

GDZ ELEKTRİK DATATHON 24

FİNAL SUNUMU

Pikachow

Anıl Öztürk – Ahmet Tarık Karakaş

SUNUM AKIŞI

VERİ ANALİZİ

Ham verideki değerlerin incelenmesi ve anlamlandırılması

VERİ KULLANIMI

Üretilmiş veriler

ÇÖZÜM MİMARİSİ

Doğrulama mimarisi

Optimizasyon mimarisi

Tahmin mimarisi

SONUÇ ve ÇIKARIMLAR

Performans analizi

Uygulanabilirlik

Potansiyel geliştirmeler

PROBLEM

İlçe bazlı günlük plansız kesinti sayıları tahmin edilebilir mi?

- 2024 Şubat'a kadar olan sürece ait bütün kesintiler verilmiş
- 2024 Şubat ayı döneminde olacak kesintilerin tahminlenmesi bekleniyor
- Hata **MAE** ile değerlendiriliyor

LİTERATÜR

MEVCUT YAKLAŞIMLAR

1. KULLANILAN ÖZGÜN VERİLER

- Geçmiş kesintiler
- Kesinti sebepleri
- Hava durumu ve hava durumu tahminleri
- Bakım tarihleri

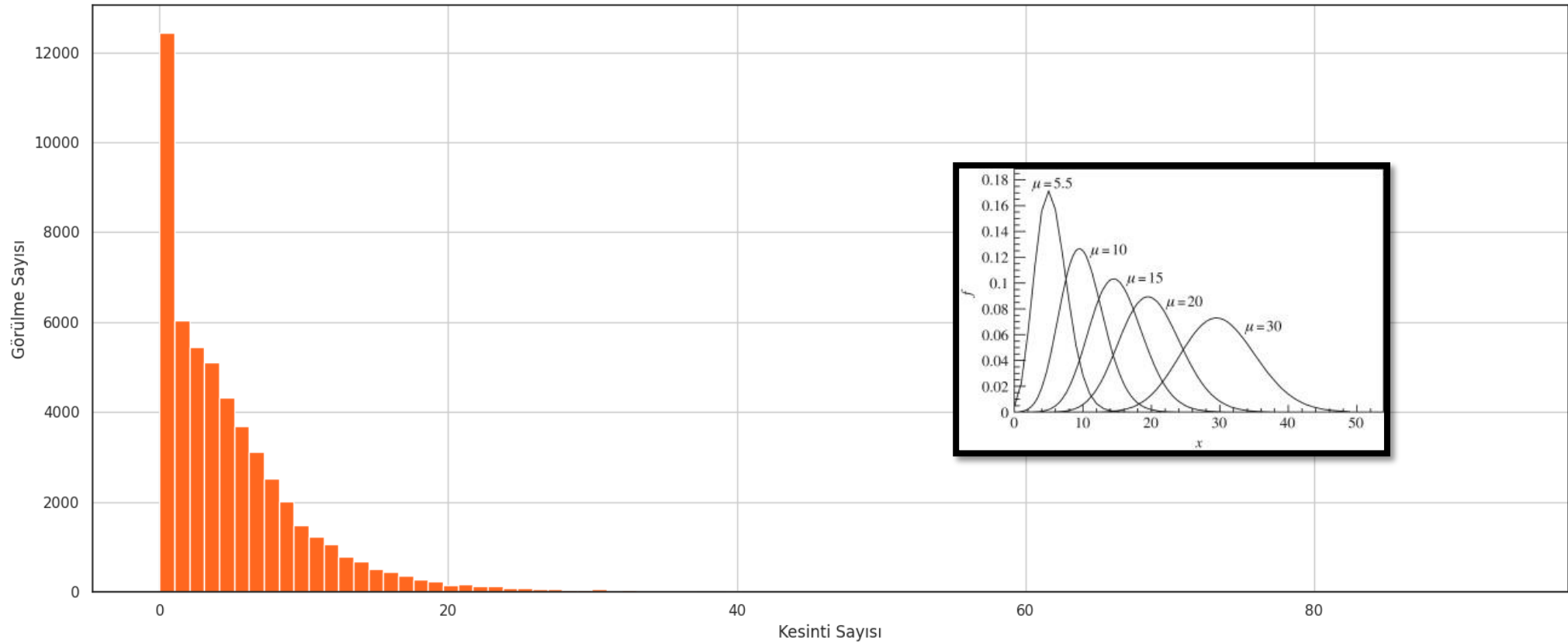
2. PROBLEMİ MODELLEME TÜRLERİ

- GBDT
- RNN
- Lineer modeller

- https://www.researchgate.net/publication/329692335_Power_outage_forecasting_Methods_results_and_uncertainty
- <https://www.ibm.com/docs/en/environmental-intel-suite?topic=solutions-outage-prediction>
- <https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/stg2.12109>
- <https://ieeexplore.ieee.org/document/10147146/>
- <https://engineering.tamu.edu/news/2024/02/predicting-power-outages.html>

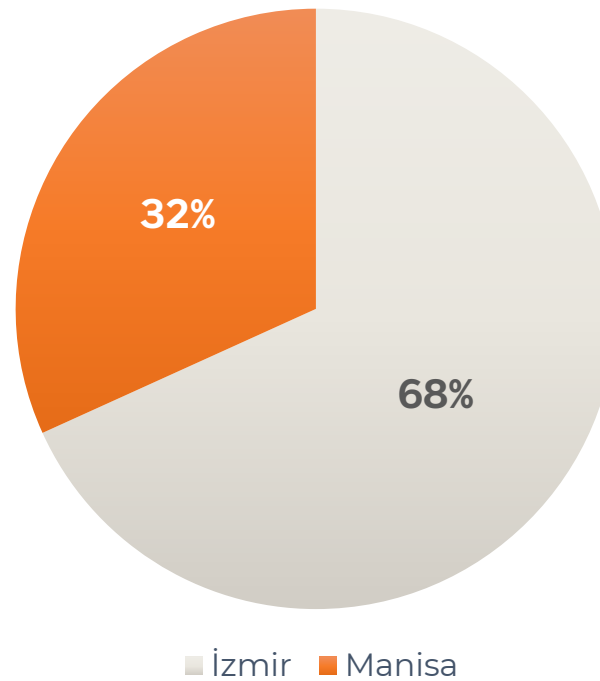
VERİ ANALİZİ

HEDEF DEĞERİN DAĞILIMI (BİLDİRİMSİZ KESİNTİ SAYISI)



