

A series of black, overlapping, irregular geometric lines that form a complex, abstract pattern in the upper left and center of the page.

GDZ DATATHON

FİNAL SUNUMU

Anıl Öztürk

TAKIM TANITIMI



Anıl Öztürk

Eğitim

- Yıldız Teknik Üniversitesi – Mekatronik Mühendisliği
- İstanbul Teknik Üniversitesi – Bilgisayar Mühendisliği (YL)



Profesyonel Deneyim

- Machine Learning Researcher (2 yıl)
- Machine Learning Engineer (1 yıl)
- Senior Machine Learning Engineer (4 ay)



SUNUM AKIŞI

VERİ ANALİZİ

Ham verideki değerlerin incelenmesi ve anlamlandırılması

VERİ KULLANIMI

Kullanılan ekstra ve sonradan üretilmiş veriler

ÇÖZÜM MİMARİSİ

Çözümün kararlılığının ve riskinin analizi

Tahmin mimarisi

SONUÇ ve ÇIKARIMLAR

Detaylı performans analizi

Problemin süreç tasarımı

Uygulanabilirlik ve potansiyel geliştirmeler

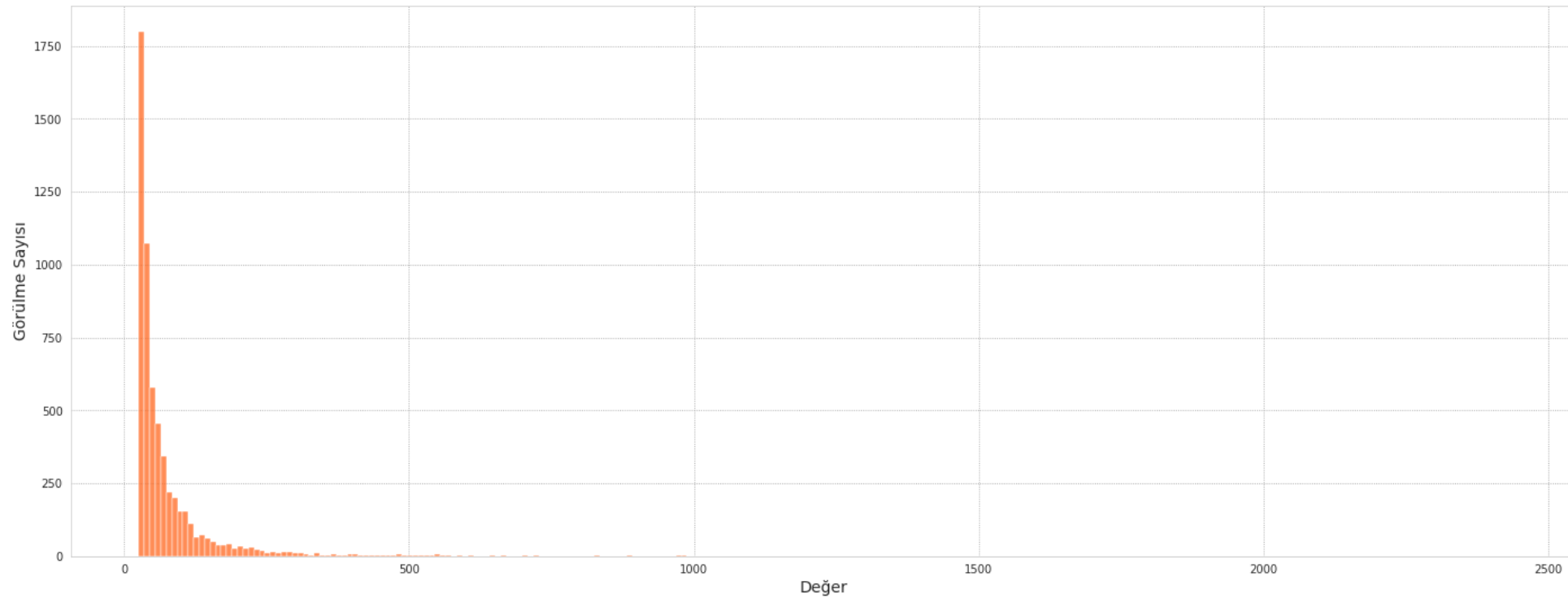
PROBLEM

Sebebi ve bölgesi verilmiş bir elektrik kesintisi için kaç çağrı alınması beklenir?

- 2022 Haziran'a kadar olan sürece ait kesintiler çağrı miktarlarıyla verilmiş
- Başarı oranı **RMSE** ile değerlendiriliyor

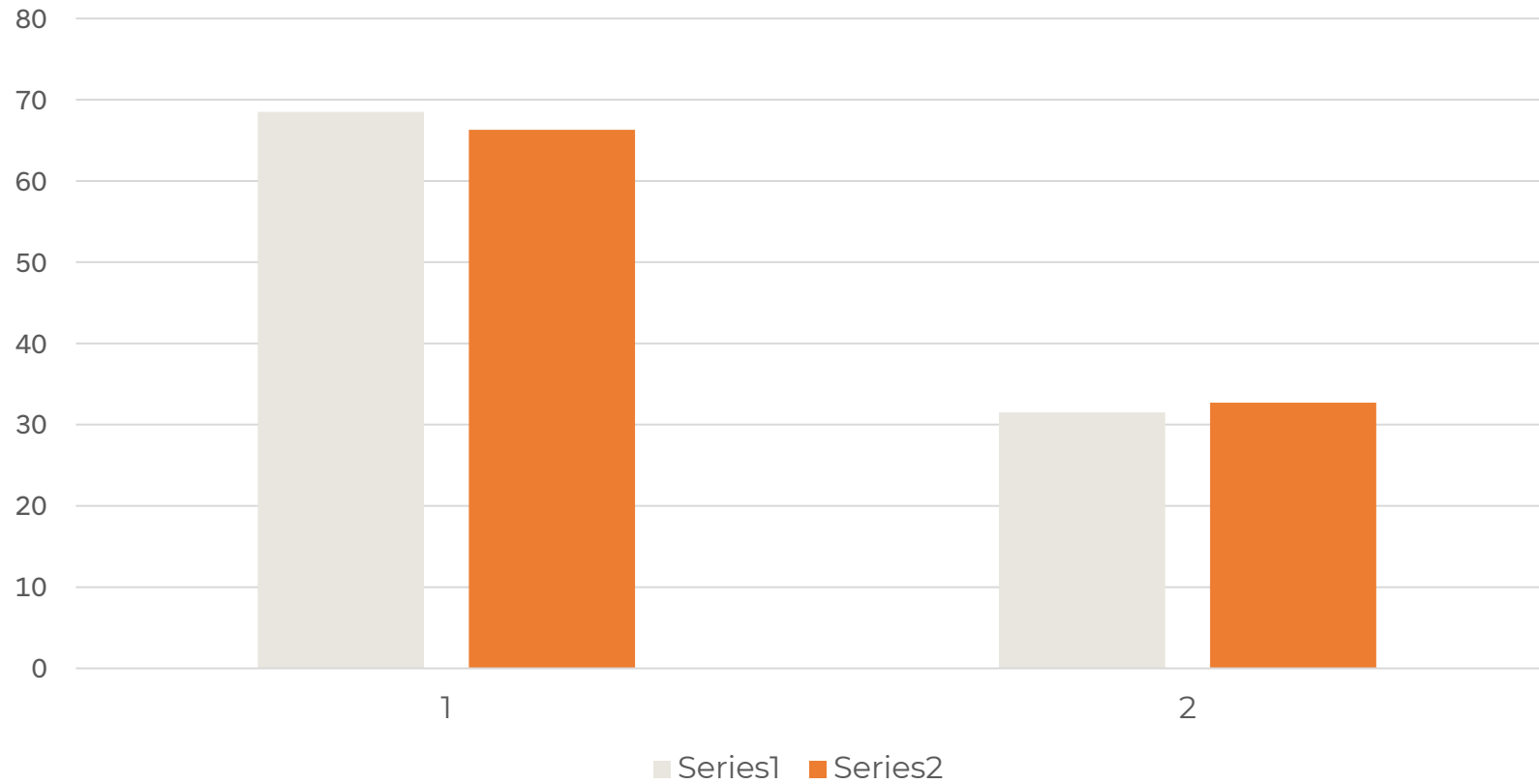
VERİ ANALİZİ

HEDEF DEĞERİN DAĞILIMI (ÇAĞRI SAYISI)



