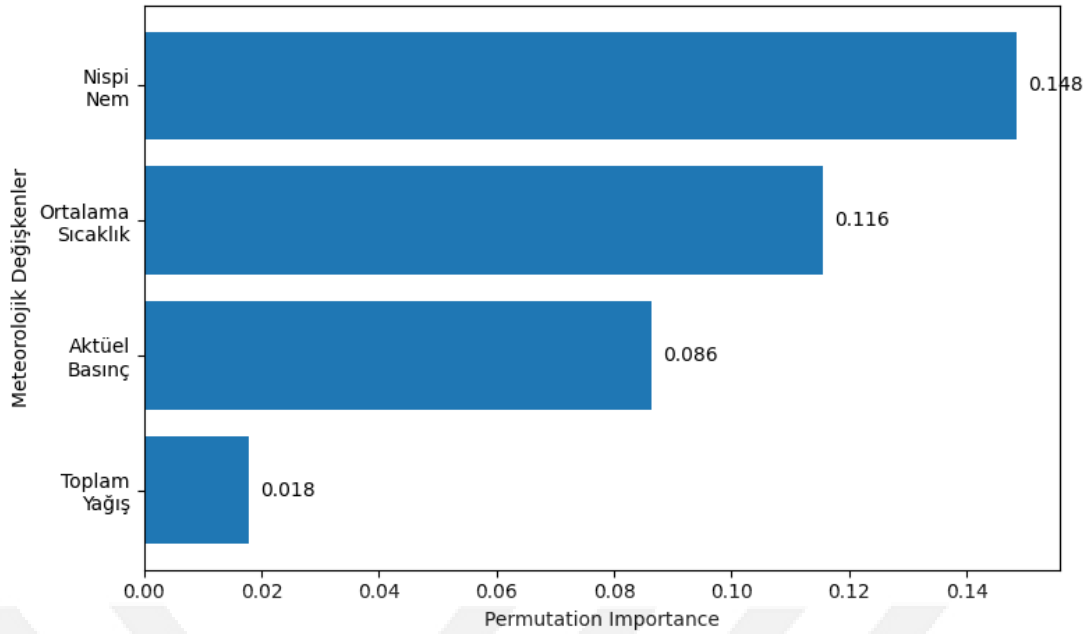
Optimize edilmiş Random Forest modelinde meteorolojik değişkenlerin önem düzeylerinin belirlenmesi amacıyla Feature Importance analizi gerçekleştirilmiş olup elde edilen sonuçlar Şekil 4.10'da verilmiştir.



Şekil 4.10 Optimize edilmiş Random Forest modeline ait Feature Importance sonuçları

Şekil 4.10 incelendiğinde, Feature Importance analizine göre en yüksek önem değerine sahip değişkenin günlük ortalama sıcaklık olduğu görülmektedir. Günlük ortalama sıcaklığı sırasıyla ortalama nispi nem, ortalama aktüel basınç ve günlük toplam yağış değişkenleri izlemektedir.

Meteorolojik değişkenlerin modele katkılarının farklı bir yaklaşımla değerlendirilmesi amacıyla Permutation Importance analizi gerçekleştirilmiş ve elde edilen sonuçlar Şekil 4.11'de sunulmuştur.



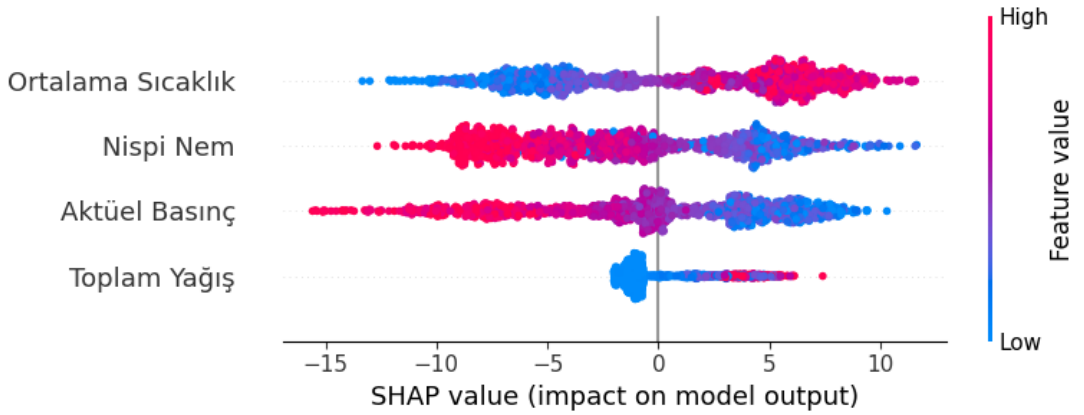
Şekil 4.11 Optimize edilmiş Random Forest modeline ait Permutation Importance sonuçları

Şekil 4.11 incelendiğinde, Permutation Importance analizine göre en yüksek önem değerine sahip değişkenin ortalama nispi nem olduğu görülmektedir. Ortalama nispi nemi sırasıyla günlük ortalama sıcaklık, ortalama aktüel basınç ve günlük toplam yağış değişkenleri takip etmektedir.

Feature Importance ve Permutation Importance analizlerinden elde edilen sonuçlar, meteorolojik değişkenlerin modele katkı düzeylerinin belirlenmesine olanak sağlamıştır.

