

BTK AKADEMİ DATATHON 2025

FİNAL SUNUMU

Unique()

Anıl Öztürk - Ahmet Tarık Karakaş

EKİP ÜYELERİ



ANIL ÖZTÜRK

Senior Data Scientist

Trendyol Group



TARIK KARAKAŞ

Data Science Tech Lead

TrendyolGo / Uber

PROBLEM

Bir e-ticaret sitesi oturumu için kullanıcının geçmişi ve oturum içi verileri kullanılarak oturumun potansiyel değeri hesaplanabilir mi?

- 21 günlük eğitim, 9 günlük test verisi var
- Model performansı **MSE** ile değerlendiriliyor

LİTERATÜR

- <https://yonsei.elsevierpure.com/en/publications/predicting-session-conversion-on-e-commerce-a-deep-learning-based>
- <https://link.springer.com/article/10.1007/s12525-020-00448-x>
- <https://www.ijnrd.org/papers/IJNRD2401271.pdf>

MEVCUT YAKLAŞIMLAR

Session seviyesinde satın alma dönüşümünün tahmin edilmesi

- Oturum içi sayfa görüntüleme feature'ları

Oturum devam ederken satın alma niyetinin erken tahmini

- Dinamik session log özellikleri
- Anonim/kayıtlı kullanıcı ayrımı
- Bağlamsal özellikler

Müşterinin yıllık/aylık/haftalık harcamasının tahmini

- Demografik veriler
- Satın alma geçmişi
- Ürün/işlem zamanı feature'ları

ÇÖZÜM AŞAMALARI

VERİ ANALİZİ & ÖN İŞLEME

Verinin incelenmesi ve formatın
standardize edilmesi

ÇÖZÜM MİMARİSİ

Feature, model, eğitim ve tahminleme
detayları

ÇIKARIMLAR

Verinin problem özelinde önemi,
yapılabilecek değişiklikler

VERİ ÖN İŞLEME

- Oturumlardaki her aksiyon ayrı bir satır olarak verilmiş
- Aynı oturumdaki bütün aksiyonlar aynı oturum değerine sahip

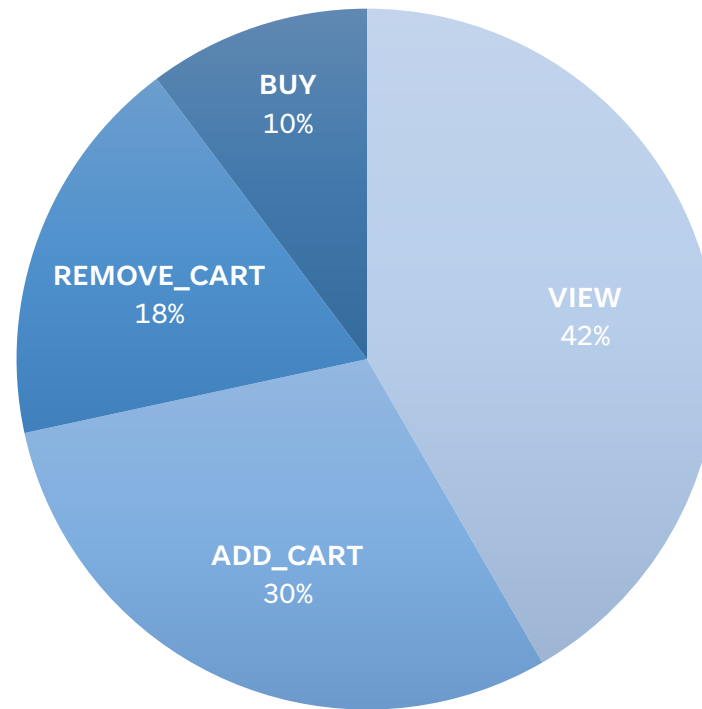
Session ID	Event	Session Value
1	BUY	10.12
1	VIEW	10.12
2	ADD_CART	5.38