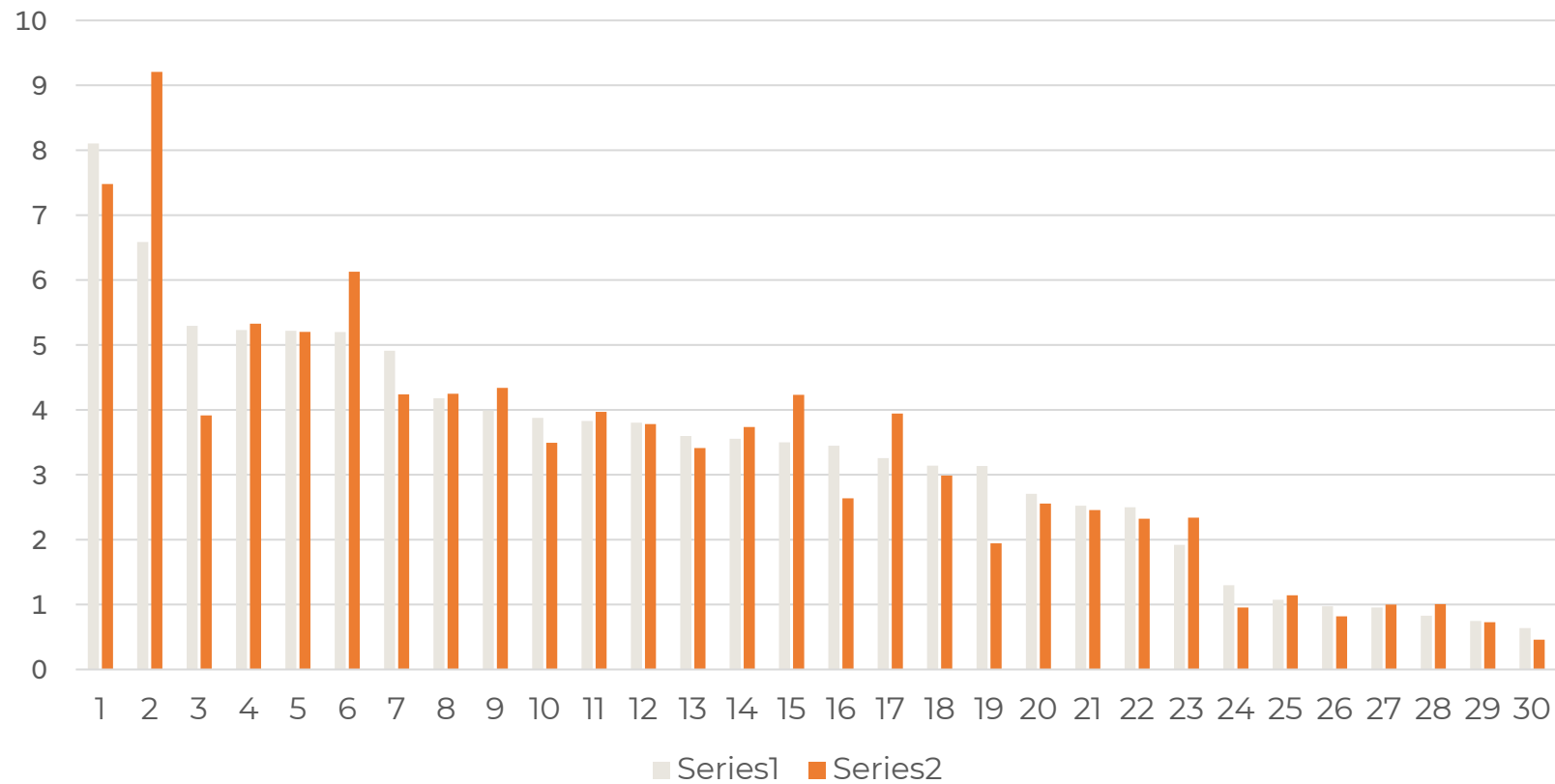
VERİ ANALİZİ

KESİNTİLERİN İLLERE GÖRE DAĞILIMI (%)



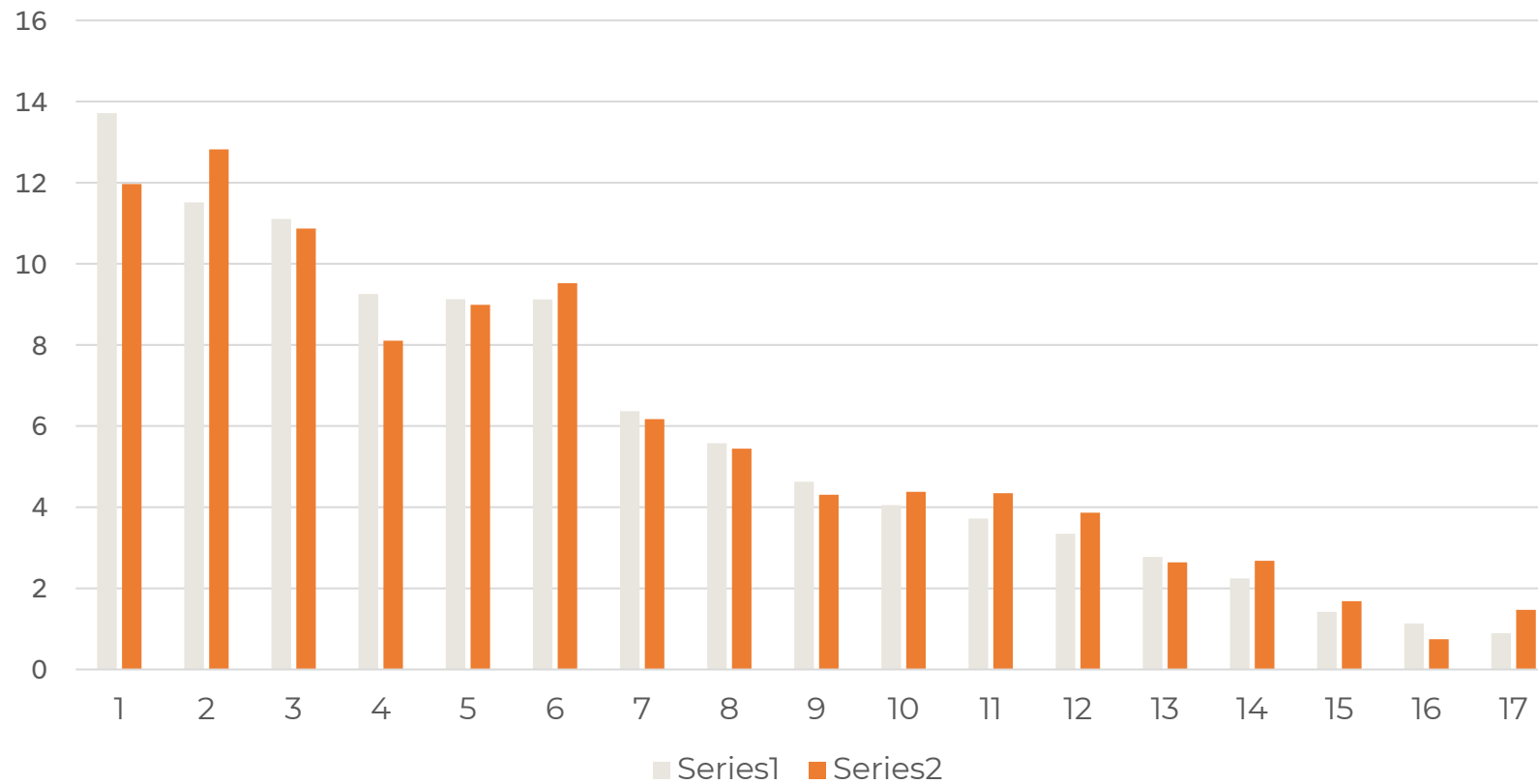
VERİ ANALİZİ

KESİNTİLERİN İLÇELERE GÖRE DAĞILIMI - İZMİR (%)



