



OCAK 2022

İSTANBUL MEDENİYET
ÜNİVERSİTESİ

VERİ MADENCİLİĞİ DERSİ FİNAL RAPORU

HAZIRLAYANLAR:

YUSUF ALTUNDAL - 19120205057

DİLARA KARADUMAN - 19120205068

BURCU TOKOL - 19120205055

02

PROJENİN AMACI VE PROBLEM

Problemin amacı: evde kalma, dışarı çıkma, aşılama oranlarının artış azalış durumuna göre günlük vakanın artış azalış oran tahmin edilmesi.

Bu problemin çözümünde 2021 verisi kullanılmıştır.

ÇÖZÜM

Problemde classification metodu kullanılmıştır. Verilen bağımsız değişkenlere artıp azalma, sabit kalma durumlarına göre etiketler verilmiştir. Bu etiketler daha sonrasında karar ağacına ve rassal orman sınırlandırıcısına aktarılıp veri tahmin sürecine girilmiştir. Karar ağacı ve rassal orman sayesinde ise tahminin doğruluk katsayısı hesaplanmıştır.

UYGULANAN YÖNTEM

“

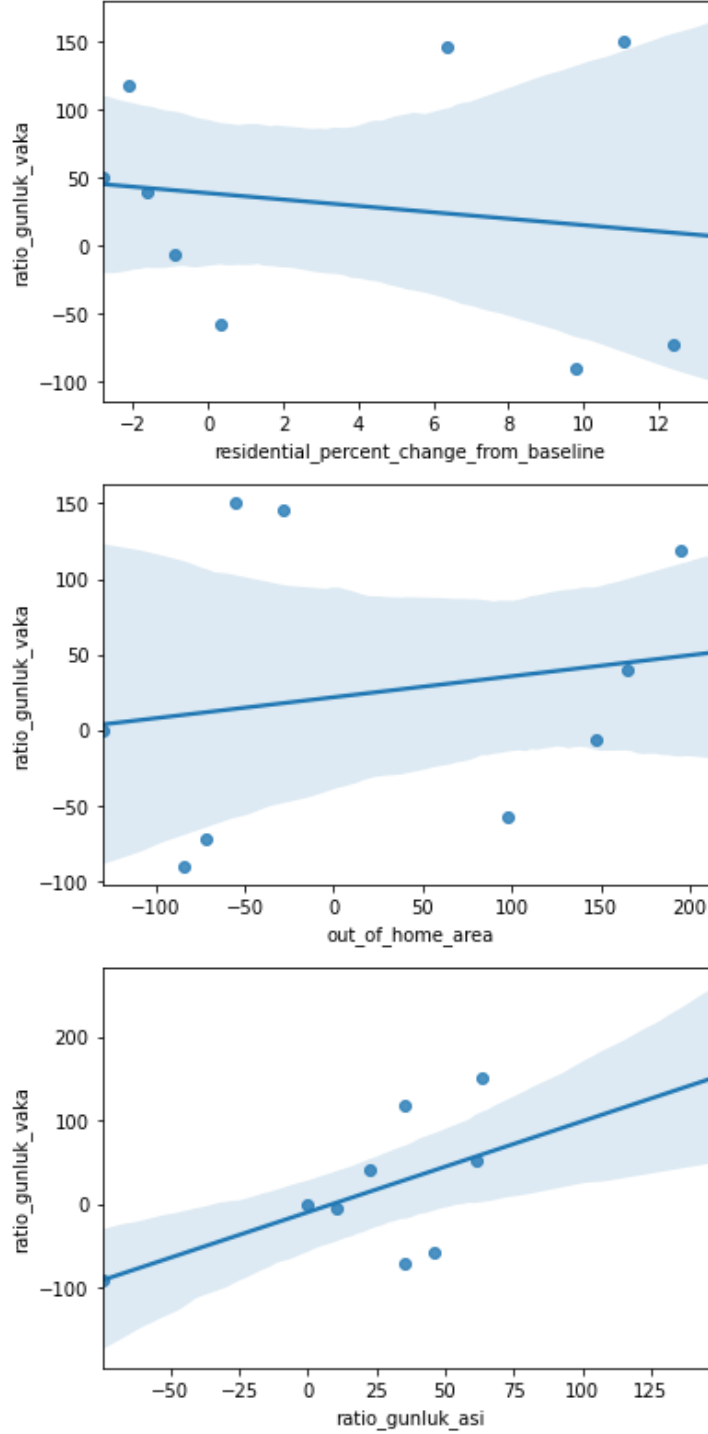
Problemin çözümü için Decision Tree (Karar Ağacı) ve Rassal Orman kullanılmıştır.

