

TİTANİC VERİ SETİ: MAKİNE ÖĞRENMESİ MODELLERİ İLE HAYATTA KALMA TAHMİNİ

Bu projede, Titanic veri seti kullanılarak yolcuların hayatta kalıp kalmayacakları tahmin edilmiştir. Farklı makine öğrenmesi modelleri eğitilerek performansları karşılaştırılmış ve en iyi model belirlenmiştir.

Lojistik Regresyon

Bağımlı değişkenimiz (Survived) kategorik bir değişken olduğundan, ikili sınıflandırma yapabilen lojistik regresyon kullanıldı.

K- En Yakın Komşu (KNN)

Titanic verisinde benzer özelliklere sahip yolcuların hayatta kalma oranlarının belirlenmesine yardımcı olabileceği için kullanıldı.

SVM (Linear Kernel)

Titanic veri setinde hayatta kalan ve ölen yolcuları en iyi şekilde ayırabilecek bir hiper düzlem bulmak için kullanıldı.

Karar Ağacı (Decision Tree)

Titanic verisinde, belirli kurallara göre hayatta kalma tahmini yapmak için kullanıldı. (örneğin: "Kadın yolcuların hayatta kalma oranı daha yüksek" gibi)

Random Forest

Titanic veri setindeki karmaşık ilişkileri daha iyi yakalayabileceği için güçlü bir modeldir. Overfitting'i azaltmak için karar ağacının geliştirilmiş versiyonu olduğu için kullanıldı.

XGBoost (Extreme Gradient Boosting)

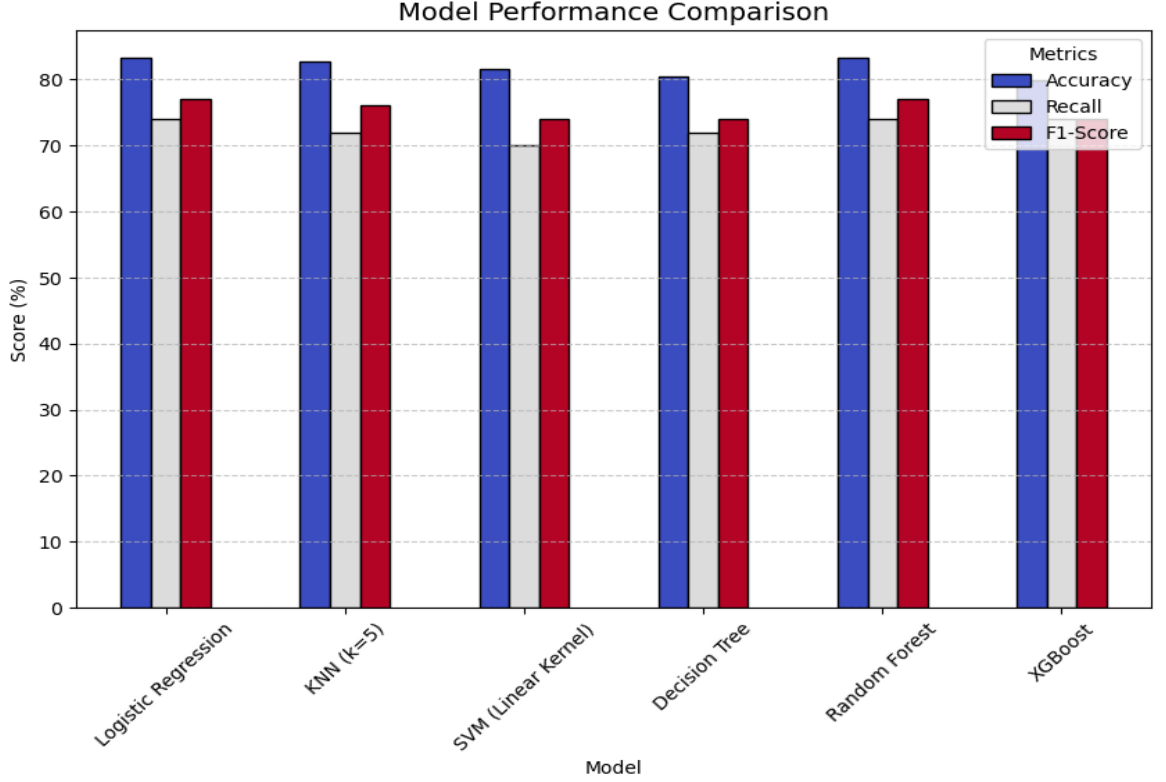
Titanic veri setinde daha yüksek doğruluk elde etmek ve hata oranını azaltmak için kullanıldı. Feature engineering gerektiren durumlarda güçlüdür