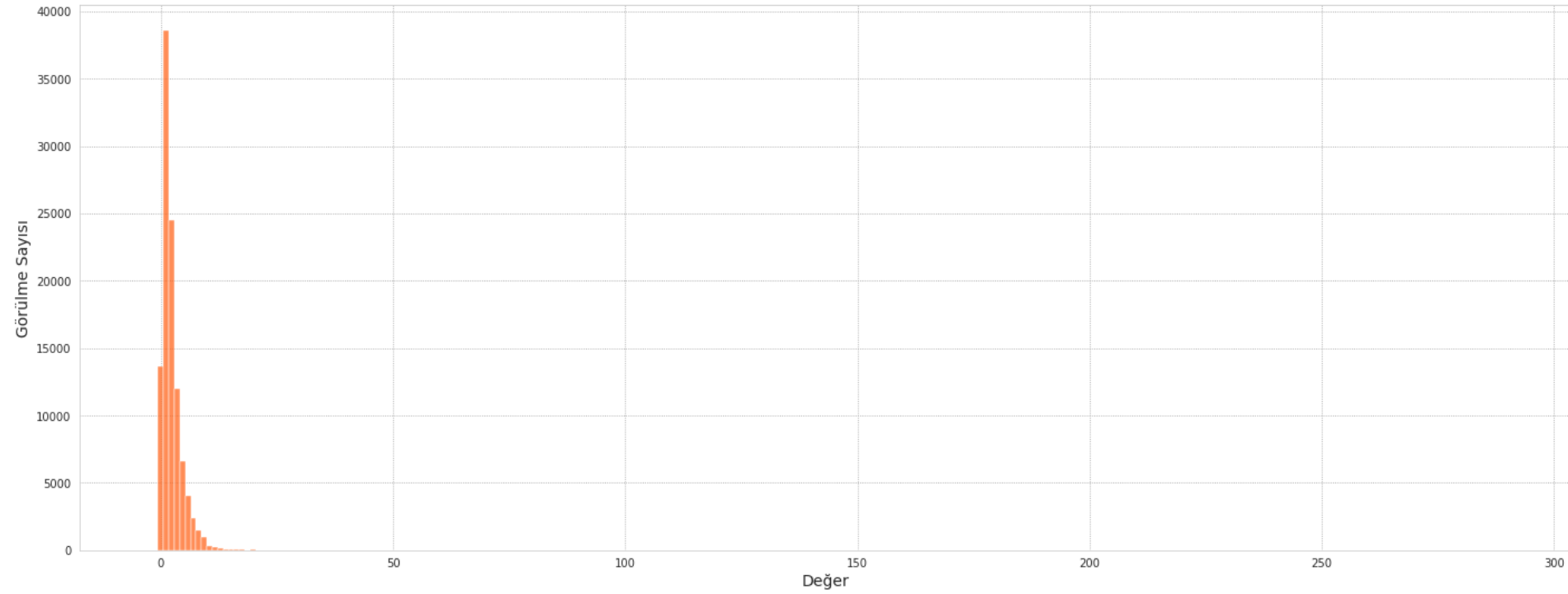
VERİ ANALİZİ

KESİNTİLERİN İLÇELERE GÖRE DAĞILIMI - MANİSA (%)



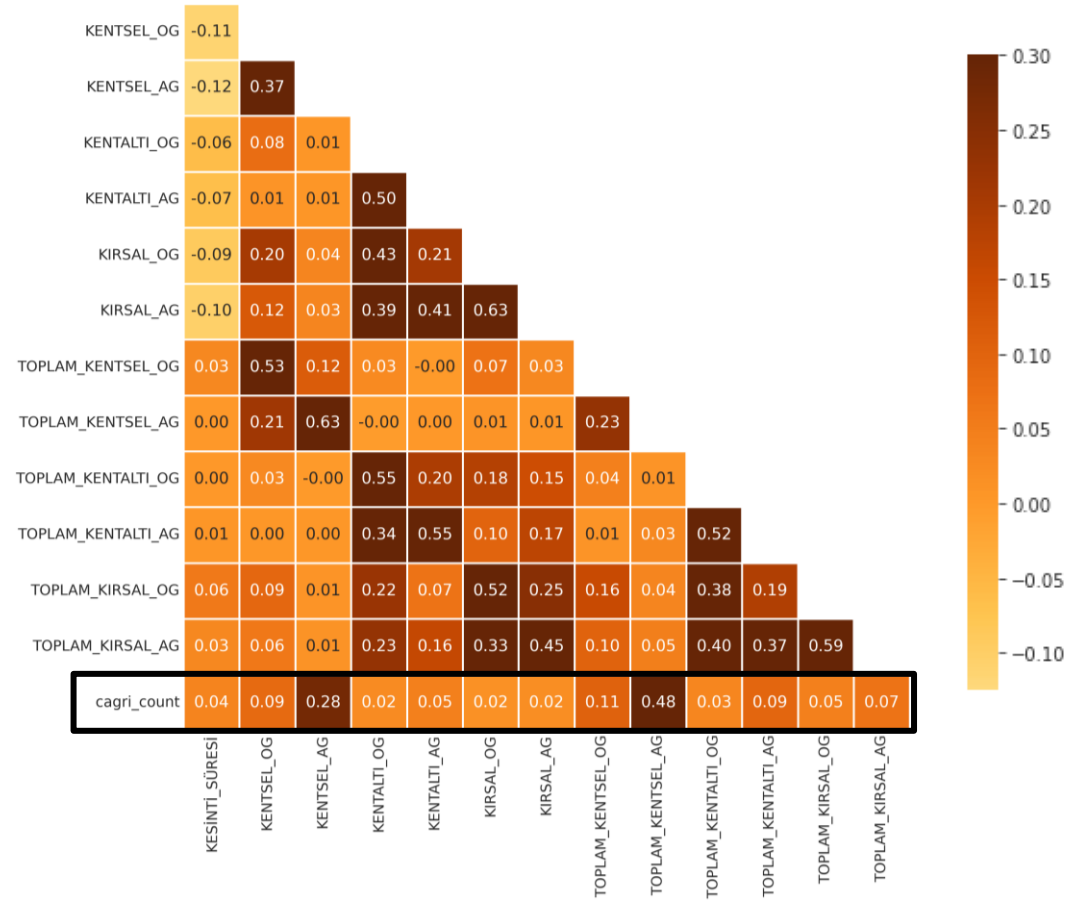
VERİ ANALİZİ

KESİNTİ SÜRESİNİN DAĞILIMI (SAAT)



