

GDZ Elektrik Datathon 2024

Sunumu

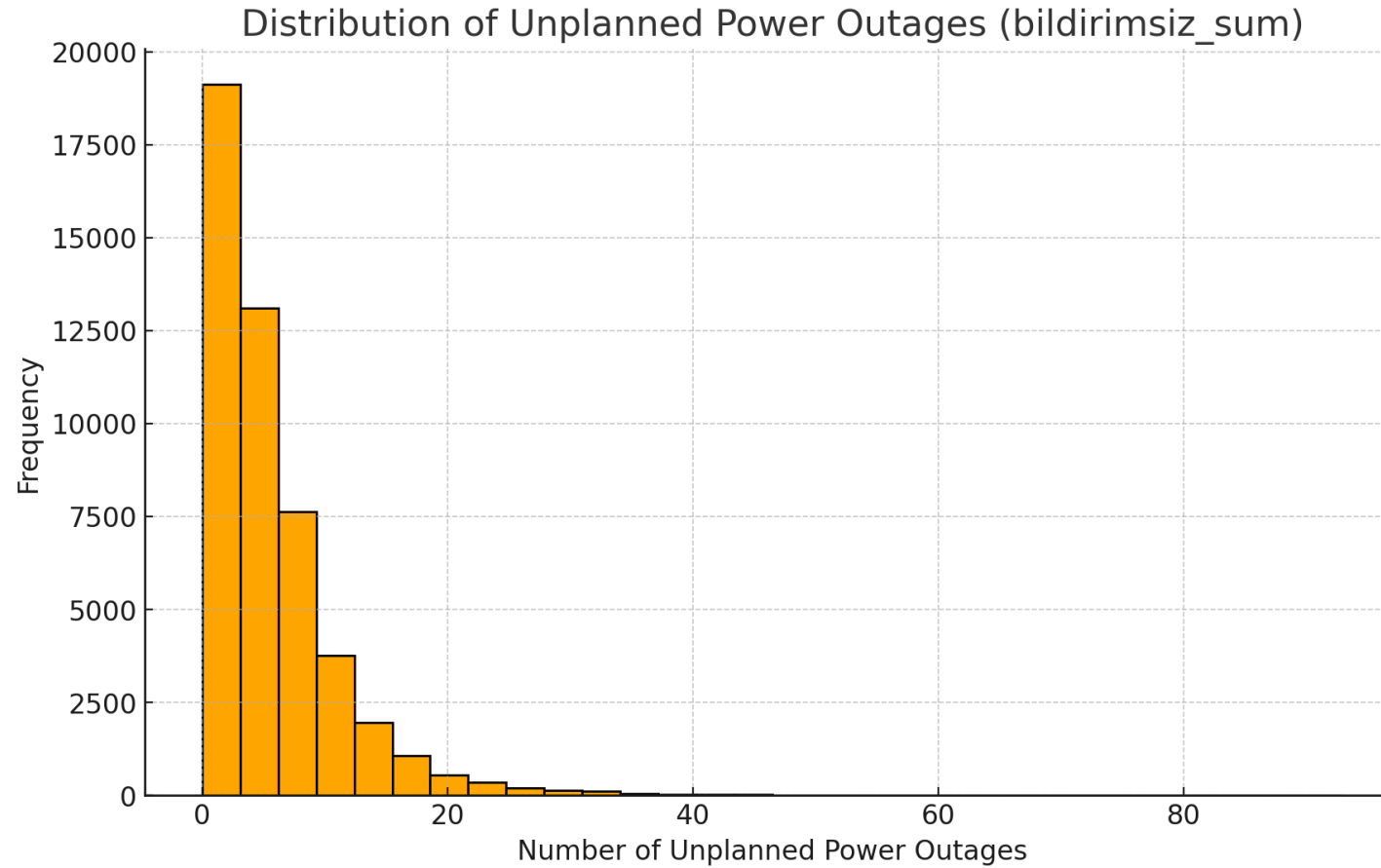
Mert Bayraktar

Sunum Akışı

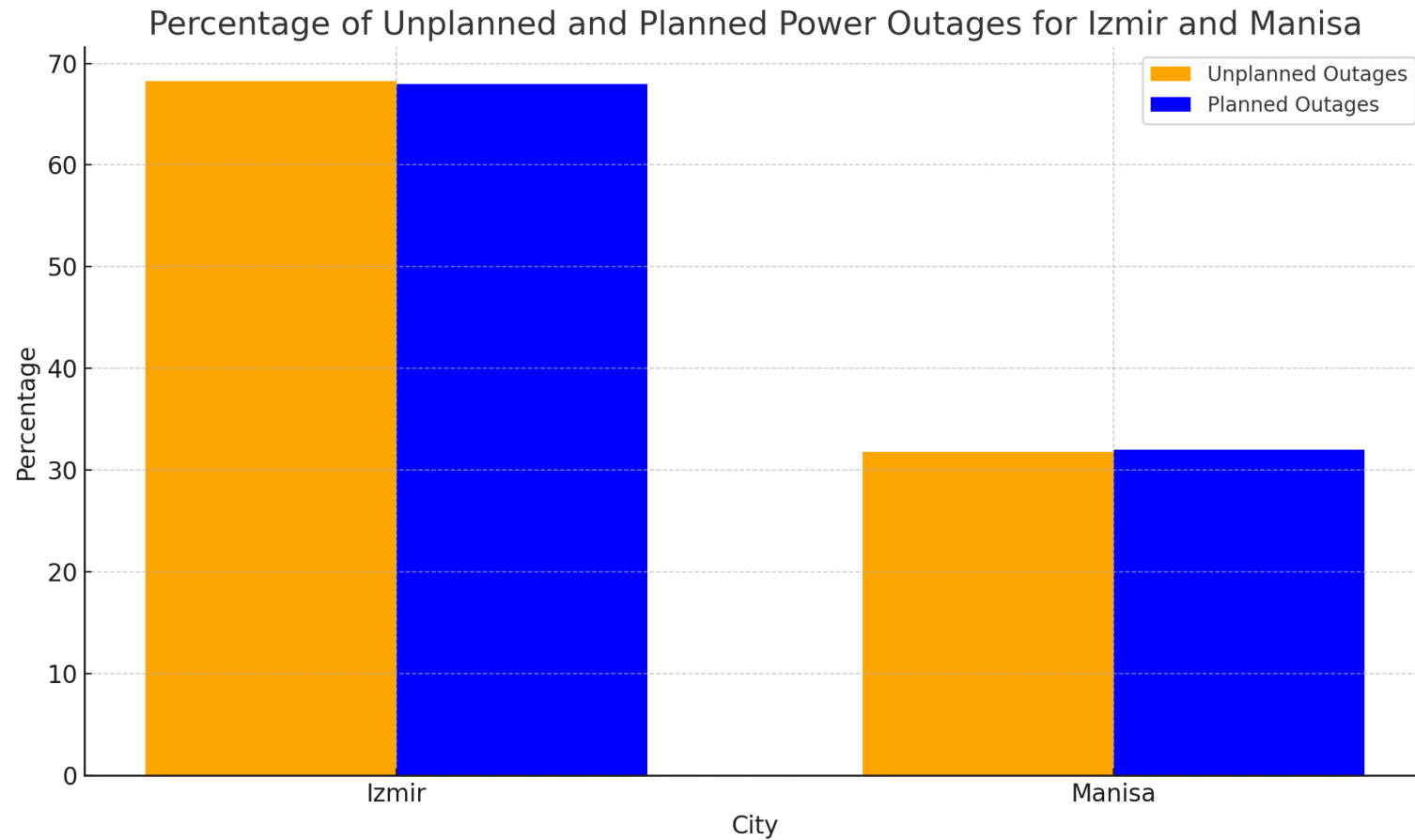
- Problem
- Veri Analizi
- Öznitelik Mühendisliği
- Modelleme Stratejisi
- Tahminleme Sonrası Sezgisel Yaklaşımlar