4.3.3 SHAP Analizi

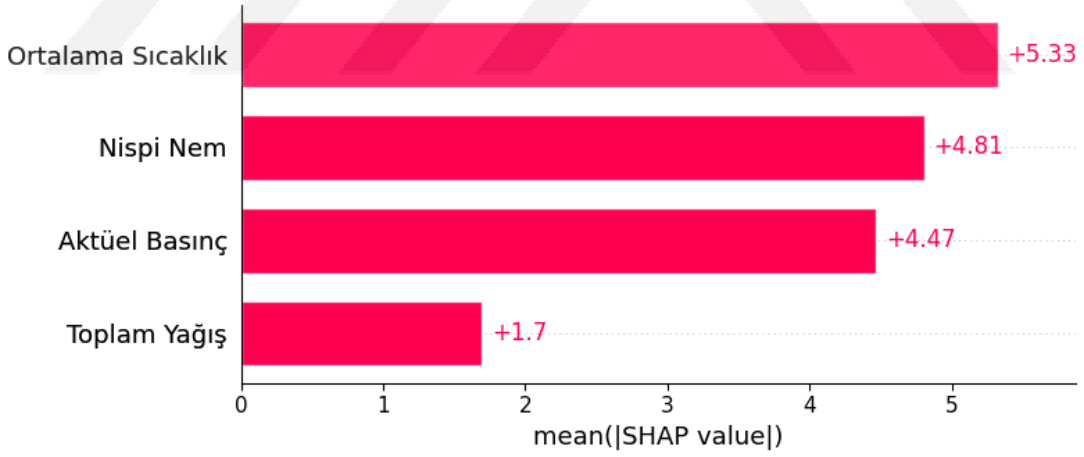
Optimize edilmiş Random Forest modelinde meteorolojik değişkenlerin tahmin sürecine olan katkılarının ve etkilerinin daha ayrıntılı olarak belirlenmesi amacıyla SHAP (SHapley Additive exPlanations) analizi gerçekleştirilmiştir. SHAP analizi, her bir değişkenin model tahminine olan katkısını hesaplayarak değişkenlerin model çıktısı üzerindeki etkilerinin görsel olarak değerlendirilmesine olanak sağlamaktadır. SHAP analizine ait özet grafik Şekil 4.12'de verilmiştir.



Şekil 4.12 Optimize edilmiş Random Forest modeline ait SHAP özet grafiği.

Şekil 4.12 incelendiğinde, SHAP analizine göre model tahminlerinde en yüksek etkiye sahip değişkenin ortalama sıcaklık olduğu görülmektedir. Ortalama sıcaklığı sırasıyla nispi nem, aktüel basınç ve toplam yağış değişkenleri izlemektedir. Ayrıca grafikte, her bir değişken için SHAP değerlerinin pozitif ve negatif yöndeki dağılımları ile değişken değerlerinin model çıktısı üzerindeki etkileri birlikte gösterilmektedir.

Meteorolojik değişkenlerin model tahminine katkılarının nicel olarak karşılaştırılabilmesi amacıyla ortalama mutlak SHAP değerleri hesaplanmış ve elde edilen sonuçlar Şekil 4.13'te sunulmuştur.



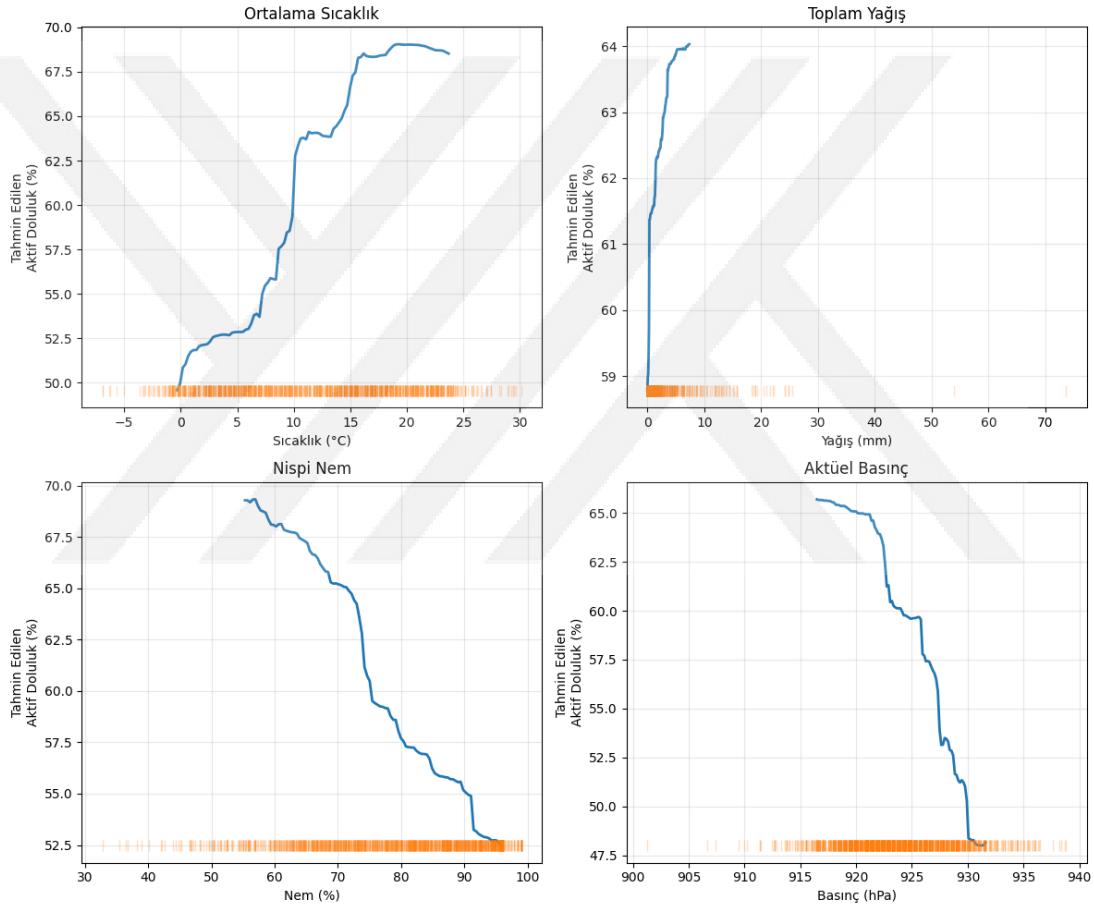
Şekil 4.13 Optimize edilmiş Random Forest modeline ait ortalama mutlak SHAP değerleri