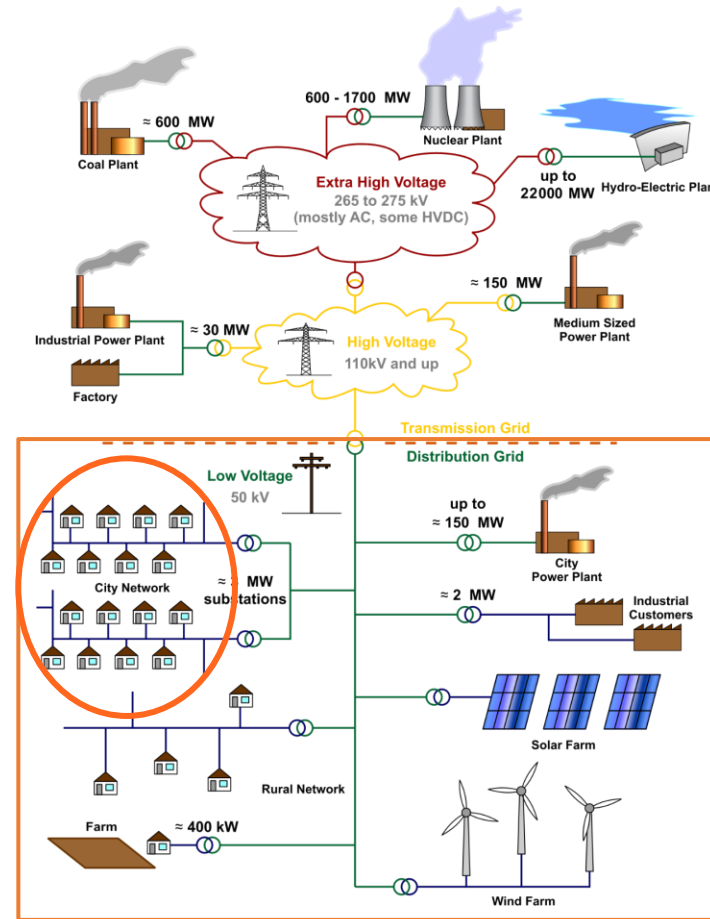
VERİ ANALİZİ

VERİ İÇİ KORELASYON



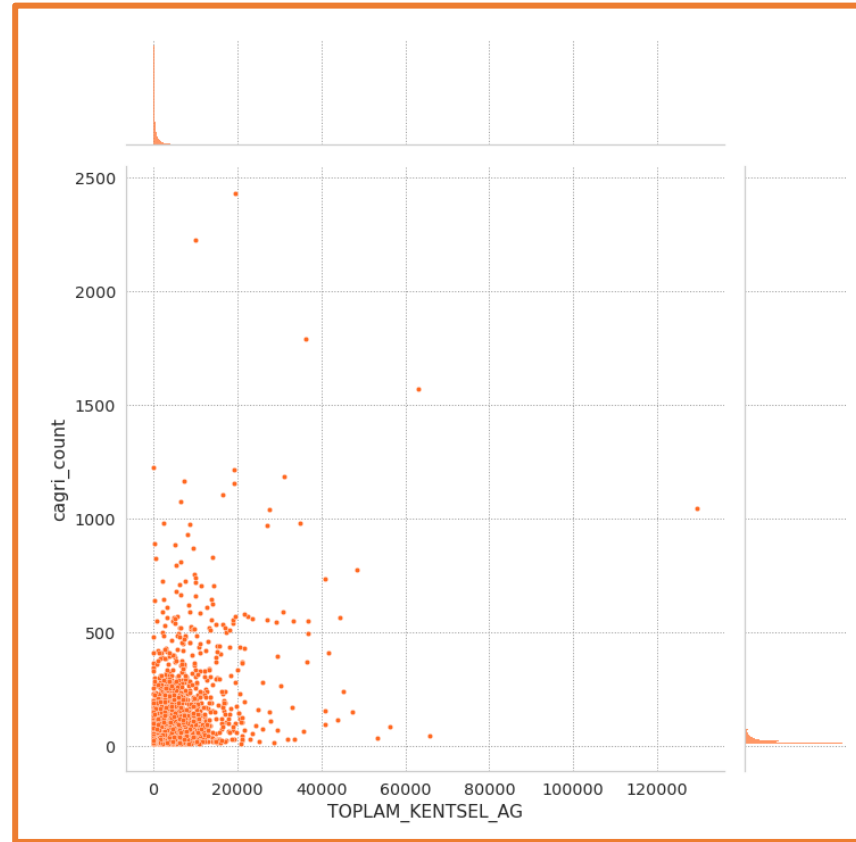
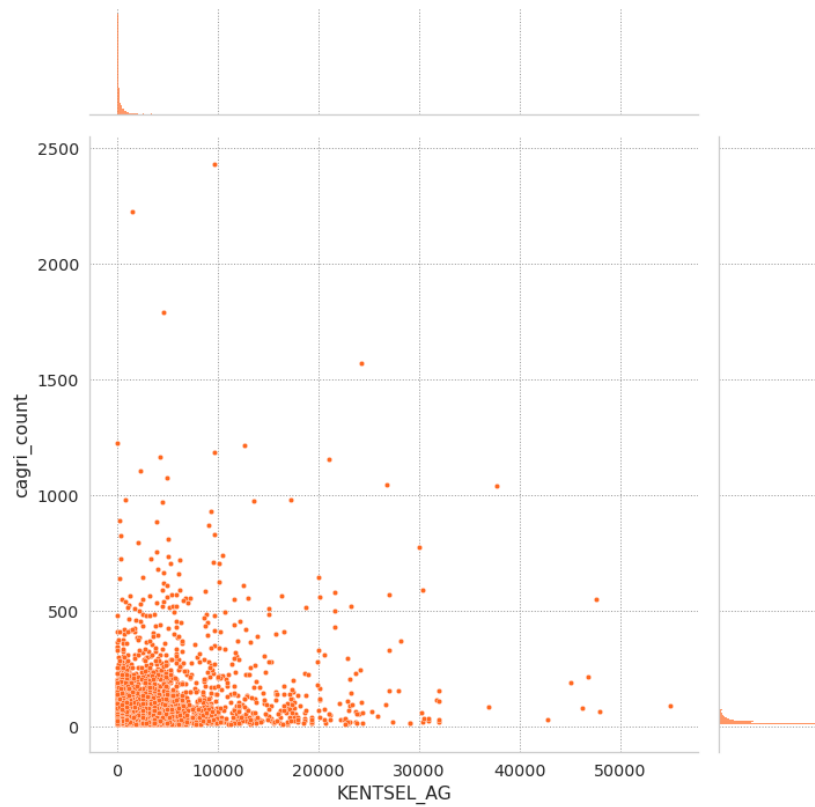
VERİ ANALİZİ

ELEKTRİK DAĞITIM SİSTEMLERİ



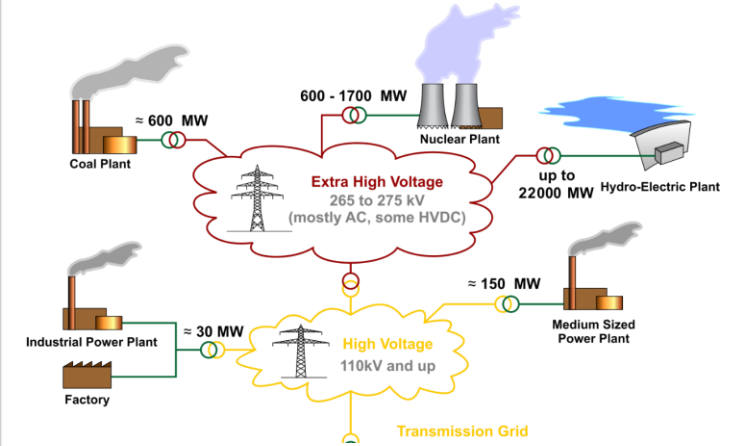
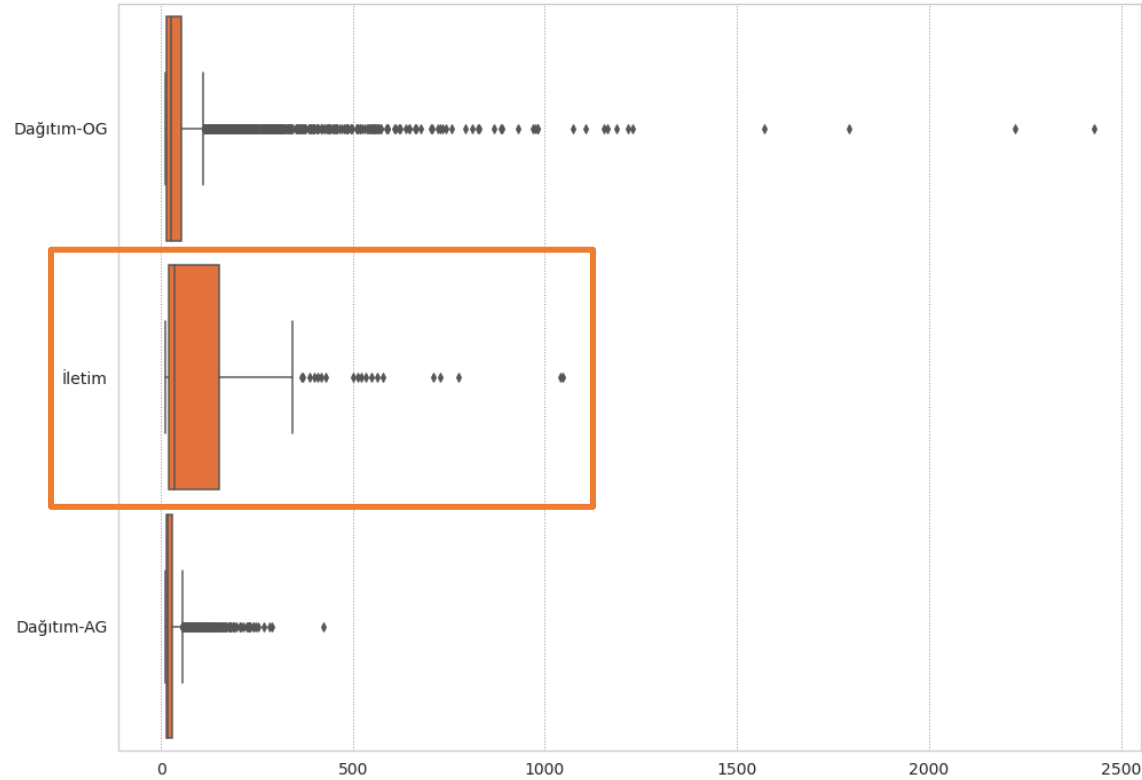
VERİ ANALİZİ

