



**T.C.
FIRAT ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
YAPAY ZEKÂ ve VERİ MÜHENDİSLİĞİ BÖLÜMÜ**

GENEL PROJE RAPORU

PROJE ADI: SUSAMAYIZ

Grup Üyeleri
Burcu YILDIRIM
Recep BÜLBÜL
Merve ÇALOĞLU
Yakup Kadri İLHAN

**Haziran 2024
Elazığ**

1.Giriş

Günümüzde, su kaynaklarının sürdürülebilir bir şekilde yönetilmesi ve içilebilir hale getirilmesi giderek daha fazla önem kazanmaktadır. Su, yaşamın temel kaynağı olmasının yanı sıra endüstri, tarım, enerji üretimi ve birçok diğer sektör için vazgeçilmez bir unsurdur. Ancak, artan nüfus ve endüstrileşme ile birlikte su kaynaklarının azalması ve kirlenmesi gibi sorunlarla karşı karşıyayız.

Bu bağlamda, bu araştırma projesi su kaynaklarının daha etkin bir şekilde kullanılması ve içilebilir hale getirilmesi amacıyla gerçekleştirilmiştir. Proje kapsamında, zaman içinde artan su ihtiyacına rağmen bu ihtiyaca yeterince cevap verilememesi sorununa odaklanılmıştır. Su kaynaklarının verimli kullanılması ve içilebilir hale getirilmesi için gereken süreçler incelenmiş ve çözüm önerileri sunulmuştur.

Proje, bir dizi adımdan oluşmuştur. İlk olarak, su bileşenlerinin suyun içilebilirliği ile olan korelasyonları incelenmiş ve bu korelasyonların görselleştirilmiş biçimleri sunulmuştur. Ardından, veri setleri üzerinde makine öğrenimi modelleri uygulanarak en performanslı model belirlenmiş ve modelin neden bu kadar etkili olduğu açıklanmıştır.

Projede ayrıca, hava kalitesi ile su kirliliği arasındaki ilişki incelenmiş ve bu ilişki bilimsel bulgular ile desteklenmiştir. Son olarak, projenin sonuçları ve elde edilen bulgular bir makale halinde sunulmuştur. Bu makale, su kaynaklarının etkin kullanımı ve içilebilir hale getirilmesi konularında farkındalık yaratmayı ve çözüm önerileri sunmayı hedeflemektedir.

2. Gereç ve Yöntem

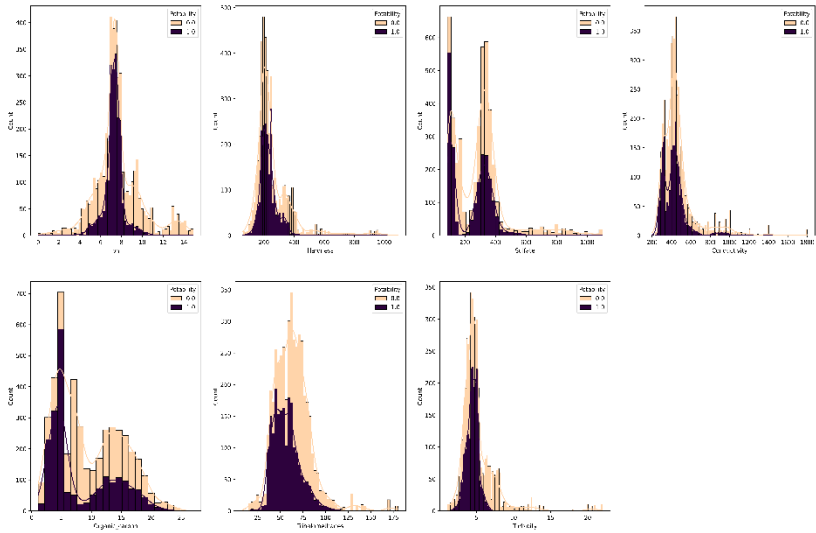
Projenin ana adımları ve kullanılan yöntemler şu şekildedir:

2.1. Veri Toplama ve Hazırlık

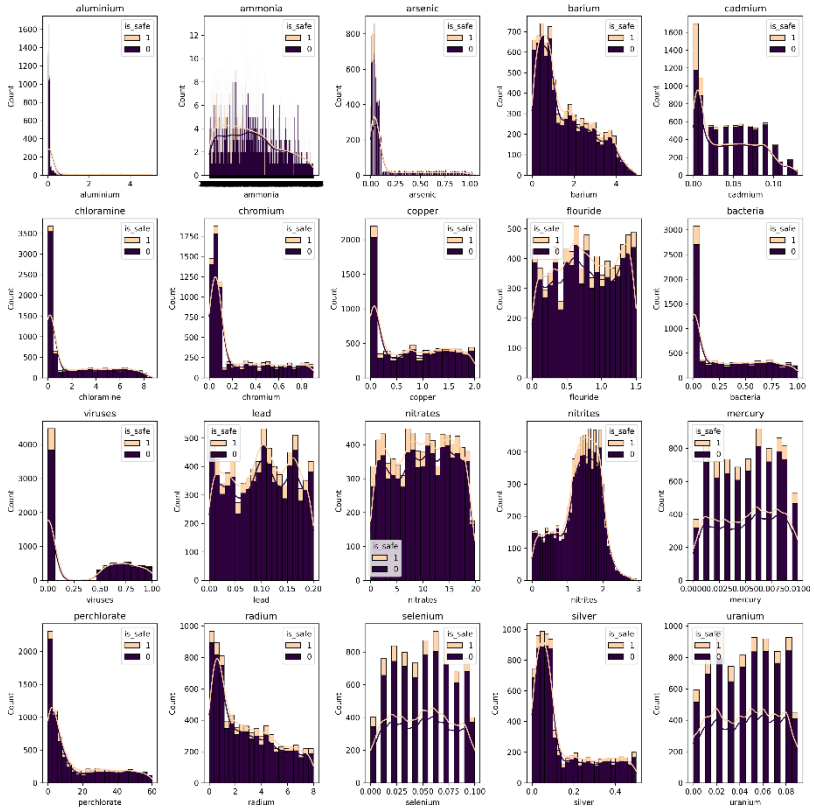
Su bileşenleri ve hava kalitesi ile ilgili veriler temin edilmiş ve bu veriler analiz için hazırlanmıştır.

2.2. Veri Görselleştirme

Elde edilen veriler, uygun veri görselleştirme teknikleri kullanılarak görselleştirilmiştir.



Şekil 2.1.Water potability veri seti parametreler arası ilişki grafiği



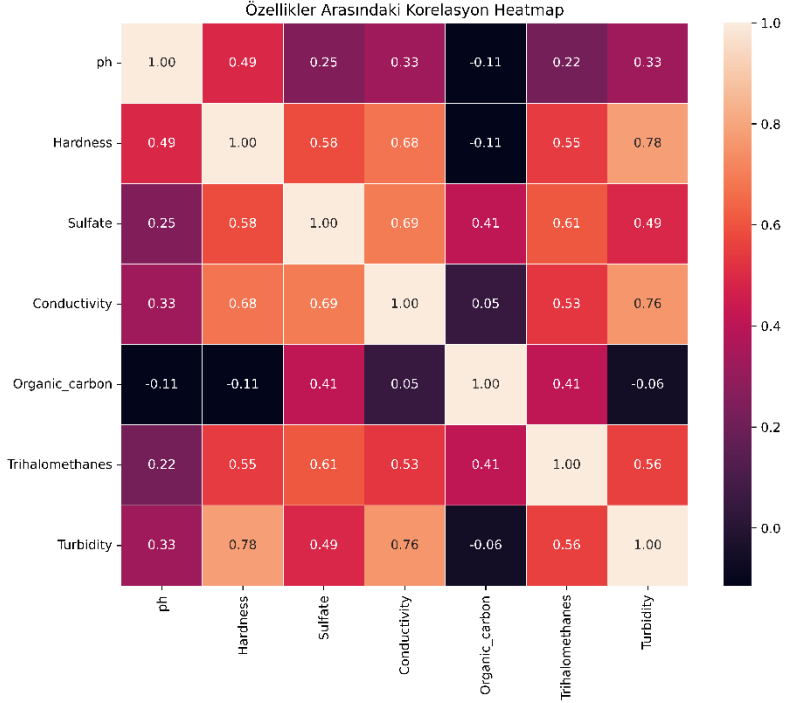
Şekil 2.2.Water quality veri seti parametreler arası ilişki grafiği