”

Öncelikle veri setinde bulunan eksik veya yanlış değerler düzeltildi ve eklendi. Covid-19 sonrası ölüm ve diğer değerler anlık olarak değişmediğinden değişimi net görebilmek için aylara göre ortalama alındı. Veri setleri tek veri üzerinde toplandı. Bağımsız değişkenlerin artma, azalma, sabit kalma durumları için belirli değerler atandı. Karar ağacına ve rassal orman algoritmasına input ve output'lar verildi. Test (%33) ve eğitim verileri oluşturuldu. Verilerin etiketlenmesi yapıldı. Karar ağacında entropy'e göre ağaç oluşturuldu. Rassal ormanda ise gini'ye göre oluşturuldu. Karar ağacı ve rassal orman, verilen etiketler aracılığı ile çizildi. Test ve eğitim verileri baz alınarak doğruluk değeri bulundu.

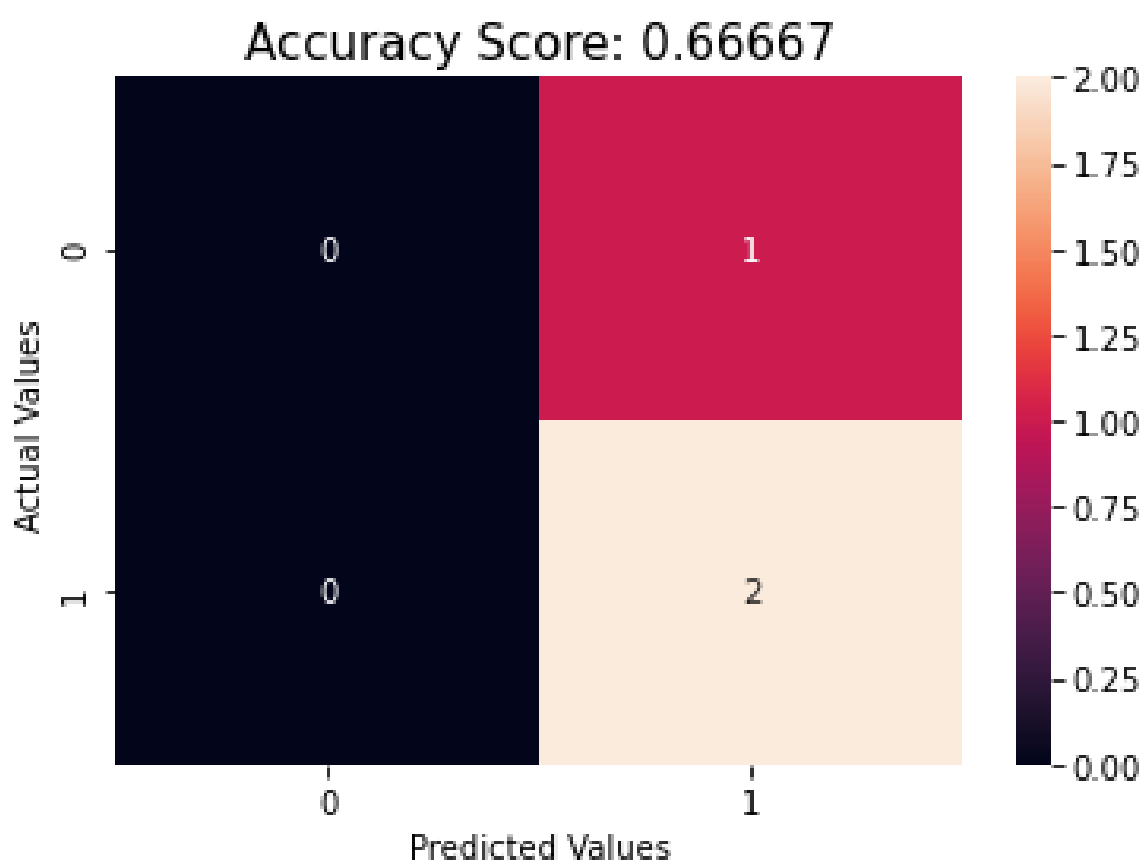
04

GRAFİKLER



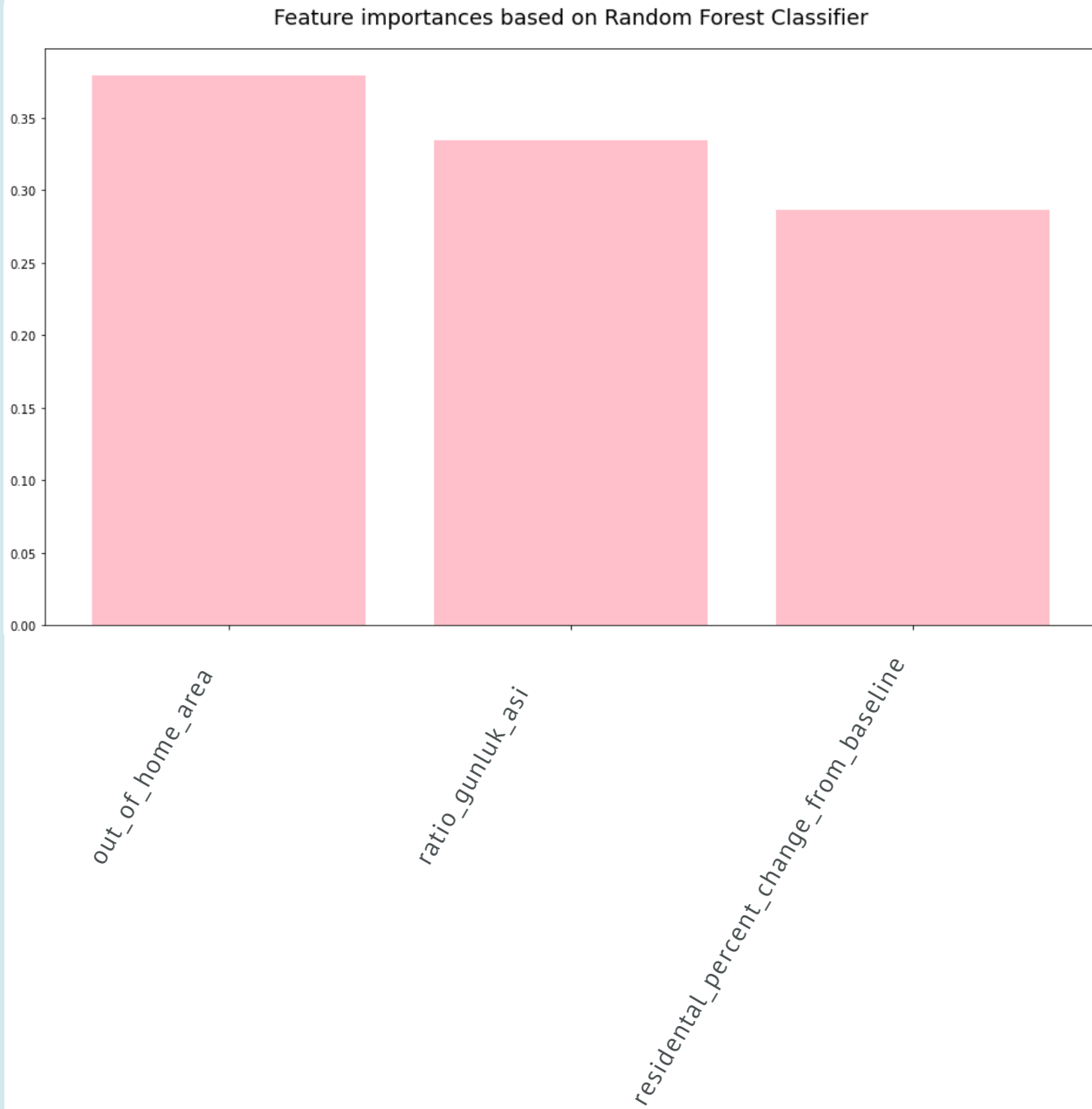
Evde kalma, dışarı çıkma ve aşı verilerinin günlük vakalarla karşılaştırılması

05



Confusion Matrix

06



Özelliklerin tahmindeki önem sıralarının grafikleştirilmesi

07

SONUÇ

Algoritma hem karar ağacında hem de rassal ormanda 0.67 başarı oranı ile tahmin sürecini tamamlamakta. Genel olarak evde kalma ve aşılama oranı arttıkça vaka oranının azaldığı, dışarı çıkma oranı arttıkça ise vaka oranının arttığı şeklinde yorumlanabilir.

