



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

**Eren Takeş
210401029**

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

- ⑩ **Survived:** Hayatta kalma durumu (0 = Hayır, 1 = Evet)
- ⑩ **Pclass:** Yolcu sınıfı (1 = 1. sınıf, 2 = 2. sınıf, 3 = 3. sınıf)
- ⑩ **Sex:** Cinsiyet (Erkek, Kadın)
- ⑩ **Age:** Yaş
- ⑩ **SibSp:** Gemideki eş veya kardeş sayısı
- ⑩ **Parch:** Gemideki ebeveyn veya çocuk sayısı
- ⑩ **Ticket:** Bilet numarası
- ⑩ **Fare:** Ücret
- ⑩ **Cabin:** Kabin numarası
- ⑩ **Embarked:** Biniş limanı (C = Cherbourg, Q = Queenstown, S = Southampton)

3. Veri Ön İşleme

- ⑩ Eksik verilerin temizlenmesi
- ⑩ Kategorik değişkenlerin dönüştürülmesi
- ⑩ Özellik mühendisliği: Aile büyüklüğü, tek başına seyahat edip etmemek gibi yeni özelliklerin oluşturulması

4. Veri Analizi

- ⑩ **Hayatta Kalma Oranları:** Cinsiyet, sınıf ve diğer faktörlerin hayatta kalma oranlarına etkisi
- ⑩ **Yaş Dağılımı:** Hayatta kalanların yaş dağılımı
- ⑩ **Sınıf Dağılımı:** Sınıflara göre hayatta kalma oranları

5. Sınıflandırma Modeli: Lojistik Regresyon

Lojistik regresyon modelini Titanic veri seti üzerinde kullanarak, yolcuların hayatta kalma durumunu etkileyen faktörleri belirledik. Model, veri setinin bir bölümünü

eđitim iin kullandık ve geri kalanını test iin ayırdık. Modeli eđitmek ve test etmek iin uygun metrikler (dođruluk, hassasiyet, geri ađırma vb.) kullandık.

Lojistik regresyon modelinin sonularına gre, cinsiyet, yař ve yolcu sınıfı gibi faktrlerin hayatta kalma zerinde nemli bir etkisi olduđunu gzlemledik. zellikle, kadınların ve 1. sınıf yolcularının hayatta kalma olasılıđının daha yksek olduđunu tespit ettik.

6. Sonular

- ⑩ Sınıflandırma modeli dođruluk oranı ve performans metrikleri
- ⑩ nemli zelliklerin sınıflandırma zerindeki etkisi

Projenin Anlatıldıđı Youtube Linki:

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

Projede Yer alanlar :

Eren Takeř: 210401029

İsmail Kerim Sakar: 210401056

Sefacan Dindar: 220401094