

KALP HASTALIĞI TAHMİNİ – PROJE ÖZETİ

1. Problem Tanımı

Bu projede UCI Heart Disease veri seti kullanılarak bireylerin çeşitli sağlık göstergelerine göre kalp hastalığı riski olup olmadığı tahmin edilmiştir. Hedef değişken **num**, kalp hastalığının olup olmadığını belirtir. Bu değişken ikili hale getirilmiş (0 = yok, 1 = var) ve sınıflandırma probleminin çözümü için çeşitli makine öğrenmesi algoritmaları uygulanmıştır.

Kalp hastalıkları, dünya genelinde en yaygın ölüm nedenlerinden biridir. Bu nedenle bu tür modellerle erken teşhis ve önleyici tedavilerin geliştirilmesi büyük önem taşımaktadır.

2. Veri Seti ve Özellikler

Veri seti 300+ gözlem ve 14 sağlık göstergesinden oluşmaktadır. Kullanılan bazı önemli değişkenler:

- age: Yaş
- sex: Cinsiyet
- cp: Göğüs ağrısı tipi
- chol: Kolesterol seviyesi
- fbs: Açlık kan şekeri
- thalach: Maksimum kalp atım hızı
- oldpeak: ST segmentindeki değişim
- slope, ca, thal: EKG ve damar tıkanıklığıyla ilişkili özellikler

İlk olarak veri temizliği yapılmış; **id**, **dataset** gibi analizde kullanılmayan sütunlar çıkarılmış ve **num** değişkeni ikili sınıfa çevrilmiştir.

3. Uygulanan Yöntemler

Aşağıdaki makine öğrenmesi modelleri karşılaştırmalı olarak uygulanmıştır:

- Lojistik Regresyon
- Random Forest Classifier

Veri standardize edildikten sonra modeller eğitilmiş ve test seti üzerinde değerlendirilmiştir.

4. Model Sonuçları

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC
Lojistik Regresyon	~0.86	~0.85	~0.88	~0.86	~0.91
Random Forest	~0.88	~0.87	~0.89	~0.88	~0.93

Random Forest modeli genel olarak daha yüksek başarı sağlamıştır.

5. Çıkarımlar

- En etkili özellikler: **cp, thalach, oldpeak, ca, thal**
- Model performansı: Random Forest modeli, ROC AUC skoruna göre kalp hastalığı tahmininde oldukça başarılıdır.
- Toplumsal fayda: Bu tür bir model, hastanelerde erken teşhis sistemlerinde kullanılabilir.