KENTSEL ALÇAK GERİLİM HAT SAYISI VERİSİ



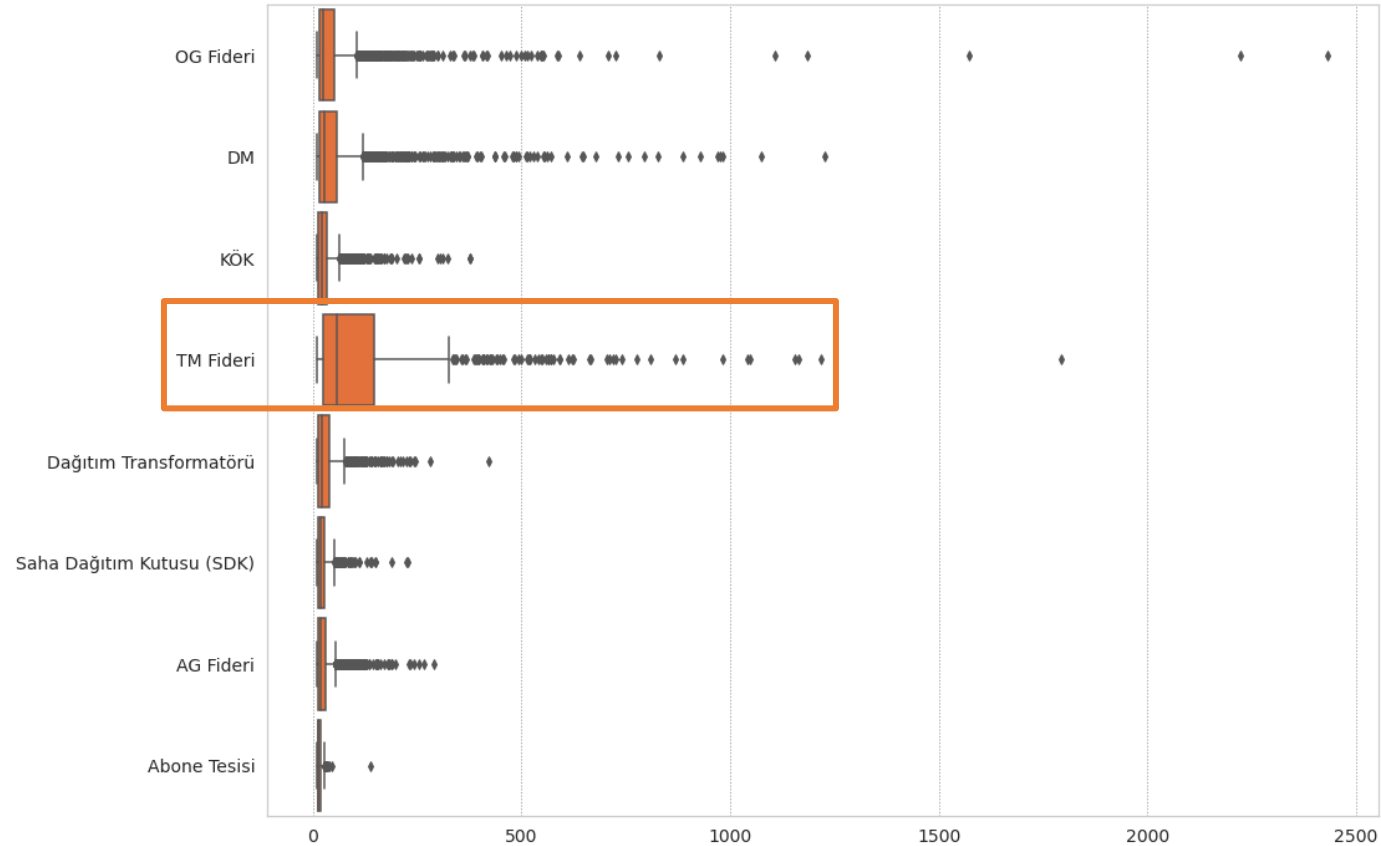
VERİ ANALİZİ

KAYNAĞA GÖRE ÇAĞRI SAYISI



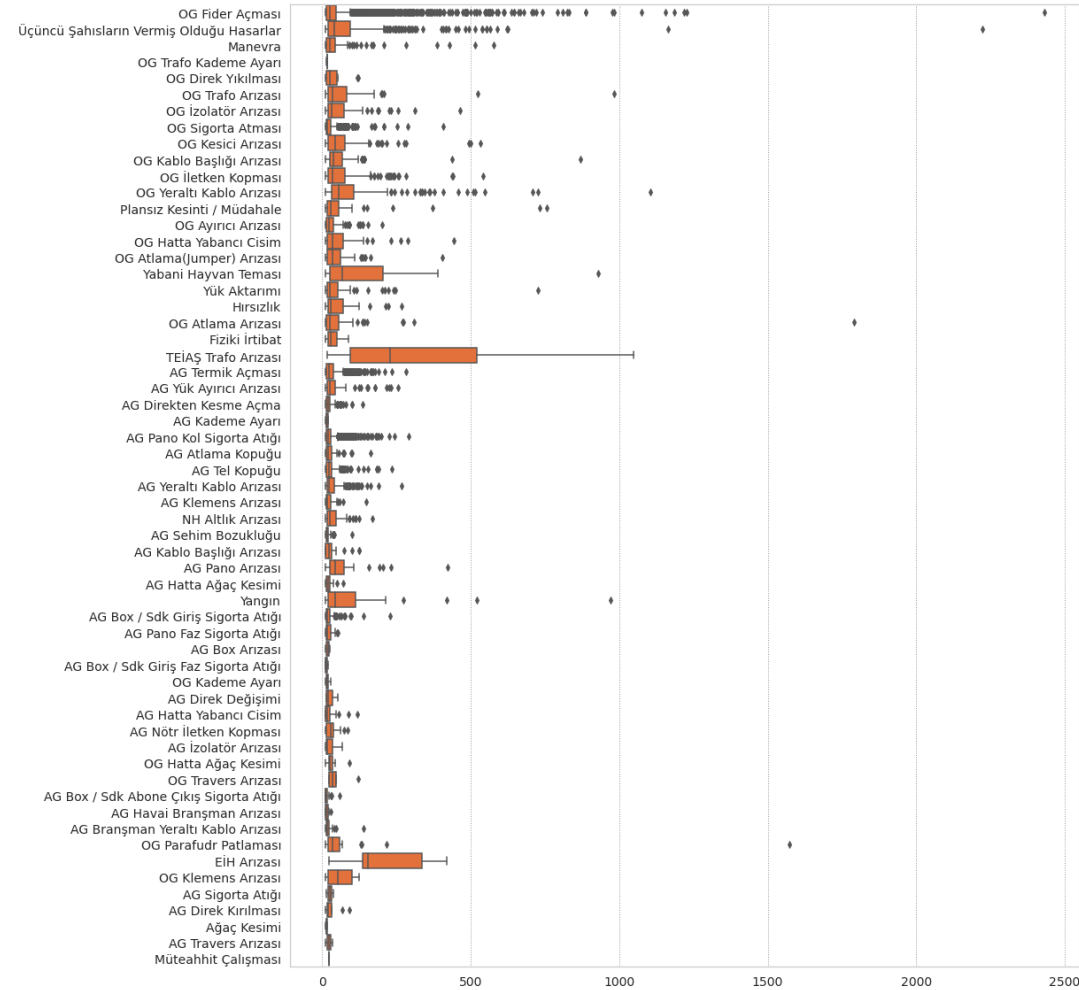
VERİ ANALİZİ

ŞEBEKE UNSURUNA GÖRE ÇAĞRI SAYISI



VERİ ANALİZİ

KESİNTİ NEDENİNE GÖRE ÇAĞRI SAYISI



KULLANILAN VERİ

SENTEZ VERİ

- İlgili trafonun o gün yaşadığı kaçınıcı kesinti olduğu
- Kentsel / Kentaltı / Kırsal toplam etkilenen OG/AG hat **sayısı / süresi**
- Toplam etkilenen OG ve AG hat **sayılarının / sürelerinin** oranı ve farkı
- Etikelenen en geniş yerleşke tipi **(Kentsel > Kentaltı > Kırsal)**
- Kesinti anında ilde ve ilçede kaç paralel kesinti yaşandığı

KULLANILAN VERİ

SENTEZ VERİ

- Trafo ismindeki karakteristik alt-dizelerin varlık durumu **(TFIDF)**
 - *Direk tipi trafo*
 - *Eski*
 - *Yeni*
 - *Kök*
 - *Hücresi*
 - *Hat Başı Ayırıcı*
 - *Sulama*
 - *Grubu*
 - *TM*
 - *TR*

KULLANILAN VERİ

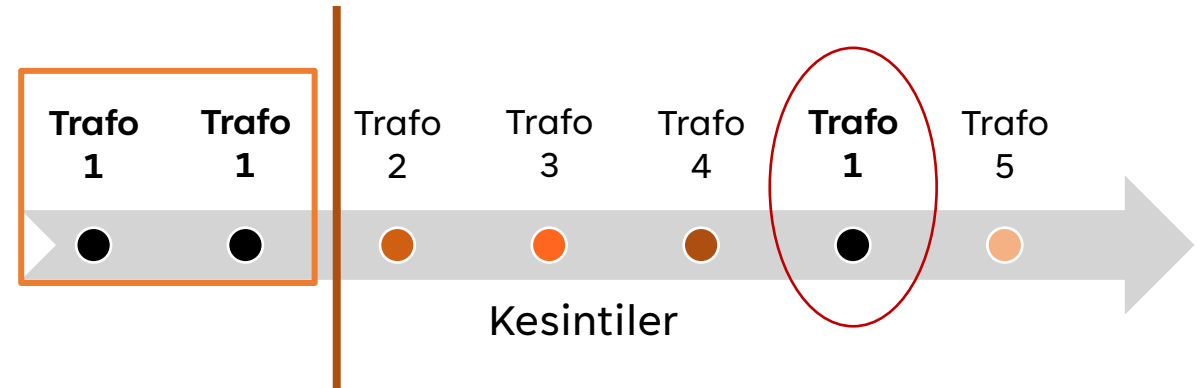
SENTEZ VERİ

- Trafo bazında zamansal istatistiksel encoding

Her trafo verisi için çağrıların o trafoya ait:

- Ortalama
- Medyan
- Standart Sapma
- Maksimum
- Skewness

değerleri çıkartıldı.



