



**Çanakkale Onsekiz Mart Üniversitesi Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü
İleri Veri Tabanı Dersi Rapor**

**İsmail Kerim Sakar
210401056**

Öğr.Gör. İsmail Kahraman

Titanic Veri Seti Rapor

1. Giriş

Titanik faciası, 1912'de Atlantik Okyanusu'nda meydana gelen büyük bir deniz kazasıdır. Bu rapor, Titanic veri setini inceleyerek, yolcuların hayatta kalma durumunu etkileyen faktörleri analiz edecek ve bir sınıflandırma modeli kullanarak hayatta kalma olasılığını tahmin edecektir.

2. Veri Seti Tanımı

- **Survived:** Hayatta kalma durumu (0 = Hayır, 1 = Evet)
- **Pclass:** Yolcu sınıfı (1 = 1. sınıf, 2 = 2. sınıf, 3 = 3. sınıf)
- **Sex:** Cinsiyet (Erkek, Kadın)
- **Age:** Yaş
- **SibSp:** Gemideki eş veya kardeş sayısı
- **Parch:** Gemideki ebeveyn veya çocuk sayısı
- **Ticket:** Bilet numarası
- **Fare:** Ücret
- **Cabin:** Kabin numarası
- **Embarked:** Biniş limanı (C = Cherbourg, Q = Queenstown, S = Southampton)

3. Veri Ön İşleme

- Eksik verilerin temizlenmesi
- Kategorik değişkenlerin dönüştürülmesi
- Özellik mühendisliği: Aile büyüklüğü, tek başına seyahat edip etmemek gibi yeni özelliklerin oluşturulması

4. Veri Analizi

- **Hayatta Kalma Oranları:** Cinsiyet, sınıf ve diğer faktörlerin hayatta kalma oranlarına etkisi
- **Yaş Dağılımı:** Hayatta kalanların yaş dağılımı
- **Sınıf Dağılımı:** Sınıflara göre hayatta kalma oranları

5. Sınıflandırma Modeli: Lojistik Regresyon

Lojistik regresyon modelini Titanik veri seti üzerinde kullanarak, yolcuların hayatta kalma durumunu etkileyen faktörleri belirledik. Model, veri setinin bir bölümünü eğitim için kullandık ve geri kalanını test için ayırdık. Modeli eğitmek ve test etmek için uygun metrikler (doğruluk, hassasiyet, geri çağırma vb.) kullandık.

Lojistik regresyon modelinin sonuçlarına göre, cinsiyet, yaş ve yolcu sınıfı gibi faktörlerin hayatta kalma üzerinde önemli bir etkisi olduğunu gözlemledik. Özellikle, kadınların ve 1. sınıf yolcularının hayatta kalma olasılığının daha yüksek olduğunu tespit ettik.

6. Sonuçlar

- Sınıflandırma modeli doğruluk oranı ve performans metrikleri
- Önemli özelliklerin sınıflandırma üzerindeki etkisi

Projenin Anlatıldığı Youtube Linki:

<https://www.youtube.com/watch?v=4Pa6HoQskvw>

Projede Yer alanlar :

Eren Takeş: 210401029

İsmail Kerim Sakar: 210401056

Sefacan Dindar: 220401094