Şekil 4.13 incelendiğinde, ortalama mutlak SHAP değerlerine göre en yüksek katkıya sahip değişkenin ortalama sıcaklık olduğu belirlenmiştir. Ortalama sıcaklığı sırasıyla nispi nem, aktüel basınç ve toplam yağış değişkenleri takip etmektedir. SHAP

analizinden elde edilen değişken önem sıralamasının, Feature Importance analizinden elde edilen sıralama ile genel olarak uyumlu olduğu görülmektedir.

4.3.4 Partial Dependence Plot (PDP) Analizi

Optimize edilmiş Random Forest modelinde meteorolojik değişkenlerin model çıktısı üzerindeki etkilerinin daha ayrıntılı olarak incelenmesi amacıyla Partial Dependence Plot (PDP) analizi gerçekleştirilmiştir. PDP analizi, her bir bağımsız değişkenin diğer değişkenler sabit tutulduğunda model tahmini üzerindeki değişimini göstermektedir. Meteorolojik değişkenlere ait kısmi bağımlılık grafikleri Şekil 4.14'te verilmiştir.



Şekil 4.14 Optimize edilmiş Random Forest modeline ait meteorolojik değişkenlerin kısmi bağımlılık (PDP) grafikleri

Şekil 4.14 incelendiğinde, günlük ortalama sıcaklık değişkenine ait PDP eğrisinin genel olarak artış eğiliminde olduğu görülmektedir. Günlük toplam yağış değişkenine ait eğri ise düşük yağış değerlerinden itibaren artış göstermekte ve yüksek yağış

değerlerinde daha yatay bir seyir izlemektedir. Buna karşılık nispi nem ve aktüel basınç değişkenlerine ait PDP eğrilerinin genel olarak azalış eğiliminde olduğu görülmektedir.

PDP grafiklerinde yer alan alt eksendeki işaretler, ilgili değişkenlere ait gözlem değerlerinin veri setindeki dağılımını göstermektedir. Böylece model tahminlerinin hangi veri aralıklarında oluşturulduğu görsel olarak değerlendirilebilmektedir.

Bu bölüm kapsamında gerçekleştirilen korelasyon analizi, Feature Importance, Permutation Importance, SHAP ve Partial Dependence Plot analizleri birlikte değerlendirildiğinde, meteorolojik değişkenlerin baraj aktif doluluk oranının tahmininde farklı düzeylerde katkı sağladığı görülmektedir. Farklı analiz yöntemleri genel olarak benzer değişken önem sıralamaları ortaya koyarken, değişkenlerin model çıktısı üzerindeki etkileri ve değişim eğilimleri de görsel olarak belirlenmiştir.

5. SONUÇLAR VE TARTIŞMA

5.1 Makine Öğrenmesi Modellerinin Performanslarının Değerlendirilmesi

Bu araştırma, Germeçtepe Barajı'nın günlük aktif doluluk oranının tahmininde farklı makine öğrenmesi algoritmalarının yüksek doğrulukta sonuçlar üretebildiğini göstermiştir. Çalışmanın en önemli çıktısı, hiperparametre optimizasyonu uygulanmış Random Forest algoritmasının diğer modeller arasında en başarılı performansı göstermesi ve rezervuar aktif doluluk oranındaki zamansal değişimleri güvenilir biçimde temsil edebilmesidir. Bu sonuç, hidrolojik süreçlerin doğrusal olmayan yapısının ağaç tabanlı topluluk öğrenmesi algoritmaları tarafından başarılı şekilde modellenebildiğini ortaya koymaktadır.