KULLANILAN VERİ

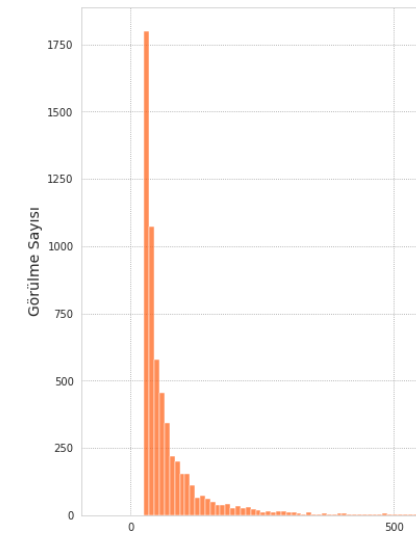
SENTEZ VERİ

- Trafo bazında zamansal istatistiksel encoding (**non-zero mode**)

Her trafo verisi için en az bir arama içeren çağruların o trafoya ait:

- Ortalama*
- Medyan*
- Standart Sapma*
- Maksimum*
- Skewness*

değerleri çıkartıldı.



KULLANILAN VERİ

SENTEZ VERİ

- Takvim tipi veriler
 - *Ay*
 - *Başlangıç saati*
 - *Bitiş saati*
 - *Başlangıç günü*
 - *Bitiş günü*

KULLANILAN VERİ

EKSTRA VERİ

- İlçe nüfusları
 - Hane halkı büyüklüğü **(Kullanılmadı)**
 - Hane halkı sayısı **(Kullanılmadı)**
-
- <https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr>
 - <https://www.nufusu.com/ilceleri/izmir-ilceleri-nufusu>
 - <http://www.manisa.gov.tr/nufus-ve-dagilim>

MODEL SEÇİMİ

- CatBoost
- LGBM
- XGBoost
- Linear Family
- Neural Nets

MODEL SEÇİMİ

- CatBoost
- **LGBM**
- XGBoost
- Linear Family
- Neural Nets



LightGBM

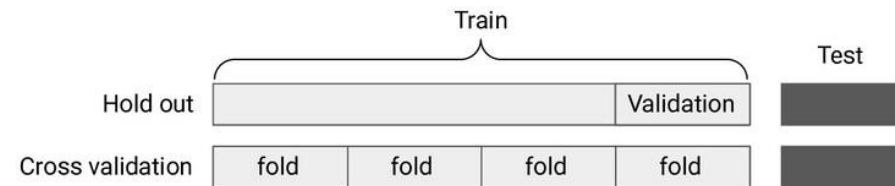
ÇÖZÜM DOĞRULAMA MİMARİSİ

EĞİTİM VE TEST VERİSİ BENZERLİĞİ

Cross Validation & Adversarial Validation

Cross validation: The data is divided into k folds; k-1 folds are used for training and the other fold is used for validation, which is done for all combinations.

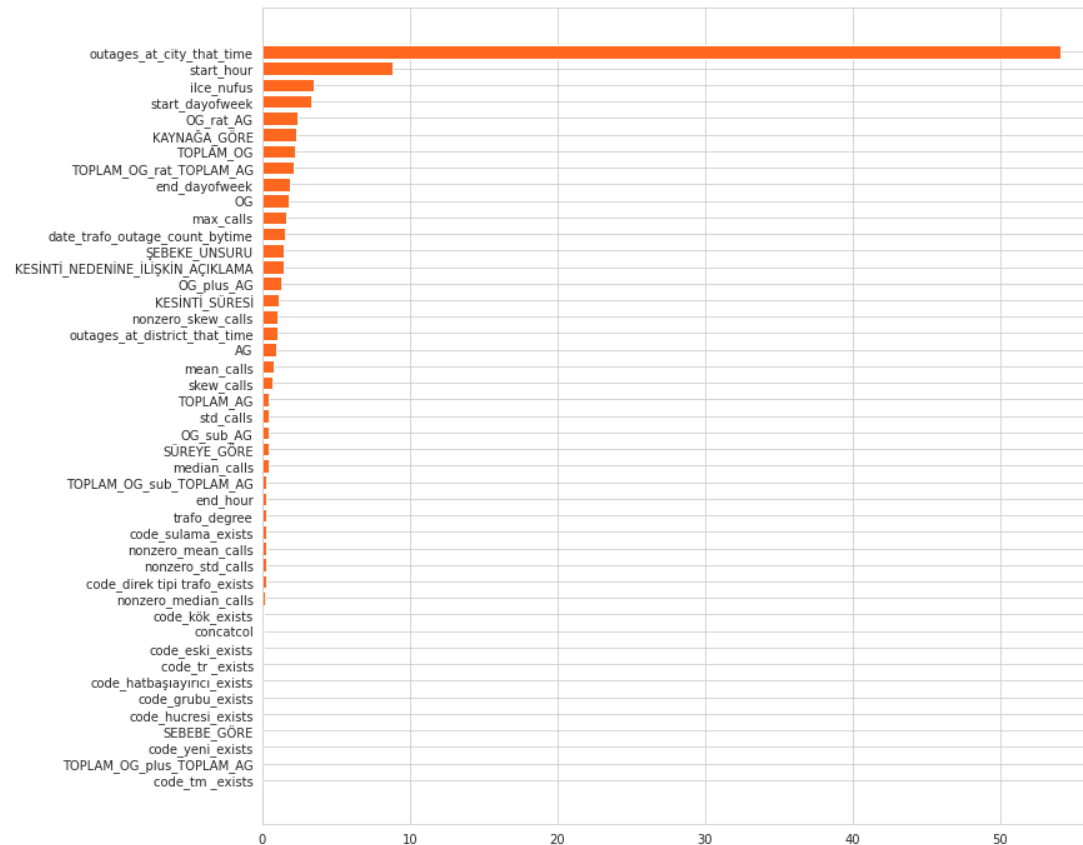
Adversarial validation: A binary classifier is trained to predict whether a sample belongs to test data or not. Training data highly similar to test data is sampled.



<https://speakerdeck.com/upura/adversarial-validation-to-select-validation-data-for-evaluating-performance-in-e-commerce-purchase-intent-prediction>

