



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

## **PROJE HAFTALIK RAPORU**

**SUSAMAYIZ  
Veri Bilimi Dersini Geçebilme Projesi**

**230291001**  
*Recep Bülbül*

**230291003**  
*Yakup Kadri İlhan*

**230291031**  
*Merve Çaloğlu*

**230291035**  
*Burcu Yıldırım*

# 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 yapılmaktadır. Proje kapsamında, zaman içinde artan su ihtiyacına rağmen bu ihtiyaca yeterince cevap verilememesi sorununa odaklanılacaktır. Su kaynaklarının verimli kullanılması ve içilebilir hale getirilmesi için gereken süreçlerin incelenmesi ve çözüm önerilerinin sunulması hedeflenmektedir.

Bu proje, bir dizi adımdan oluşmaktadır. İlk olarak, su bileşenlerinin suyun içilebilirliği ile olan korelasyonlarının incelenmesi ve bu korelasyonların görselleştirilmiş biçimlerinin sunulması planlanmaktadır. Ardından, veri setleri üzerinde makine öğrenimi modelleri uygulanarak en performanslı modelin belirlenmesi ve modelin neden bu kadar etkili olduğunun açıklanması amaçlanmaktadır.

Projede ayrıca, hava kalitesi ile su kirliliği arasındaki ilişkinin incelenmesi ve bu ilişkinin bilimsel bulgular ile desteklenerek ülkeler bazında dünya haritasında görselleştirilmesi planlanmaktadır. Son olarak, projenin sonuçları ve elde edilen bulguların bir makale halinde sunulması amaçlanmaktadır. Bu makale, su kaynaklarının etkin kullanımı ve içilebilir hale getirilmesi konularında farkındalık yaratmayı ve çözüm önerileri sunmayı hedeflemektedir.

## 2.Yapılan Çalışmalar

### Veri Seti 1:

#### Veri Temizliği ve Aykırı Değerlerin İşlenmesi

Veri setlerindeki aykırı değerlerin belirlenmesi ve incelenmesi, gerektiğinde bu aykırı değerlerin işlenmesi veya ayıklanması önemlidir. Veri setlerinin tutarlılığını sağlamak için hatalı veya anormal verilerin düzeltilmesi veya kaldırılması gerekebilir.

#### Korelasyon Analizi İçin Değişken Seçimi

Su içilebilirliğini etkileyen temel bileşenlerin belirlenmesi ve bu bileşenler arasından analiz için uygun olan değişkenlerin seçilmesi gerekmektedir.

#### Korelasyonların Hesaplanması

Belirlenen su bileşenleri arasındaki korelasyon katsayılarının hesaplanması önemlidir. Bu korelasyon katsayılarının anlamlılığı p-değerleri ile test edilmelidir.

#### Korelasyonların Görselleştirilmesi

Korelasyon matrisinin oluşturulması ve görselleştirilmesi gerekmektedir. Bu görselleştirmeler arasında scatter plotlar, heatmap'ler veya diğer görselleştirme tekniklerinin kullanılması korelasyonların daha anlaşılır hale gelmesini sağlar.

#### Korelasyon Analizinin Yorumlanması

Elde edilen korelasyon sonuçlarının projenin genel amacı ve hedefleriyle ilişkilendirilmesi önemlidir. Su bileşenlerinin su içilebilirliği ile en güçlü veya anlamlı korelasyonlarının belirlenmesi ve yorumlanması gerekmektedir.

## **İleri Analiz ve Derinlemesine İnceleme**

Güçlü veya anlamlı korelasyonlar bulunan değişkenlerin daha derinlemesine incelenmesi önemlidir. Ayrıca su içilebilirliğini etkileyen diğer faktörlerin (örneğin, çevresel faktörler) göz önüne alınması gerekmektedir.

### **Makine Öğrenimi Algoritmalarının Seçimi**

Proje hedeflerine ve veri setine uygun makine öğrenimi algoritmalarının belirlenmesi ve bu algoritmaların avantajları ve dezavantajları hakkında kısa bir değerlendirme yapılması gerekmektedir.

### **Veri Setinin Hazırlanması**

Makine öğrenimi modelleri için giriş özelliklerinin ve hedef değişkenin belirlenmesi, veri setinin eğitim ve test veri kümelerine ayrılması gerekmektedir.

### **Model Eğitimi**

Seçilen makine öğrenimi algoritmalarının eğitimi ve eğitim veri kümesi üzerinde modelin uyumlaştırılması gerekmektedir.