Random Forest algoritmasının öne çıkması, yalnızca elde edilen performans ölçütleriyle değil, algoritmanın yapısal özellikleriyle de açıklanabilir. Bootstrap örnekleme ve rastgele değişken seçimine dayanan topluluk öğrenmesi yaklaşımı, tek karar ağaçlarında görülebilen aşırı uyum ve yüksek varyans problemlerini azaltarak daha kararlı tahminler üretebilmektedir (Breiman, 2001). Çalışmada hiperparametre optimizasyonu sonrasında model doğruluğunun artması da uygun model ayarlarının hidrolojik tahminlerde önemli bir rol oynadığını göstermektedir. Bu bulgu, Azad vd. (2025), Pedregosa vd. (2011) ve Wee vd. (2021) tarafından bildirilen sonuçlarla uyumludur.

Araştırmada dikkat çeken bir diğer sonuç, Random Forest ve XGBoost algoritmalarının birbirine oldukça yakın performans göstermesine rağmen Random Forest modelinin daha düşük hata değerleri üretmesidir. Bu durum, kullanılan veri setinin uzun dönemli, düzenli ve yüksek zamansal sürekliliğe sahip olmasının yanı sıra modele dahil edilen Lag-1 değişkeninin zaman serisindeki güçlü otokorelasyonu başarıyla yansıtmasıyla ilişkilendirilebilir. Benzer şekilde Doğrusal Regresyon modelinin beklenenden yüksek performans göstermesi de rezervuar doluluk oranının günlük ölçekte yüksek süreklilik göstermesinin bir sonucudur. Bununla birlikte, yalnızca geçmiş gözlemlere dayalı modellerin olağan dışı hidrolojik olayları temsil etmekte yetersiz kalabileceği dikkate alındığında, meteorolojik değişkenlerin modele

dahil edilmesi daha gerçekçi ve güvenilir tahminlerin elde edilmesine katkı sağlamaktadır.

Elde edilen bulgular, Random Forest algoritmasının rezervuar modelleme çalışmalarında başarılı sonuçlar verdiğini bildiren Ouma vd. (2022), Rehamnia ve Mahdavi-Meymand (2025) ile Yang vd. (2017) uyumludur. Buna karşılık bazı araştırmalarda LSTM, MLP veya NARX gibi farklı algoritmaların daha yüksek performans gösterebildiği rapor edilmiştir (García-Feal vd., 2022; Özdoğan-Sarıkoç vd., 2023; Pathan vd., 2024). Bu farklılıklar, tek bir makine öğrenmesi algoritmasının tüm rezervuar sistemleri için en uygun yöntem olarak değerlendirilemeyeceğini göstermektedir. Model başarısı; veri kalitesi, gözlem süresi, giriş değişkenleri, hidrolojik özellikler ve hiperparametre optimizasyonu gibi birçok faktörün ortak etkisi altında şekillenmektedir. Dolayısıyla model seçiminde çalışma alanına özgü veri yapısının dikkate alınması gerekmektedir.

Bu çalışmanın önemli katkılarından biri, farklı makine öğrenmesi algoritmalarını aynı veri seti, aynı eğitim-test bölünmesi ve aynı performans ölçütleri kullanılarak sistematik biçimde karşılaştırmasıdır. Ayrıca hiperparametre optimizasyonunun model performansına etkisi ortaya konulmuş ve sonraki bölümde gerçekleştirilen açıklanabilir yapay zekâ analizleri ile yalnızca en başarılı model belirlenmemiş, aynı zamanda model karar mekanizmasının da yorumlanmasına olanak sağlanmıştır. Bu yönüyle çalışma, rezervuar aktif doluluk oranının tahmininde yüksek doğruluğun yanı sıra yorumlanabilirliği de ön plana çıkararak literatürdeki benzer araştırmalara önemli bir katkı sunmaktadır.

5.2 Meteorolojik Değişkenlerin Baraj Doluluk Oranı Üzerindeki Etkilerinin Değerlendirilmesi

Bu çalışmanın temel sonuçlarından biri, Germeçtepe Barajı'nın günlük aktif doluluk oranının yalnızca aynı güne ait meteorolojik değişkenlerle yeterli düzeyde açıklanamayacağının ortaya konulmasıdır. Meteorolojik değişkenlerin modele belirli ölçüde katkı sağladığı görülmekle birlikte, rezervuar doluluğunun esas olarak geçmiş depolama koşulları, havza ölçeğindeki hidrolojik süreçler ve işletmeye ilişkin

faktörlerin bütünleşik etkisi altında şekillendiği anlaşılmıştır. Bu sonuç, rezervuarların atmosferik koşullara doğrudan ve anlık tepki veren sistemler olmadığını; yağışın akışa dönüşmesi, yeraltı suyu katkısı, su çekimleri ve işletme kararları gibi süreçlerin zaman gecikmeli etkiler taşıdığını göstermektedir. Literatürde de yalnızca meteorolojik değişkenlerin kullanıldığı modellerin sınırlı performans gösterdiği, geçmiş su seviyesi, rezervuar hacmi ve debi gibi hidrolojik değişkenlerin modele eklenmesiyle tahmin başarısının arttığı bildirilmektedir (Azad vd., 2025; Büyükyıldız vd., 2014; Ouma vd., 2022; Pathan vd., 2024).