Session ID	Features	Session Value
1	...	10.12
2	...	5.38

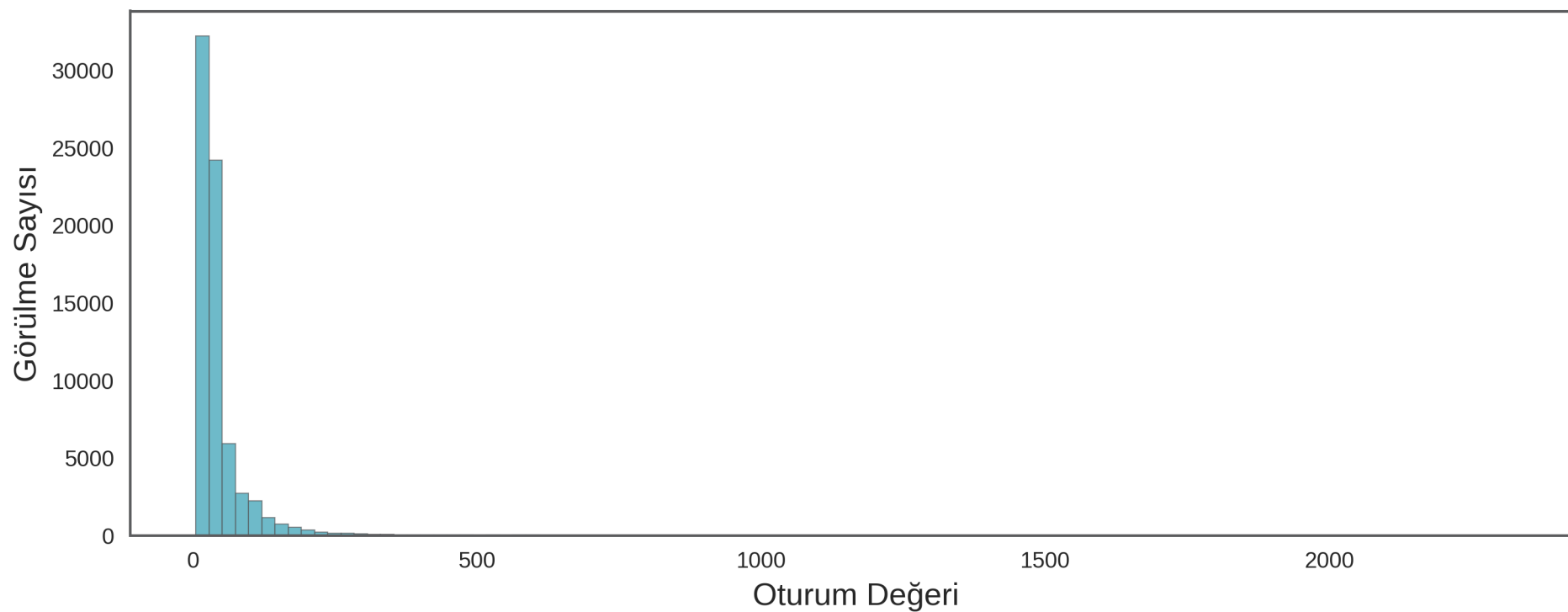
VERİ ANALİZİ

OTURUM AKSİYON DAĞILIMI



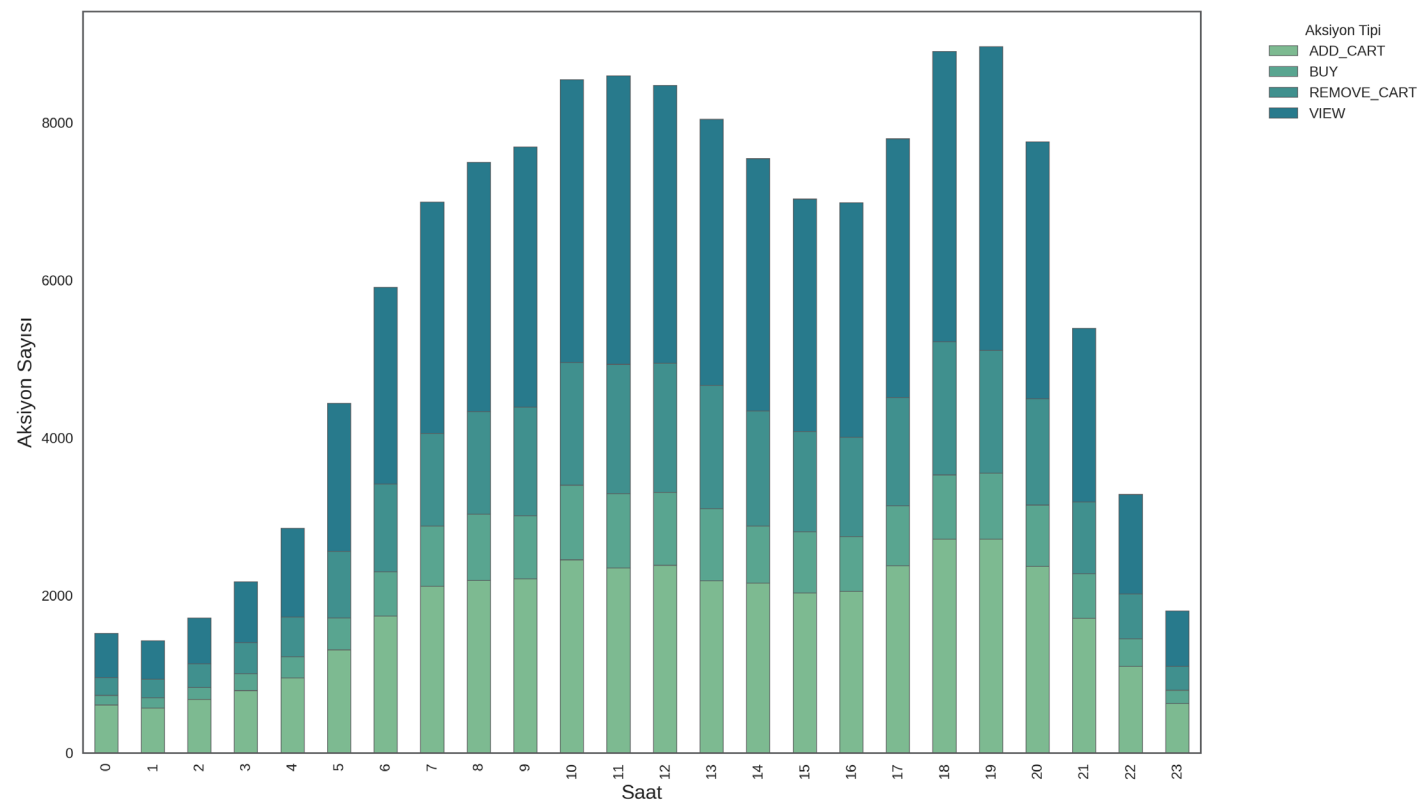
VERİ ANALİZİ

OTURUM DEĞERİ DAĞILIMI



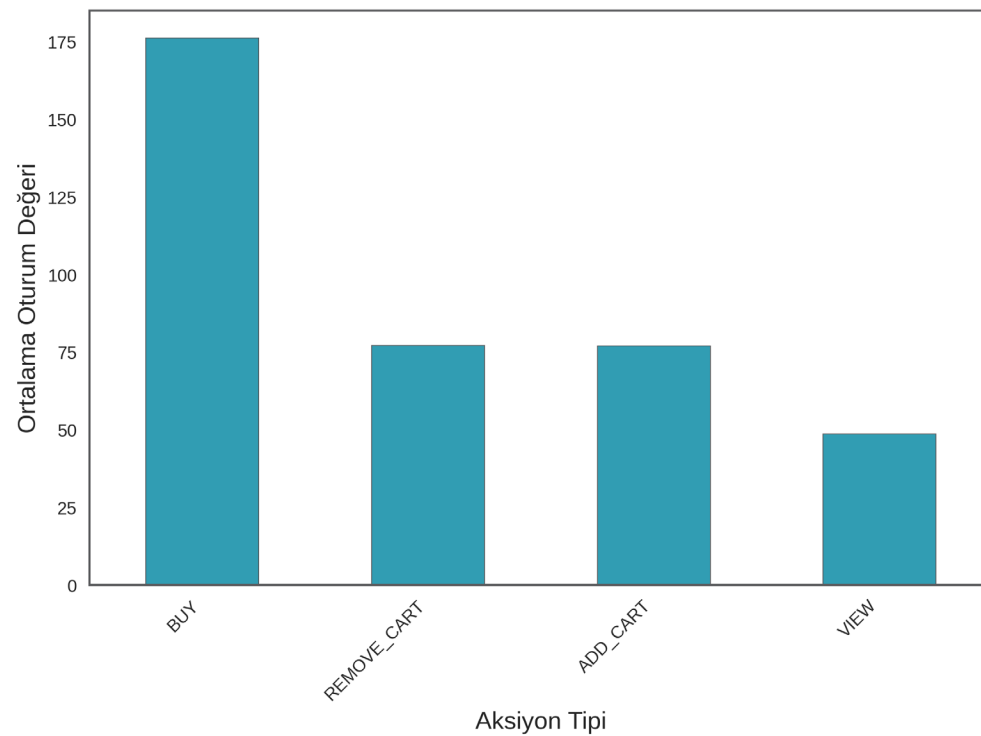
VERİ ANALİZİ

SAATE GÖRE AKSİYON DAĞILIMI



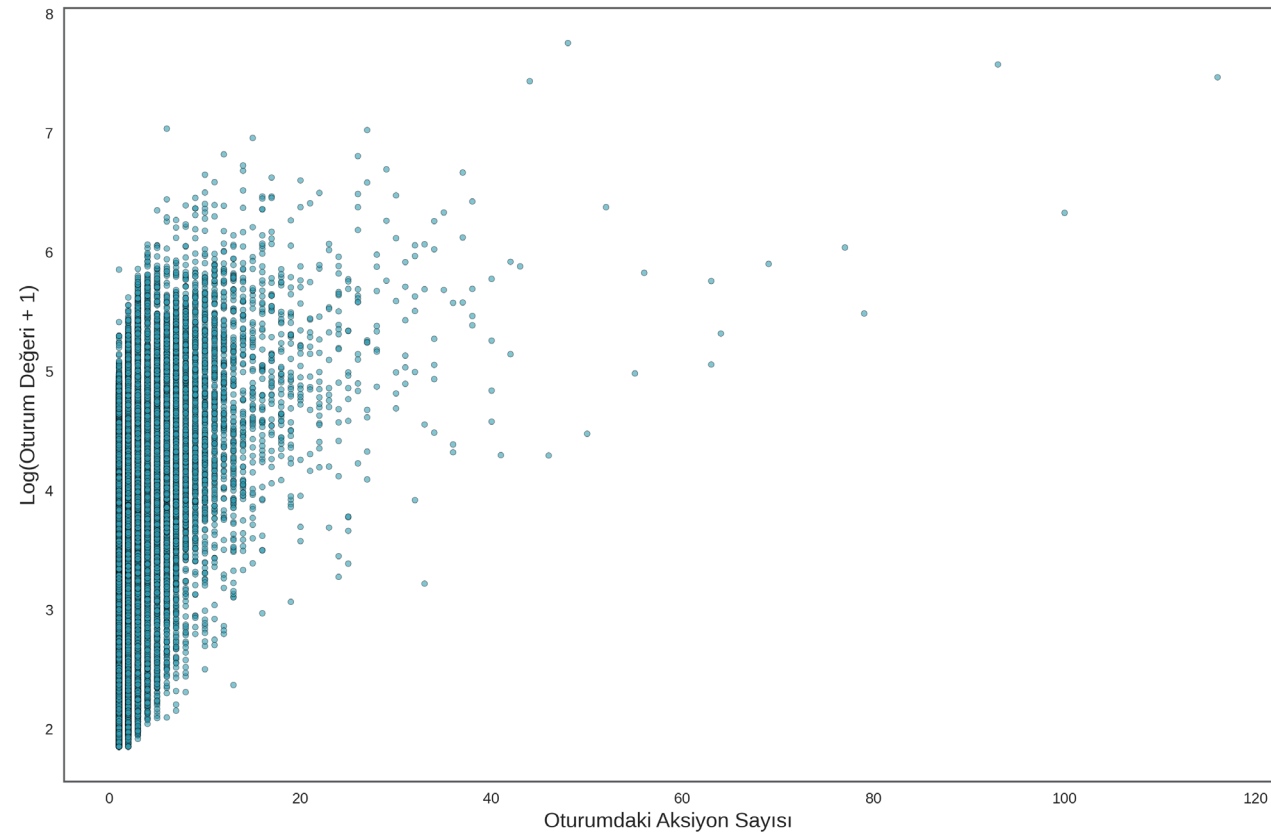
VERİ ANALİZİ

AKSİYONA GÖRE ORTALAMA OTURUM DEĞERLERİ