Problem

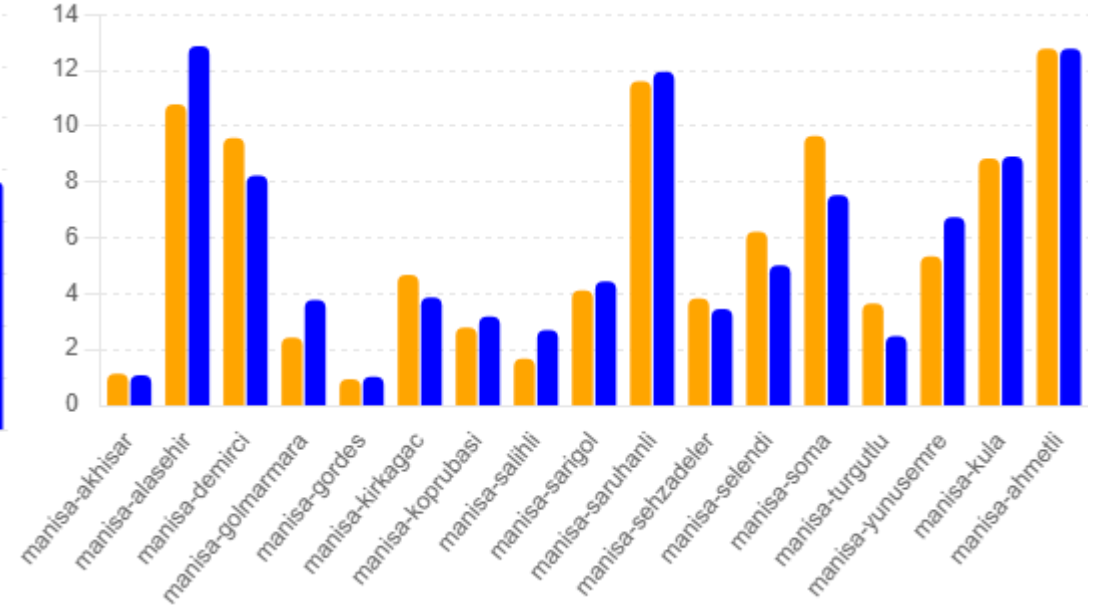
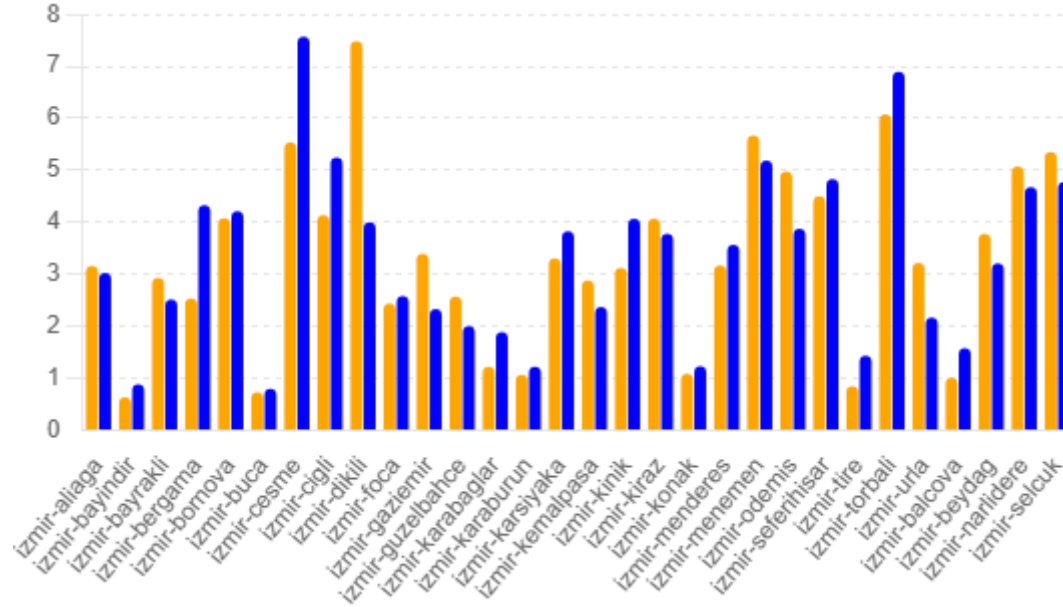
- GDZ'nin hizmet verdiği bölgelerde ilçe bazlı ileriye yönelik günlük plansız kesinti sayılarının tahminlenmesi.
 - Eğitim verisi 1 Ocak 2021 - 31 Ocak 2024 arası.
 - Test verisi 1 Şubat – 29 Şubat 2024 arası.
 - Başarı metriği **Mean Absolute Error (MAE)**.