2.3. Korelasyon Analizi

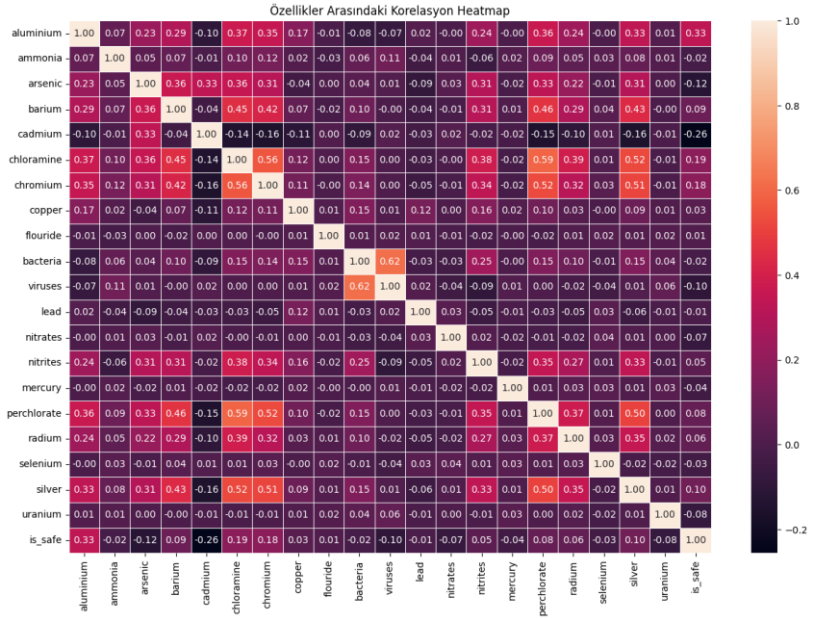
Suyun içilebilirliğini etkileyen temel bileşenler ve hava kalitesi değişkenleri arasındaki korelasyonlar hesaplanmıştır. Korelasyon katsayıları belirlenmiş ve bu katsayıların anlamlılığı test edilmiştir.

2.4 Korelasyonların Görselleştirilmesi

Korelasyon matrisi ve heatmap kullanılarak elde edilen korelasyon sonuçları görselleştirilmiştir.



Şekil 2.3. Water potability veri seti özellikleri arasındaki korelasyon ısı haritası



Şekil 2.4.Water quality veri seti özellikleri arasındaki korelasyon ısı haritası

2.5.Korelasyon Analizlerinin Değerlendirilmesi

Korelasyon sonuçları yorumlanmış ve su içilebilirliği üzerinde en fazla etkiye sahip değişkenler belirlenmiştir. Güçlü korelasyonlar, daha detaylı analizler için temel oluşturmuştur.

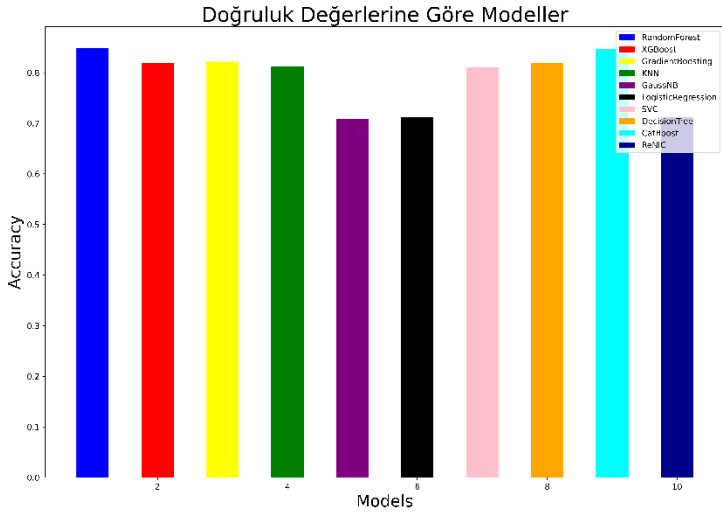
2.6. Makine Öğrenimi Algoritmalarının Seçimi

Proje hedeflerimize ve veri setimize en uygun makine öğrenimi algoritmalarını belirlemek için çeşitli algoritmalar test edilmiştir. Test edilen algoritmalar şunlardır:

