

## Sisterslab Yapay Zeka Akademisi Bitirme Projesi

# PROJE ÖZETİ

### Proje Başlığı: Kalp Hastalığı Risk Analizi (Heart Disease UCI)

#### 1. Problem Tanımı

Hangi veri seti seçildi?

Bu projede, kalp hastalığı riskini tahmin etmeye yönelik **Heart Disease - UCI** veri seti kullanılmıştır. Veri seti **Kaggle** platformundan alınmış olup, URL'si aşağıda verilmiştir:

<https://www.kaggle.com/datasets/mragpavank/heart-diseaseuci>

Hedef değişken: **target**

- target = 1 → Kişi kalp hastalığı riski taşıyor
- target = 0 → Kişi risk taşımıyor

Bu tahmin ile: Erken teşhis sayesinde doktorlar önleyici tedbirler alabilir, sağlık sistemi üzerindeki yük hafifletilebilir ve toplum sağlığı korunabilir.

#### 2. Veri Analizi ve Görselleştirme (EDA)

**df.info() & df.describe():** Veri setinde 303 gözlem ve 14 sütun bulunuyor.

**Eksik Veri:** Eksik veri bulunmamaktadır.

**Hedef değişkenin dağılımı:** Dengeli şekilde 0 ve 1 sınıfları içeriyor.

**Veri görselleştirmeleri:** age, sex, cp, trestbps, chol, thalach gibi değişkenlerin dağılımları ve korelasyonları incelendi.

#### 3. Veri Ön İşleme

**Kodlama:** sex, fbs, exang → LabelEncoder; cp, restecg → OneHotEncoder

**Ölçeklendirme:** Sayısal değişkenler StandardScaler ile normalize edildi.

**Eğitim/Test Ayrımı:** train\_test\_split(test\_size=0.2) ile %80 eğitim, %20 test

#### 4. Heart Disease UCI seçildiyse (Sınıflandırma)

İki farklı model eğitildi: Logistic Regressin ve Random Forest

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.85	0.90	0.81	0.85	0.93
Random Forest	0.82	0.84	0.81	0.83	0.92

Her iki model de **Confusion Matrix** ve **ROC eğrisi** ile değerlendirildi.

Logistic Regression modeli genel olarak daha dengeli ve yüksek başarı gösterdi.

## 5. Model Karşılaştırması

**En başarılı model:** Logistic Regression

**Öne çıkan metrik:** Precision

**Aşırı/yetersiz öğrenme:** Belirgin overfitting veya underfitting gözlenmedi. Test performansları yüksek ve dengeliydi.

## 6. Sonuç ve Yorumlar

Logistic Regression, doğruluk ve ROC-AUC gibi metriklerde daha başarılıydı.

Forest modeli de etkili çalıştı ve destekleyici bir model olarak değerlidir.

Bu proje, sağlık alanında erken teşhis sistemlerine katkı sağlayabilir.

Geliştirme için öneriler: Hiperparametre optimizasyonu (GridSearchCV), feature importance analizi, veri genişletme.

**NOT:** Kodlar, analizler,görseller zeynepcol\_bitirmeprojesi\_sisterslab.ipynb adlı Colab dosyasında çalıştırılmıştır.

## 🌸 TEŞEKKÜRLER

Yapay Zeka Akademisi boyunca edindiğim bilgi, deneyim ve ilham için en içten teşekkürlerimi sunmak isterim. Başta değerli eğitmenimiz Yasemin Arslan olmak üzere, tüm SistersLab proje asistanlarına ve bu sürecin her adımında emeği geçen herkese gönülden teşekkür ederim.

Sadece teknik bilgi değil, aynı zamanda motivasyon ve destek dolu bir ortam sunduğunuz için minnettarım.

Her birinize emekleriniz ve ilginiz için ayrı ayrı teşekkür ederim.

İyi ki varsınız!

Zeynep ÇÖL