VERİ ANALİZİ

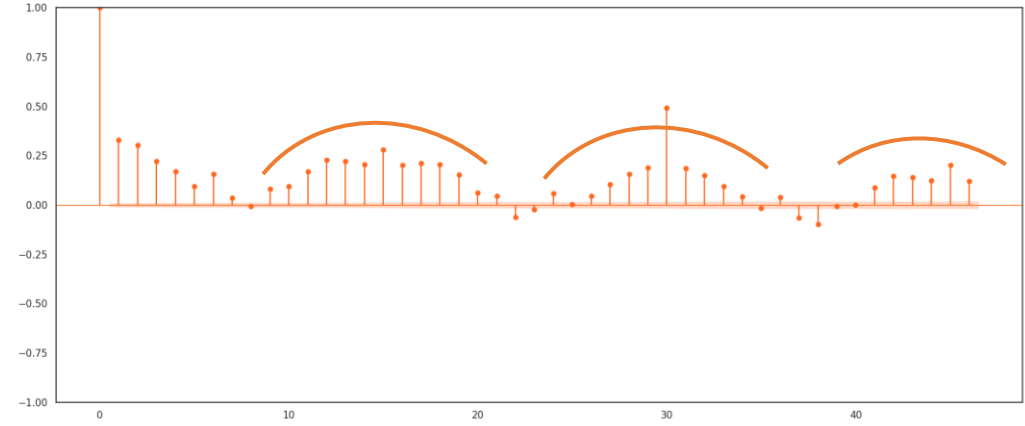
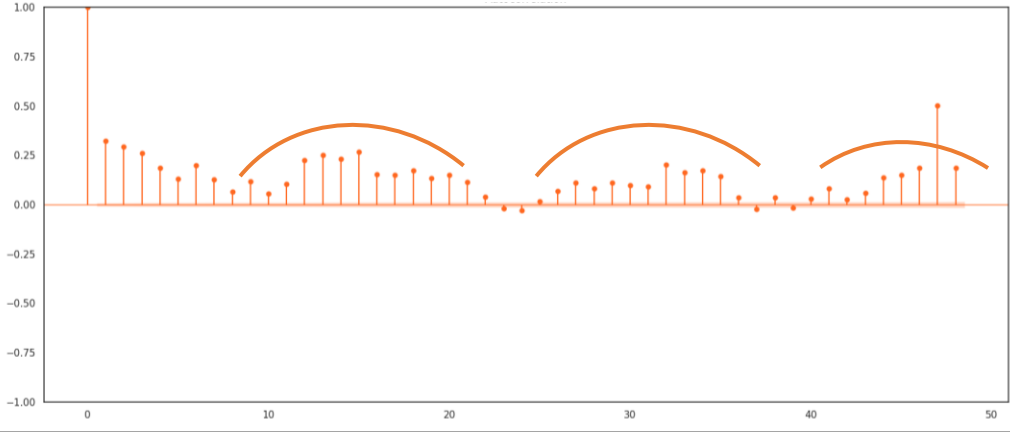
BİLDİRİMSİZ KESİNTİLERİN İLLERE GÖRE DAĞILIMI (%)



