

Ara Rapor: Kalp Hastalığı Risk Tahmini Projesi

Proje Ekibi:

- Muhammed Gülistan - Veri Seti ve Ön İşleme**
- Mehmet Akif Erol - Kullanılan Algoritmalar ve Performans Değerlendirmesi**
- Oğulcan Esendir - Görselleştirmeler ve Analizler**

Proje Konusu:

Bu proje, kalp hastalığı riskini tahmin eden bir makine öğrenmesi modeli geliştirmeyi amaçlamaktadır. Projede çeşitli makine öğrenmesi algoritmaları kullanılarak farklı modellerin performansları karşılaştırılmıştır.

1. Veri Seti ve Ön İşleme

Kullanılan veri seti: *heart_disease_risk_dataset_earlymed.csv*

- Eksik veriler temizlenmiş,
- Özellikler ölçeklendirilmiş,
- Eğitim ve test setlerine ayrılmıştır (%80 eğitim, %20 test).

2. Kullanılan Algoritmalar ve Performans Değerlendirmesi

Aşağıdaki üç model eğitilmiş ve test edilmiştir:

- Lojistik Regresyon**
- Rastgele Orman (Random Forest)**
- Destek Vektör Makineleri (SVM)**

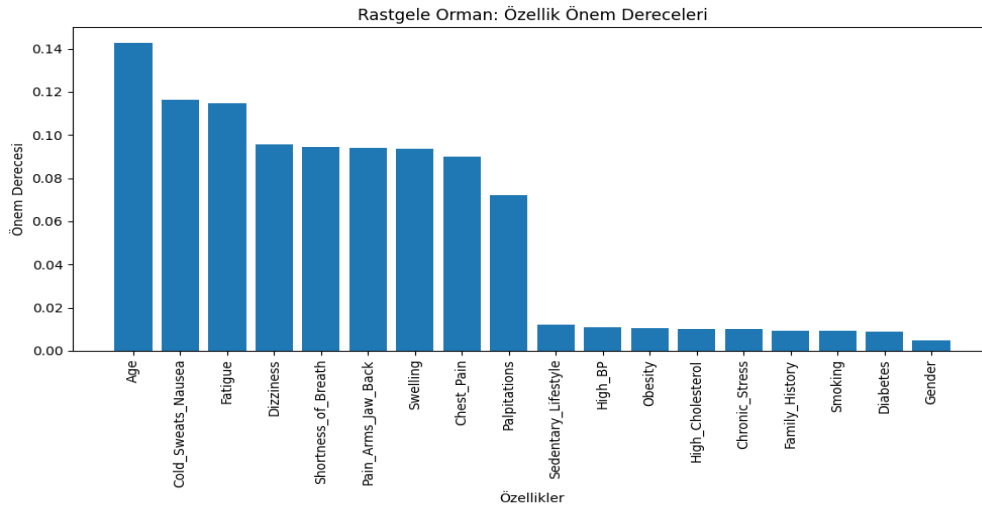
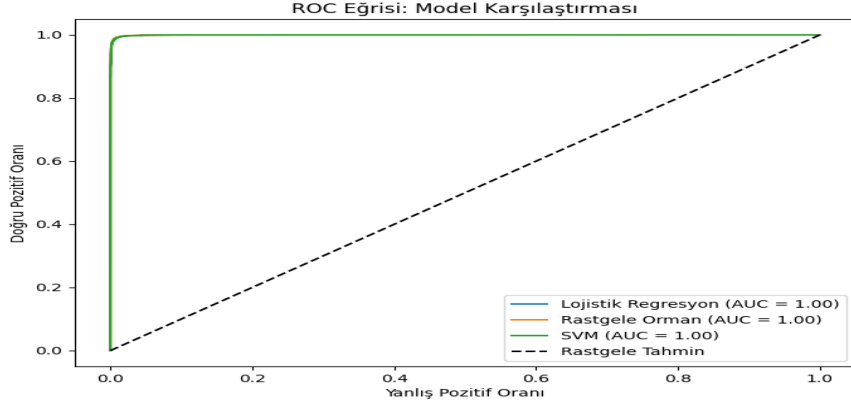
Her model doğruluk, hassasiyet, duyarlılık, F1 skoru ve ROC AUC metriği ile değerlendirilmiştir.

3. Görselleştirmeler ve Analizler

- ROC Eğrisi:** Modellerin ROC eğrileri karşılaştırılmıştır.
- Karmaşıklık Matrisi:** Lojistik regresyon modeli için hata analizi yapılmıştır.
- Özellik Önemi:** Rastgele orman modelinde en önemli değişkenler belirlenmiştir.

4. Sonuç ve Gelecek Adımlar

- Verinin dengeli olup olmadığı incelenmiştir.
- Daha fazla özellik mühendisliği ve hiperparametre optimizasyonu ile model geliştirilmesi planlanmaktadır.
- Kullanıcı dostu bir arayüz ile modelin entegrasyonu hedeflenmektedir.



Lojistik Regresyon - Karmaşıklık Matrisi