Meteorolojik değişkenler arasında günlük ortalama sıcaklığın öne çıkması, sıcaklığın rezervuar sistemini etkileyen çok sayıda hidrolojik sürecin dolaylı bir göstergesi olmasıyla açıklanabilir. Sıcaklık; buharlaşma, evapotranspirasyon, kar erimesi, toprak nemi ve akış oluşumu üzerinde etkili olduğundan yalnızca atmosferik bir değişken değil, aynı zamanda havza tepkisini temsil eden bileşik bir gösterge niteliğindedir. Özellikle karasal iklim özelliklerinin etkili olduğu Kastamonu çevresinde ilkbahar sıcaklıklarının kar erimesi ve rezervuara ulaşan akımlar, yaz sıcaklıklarının ise buharlaşma kayıpları üzerindeki etkisi önem taşımaktadır. Bu değerlendirme, sıcaklık ve diğer meteorolojik değişkenlerin rezervuar su seviyesi üzerinde belirleyici etkiler oluşturabildiğini bildiren Anwar (2023) ve Ouma vd. (2022) ile uyumludur.

Çalışmada kullanılan farklı açıklanabilir yapay zekâ yöntemlerinin değişken önem sıralamalarında tamamen aynı sonuçları üretmemesi, model yorumlamasında tek bir yönteme bağlı kalınmaması gerektiğini göstermiştir. Feature Importance ve SHAP analizlerinde sıcaklık öne çıkarken, Permutation Importance analizinde nispi nemin daha yüksek önem kazanması, her yöntemin değişken katkısını farklı bir hesaplama yaklaşımıyla değerlendirmesinden kaynaklanmaktadır. Karar ağacı bölünmelerine dayanan Feature Importance, model doğruluğundaki değişimi esas alan Permutation Importance ve gözlem bazında katkı hesaplayan SHAP yöntemi birlikte değerlendirildiğinde, model karar mekanizmasının daha kapsamlı ve güvenilir biçimde yorumlanması mümkün olmuştur. Bu sonuç, açıklanabilir makine öğrenmesi uygulamalarında tamamlayıcı yöntemlerin birlikte kullanılmasının önemine işaret etmektedir (Molnar, 2022).

SHAP analizinin çalışmaya sağladığı temel katkı, meteorolojik değişkenlerin yalnızca göreli önemini değil, tahminleri hangi yönde ve hangi büyüklükte etkilediğini de ortaya koymasıdır. Geleneksel korelasyon analizi değişkenler arasındaki genel doğrusal ilişkiyi gösterirken, SHAP yaklaşımı doğrusal olmayan model yapısında her bir değişkenin tahmine yaptığı katkıyı gözlem bazında değerlendirebilmektedir. Böylece sıcaklık, yağış, nispi nem ve aktüel basınç değişkenlerinin model üzerindeki etkilerinin tek yönlü ve sabit olmadığı; farklı koşullar altında değişebildiği belirlenmiştir. Bu özellik, karmaşık makine öğrenmesi modellerinin karar süreçlerinin açıklanmasında SHAP yönteminin sağladığı temel üstünlüklerden biridir (Lundberg ve Lee, 2017).

Partial Dependence Plot (PDP) analizleri de meteorolojik değişkenlerin aktif doluluk oranı üzerindeki etkilerinin doğrusal ve sabit olmadığını göstermiştir. Değişkenlerin belirli değer aralıklarında farklı yön ve büyüklükte etkiler oluşturması, rezervuar sisteminin meteorolojik koşullara verdiği tepkinin eşik değerleri ve mevsimsel süreçlerle bağlantılı olduğunu düşündürmektedir. Bu sonuç, klasik doğrusal modellerin her koşulda sabit bir ilişki varsayımına karşılık, Random Forest gibi ağaç tabanlı algoritmaların karmaşık ve doğrusal olmayan hidrolojik ilişkileri daha esnek biçimde temsil edebildiğini ortaya koymaktadır.

Çalışmanın özgün katkılarından biri, korelasyon analizi ile Feature Importance, Permutation Importance, SHAP ve PDP yöntemlerinin aynı değerlendirme çerçevesinde birlikte kullanılmasıdır. Rezervuar doluluk tahmini çalışmalarının önemli bir kısmında model başarısı performans ölçütleri üzerinden değerlendirilmekte, modelin hangi değişkenlerden ve hangi mekanizmalarla etkilendiği sınırlı düzeyde açıklanmaktadır. Bu çalışmada ise yalnızca yüksek doğruluk sağlayan bir model geliştirilmemiş; model tahminlerinin arkasındaki değişken katkıları, etki yönleri ve doğrusal olmayan ilişkiler de ortaya konulmuştur. Bu yönüyle araştırma, açıklanabilir yapay zekâ yaklaşımlarının rezervuar modellemesine uygulanmasına metodolojik bir katkı sağlamaktadır (Azad vd., 2025; Lundberg ve Lee, 2017; Molnar, 2022).

Genel olarak değerlendirildiğinde, meteorolojik değişkenler rezervuar aktif doluluk oranının değişimini açıklamada tek başına yeterli olmamakla birlikte sistem

davranışının yorumlanmasında önemli tamamlayıcı bilgiler sunmaktadır. Çalışmanın en önemli çıktısı, rezervuar doluluğunun meteorolojik koşullar, geçmiş doluluk bilgisi, havza hidrolojisi ve işletme süreçlerinin ortak etkisiyle şekillenen çok bileşenli bir yapı olduğunun gösterilmesidir. Açıklanabilir yapay zekâ analizleri sayesinde bu karmaşık yapı yalnızca tahmin edilmemiş, aynı zamanda bilimsel olarak yorumlanabilir hâle getirilmiştir. Bu sonuç, geliştirilen modelin ileride rezervuar işletmesi, kuraklık izleme ve su kaynakları yönetimine yönelik karar destek sistemlerinde kullanılabilmesi açısından önemli bir temel oluşturmaktadır.