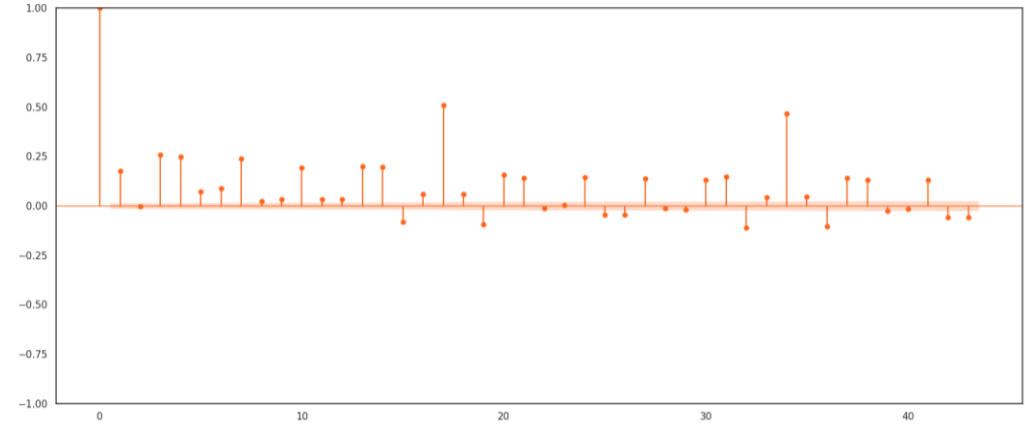
VERİ ANALİZİ

OTOKORELASYON - BİLDİRİMSİZ KESİNTİ SAYISI

Bütün Veri



İzmir

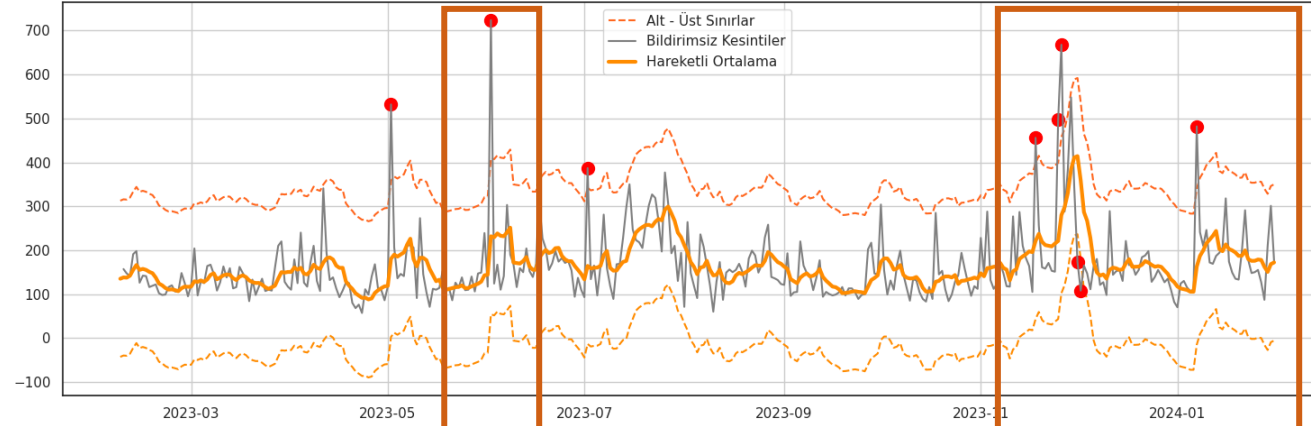


Manisa

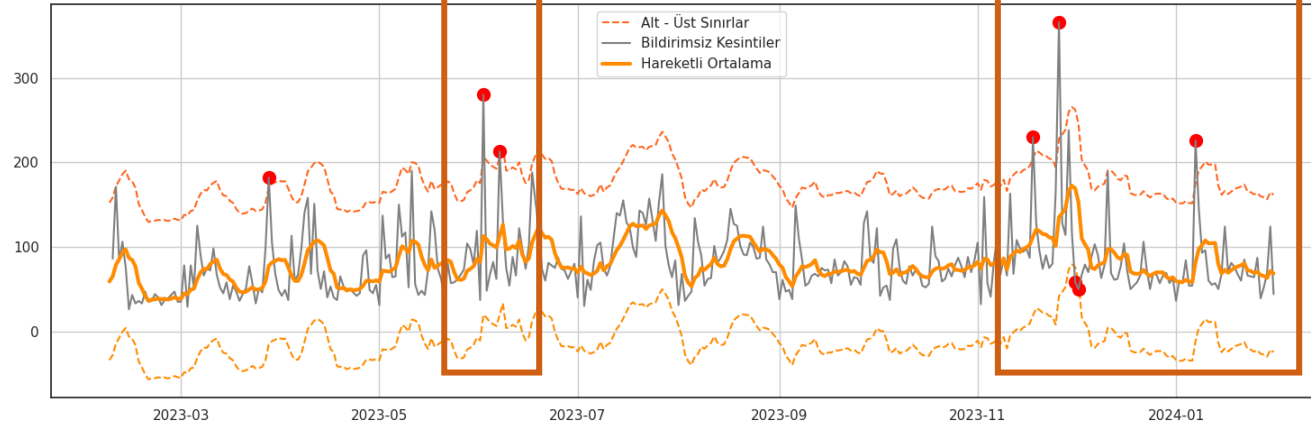
VERİ ANALİZİ

İL BAZINDA SON 1 YILDAKİ BİLDİRİMSİZ KESİNTİ SAYISI

İzmir

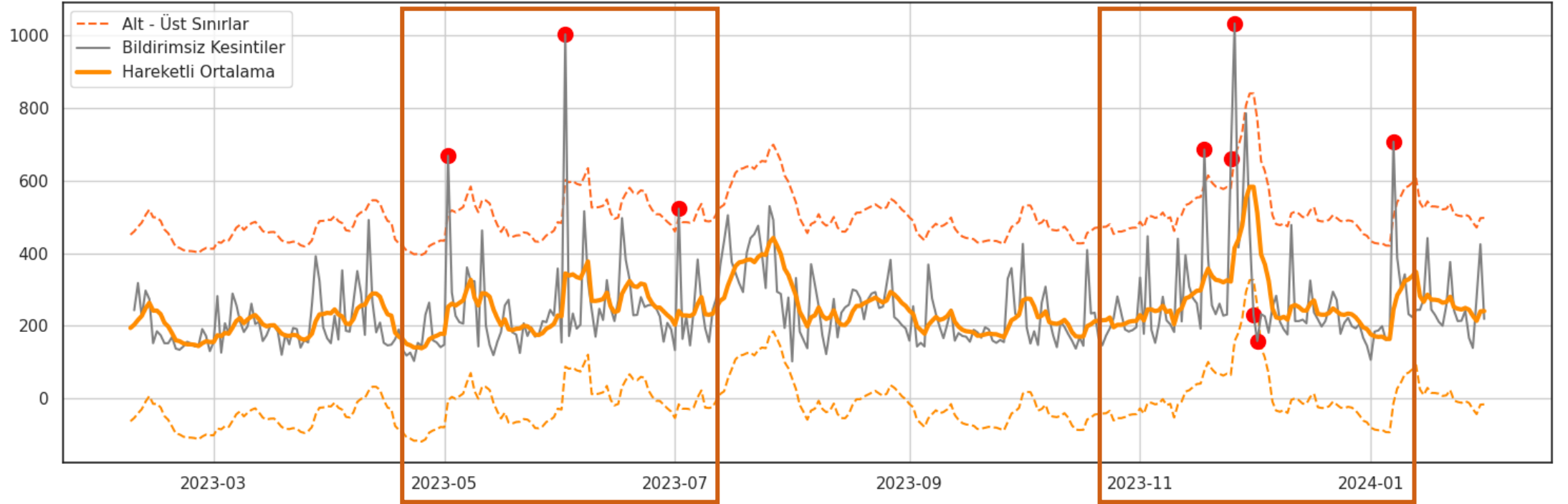


Manisa



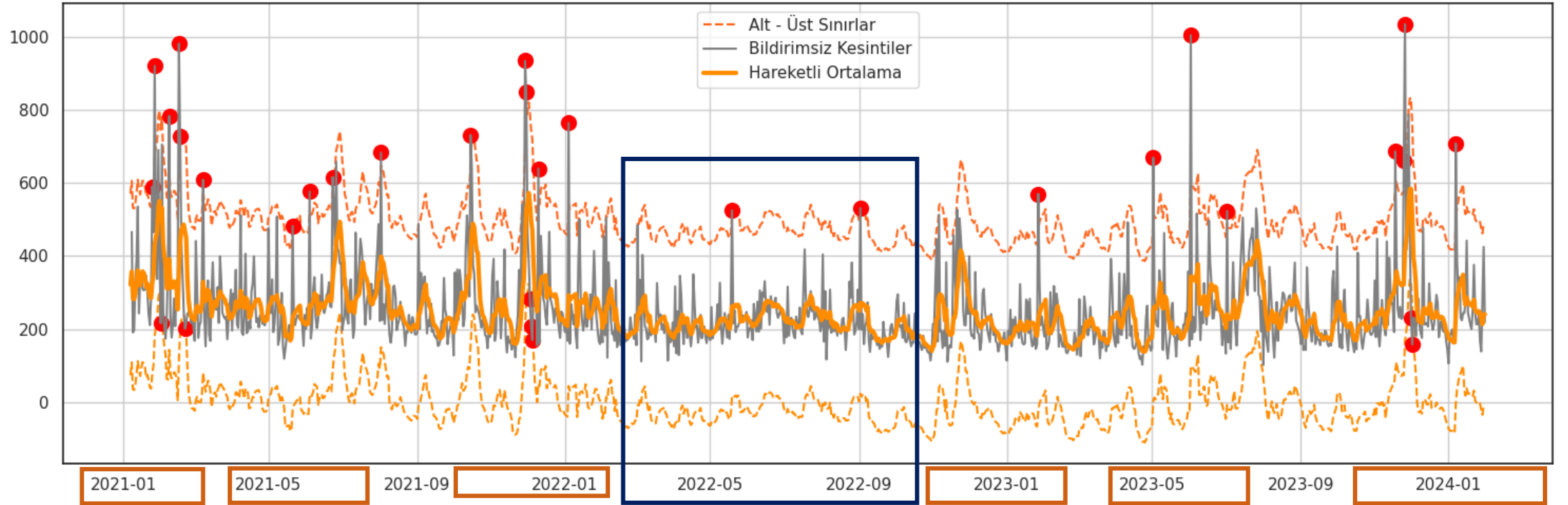
VERİ ANALİZİ

BÜTÜN VERİDE SON 1 YILDAKİ BİLDİRİMSİZ KESİNTİ SAYISI



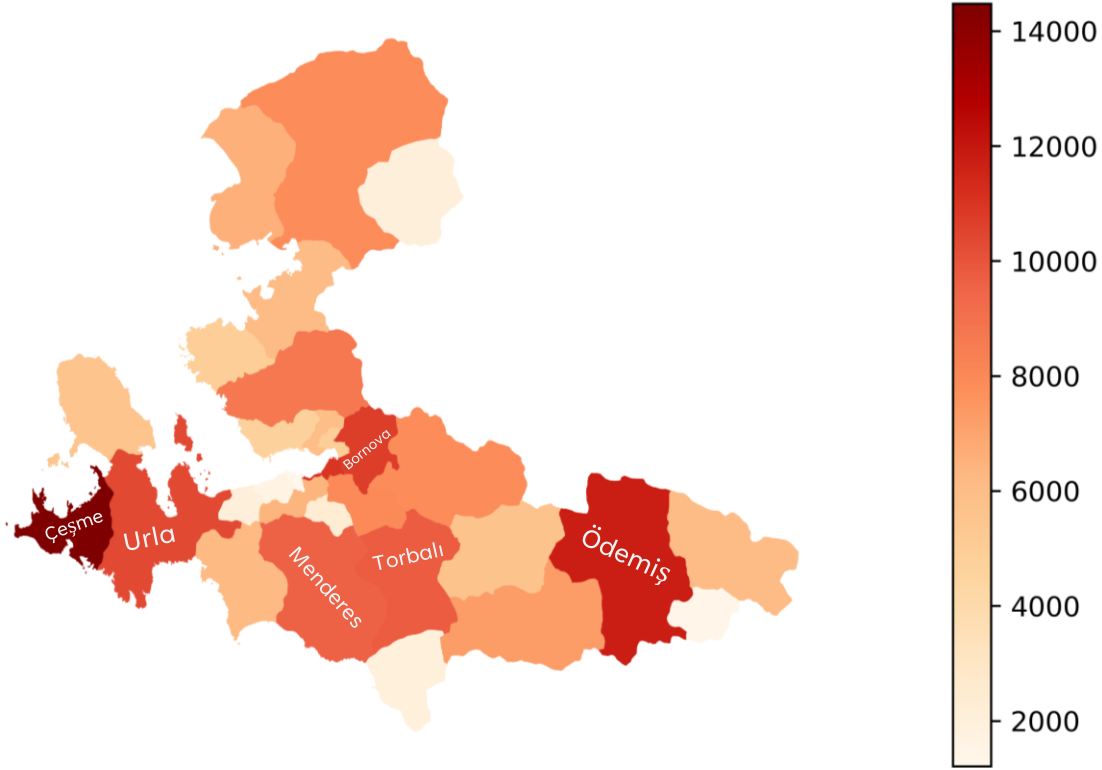
VERİ ANALİZİ