VERİ ANALİZİ

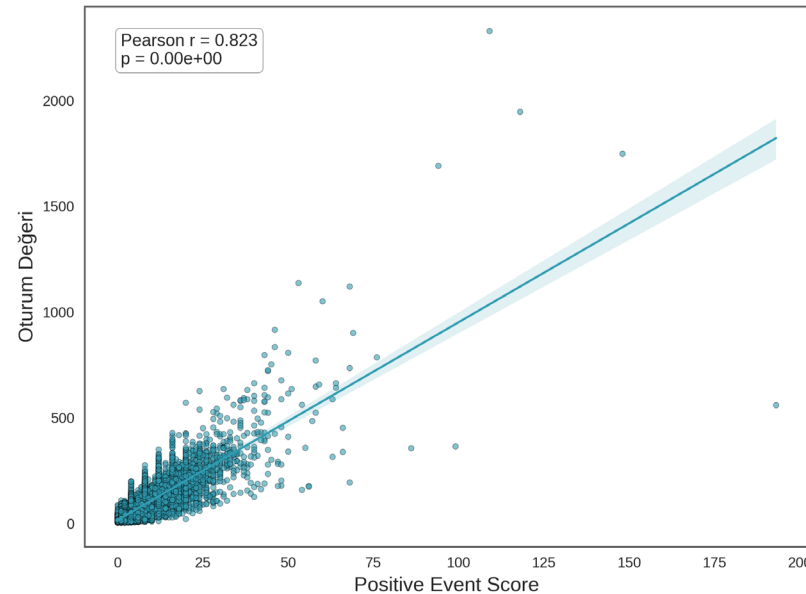
OTURUMUN DEĞERİ VE AKSİYON SAYISI İLİŞKİSİ



VERİ ANALİZİ

OTURUMUN DEĞERİ VE AĞIRLIKLANDIRILMIŞ TOPLAM AKSİYON SAYISI İLİŞKİSİ

$$session_value_i = \beta_0 + \beta_1 \times buy_count_i + \beta_2 \times add_cart_count_i + \beta_3 \times view_count_i$$



$$positive_event_score_i = 4 \times buy_count_i + 2 \times add_cart_count_i + view_count_i$$

ÇÖZÜM MİMARİSİ

VALİDASYON ŞEMASI

- Kullanıcı Bazında Quantile-Stratified Group KFold
- Kullanıcı Bazında Group KFold
- Time-series Split

ÇÖZÜM MİMARİSİ

VALİDASYON ŞEMASI

- **Kullanıcı Bazında Quantile-Stratified Group KFold**
- Kullanıcı Bazında Group KFold
- Time-series Split

ÇÖZÜM MİMARİSİ

FEATURE ENGINEERING

Oturum bazında:

- Toplam harcanan süre
- Aksiyon sıklığı
- Aksiyon sayıları (*satın alım, tıklama, toplam vb.*)
- İlk/son aksiyon
- Farklı aksiyon sayısı
- View/cart, cart/buy, view/buy vb. dönüşüm oranları
- Sepet büyüklüğü, çeşitliliği ve çoklu-tek ürün alışveriş durumları
- Saat, gün
- Funnel derinliği (*view-buy yolculuğunun tamamlanma oranı*)
- Ürün kategorisi çeşitliliği
- ***session_positive_score***: Ağırlıklandırılmış toplam aksiyon sayısı
- *Leave-One-Out Product Category Mean Encoding*