6. ÖNERİLER

Gelecekte gerçekleştirilecek çalışmalarda, rezervuar aktif doluluk oranını etkileyen süreçlerin daha kapsamlı biçimde temsil edilebilmesi için veri altyapısının geliştirilmesi önerilmektedir. Bu kapsamda yalnızca meteorolojik değişkenler ve geçmiş doluluk bilgisiyle sınırlı kalınmamalı; rezervuara gelen giriş debisi, mansaba bırakılan su miktarı, sulama amaçlı su çekimleri, buharlaşma, kar örtüsü, yeraltı suyu katkısı, rezervuar işletme kayıtları ve havzadaki arazi kullanım değişimleri de düzenli olarak ölçülerek veri setine dahil edilmelidir. Günlük verilerin yanında saatlik veya daha yüksek zamansal çözünürlüğe sahip gözlemlerin kullanılması, özellikle ani yağış, taşkın ve hızlı doluluk değişimlerinin daha başarılı biçimde modellenmesine katkı sağlayacaktır. Ayrıca farklı iklim ve hidrolojik özelliklere sahip barajlardan elde edilecek uzun dönemli ve kesintisiz veri serileri, geliştirilen modellerin genellenebilirliğinin sınanmasına olanak sağlayacaktır.

Model geliştirme açısından, farklı gecikme sürelerini temsil eden Lag değişkenlerinin, hareketli ortalamaların ve mevsimsel göstergelerin birlikte değerlendirilmesi önerilmektedir. Random Forest ve XGBoost gibi topluluk öğrenmesi algoritmalarına ek olarak LSTM, GRU ve Transformer tabanlı derin öğrenme modelleri ile fiziksel süreç bilgisi içeren hibrit veya fizik bilgili makine öğrenmesi yaklaşımlarının uygulanması, farklı tahmin ufuklarında model performansının geliştirilmesine katkı sağlayabilir. Bunun yanında Grid Search dışındaki Bayesian Optimization ve benzeri hiperparametre optimizasyon yöntemleri kullanılmalı; model başarısı yalnızca MAE, RMSE ve R^2 ölçütleriyle değil, belirsizlik analizi, ekstrem dönem performansı, zamansal çapraz doğrulama ve farklı barajlarda dış doğrulama yoluyla da değerlendirilmelidir. Model yorumlanabilirliğinin artırılması amacıyla SHAP ve PDP analizlerinin yanı sıra ALE, LIME ve karşı-olgusal açıklama yöntemlerinin kullanılması da yararlı olacaktır.

Bu çalışmada geliştirilen optimize edilmiş Random Forest modeli, gerekli veri akışı ve teknik altyapının sağlanması durumunda Germeçtepe Barajı işletmesinde operasyonel bir karar destek aracı olarak kullanılabilir. Modelin günlük olarak

güncellenen meteorolojik ve hidrolojik verilerle çalıştırılması, kısa dönemli aktif doluluk tahminlerinin üretilmesine ve farklı işletme senaryolarının önceden değerlendirilmesine olanak sağlayabilir. Böyle bir sistem; kurak dönemlerde doluluk oranındaki olası azalışların erken belirlenmesi, sulama suyu tahsislerinin planlanması, kritik doluluk eşiklerine göre su kullanım önceliklerinin düzenlenmesi ve olası su açığına yönelik tedbirlerin önceden alınması amacıyla kullanılabilir. Benzer şekilde yüksek yağış ve hızlı doluluk artışı beklenen dönemlerde rezervuar işletme kararlarının desteklenmesine katkı sağlayabilir. Modelin gerçek zamanlı ölçüm sistemleri, meteorolojik tahminler ve iklim değişikliği senaryolarıyla bütünleştirilmesi durumunda, geleceğe yönelik kuraklık ve su arzı değerlendirmeleri üreten operasyonel bir erken uyarı ve karar destek sistemine dönüştürülmesi mümkün olacaktır.

KAYNAKLAR

Almubaidin, M. A., Ahmed, A. N., Winston, C. A. A., & El-Shafie, A. (2023). Application of machine learning for daily forecasting dam water levels. *Tikrit Journal of Engineering Sciences,* 30(4), 74–87. https://doi.org/10.25130/tjes.30.4.9

Anwar, A. (2023). Doğancı Barajı'nda meteorolojik veriler ile su kalitesi ve su seviyesi arasındaki ilişkinin yapay sinir ağları ile modellenmesi. Yüksek Lisans Tezi, *Bursa Uludağ Üniversitesi Fen Bilimleri Enstitüsü*. Bursa.

Arslan, H., Üneş, F., Demirci, M., Taşar, B., & Yılmaz, A. (2020). Keban Baraj Gölü seviye değişiminin ANFİS ve destek vektör makineleri ile tahmini. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi,* 3(2), 71–77.