TÜM ZAMANLARDA BİLDİRİMSİZ KESİNTİ SAYISI

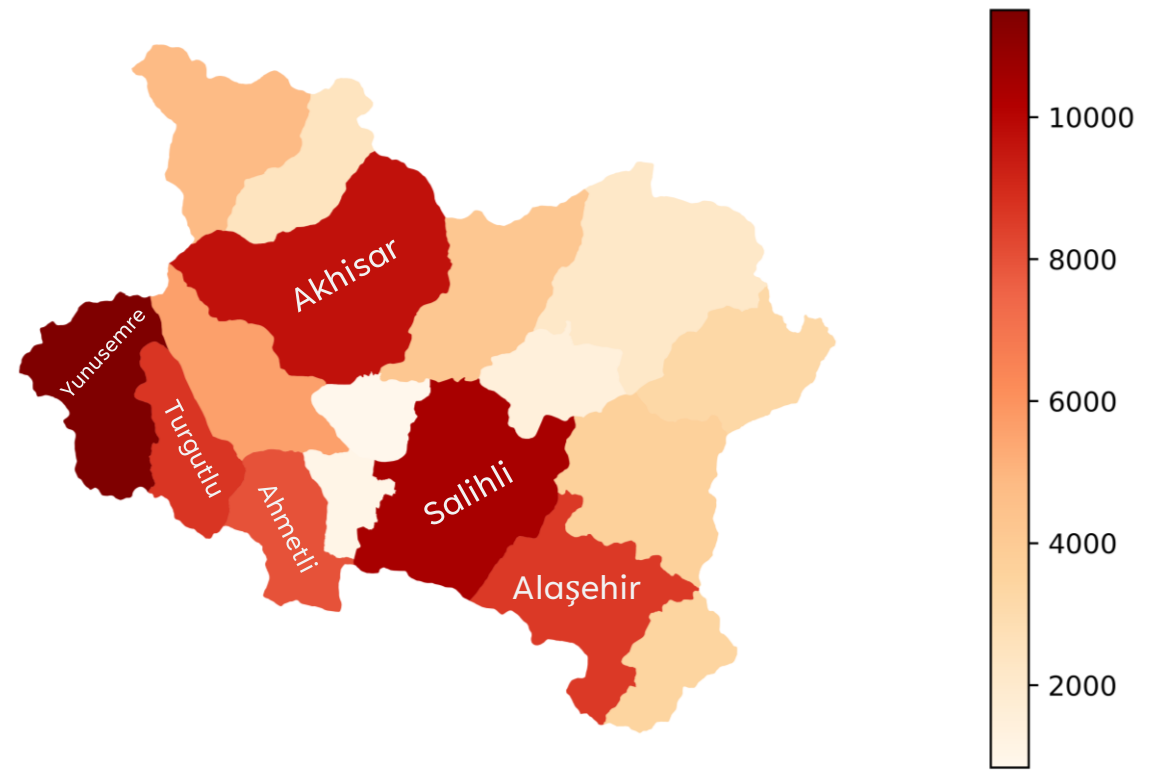


VERİ ANALİZİ

BİLDİRİMSİZ KESİNTİLERİN İLÇELERE GÖRE DAĞILIMI



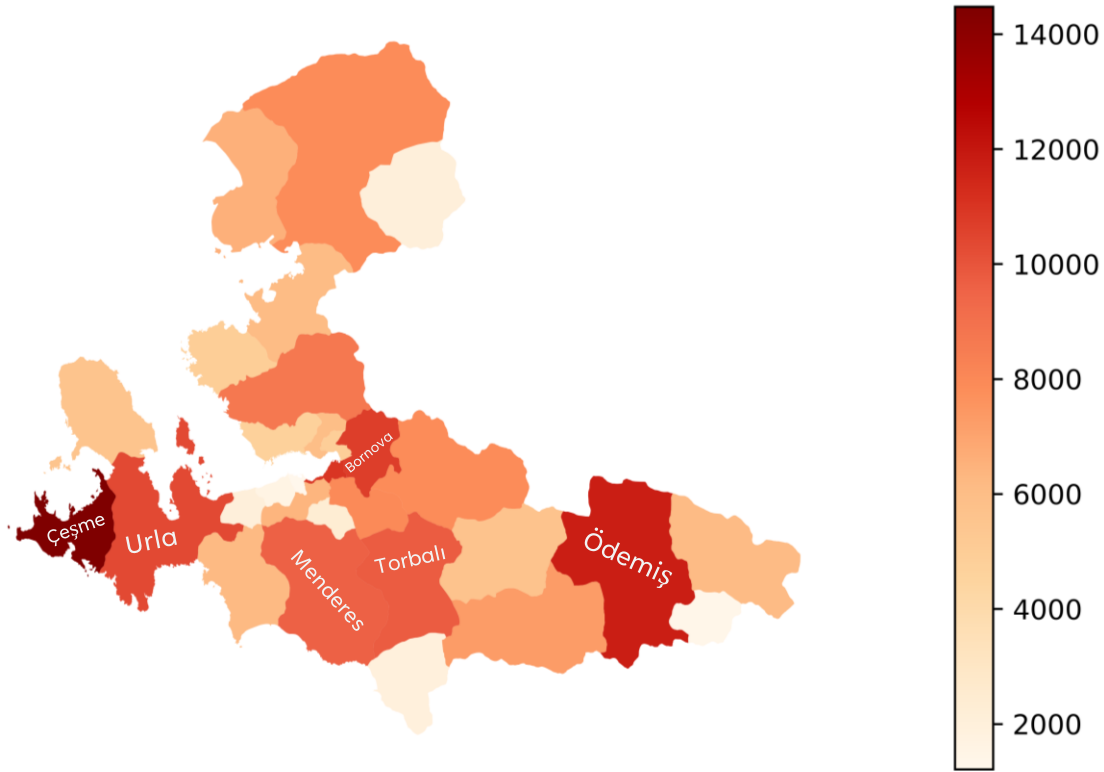
İzmir



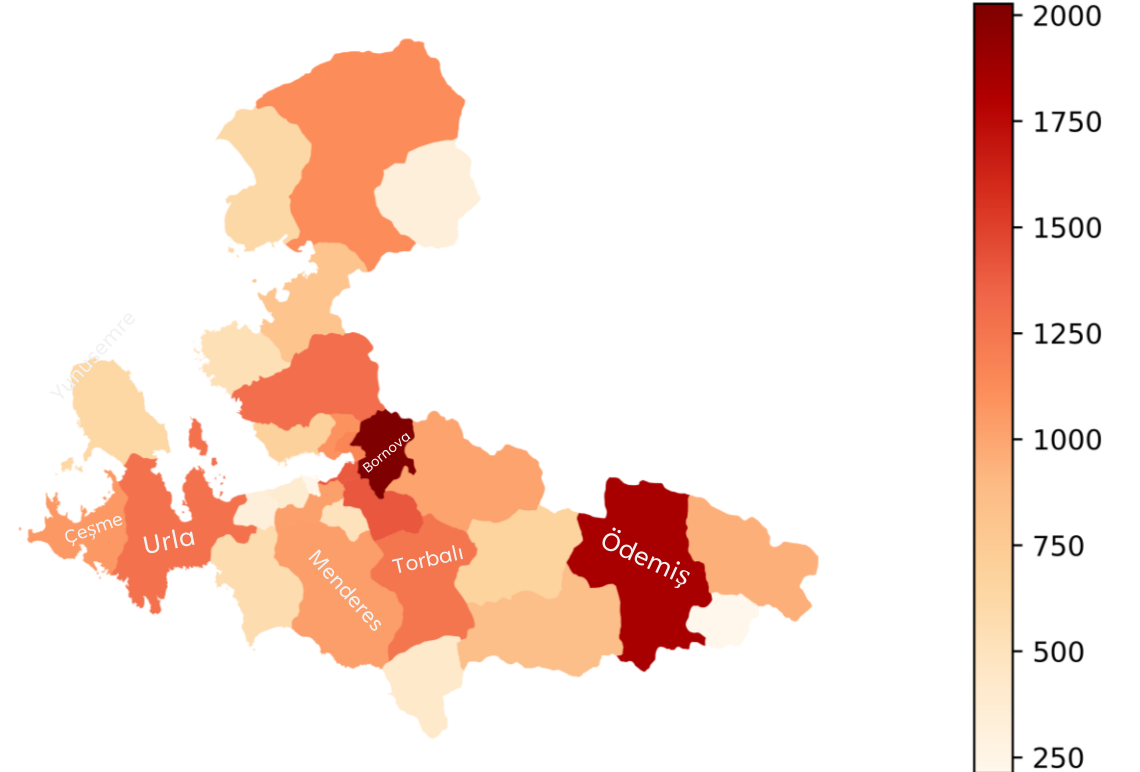
Manisa

VERİ ANALİZİ

KESİNTİLERİN İZMİR'DEKİ DAĞILIMI



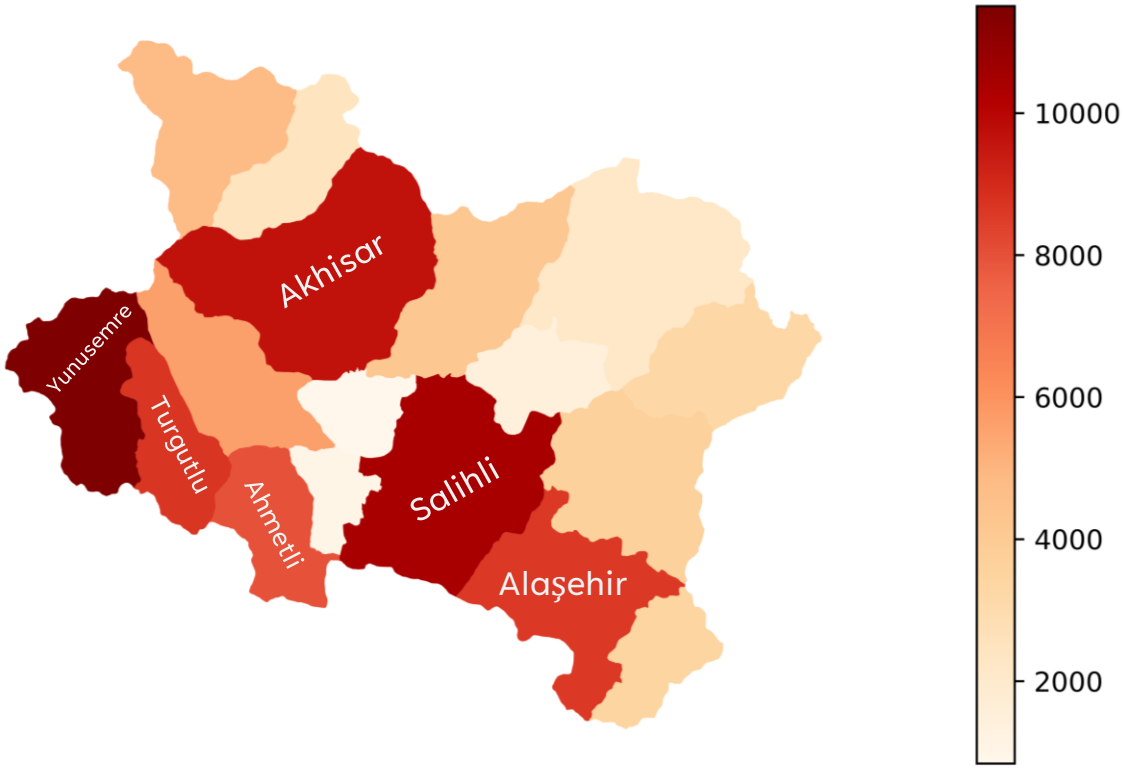
Bildirimsiz Kesinti



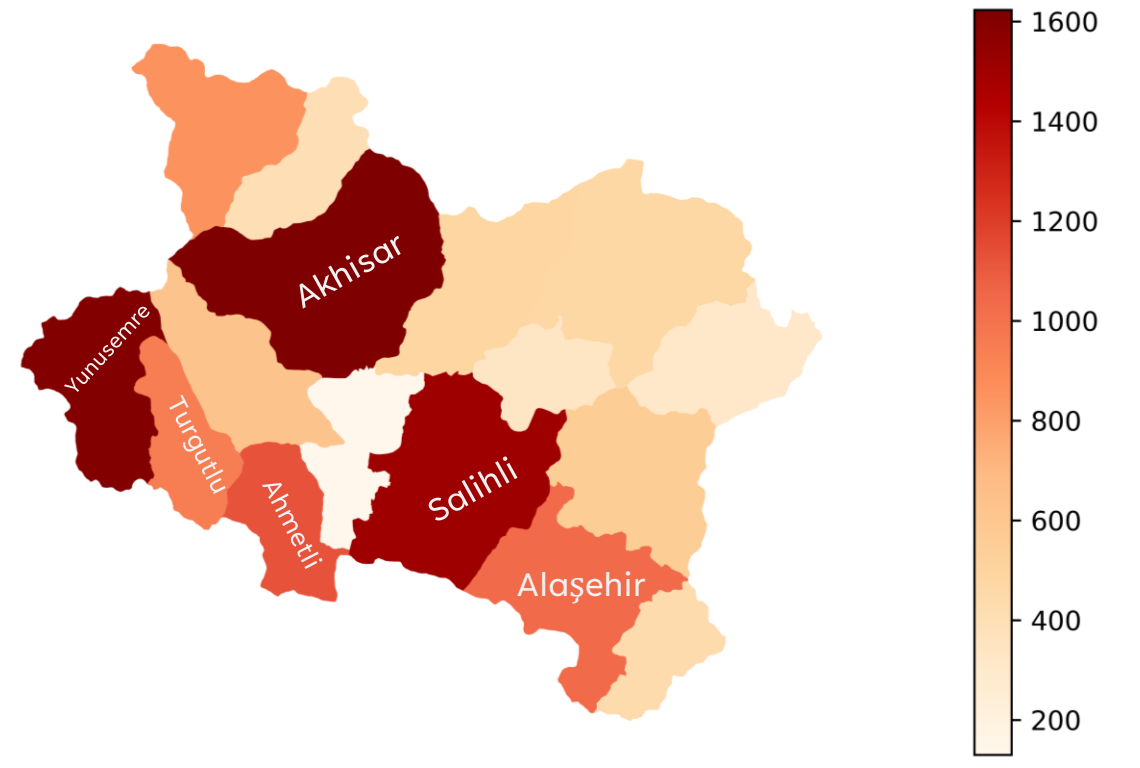
Bildirimli Kesinti

VERİ ANALİZİ

KESİNTİLERİN MANİSA'DAKİ DAĞILIMI



