



**FEN FAKÜLTESİ**  
**İSTATİSTİK BÖLÜMÜ**  
**İST4004 ARAŞTIRMA PROJESİ 2**

**LOJİSTİK REGRESYON VE SINIFLANDIRMA AĞAÇLARI İLE BİR UYGULAMA**

**Zehra Helin TEKİN**

**C181120078**

**Akademik Danışman: Doç. Dr. Hatice Kübra AKDUR**

**HAZİRAN 2024**

## İÇİNDEKİLER

1.ÖZET .....	4
2. GİRİŞ .....	Error! Bookmark not defined.
2.1. REGRESYON .....	Error! Bookmark not defined.
2.1.1 REGRESYON ANALİZİNDE KULLANILAN BAZI TEMEL YÖNTEMLER .....	Error! Bookmark not defined.
2.2. SINIFLANDIRMA .....	Error! Bookmark not defined.
2.2.1. SINIFLANDIRMA YÖNTEMLERİ .....	Error! Bookmark not defined.
2.3. SINIFLANDIRMA AĞAÇLARI (CART) METODOLOJİSİ .....	Error! Bookmark not defined.
2.3.1. SINIFLANDIRMA AĞACI .....	Error! Bookmark not defined.
2.3.2 GUIDE .....	Error! Bookmark not defined.
2.4. SINIFLANDIRMA VE REGRESYON AĞAÇLARI AVANTAJLARI VE DEZAVANTAJLARI .....	Error! Bookmark not defined.
2.4.1.AVANTAJLARI .....	Error! Bookmark not defined.
2.4.2. DEZAVANTAJLARI .....	Error! Bookmark not defined.
2.5. RANDOM FOREST .....	Error! Bookmark not defined.
2.5.1. MİRAS ALDIĞI CART ÖZELLİKLERİ .....	Error! Bookmark not defined.
2.5.2. MİRAS ALMADIĞI CART ÖZELLİKLERİ .....	Error! Bookmark not defined.
2.5.3. AVANTAJLAR .....	Error! Bookmark not defined.
2.5.4 DEĞERLENDİRME BİLEŞENLERİ .....	Error! Bookmark not defined.
2.5.4.1. VARIABLE IMPORTANCE .....	Error! Bookmark not defined.
2.5.4.2 YAKINLIKLAR .....	Error! Bookmark not defined.
2.5.5. TRAIN VE TEST VERİLERİ .....	Error! Bookmark not defined.
2.5.6. "randomForest" PAKETİ .....	Error! Bookmark not defined.
2.6. BAGGING .....	Error! Bookmark not defined.
2.7. EĞİTİM SETİ .....	Error! Bookmark not defined.
2.8. BOOSTING .....	Error! Bookmark not defined.
2.8.1. BOOSTING'IN TEMEL İŞLEVLERİ .....	Error! Bookmark not defined.
2.9. BAGGING VE BOOSTING İŞE YARAR MI? .....	Error! Bookmark not defined.
2.10. LOJİSTİK REGRESYON MODELİ .....	Error! Bookmark not defined.
2.10.1. Lojistik Yanıt Fonksiyonuna Giriş .....	Error! Bookmark not defined.
2.10.2. Çoklu Lojistik Regresyon Modelinin Uyarlanması .....	Error! Bookmark not defined.
2.10.3. BETA TAHMİNLERİNİN YORUMLANMASI .....	Error! Bookmark not defined.
2.11. ROC EĞRİSİ OLUŞTURMA .....	Error! Bookmark not defined.
3. UYGULAMA .....	Error! Bookmark not defined.