Linear Regression

Titanic veri setinde bağımlı değişkenimiz kategorik olmasına rağmen, regresyon modellerinin tahmin gücünü görmek için test ettik.

GRAFİKLER

BAR CHART (SÜTUN GRAFİĞİ)

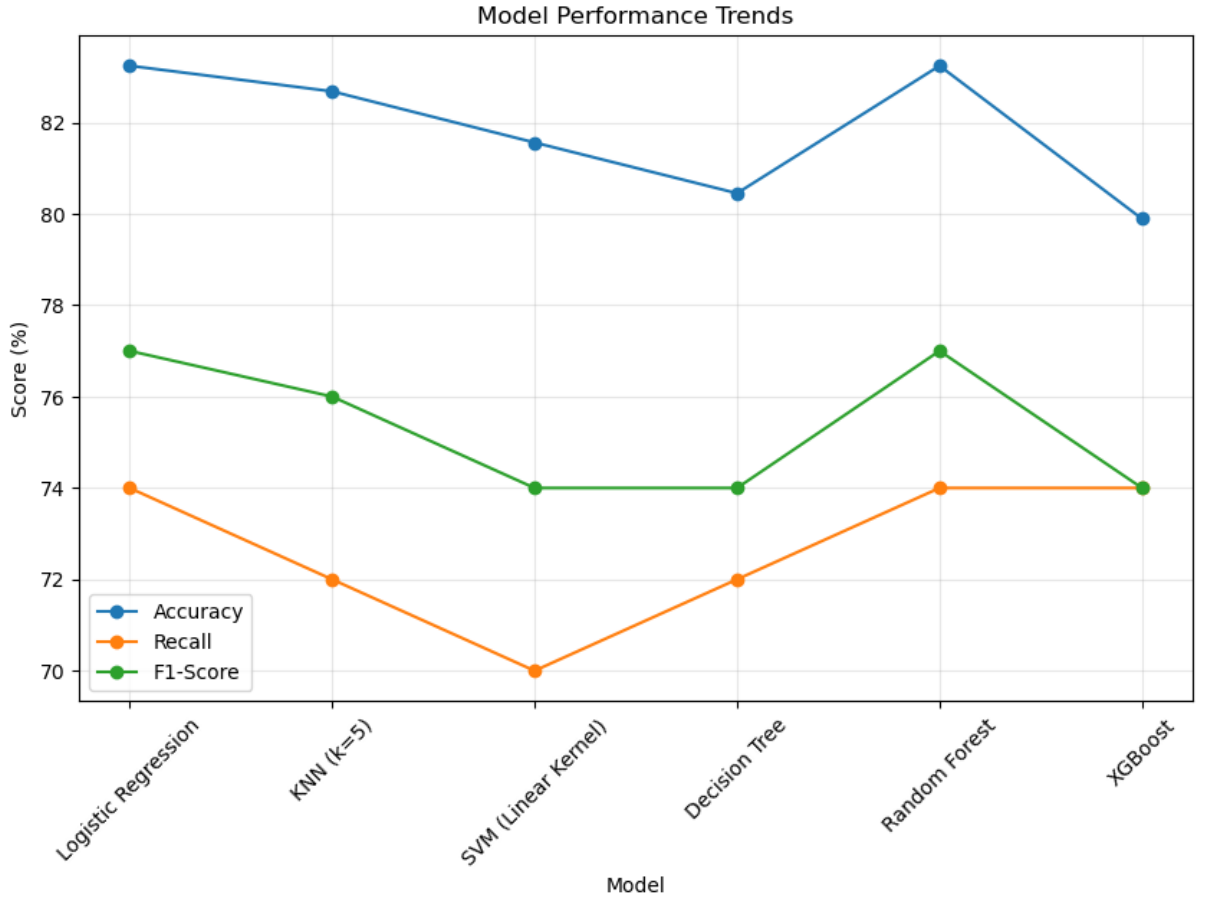


- Bu grafikte Accuracy (Doğruluk), Recall ve F1-Score değerlerini modeller arasında karşılaştırıyoruz.
- En yüksek doğruluk (Accuracy) oranı Lojistik Regresyon ve Random Forest modellerinde görülüyor (%83.24).
- XGBoost modeli, diğerlerine kıyasla en düşük doğruluk oranına sahip (%79.89).
- Recall açısından bakıldığında, tüm modellerde %70-%74 aralığında değerler mevcut.
- F1-Score açısından en iyi performans %77 ile Lojistik Regresyon ve Random Forest modellerinde olduğu görülüyor.

Sonuç:

Lojistik Regresyon ve Random Forest, hem doğruluk hem de F1-Score açısından en başarılı modeller olarak öne çıkıyor. XGBoost'un doğruluk skoru diğer modellere göre daha düşük, bu yüzden bu modelin performansı diğerlerinden biraz geride kalıyor.

LİNE CHART (ÇİZGİ GRAFİĞİ)



Bu grafik, modellerin performansının metriklere göre nasıl değiştiğini gösteriyor.