ÇÖZÜM MİMARİSİ

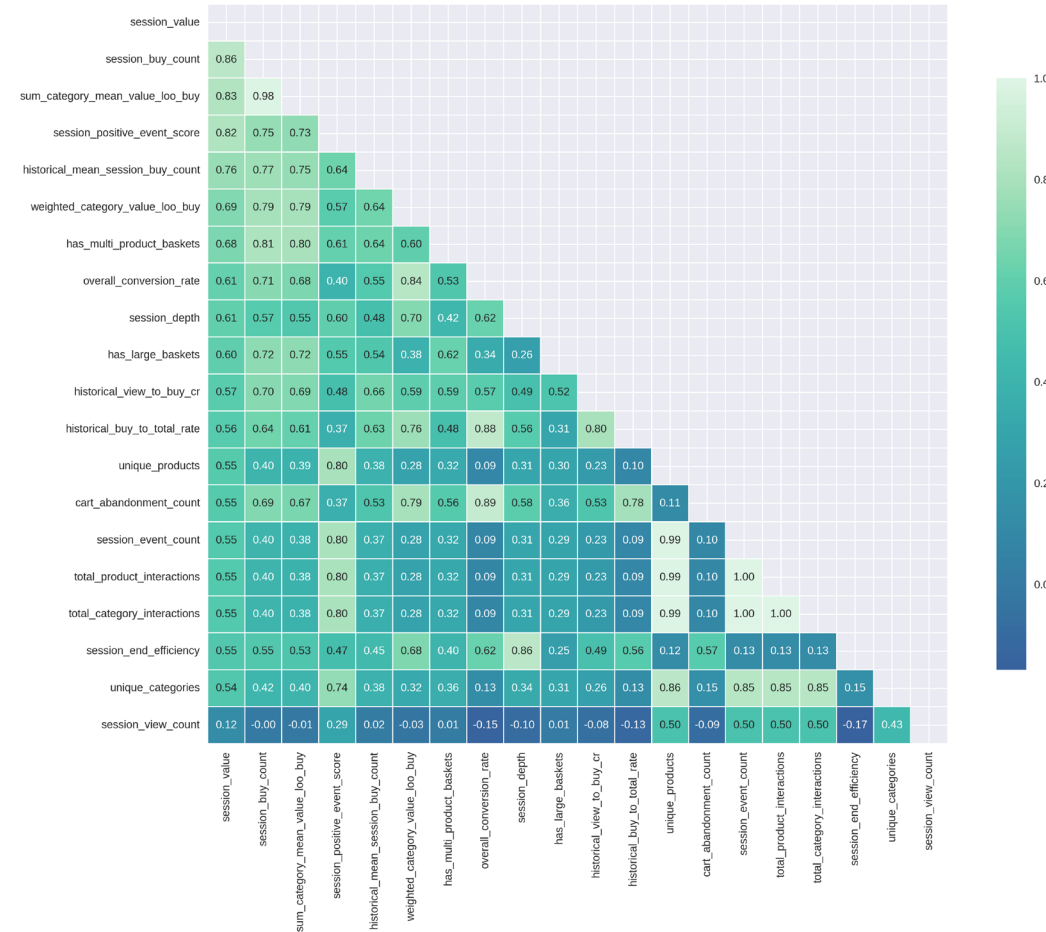
FEATURE ENGINEERING

Kullanıcı bazında:

- Kümülatif toplam aksiyon sayısı (*satın alım, tıklama, toplam, ağırlıklandırılmış vb.*)
- Ortalama toplam aksiyon sayısı (*satın alım, tıklama, toplam, ağırlıklandırılmış vb.*)
- Kümülatif ortalama dönüşüm oranları
- Kümülatif oturum sayısı
- Tutarlılık (mevcut/ortalama) (*oturum aksiyon sayısı, süresi, kategori çeşitliliği*)
- Geçmişe ait çoklu-tek ürün siparişi istatistikleri
- Momentum / trend (*aksiyon sayısı ve süresi*)

ÇÖZÜM MİMARİSİ

KORELASYON ANALİZİ



Unique() – Final Sunumu

ÇÖZÜM MİMARİSİ

FEATURE SELECTION

- CatBoost
- SHAP Values
- 10-Fold CV
- 5-step

ayarları kullanılarak feature-selection yapıldı.