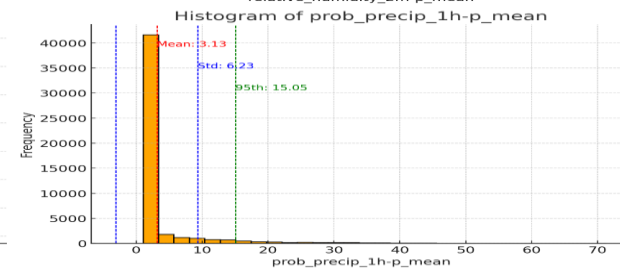
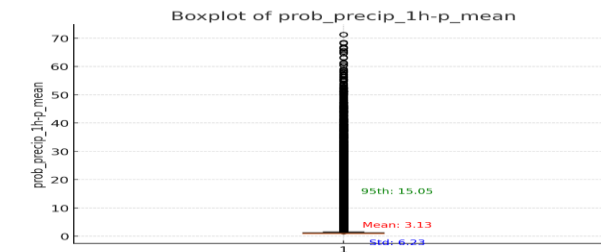
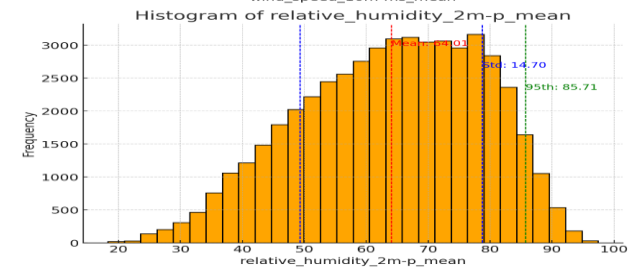
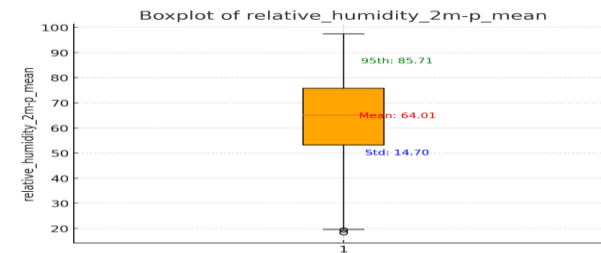
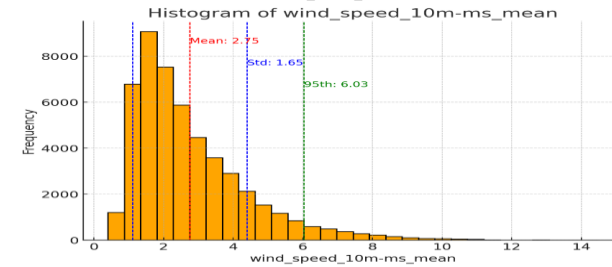
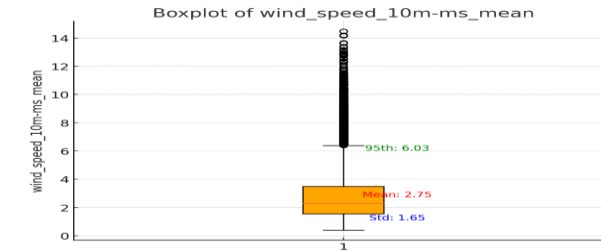
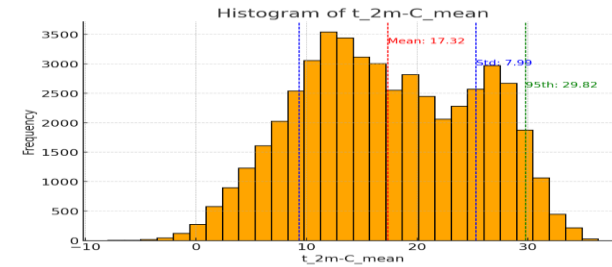
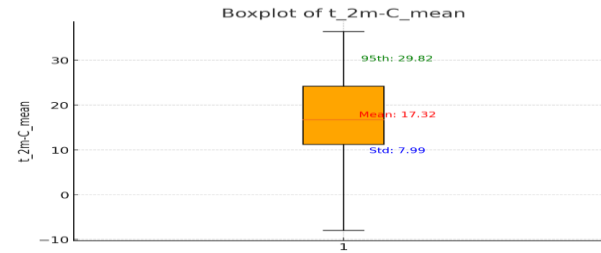
Veri Analizi