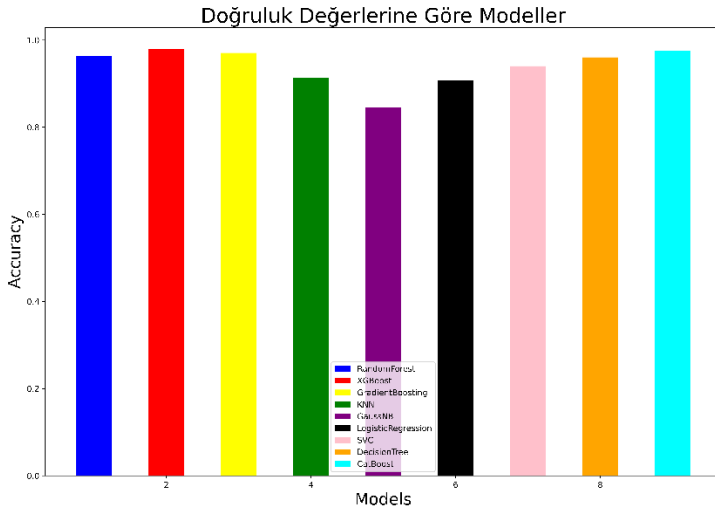
- Random Forest
- XGBoost
- Gradient Boosting
- K-En Yakın Komşu (KNN)
- Gauss Naive Bayes (GaussNB)
- Logistic Regression
- Destek Vektör Makineleri (SVC)
- Karar Ağaçları (Decision Tree)
- CatBoost

Veri setimizin özellikleri ve hedef değişkeni belirlenmiş, ardından veri setimiz eğitim ve test veri kümelerine ayrılmıştır. Seçilen makine öğrenimi algoritmaları eğitilmiş ve her bir algoritma eğitim veri kümesi üzerinde uyumlaştırılmıştır. Eğitilen modeller test veri kümesi üzerinde değerlendirilmiş; doğruluk, hassasiyet, özgünlük, F1 skoru gibi performans metrikleri hesaplanmış ve karşılaştırılmıştır.

Ayrıca, Water Potability veri seti için kendi geliştirdiğimiz model olan **Renic** eğitilmiş ve bu model Gauss Naive Bayes (GaussNB) modelini geçecek doğruluk oranı sağlamıştır.



Şekil 2.5.Water potability veri seti makine öğrenim model doğruluk sütun grafiği



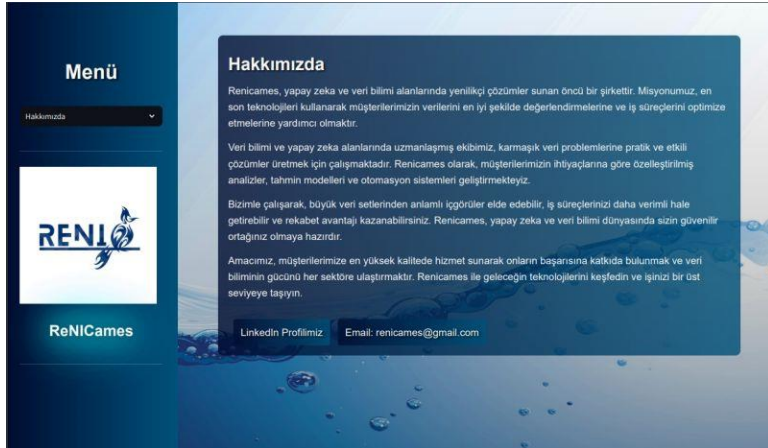
Şekil 2.6.Water quality veri seti makine öğrenim model doğruluk sütun grafiği

2.7.Web Sitesi Kurulumu

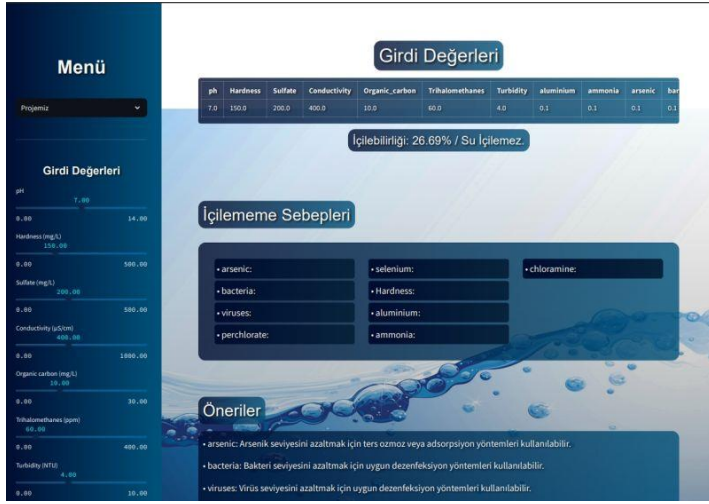
Projenin sonuçlarını ve bulgularını paylaşmak için bir web sitesi kurulmuş ve bu sitede proje ile ilgili tüm veriler ve analizler sunulmuştur.