ÇÖZÜM DOĞRULAMA MİMARİSİ

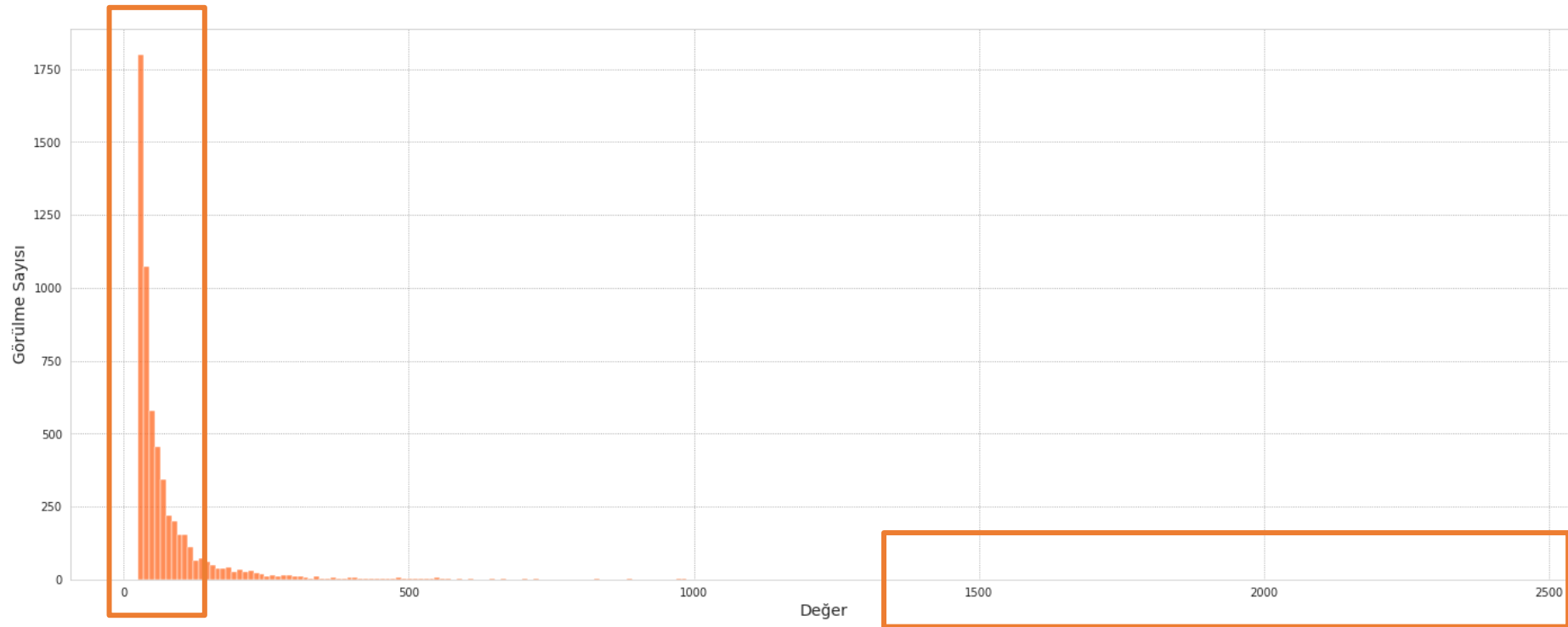
EĞİTİM VE TEST VERİSİ BENZERLİĞİ



ROC-AUC Score: 0.5686

ÇÖZÜM DOĞRULAMA MİMARİSİ

ETİKET DEĞERİNİN BENZER DAĞITILMASI

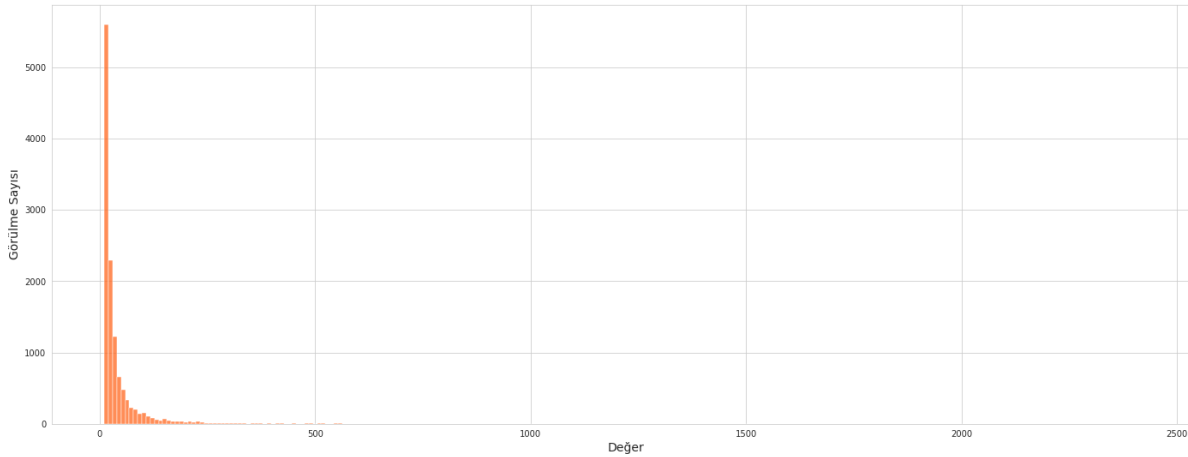


$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

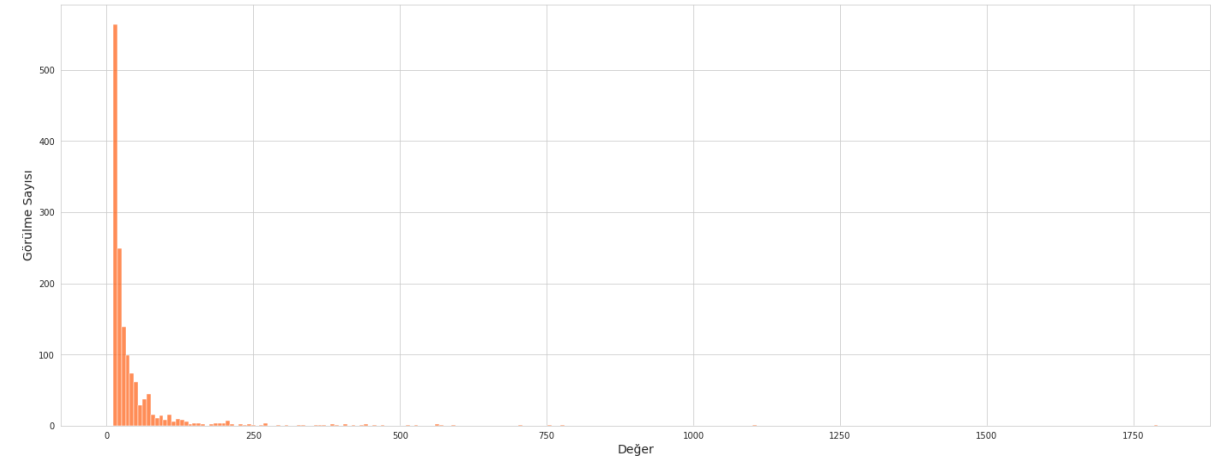
ÇÖZÜM DOĞRULAMA MİMARİSİ

ETİKET DEĞERİNİN BENZER DAĞITILMASI

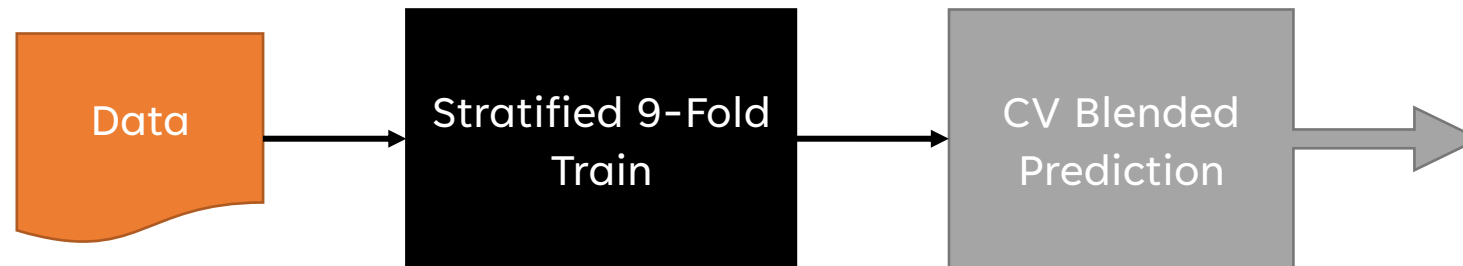
Eğitim



Validasyon

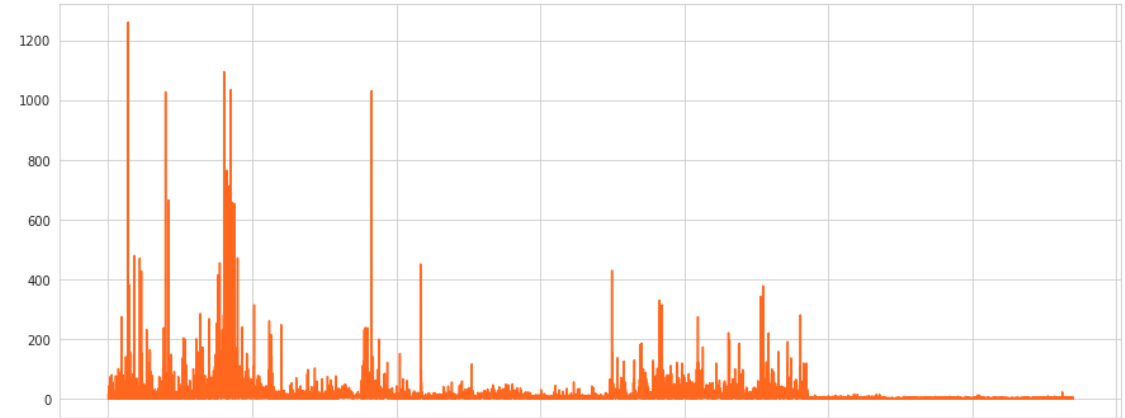


TAHMİN MİMARİSİ



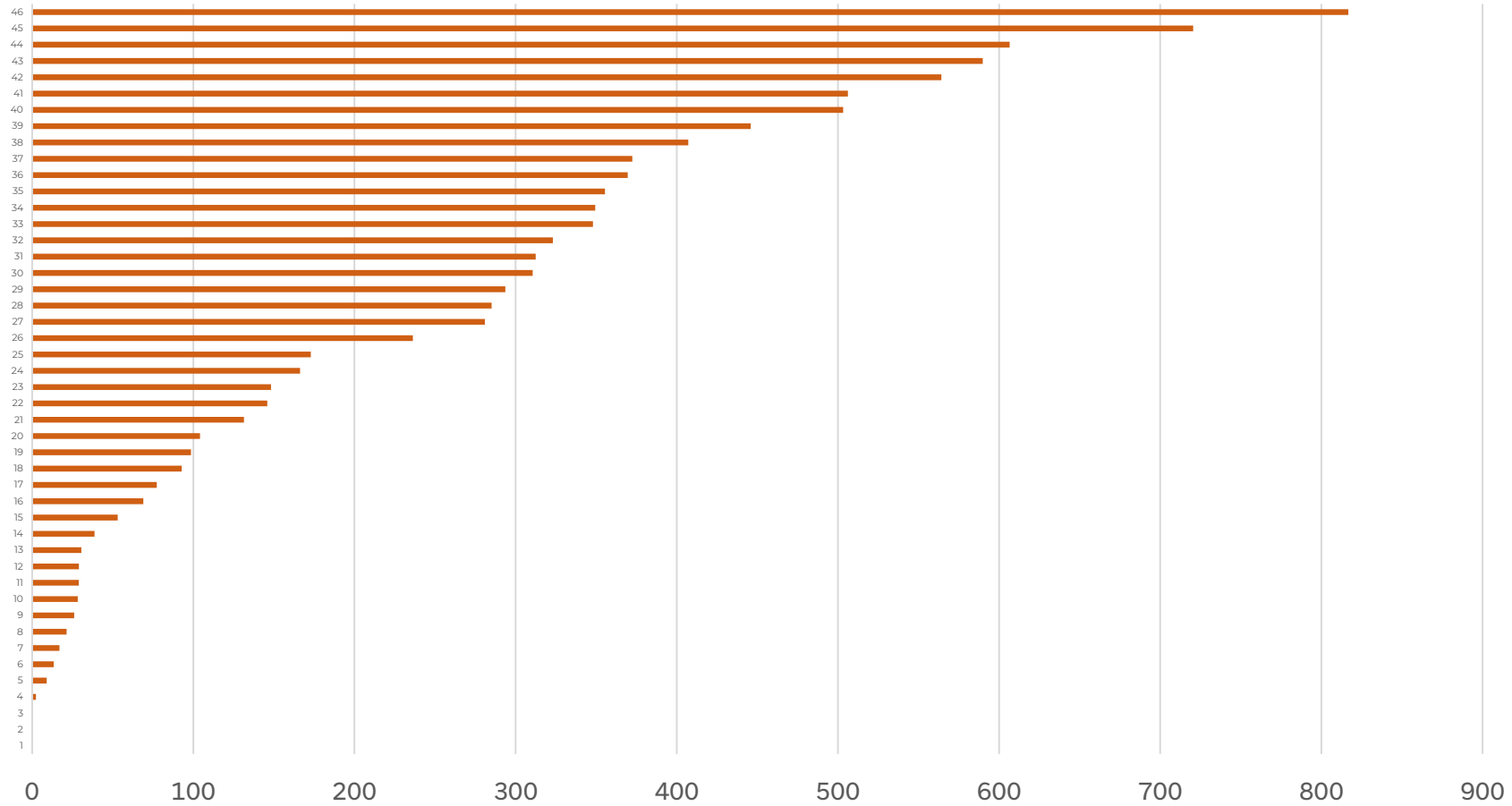
SONUÇ

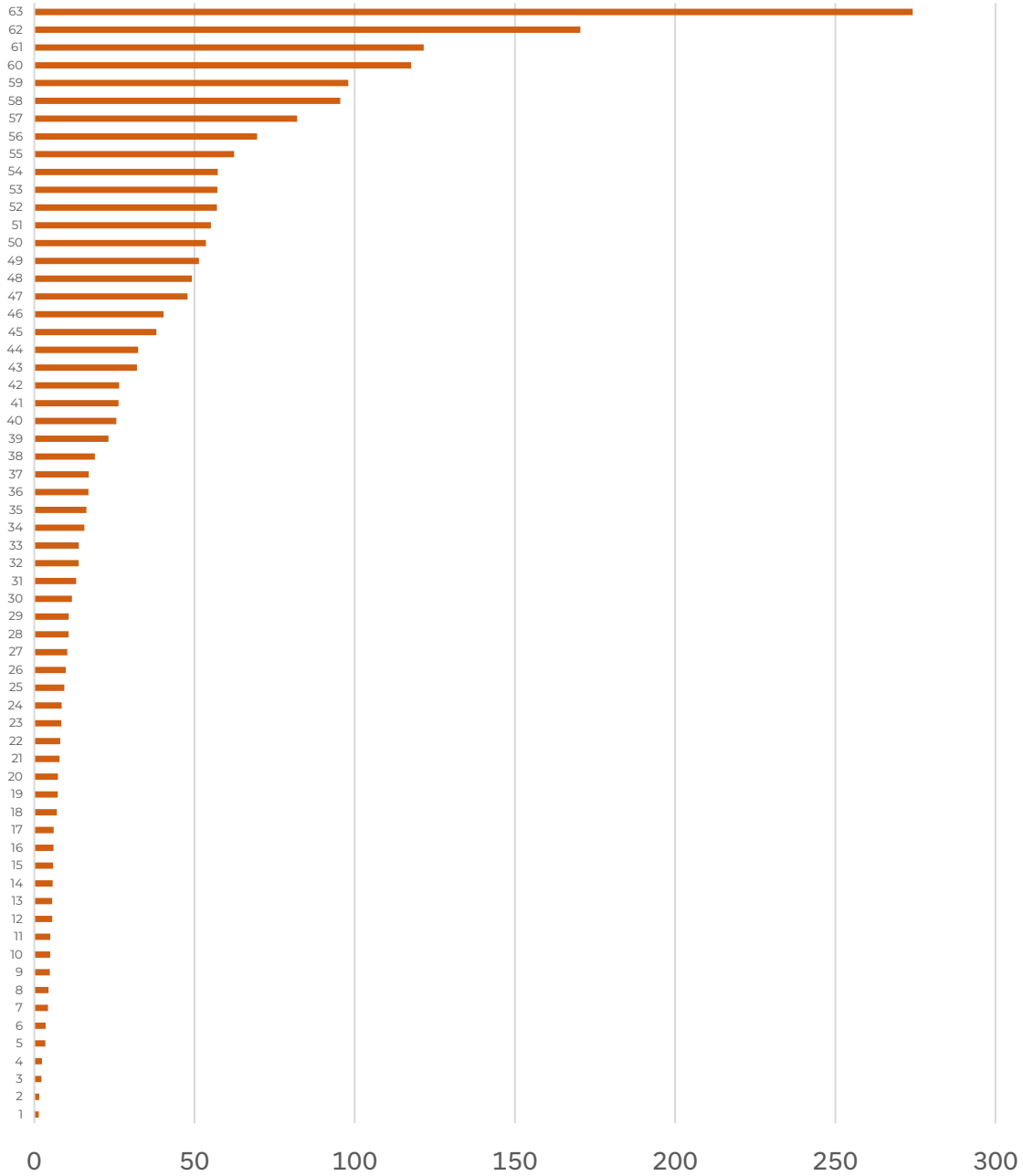
25.87 – 27.74 RMSE başarı oranı



SONUÇ

ÖNEMLİ PARAMETRELER (FEATURE IMPORTANCE)





SONUÇ

ÇÖZÜMÜN ALT GRUPLARDAKİ HATASI (RMSE)

SONUÇ

ÇÖZÜMÜN UYGULANABİLİRLİĞİ

- Ensemble çözüm
- Hafif modeller
- Canlıya alınabilir çözüm yapısı
- Genelleştirilmiş feature-set



SONUÇ

KESİNTİ SEBEPLERİNİN DARALTILMASI

- Afet kategorileri (*patlama, yangın*)
- Dış müdahale kategorileri (*yabancı cisim, yabani hayvan*)
- K-Means veya PCA gibi kümeleme-encoding algoritmaları ile unsupervised gruplar yaratılarak kesintiler gruplanabilir

SONUÇ

POTANSİYEL GELİŞTİRMELER

- Trafo veya muhit hakkında kesinti saatine ait güç tüketim ortalamalarının kullanılması
- Daha fazla veri ile lineer model eğitimi



TEŐEKKÜRLER