Veri Analizi



Veri Analizi



Öznitelik Mühendisliği

- **Hava durumu verileri**
 - İlçe ve tarih bazında günlük ortalama, maks., min., std., toplam değerler
 - Aşırı sıcak, aşırı soğuk, yüksek nemli, yüksek rüzgarlı günler
 - Diğer kötü hava koşulları (bulutluluk-yağış, rüzgar hızı-yağış).
- **Takvim öznitelikleri**
 - Yıl, ay, çeyrek
 - Haftanın günü, yılın günü, ayın günü
 - Ayın başı, sonu olma durumu
 - Çeyreğin başı, sonu olma durumu
 - Yılın başı, sonu olma durumu
 - Ay ve yılın günü döngüsel öznitelikler
- **Gecikme öznitelikleri**
 - 5 gün arayla 50 günlük planlı kesinti öznitelikleri
 - 3 günlük hava durumu öznitelikleri
- **Grup bazlı öznitelikler**
 - Sezonluk, yıllık, aylık ilçe bazında planlı kesinti öznitelikleri
- **Kümeleme öznitelikleri**
 - Coğrafi olarak yakın ilcelerin ortalama hava durumu verileri

Kaynakça:

<https://ieeexplore.ieee.org/document/9046816?denied=>

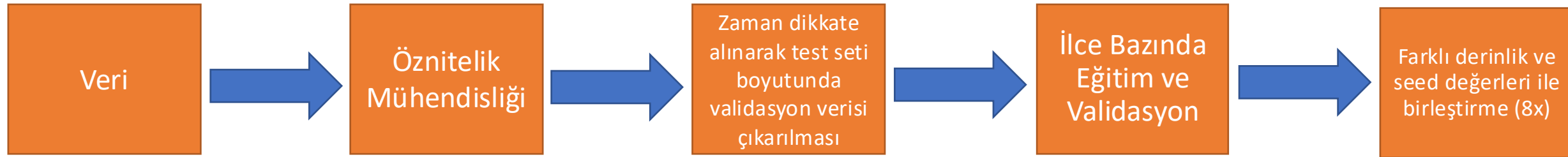
<https://ieeexplore.ieee.org/document/8656482?denied=>

<https://www.sciencedirect.com/science/article/pii/S2212094722000664>

Modelleme

- **Catboost**
- LGBM
- XGBoost
- Prophet
- AutoGluon (PatchTST, ARIMA, Theta etc.)