3.1 VERİ SETİ.....	Error! Bookmark not defined.
3.2 VERİ HAZIRLAMA .....	Error! Bookmark not defined.
3.3.BETİMLEYİCİ İSTATİSTİKLER VE VERİ GÖRSELLEŞTİRME .....	Error! Bookmark not defined.
3.3.1 Frekans Değerleri.....	Error! Bookmark not defined.
3.3.2. Hayatta Kalma Değişkeni Frekansı.....	Error! Bookmark not defined.
3.3.3. Sınıf Değişkeni Frekansı .....	Error! Bookmark not defined.
3.3.4. Cinsiyet Değişkeni Frekansı .....	Error! Bookmark not defined.
3.3.5. Embarked Değişkeni Frekans.....	Error! Bookmark not defined.
3.3.6. Sürekli Değişkenlerin Özet İstatistikleri.....	Error! Bookmark not defined.
3.4. VERİ GÖRSELLEŞTİRME .....	Error! Bookmark not defined.
3.4.1. Tüm Değişkenlerin Tek Başına Görselleştirilmesi .....	Error! Bookmark not defined.
3.5. İKİLİ İLİŞKİLERİN ANALİZİ .....	Error! Bookmark not defined.
3.6.LOJİSTİK REGRESYON MODELİ .....	Error! Bookmark not defined.
3.6.1. MODEL ÇIKTISI.....	Error! Bookmark not defined.
3.6.2.BETA TAHMİNLERİNİN YORUMLANMASI .....	Error! Bookmark not defined.
3.7. RANDOM FOREST TITANIC VERİ SETİ UYGULAMASI .....	Error! Bookmark not defined.
3.7.1 VERİ HAZIRLIĞI .....	Error! Bookmark not defined.
3.7.2. TRAİN VE TEST VERİ SETLERİNE AYIRMA .....	Error! Bookmark not defined.
3.7.3. RANDOM FOREST MODELİ UYGULAMA .....	Error! Bookmark not defined.
3.7.4. TEST VERİSİ İÇİN TAHMİNLER YAPMA .....	Error! Bookmark not defined.
3.7.5 DOĞRULUK ORANINI HESAPLAMA .....	Error! Bookmark not defined.
3.7.6. KARAR AĞACI OLUŞTURMA .....	Error! Bookmark not defined.
3.8. ROC EĞRİSİ ÇİZDİRME.....	Error! Bookmark not defined.
3.9. GUIDE REGRESYON AĞACI .....	Error! Bookmark not defined.
4.SONUÇ VE DEĞERLENDİRME .....	Error! Bookmark not defined.
EKLER.....	Error! Bookmark not defined.
KAYNAKÇA .....	Error! Bookmark not defined.

## 1.ÖZET

Bu rapor, Titanic faciasında hayatta kalma olasılığını etkileyen faktörleri incelemek ve bu faktörlerin hayatta kalma üzerindeki etkilerini analiz etmek amacıyla gerçekleştirilmiştir. Veri seti, geminin yolcularına ait bilgileri içermekte olup, bu veriler üzerinden istatistiksel analizler yapılarak hayatta kalma durumu ile diğer değişkenler arasındaki ilişkiler incelenmiştir.

Raporun ilk aşamasında, veri setinin genel yapısı ve değişkenler arasındaki ilişkiler görsel olarak incelenmiştir. Değişkenlerin dağılımı ve birbiriyle ilişkisi, çeşitli grafikler ve görsel teknikler kullanılarak detaylı bir şekilde analiz edilmiştir. Değişkenler tek başına görselleştirilerek analiz edilmiş daha sonra yolcuların yaş, cinsiyet, sınıf gibi özellikleri ile hayatta kalma durumu arasındaki ilişkiler gözlemlenmiş ve yorumlanmıştır.

Ardından, korelasyon analizi ile değişkenler arasındaki ilişkilerin gücü incelenmiştir. Özellikle hayatta kalma değişkeni ile biniş yeri ve ücret arasındaki ilişkiler değerlendirilmiş ve bu ilişkilerin ne derecede güçlü olduğu belirlenmeye çalışılmıştır.

Lojistik regresyon modeli oluşturularak hayatta kalma durumunun diğer değişkenler üzerindeki etkileri sayısal olarak incelenmiştir. Bu model ile hangi faktörlerin hayatta kalma olasılığını daha fazla arttırdığı veya azalttığı belirlenmiş ve bu faktörlerin önemi değerlendirilmiştir.

Raporun ikinci aşamasında, random forest sınıflandırma algoritması çalıştırılmış, ağaç çıktısı elde edilip yorumlanmıştır. OOB hata tahmini hesaplanmıştır. Ayrıca ROC eğrisi üzerinden sınıflandırma modeli performansı test edilmiştir.

Son olarak GUIDE algoritması ile regresyon sınıflandırma ağacı elde edilip yorumlanmıştır.

Bu çalışma, Titanic faciasında hayatta kalma üzerinde etkili olan faktörlerin belirlenmesi, sınıflandırılması ve analiz edilerek daha iyi anlaşılmasına katkı sağlamayı hedeflemektedir. Elde edilen bulgular, benzer durumların analizinde ve benzer olaylardan çıkarılacak derslerde faydalı olabilir.