Yandex
CatBoost



Herhangi bir feature setinin çıkarılması performansa olumlu etki etmedi.

ÇÖZÜM MİMARİSİ

MODEL SEÇİMİ

- CatBoost
- LGBM
- XGBoost
- AutoGluon

ÇÖZÜM MİMARİSİ

MODEL SEÇİMİ

- **CatBoost**
- **LGBM**
- XGBoost
- AutoGluon



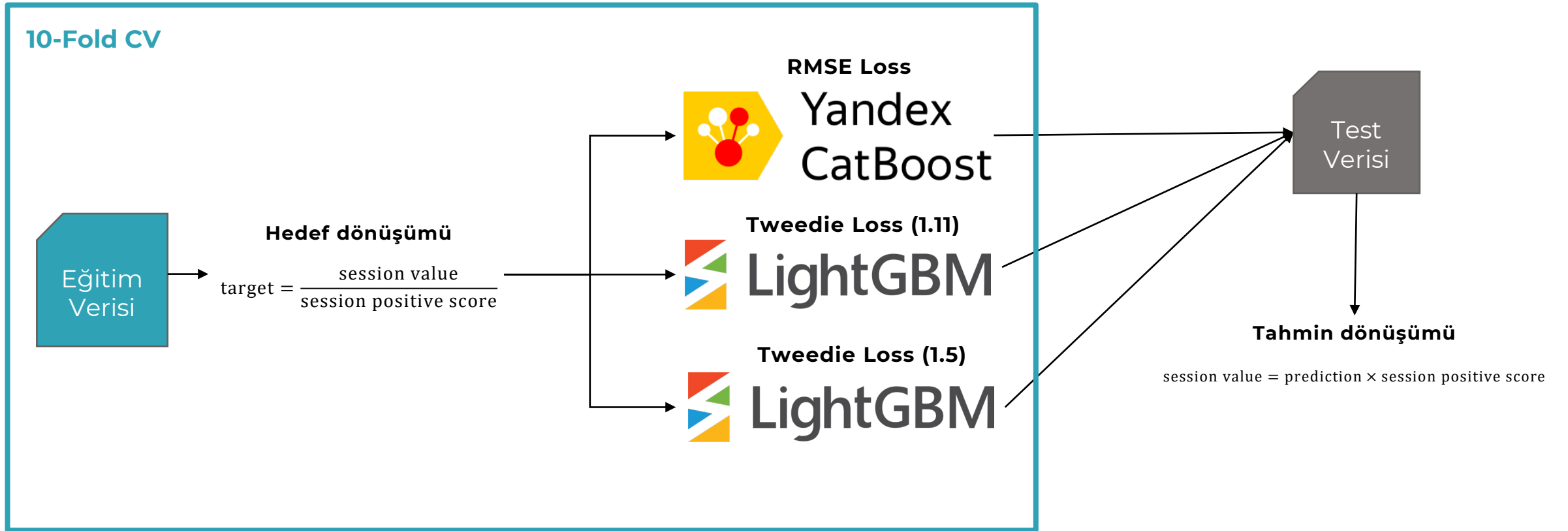
Yandex
CatBoost



LightGBM

ÇÖZÜM MİMARİSİ

EĞİTİM VE TAHMİNLEME



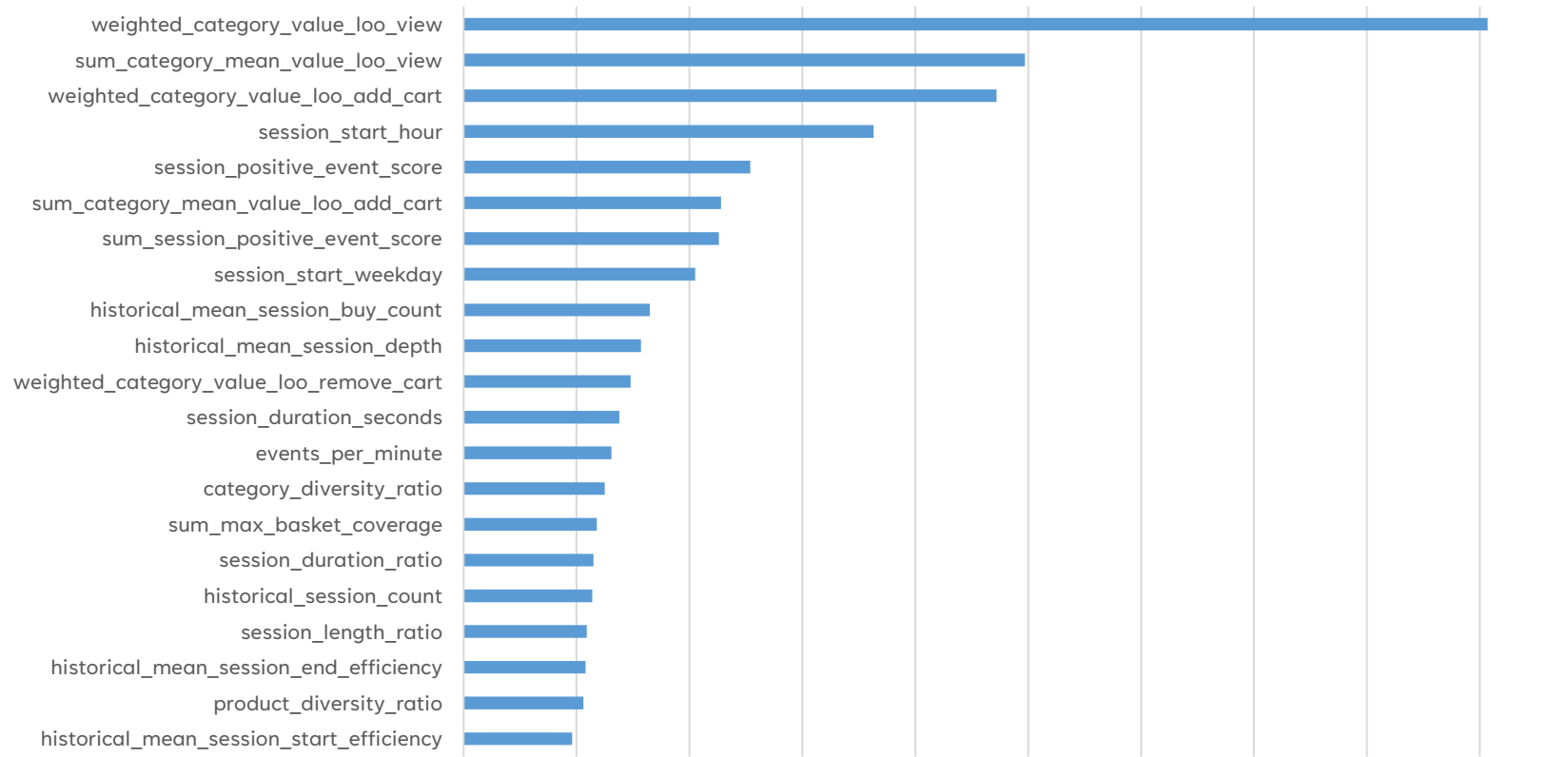
ÇIKARIMLAR

METOTLARIN ETKİSİ

Yöntem	Private LB	Public LB
Baseline Session Features	577	1364
Shared Session Override	499	680
Historical User Features	438	587
Model Blending	399	508
Target Transform	375	471
Category LOO Encode (null-override)	367	453

ÇIKARIMLAR

FEATURE IMPORTANCE



ÇIKARIMLAR

ÖRNEK SENARYOLAR

- Oturum değerinde yapılabilecek dönüşümler
- Ortalama oturum değeri yüksek kategoriler üzerinden kampanyalar
- Sepetten silme oranı yüksek ürün ve kategoriler için özel analiz



TEŞEKKÜRLER