Şekil 2.7.Web sitesi ana sayfa görünümü



Şekil 2.8.Web sitesi hakkımızda bölümünün sayfa görünümü



Şekil 2.9.Web sitesi proje içeriği sayfa görünümü

2.8.Makale Yazımı

Projenin sonuçları ve elde edilen bulgular detaylı bir makale halinde sunulmuştur.

3. Bulgular

3.1. Su Bileşenleri Arasındaki Korelasyonlar

Su bileşenleri arasındaki korelasyonlar incelenmiş ve suyun içilebilirliği üzerinde en fazla etkiye sahip bileşenler belirlenmiştir.

3.2. Hava Kalitesi ve Su Kirliliği Değişkenleri Arasındaki Korelasyonlar

Hava kalitesi ve su kirliliği arasındaki ilişki incelenmiş ve bu ilişkinin su içilebilirliği üzerindeki etkileri değerlendirilmiştir.

3.3.Makine Öğrenimi Sonuçları

Test edilen modeller arasında en yüksek doğruluk oranını veren modeller Random Forest ve XGBoost olarak belirlenmiştir. Bu iki model, diğer modellerle karşılaştırılmış ve en yüksek doğruluk oranını elde ettikleri için seçilmiştir.

Random Forest ve XGBoost modellerinin diğer modellere göre neden daha iyi performans gösterdiği detaylı bir şekilde analiz edilmiştir:

Random Forest: Yüksek çeşitlilik ve dengeli birleştirme, overfitting'i azaltma, daha az hiperparametre ayarı gerektirmesi.

XGBoost: Gradient boosting algoritması, hata odaklı öğrenme, ölçeklenebilirlik ve performans.

Water quality veri seti için CatBoost'un performansı ayrıca analiz edilmiştir:

CatBoost: Kategorik değişkenlere otomatik işleme, overfitting'in azaltılması, yüksek performans.

Ayrıca, Water Potability veri seti için geliştirilen **Renic** modeli, Gauss Naive Bayes (GaussNB) modelini geçecek bir doğruluk oranı sağlamıştır.

3.4.Hava Kalitesi ve Su Kirliliğinin Değerlendirilmesi

Hava kalitesi ve su kirliliği arasındaki ilişkiler incelenmiş ve bu ilişkilerin suyun içilebilirliği üzerindeki etkileri değerlendirilmiştir.

4. Sonuç

Proje kapsamında yapılan analizler, suyun içilebilirliğini etkileyen çeşitli faktörler arasındaki ilişkileri ortaya koymuştur. Korelasyon analizleri, su bileşenleri ve hava kalitesi değişkenleri arasındaki ilişkileri anlamamıza yardımcı olmuştur. Elde edilen bulgular, su kaynaklarının daha etkin bir şekilde yönetilmesi ve su kalitesinin iyileştirilmesi için stratejiler geliştirilmesine katkı sağlamıştır.

Makine öğrenimi modelleri kullanılarak yapılan analizlerde, Random Forest ve XGBoost modelleri en yüksek doğruluk oranlarını vermiştir. Bu modellerin neden diğerlerinden daha iyi performans gösterdiği detaylı bir şekilde analiz edilmiştir. CatBoost modeli, özellikle kategorik değişkenlerin bulunduğu veri setlerinde üstün performans göstermiştir. Ayrıca, 1. veri seti (Water Potability) için geliştirilen **Renic** modeli, Gauss Naive Bayes (GaussNB) modelini geçerek yüksek doğruluk oranı sağlamıştır.

Proje sonuçlarını ve bulgularını paylaşmak amacıyla bir web sitesi kurulmuş, bu sitede suyun içilebilirliğini test eden ve projeye ilgili tüm verilerin yer aldığı içerikler sunulmuştur. Ayrıca, projenin sonuçları ve elde edilen bulgular, detaylı bir makale halinde yazılarak bilimsel topluluk ve ilgili paydaşlarla paylaşılmıştır.

Bu proje, su kaynaklarının korunması ve sürdürülebilir bir şekilde yönetilmesi için önemli bir adım olup, gelecekteki

alıřmalar iin saėlam bir temel oluřturmuřtur. Elde edilen bulgular, ilgili paydařlara veriye dayalı stratejik kararlar alma konusunda yardımcı olabilir.