Modelleme



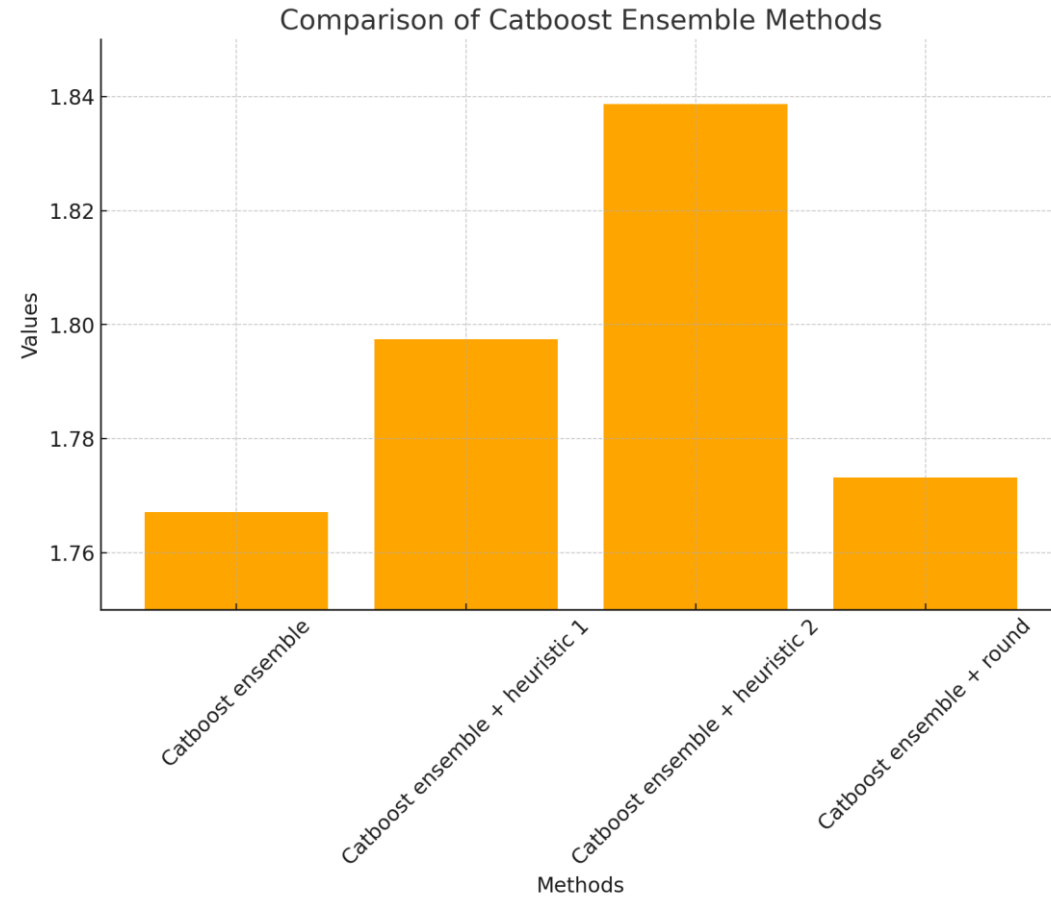
Başarı Metriği

- İlk aşamada başarı metriği ve kayıp fonksiyonu olarak MAE kullanıldı.
- İlk aşamada plansız kesintinin yüksek olduğu ilçelerde MAE skoru yüksek geliyordu (Ocak ayında yüksek kesinti olduğu varsayıldı).
- İkinci aşamada modelin düşük sayıdaki kesintilerde daha iyi sonuç vermesi için başarı metriği ve kayıp fonksiyonu olarak SMAPE ve MAPE, MAE kombinasyonları denendi. En iyi sonuç her ikisinde de MAPE ile elde edildi.

Tahminleme Sonrası

- Model her ilçe için ayrı olarak eğitildi. Düşük skorlu ilçelerde bir takım sezgisel yaklaşımlar denendi. Aynı yaklaşımlar az validasyon verisi bulunan ilçeler için de denendi. Bunlar:
 - Model tahminlerine ek olarak, belirli bir eşik değerinin üstünde hatalı tahmin yapılan ilçelerde sırasıyla özyinelemeli olarak son 10, 15, 20, 25, 30, 40 gün için ortalama değerler alındı. Bu değerler ilgil ilçeler için tahmin değerleri olarak atandı.
 - Model tahminlerine ek olarak, belirli bir eşik değerinin üstünde hatalı tahmin yapılan ilçelerde en sık rastlanan değerler tahmin değeri olarak atandı.
- Tahminlenen değerler tam sayı olarak alındı.

Karşılaştırma



SORU CEVAP