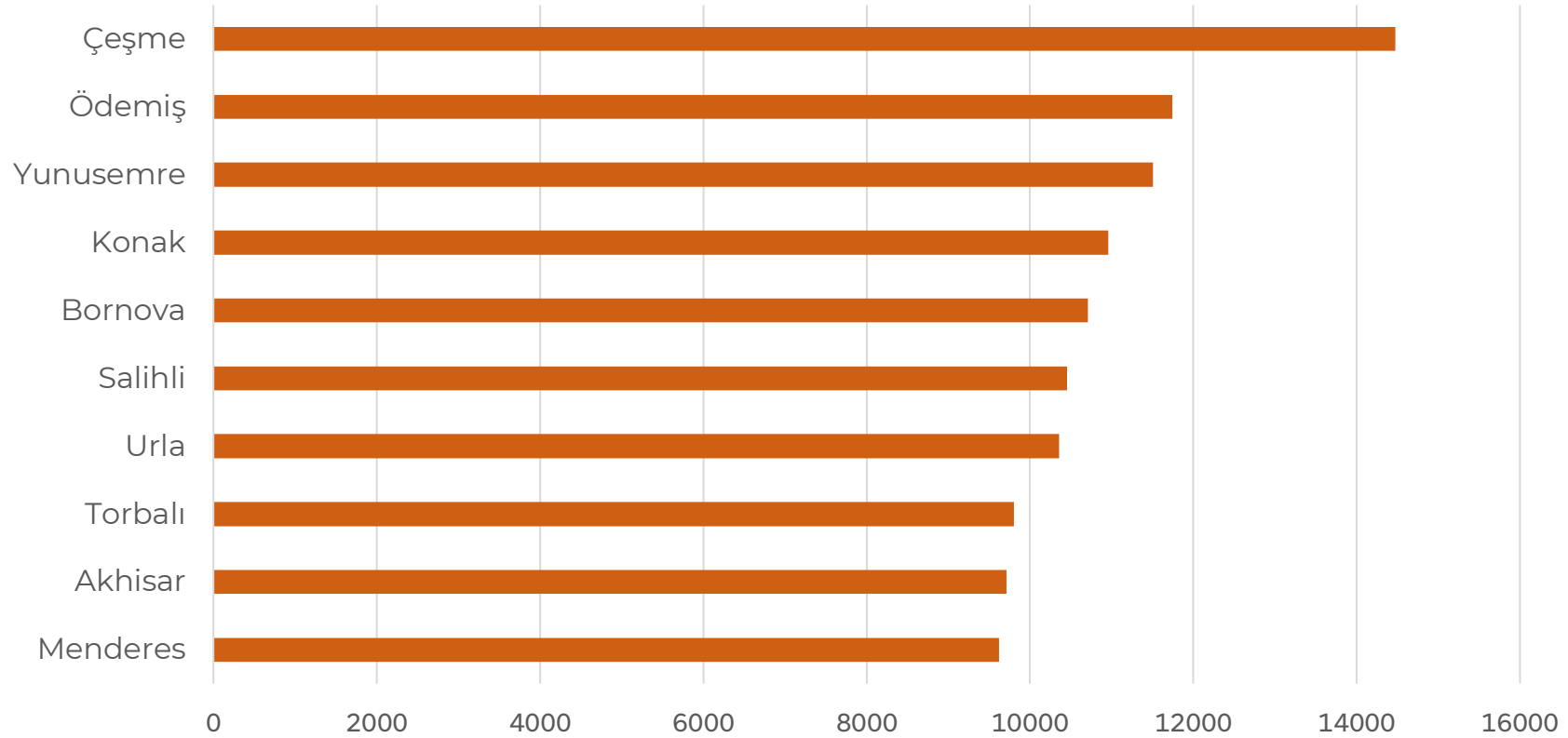
Bildirimsiz Kesinti



Bildirimli Kesinti

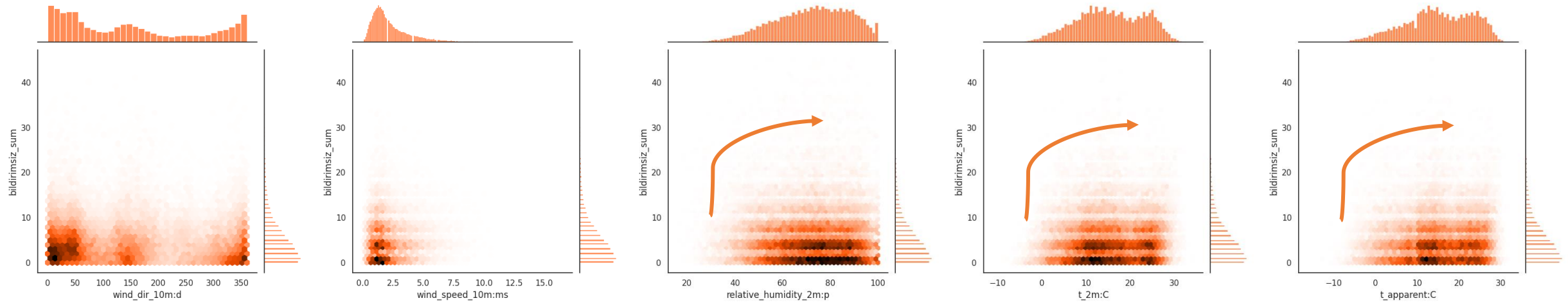
VERİ ANALİZİ

BİLDİRİMSİZ KESİNTİLERİN EN YÜKSEK OLDUĞU 10 İLÇE



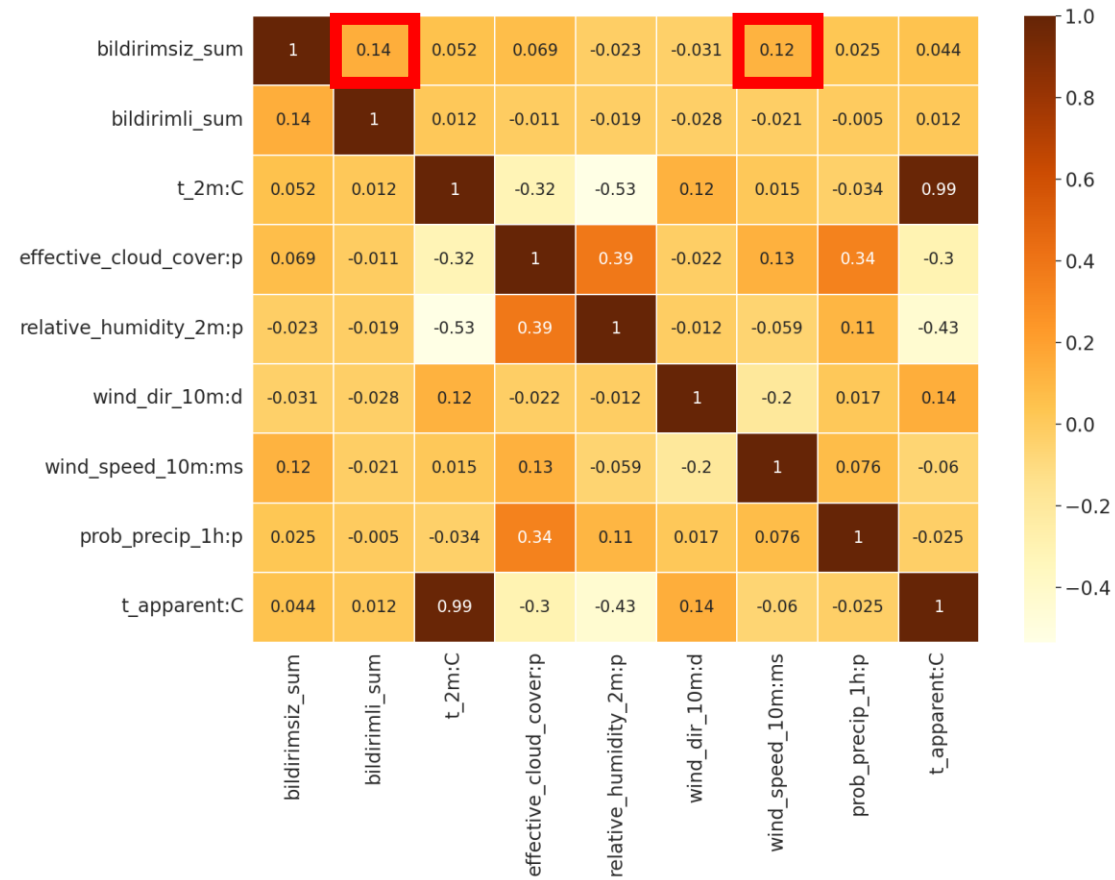
VERİ ANALİZİ

HAVA DURUMU İLE KESİNTİLERİN İLİŞKİSİ



VERİ ANALİZİ

VERİ İÇİ KORELASYON



KULLANILAN VERİ

SENTEZ VERİ

- **İlçeler** ve **bütün veri** bazında:
 - Günlük ortalama, maksimum, ve quantile hava durumu değerleri
 - Geçen aya ait toplam bildirimsiz kesinti sayısı