Azad, A. S., Islam, N., Nabi, M. N., Khurshid, H., & Siddique, M. A. (2025). Developments and trends in water level forecasting using machine learning models—A review. IEEE Access, 13, 63048–63065. https://doi.org/10.1109/ACCESS.2025.3557910

Breiman, L. (2001). Random forests. *Machine Learning,* 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees.* Wadsworth.

Büyükyıldız, M., Kumcu, S. Y., & Karadağ, A. A. (2014). Estimation of monthly water level fluctuations of Lake Beyşehir using artificial intelligence methods. *Water Resources Management,* 28(13), 4707–4723. https://doi.org/10.1007/s11269-014-0784-2

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific Model Development,* 7(3), 1247-1250. https://doi.org/10.5194/gmd-7-1247-2014

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792. https://doi.org/10.1890/07-0539.1

Damla, Y., Temiz, T., & Keskin, E. (2020). Yapay sinir ağı kullanılarak su seviyesinin tahmin edilmesi: Yalova Gökçe Barajı örneği. *Kırklareli University Journal of Engineering and Science,* 6(1), 32–49. https://doi.org/10.34186/klujes.754845

Das, M., Ghosh, S. K., Chowdary, V. M., Saikrishnaveni, A., & Sharma, R. K. (2016). A probabilistic nonlinear model for forecasting daily water level in reservoir. *Water Resources Management*, 30(9), 3107-3122. https://doi.org/10.1007/s11269-016-1334-6

DSİ (2023). *Toprak ve su kaynakları*. Tarım ve Orman Bakanlığı. https://www.dsi.gov.tr/sayfa/detay/754

DSİ (2025). *Germeçtepe Barajı teknik özellikleri ve işletme bilgileri.* T.C. Tarım ve Orman Bakanlığı, Devlet Su İşleri Genel Müdürlüğü. https://bolge23.dsi.gov.tr/Sayfa/Detay/1235?utm_source=chatgpt.com

Egawa, T., Suzuki, K., Ichikawa, Y., Iizaka, T., Matsui, T., & Shikagawa, Y. (2011). A water flow forecasting for dam using neural networks and regression models. In *2011 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 3048–3053). IEEE. https://doi.org/10.1109/ICSMC.2011.6084143

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177), 1–81.

Food and Agriculture Organization of the United Nations. (2021). *The state of the world's land and water resources for food and agriculture: Systems at breaking point (SOLAW 2021).* Food and Agriculture Organization of the United Nations. https://doi.org/10.4060/cb7654en

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

García-Feal, O., González-Cao, J., Fernández-Nóvoa, D., Astray Dopazo, G., & Gómez-Gesteira, M. (2022). Comparison of machine learning techniques for reservoir outflow forecasting. *Natural Hazards and Earth System Sciences*, 22(12), 3859–3874. https://doi.org/10.5194/nhess-22-3859-2022

Güneş Şen, S. (2025). Machine learning-based water level forecast in a dam reservoir: Karaçomak Dam in the Kızılırmak Basin, Türkiye. *Sustainability,* 17(18), 8378. https://doi.org/10.3390/su17188378

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature,* 585(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hipni, A., El-shafie, A., Najah, A., Karim, O. A., Hussain, A., & Mukhlisin, M. (2013). Daily forecasting of dam water levels: comparing a support vector machine (SVM) model with adaptive neuro fuzzy inference system (ANFIS). *Water resources management,* 27(10), 3803-3823. https://doi.org/10.1007/s11269-013-0382-4

Ibañez, S. C., Dajac, C. V. G., Liponhay, M. P., Legara, E. F. T., Esteban, J. M. H., & Monterola, C. P. (2022). Forecasting Reservoir Water Levels Using Deep Neural Networks: A Case Study of Angat Dam in the Philippines. *Water*, 14(1), 34. https://doi.org/10.3390/w14010034

Intergovernmental Panel on Climate Change. (2023). *Climate change 2021: The physical science basis. Working Group I contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press. https://doi.org/10.1017/9781009157896

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science,* 349(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kisi, O., Shiri, J., & Tombul, M. (2012). Modeling lake level fluctuations using different machine learning techniques. *Journal of Hydrology,* 438–439*,* 65–73. https://doi.org/10.1016/j.jhydrol.2012.03.006

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)* (Vol. 2, pp. 1137–1145). International Joint Conferences on Artificial Intelligence Organization.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill/Irwin.

Liang, C., Li, H., Lei, M., & Du, Q. (2018). Dongting Lake water level forecast and its relationship with the Three Gorges Dam based on a long short-term memory network. *Water*, 10(10), 1389. https://doi.org/10.3390/w10101389

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News,* 2(3), 18-22.

Loh, W. Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery,* 1(1), 14–23.

Lundberg, S.M. & Lee, S.-I. (2017) A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc.

McKinney, W. (2010). Data structures for statistical computing in Python. *Scipy*, 445(1), 51-56.