### **Model Performansının Değerlendirilmesi**

Eğitilen modellerin test veri kümesi üzerinde değerlendirilmesi ve performans metriklerinin hesaplanması ve karşılaştırılması (örneğin: doğruluk, hassasiyet, özgünlük, F1 skoru) önemlidir.

### **Modelin Yanıltıcı Olan Varyansı Hesaba Katılması**

Yanıltıcı olan varyansın belirlenmesi ve modeller arasında karşılaştırma yapılırken varyansın etkisinin değerlendirilmesi gerekmektedir.

### **En Performanslı Modelin Belirlenmesi**

Modellerin performanslarının karşılaştırılması ve en iyi modelin belirlenmesi önemlidir. En iyi modelin neden diğerlerinden daha iyi performans gösterdiğinin analiz edilmesi gerekmektedir.

## **Veri Seti 2:**

### **Veri Setinin Yüklenmesi ve İncelemesi**

Bu hafta, su içilebilirliğini etkileyen faktörleri inceledik. İlk adım olarak, veri setimizi yükledik ve yapısını analiz ettik. Veri setinin sütunlarını inceledik ve içeriğini anlamaya çalıştık.

### **Eksik Verilerin İncelenmesi ve İşlenmesi**

Daha sonra, veri setimizdeki eksik verileri belirledik ve sayılarını hesapladık. Eksik verileri uygun bir şekilde işlemek için doldurma veya çıkarma yöntemlerini kullandık.

### **Değişken Seçimi ve Korelasyon Analizi için Hazırlık**

Su içilebilirliğini etkileyen temel bileşenleri belirledik ve bu bileşenlere odaklanarak korelasyon analizi için uygun değişkenleri seçtik.

### **Korelasyonların Hesaplanması**

Seçilen su bileşenleri arasındaki korelasyon katsayılarını hesapladık ve bu katsayıların anlamlılığını test ettik.

### **Korelasyonların Görselleştirilmesi**

Son olarak, elde ettiğimiz korelasyon sonuçlarını görselleştirdik. Korelasyon matrisi oluşturarak ve heatmap kullanarak, korelasyonları daha iyi anlamaya çalıştık.

### 3.Yapılması Planlanan Çalışmalar

#### Veri Seti 1:

##### Sonuçların Belirlenmesi

Bu aşamada, yapılan korelasyon analizinin sonuçları özetlenecek ve belgelenecektir. Hangi bileşenlerin su içilebilirliği üzerinde en fazla etkiye sahip olduğu belirlenecek ve projedeki diğer iş paketleriyle nasıl bir bağlantı kurulacağı, alınacak ileri adımlar belirlenecektir.

#### Veri Seti 2:

##### Korelasyon Analizinin Yorumlanması

Elde edilen korelasyon sonuçları yorumlanacak ve hangi su bileşenlerinin su içilebilirliği üzerinde daha etkili olduğu belirlenecektir.

##### İleri Analiz ve Derinlemesine İnceleme

Güçlü veya anlamlı korelasyonlar bulunan değişkenler daha derinlemesine incelenecek ve su içilebilirliğini etkileyen diğer faktörler (örneğin, çevresel faktörler) göz önüne alınacaktır.

##### Sonuçların Belirlenmesi

Korelasyon analizinin sonuçları özetlenecek ve belgelenecek, hangi bileşenlerin su içilebilirliği üzerinde en fazla etkiye sahip olduğu belirlenecektir.

### 4. Sonuç ve Değerlendirme

Anlaşılan o ki, bu hafta yaptığımız analizler sonucunda veri setlerimizde beklediğimizden daha az kuvvetli korelasyonlar gözlemledik. Bu durum, su içilebilirliğini etkileyen faktörler arasındaki ilişkilerin daha karmaşık veya belki de daha az doğrusal olduğunu gösteriyor olabilir.

Özellikle, su kalitesini etkileyen faktörlerin birbirleriyle olan ilişkileri incelendiğinde, beklenen güçlü korelasyonların bulunamaması, su kalitesini etkileyen faktörlerin karmaşıklığını ve etkileşimlerini daha iyi anlamamız gerektiğini gösteriyor olabilir. Bu durum, ileriki çalışmalarda daha derinlemesine analizlere ve farklı analiz yöntemlerine başvurmayı gerektirebilir.

Bununla birlikte, elde ettiğimiz sonuçlar bize, su kalitesini etkileyen faktörler arasındaki ilişkilerin daha iyi anlaşılması için daha fazla araştırma ve analiz yapılması gerektiğini gösteriyor. Bu, su kaynaklarının korunması ve su kalitesinin iyileştirilmesi için daha etkili stratejiler geliştirmek adına önemli bir adımdır.