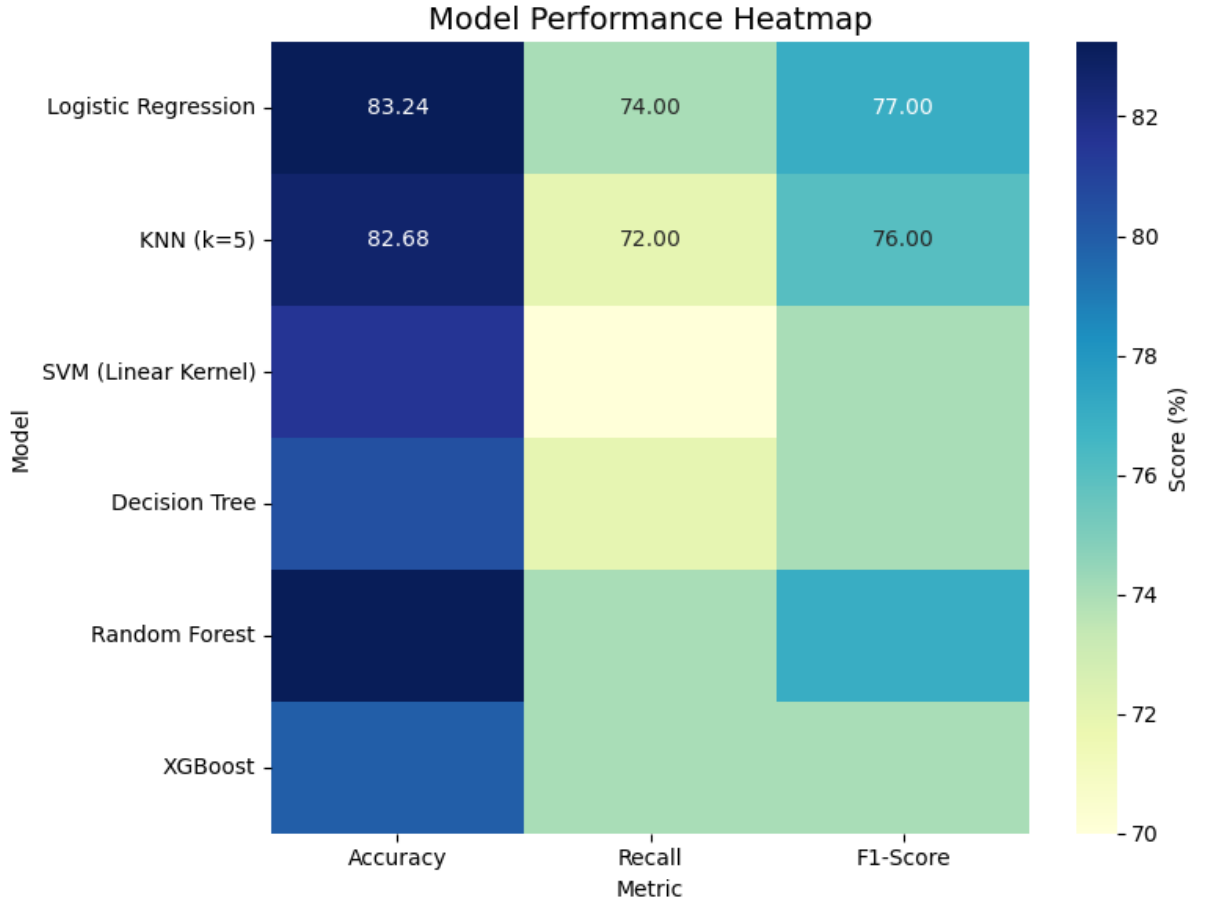
- Genel eğilim incelendiğinde, Accuracy değerleri en üst seviyede (80-83 aralığında) kalıyor.
- Recall değerleri ise genellikle daha düşük (70-74 aralığında), bu da bazı modellerin hatalı negatifleri fazla ürettiğini gösteriyor.
- F1-score, genellikle Recall ile paralel hareket ediyor.

Sonuç:

Random Forest ve Lojistik Regresyon, her üç metrikte de oldukça istikrarlı bir performans gösteriyor.

XGBoost, doğruluk açısından biraz düşük kalsa da Recall açısından başarılıdır.

ISI HARİTASI (HEAT MAP)



Bu ısı haritası, her modelin performans metriklerini (Accuracy, Recall, F1-Score) tek bir tabloda renk tonlarıyla gösterir. Her hücrede metrik değeri bulunur; renk yoğunluğu ise değerin büyüklüğünü ifade ediyor.

- Daha koyu renkler yüksek performans değerlerini (örneğin, %83.24 gibi) temsil ediyor.
- Daha açık renkler ise düşük performans değerlerini (örneğin, %79.89 gibi) gösteriyor.
- Sağdaki renk barı, değerlerin yüzde cinsinden büyüklüğünü görsel olarak anlamamıza yardımcı oluyor.

Sonuç:

Isı haritası, modellerin performansını kıyaslamak için çubuklu grafik gibi geleneksel yöntemlere alternatif olarak, görsel olarak daha bütüncül bir bakış sunar. Bu grafik sayesinde, hangi modellerin genel olarak daha yüksek metriklere sahip olduğu net bir şekilde ortaya konulabilir.

GENEL SONUÇ

En iyi model seçiminde amaç önemlidir.

- Eğer genel başarı (Accuracy & F1-score) isteniyorsa, Random Forest veya Lojistik Regresyon seçilebilir.
- Eğer hatalı negatiflerin minimum olması isteniyorsa (Recall önemliyse), XGBoost veya Random Forest iyi bir seçenek olabilir.
- Daha kompleks modeller (XGBoost ve SVM), basit modellere göre daha düşük performans gösteriyor.