Molnar, C. (2022). *Interpretable machine learning: A guide for making black box models explainable* (2nd ed.). Leanpub. https://christophm.github.io/interpretable-ml-book

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). Wiley.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900. doi: 10.13031/2013.23153

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water, 10*(11), 1536. https://doi.org/10.3390/w10111536

Nwobi-Okoye, C. C., & Igboanugo, A. C. (2013). Application of artificial neural networks for water level prediction. *Applied Mechanics and Materials, 204–210*, 2158–2164. https://doi.org/10.4028/www.scientific.net/AMR.204-210.2158

Ouma, Y. O., Moalafhi, D. B., Anderson, G., Nkwae, B., Odirile, P., Parida, B. P., & Qi, J. (2022). Dam water level prediction using Vector AutoRegression, Random Forest Regression and MLP-ANN models based on land-use and climate factors. *Sustainability*, 14(22), 14934. https://doi.org/10.3390/su142214934

Özdoğan-Sarıkoç, G., Sarıkoç, M., Çelik, M., & Dadaser-Çelik, F. (2023). Reservoir volume forecasting using artificial intelligence-based models: Artificial neural networks, support vector Regresyon, and long short-term memory. *Journal of Hydrology,* 616, 128766. https://doi.org/10.1016/j.jhydrol.2022.128766

Pathan, A. I., Sidek, L. B. M., Basri, H. B., Hassan, M. Y., Khebir, M. I. A. B., Omar, S. M. B. A., Khambali, M. H. B. M., Torres, A. M., & Ahmed, A. N. (2024). Comparative assessment of rainfall-based water level prediction using machine learning techniques. *Ain Shams Engineering Journal,* 15, 102854. https://doi.org/10.1016/j.asej.2024.102854

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 31). Curran Associates, Inc.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.

Rehamnia, I., & Mahdavi-Meymand, A. (2025). Advancing reservoir water level predictions: Evaluating conventional, ensemble and integrated swarm machine learning approaches. *Water Resources Management,* 39, 779–794. https://doi.org/10.1007/s11269-024-03990-x

Sapitang, M., Ahmed, A. N., El-Shafie, A., Sidek, L. M., & Huang, Y. F. (2020). Machine learning application in reservoir water level forecasting for sustainable water resources management. *Stochastic Environmental Research and Risk Assessment,* 34, 1457–1473.

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research,* 54(11), 8558–8593. https://doi.org/10.1029/2018WR022643

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, M. Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology,* 82(12), 2635–2670. https://doi.org/10.2166/wst.2020.369

Soria-Lopez, A., Sobrido-Pouso, C., Mejuto, J. C., & Astray, G. (2023). Assessment of different machine learning methods for reservoir outflow forecasting. *Water,* 15(19), 3380. https://doi.org/10.3390/w15193380

Stojanovic, B., Milivojevic, M., Ivanovic, M., Milivojevic, N., & Divac, D. (2013). Adaptive system for dam behavior modeling based on linear regression and genetic algorithms. *Advances in Engineering Software*, 65, 182–190. https://doi.org/10.1016/j.advengsoft.2013.06.019

Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A brief review of random forests for water scientists and practitioners and their recent history in water resources. *Water*, 11(5), 910. https://doi.org/10.3390/w11050910

UNESCO World Water Assessment Programme. (2023). *The United Nations World Water Development Report 2023: Partnerships and cooperation for water*. UNESCO. https://doi.org/10.54679/QQJG4824

VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data.* O'Reilly Media, Inc.

Wang, Q., Peng, Y., Li, M., Zhang, J., Sun, Y., Ding, W., & Zhu, F. (2025). Dynamic Control of the Flood Limited Water Level in Reservoir Considering the Uncertainty Error Domain of Forecast Information. *Water Resources Management,* 39(14), 7491–7514.

Wee, S. Y., Wahab, A. K. A., & Shahid, S. (2021). Reservoir water level forecasting using machine learning models: A review. *International Journal of Hydrology Science and Technology,* 12(4), 347–370.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82. https://doi.org/10.3354/cr030079

Wong, J. K., Alraih, M., Ahmed, A. N., Fai, C. M., El-Shafie, A., & El-Shafie, A. M. (2019). Daily forecasting of dam water levels using machine learning. *International Journal of Civil Engineering and Technology, 10*(6), 314–323.

Yang, T., Asanjan, A. A., Welles, E., Gao, X., Sorooshian, S., & Liu, X. (2017). Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. *Water Resources Research*, 53(4), 2786–2812. https://doi.org/10.1002/2017WR020482

Yang, Y., Zhang, M., Sun, Z., Han, J., & Wang, J. (2018). The relationship between water level change and river channel geometry adjustment in the downstream of the Three Gorges Dam. *Journal of Geographical Sciences*, 28(12), 1975–1993. https://doi.org/10.1007/s11442-018-1575-9

Yılmaz, F., Ulusoy, İ., & Toros, H. (2020). İstanbul Baraj Doluluk Oranlarının Zamansal İncelenmesi ve Çözüm Önerileri. *Journal of Research in Atmospheric Science,* 2(2), 51-55.

Yi, S., & Yi, J. (2024). Reservoir-based flood forecasting and warning: Deep learning versus machine learning. *Applied Water Science,* 14, 237. https://doi.org/10.1007/s13201-024-02298-w

Zhou, F., & Li, L. (2021). Forecasting reservoir inflow via recurrent neural ODEs. *Proceedings of the AAAI Conference on Artificial Intelligence,* 35(17), 15025–15032. https://doi.org/10.1609/aaai.v35i17.17763