

1. Proje Başlığı ve Amaç

Proje Adı: Kalp Hastalığı Teşhisi

Amaç: Kalp hastalığı riski taşıyan bireylerin önceden belirlenebilmesi için, veri madenciliği ve makine öğrenmesi yöntemlerini kullanarak etkili bir tespit modeli oluşturmak.

2. Problemin Aka Planı ve Önemi

Kalp hastalıkları, dünya genelinde ölüm nedenlerinin başında gelmektedir. Erken teşhis, hem bireyin yaşam kalitesini arttırmak hem de sağlık sistemlerindeki yükü azaltmak açısından kritik öneme sahiptir. Bu nedenle, veri bilimi ve yapay zeka kullanarak erken teşhis modelleri oluşturmak, sağlık teknolojilerinde önemli katkılar sağlayabilir.

3. Veri Seti Tanıtımı

Kullanılan Veri Seti: UCI Heart Disease veri seti

- Gözlem Sayısı: 920
- Sütun Sayısı: 16 (1 hedef: target)

Hedef Değişken (target):

- 0: Kalp hastalığı yok
- 1-4: Farklı seviyelerde kalp hastalığı var

4. Veri Temizleme: Zorluklar ve Stratejiler

Karşılaşılan Zorluklar:

- Birçok sütunda eksik değer mevcuttu:
 - ca: %66 eksik
 - thal: %52 eksik
 - slope: %33 eksik

Uygulanan Stratejiler:

- Sayısal sütunlar için medyan ile doldurma
- Kategorik sütunlar için en sık görülen değerle doldurma
- Eksik veriler başarıyla giderildi
- Veri seti eğitim (%80) ve test (%20) olarak ayrıldı.

5. Model Seçimi ve Değerlendirme

Kullanılan Modeller:

- Logistic Regression
- Random Forest Classifier

Model Seçimi Kriterleri:

- Basitlik ve yorumlanabilirlik (Log. Reg.)
- İyi genel performans ve overfitting kontrolü (RF)

Sonuçlar:

| Model | Doğruluk |
|---------------------|----------|
| Random Forest | %0.8370 |
| Logistic Regression | %0.8261 |

6. Sonuçlar ve Görsel Bulgular

- Kalp hastalığı tahmini için yapay zeka modelleri umut vaat etmektedir.
- Eksik verilerin doğru yönetimi ve görsel analizler, modelin başarısını desteklemiştir.

Öne Çıkan Görseller:

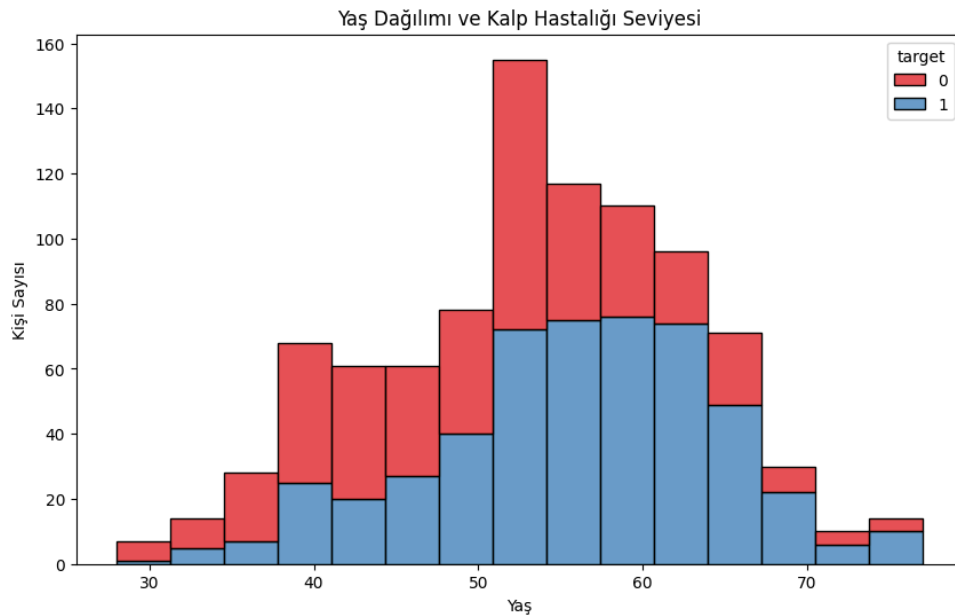
- Korelasyon Matrisi



target (hedef değişken) ile korelasyonlar:

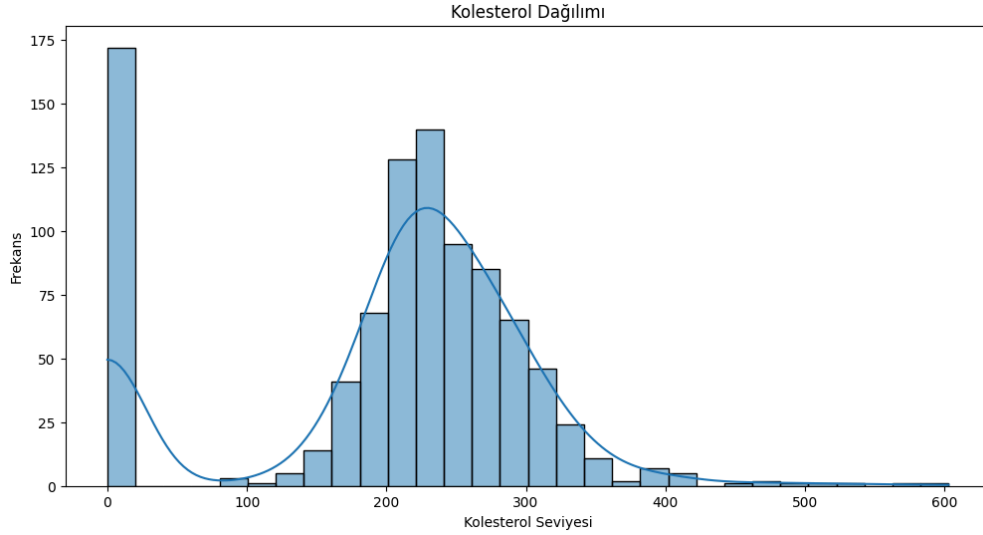
- **thalach (maksimum kalp atış hızı): -0.38**
Kalp atış hızı arttıkça hastalık riski azalıyor olabilir.
- **oldpeak (ST depresyonu): 0.37**
ST depresyonu arttıkça hastalık riski de artıyor olabilir.
- **age (yaş): 0.28**
Yaş arttıkça kalp hastalığı riski artabilir.
- **chol (kolesterol): -0.23**
Düşük seviyede negatif ilişki vardır.

● Yaş Dağılımı ve Kalp Hastalığı Seviyesi



- Genellikle 50-60 yaş arası bireyler veri setinde yoğunlukta.
- Kalp hastalığı seviyeleri yaşla birlikte değişim gösteriyor:
 - Genç yaşlarda hastalık seviyesi daha düşük (sıfıra yakın),
 - Yaş ilerledikçe bazı seviyelerde artış gözleniyor.

- Kolesterol Dağılım Grafiği



- Kolesterol değerleri yaklaşık 125 ile 560 arasında değişiyor.
- Dağılım **asimetrik** görünüyor; uç değerler (outlier'lar) sağ tarafa doğru yoğunlaşmış.
- KDE eğrisi sağa çarpık bir yapı gösteriyor → bu, bazı bireylerin çok yüksek kolesterol seviyelerine sahip olduğunu gösterir.