- Bildirimsiz kesinti sayısı üzerinde:
 - Expanding
 - Rolling (3, 7, 14, 29, 58)İstatistikler
- Takvim değerleri (*ay, haftanın günü, çeyrek, yılın haftası*)
- İlgili gün için ilçelerin saatlik hava durumu değerleri
- İleriki 3-7-15 gün içerisinde bayram olma durumu
- İlçe bazında son ay kesintili ve kesintisiz gün sayısı

MODEL SEÇİMİ

- CatBoost
- LightGBM
- XGBoost
- Local Modeller (ARIMA, Croston, EMA vs.)
- Global NN Modeller (RNN, TFT, TimesNet vs.)

MODEL SEÇİMİ

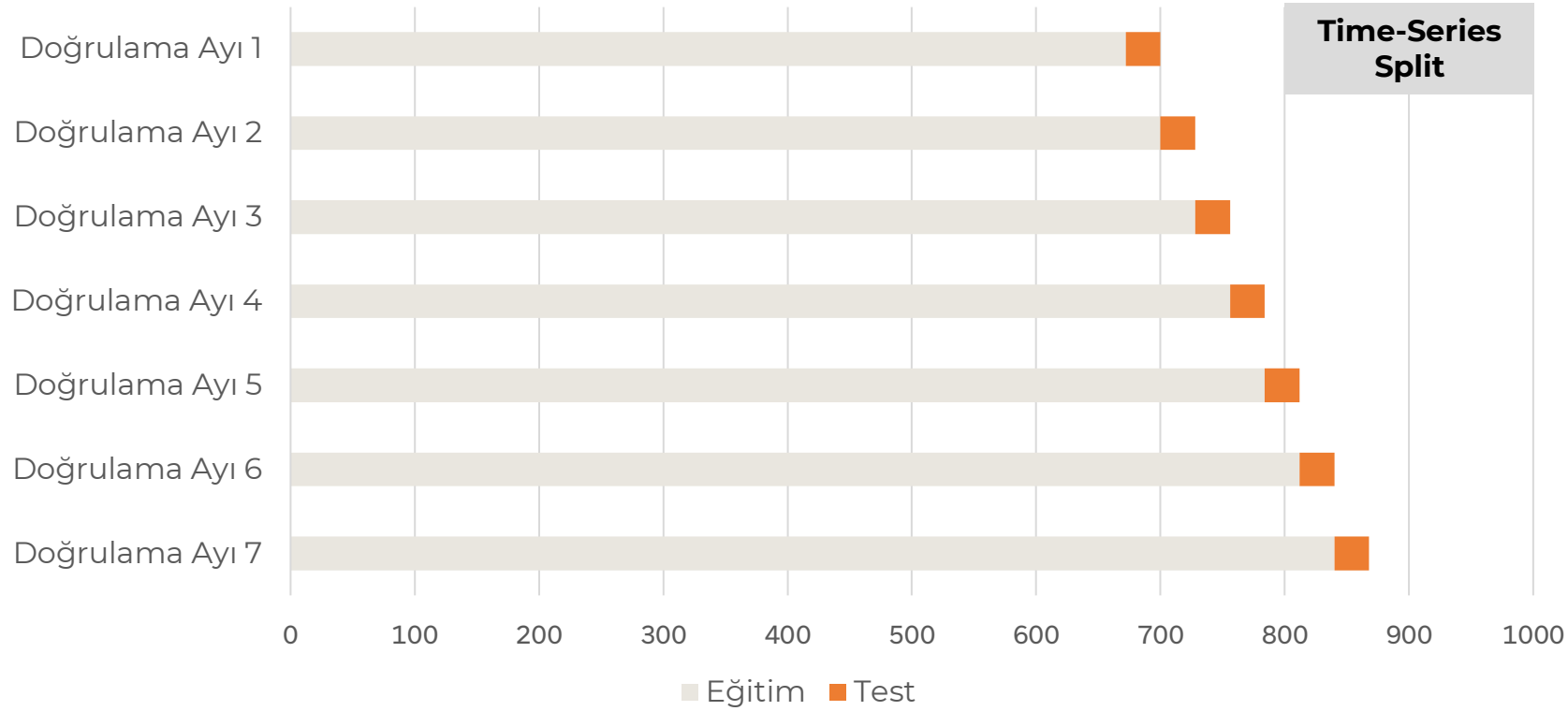
- CatBoost
- **LightGBM**
- XGBoost
- Local Modeller (ARIMA, Croston, EMA vs.)
- Global NN Modeller (RNN, TFT, TimesNet vs.)



LightGBM

ÇÖZÜM DOĞRULAMA MİMARİSİ

VALİDASYON ŞEMASI



Fold hatalarının:

- Ortalaması
- Standart sapması

OPTİMİZASYON MİMARİSİ

FEATURE SELECTION

- Basit bir CatBoost modeli
- **Kıstas:** SHAP value
- 7 fold
- 15 adım
- 25 feature

Başlangıç: 490 feature

Sonuç: 97 feature



Yandex
CatBoost



Shap

OPTİMİZASYON MİMARİSİ

HYPER-PARAMETER TUNING

Tuner:

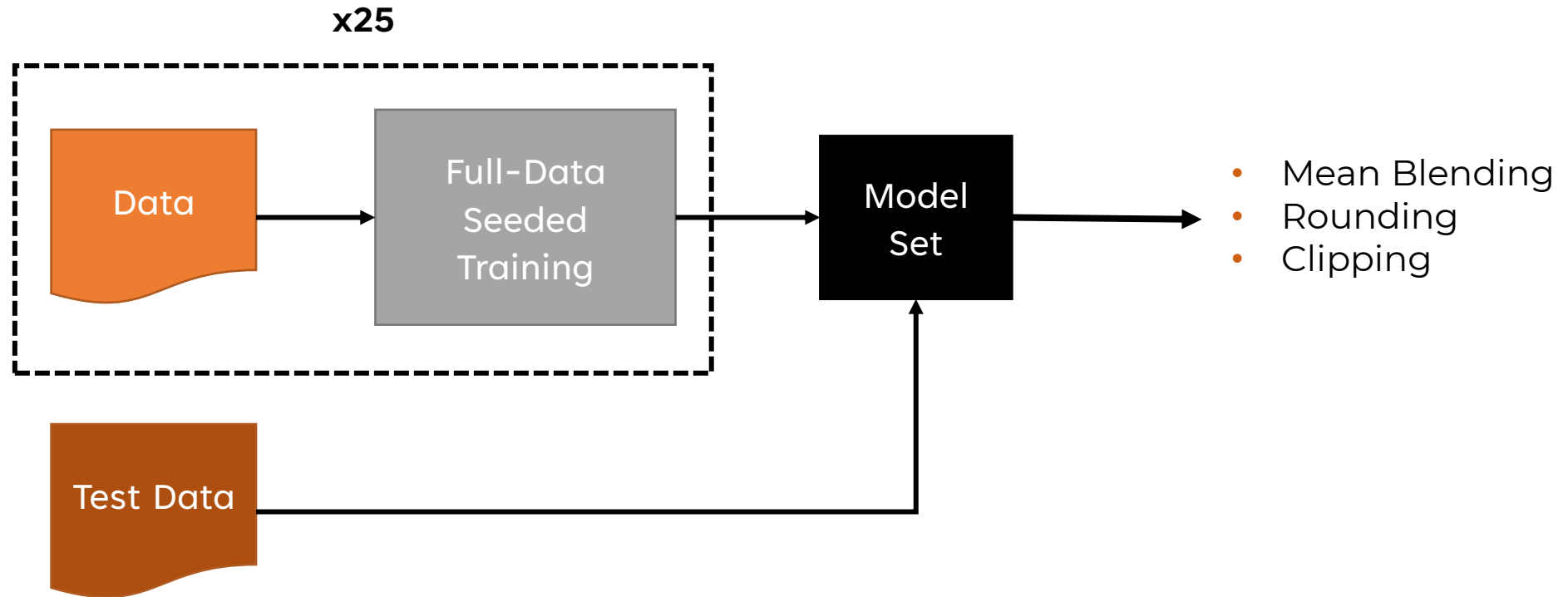
- Optuna
- TPESampler

Parametreler:

- objective
- learning_rate
- num_leaves
- max_depth
- bagging_fraction
- bagging_freq
- lambda_l1
- feature_fraction
- max_bin
- boosting_type



TAHMİN MİMARISI



SONUÇ

FİNAL SKORU

Local Mean

2.77127

Local Std

0.58115

Leaderboard

1.70799

Local Mean

2.79808



submission.csv

Complete (after deadline) · Anil Ozturk · 8d ago

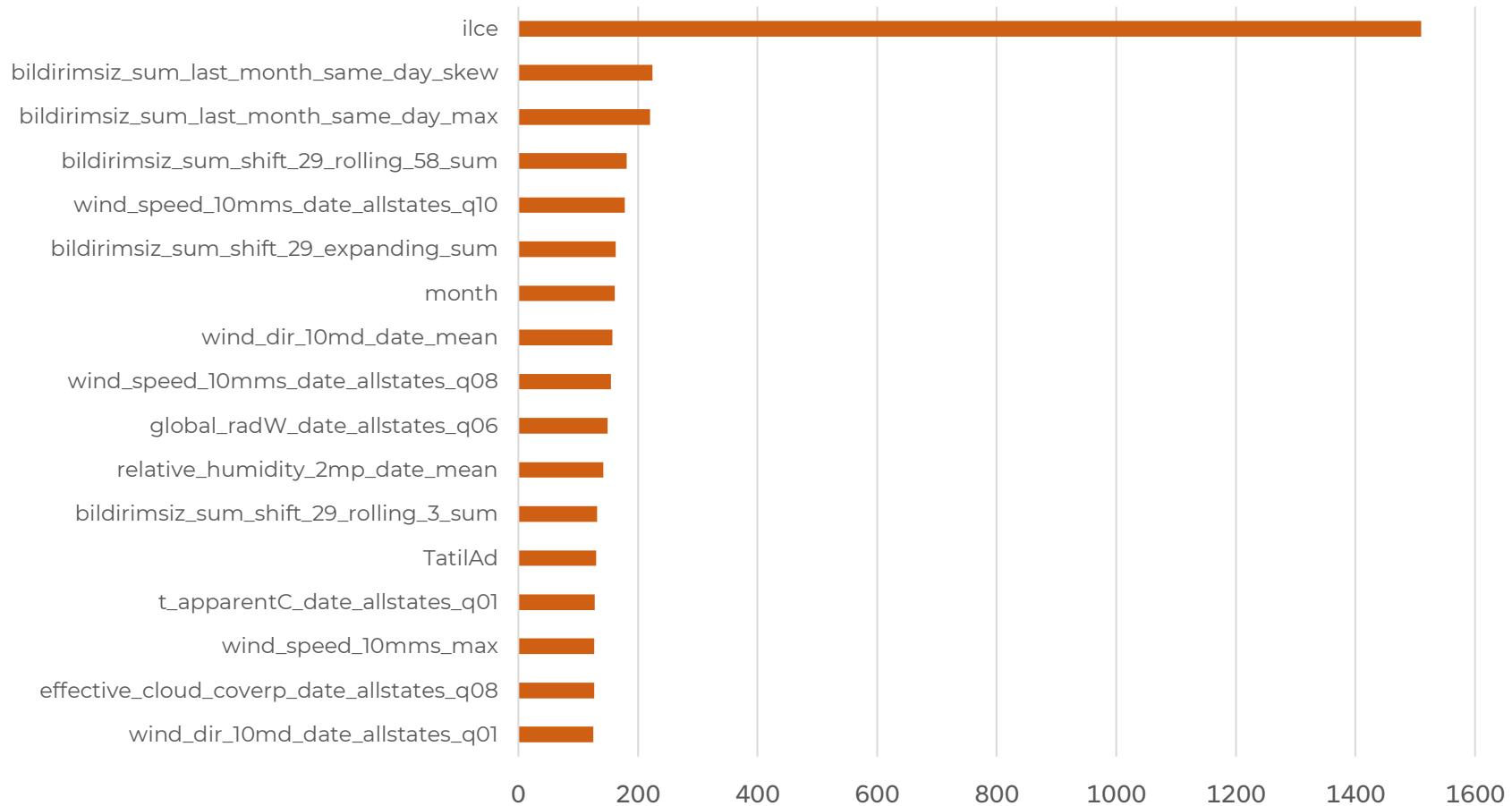
1.69479

Leaderboard

1.69479

SONUÇ

ÖNEMLİ PARAMETRELER (FEATURE IMPORTANCE)



SONUÇ

ÇÖZÜMÜN UYGULANABİLİRLİĞİ

- Ensemble ve açıklanabilir çözüm
- Regularized pipeline
- Daraltılmış feature-set
- Küçük model kümesi (~1MB x 25, ~25MBs)
- Yeni veri ile eğitilebilir

SONUÇ

POTANSİYEL GELİŞTİRMELER

- Feature üretimi hızlandırılabilir (*Pandas > Polars*)
- Diğer tip modeller ile ensemble denenebilir (*İsabet ↑ , Açıklanabilirlik ↓*)
- Bildirimsiz kesinti tiplerinin ayrıştırılması ya da kesinti ölçeğinin tahmin edilmesi
- Metrik değişikliği

A series of thin, light-orange lines forming an abstract geometric pattern in the top-left corner of the slide. The lines intersect to create various polygons and shapes, adding a modern, architectural feel to the design.

TEŞEKKÜRLER