



MIGROS

 coderspace

MIGROS DATATHON

FINAL SUNUMU

Anıl Öztürk – Tarık Karakaş

Spray N Pray

SUNUM AKIŐI

VERİ ANALİZİ

Ham verideki deęerlerin incelenmesi ve anlamlandırılması

SENTEZ VERİ

Kullanılan ekstra ve sonradan üretilmiş veriler

Feature seçimi

ÇÖZÜM MİMARİSİ

Tahmin mimarisi

Doęrulama mimarisi

Model seçimi

SONUÇ ve ÇIKARIMLAR

Öne çıkan featurelar

Performans analizi

Denenen alternatifler

Uygulanabilirlik ve potansiyel geliřtirmeler

PROBLEM

Belirli bir kampanya dönemi içerisinde müşterilerin geçmiş harcama verileri kullanılarak kampanyalara geri dönüş yapma durumları tahmin edilebilir mi?

- Aralık 2020 - Aralık 2021 arasındaki kullanıcı işlemleri verilmiş
- **Metrik:** F1-Score

LİTERATÜR

Ana Kavramlar

- Müşteri Serüveni
- Kampanya Katılım Tahmini
- Kampanya Önerimi

Sanal Market [1]:

- Site – uygulamada geçirilen süre
- Müşteri segmenti
- En çok ziyaret edilen ürünler
- CTR verileri

Tahmin Modelleri

- Karar ağaçları
- Yapay sinir ağları
- SVM

Ana Odak: Sanal alışverişler

Fiziksel Mağazalar [2]:

- Mağaza konumuna ait demografik bilgiler
- Müşteri memnuniyetine dair anket sonuçları

[1] <https://doi.org/10.1016/j.elerap.2021.101058>

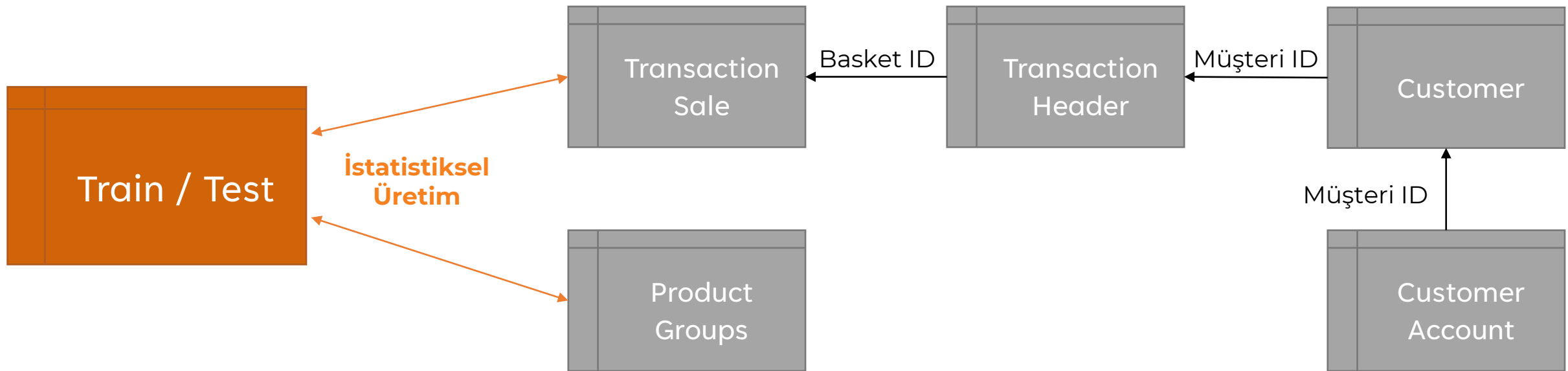
[2] <https://www.mdpi.com/0718-1876/17/1/9>

HAM VERİ

- Cinsiyet
- Şehir Kodu
- Doğum Yılı
- İşlem Geçmişi
 - Ürün Alt Kategorileri
 - Uygulanan İndirimler
 - Ürün Tutarı & Adedi
 - İşlemin sanal olma durumu
- Sorgulanan Kampanya
 - Kampanya Kodu
 - Gereken Minimum Harcama
 - Elde Edilecek Ödül Miktarı

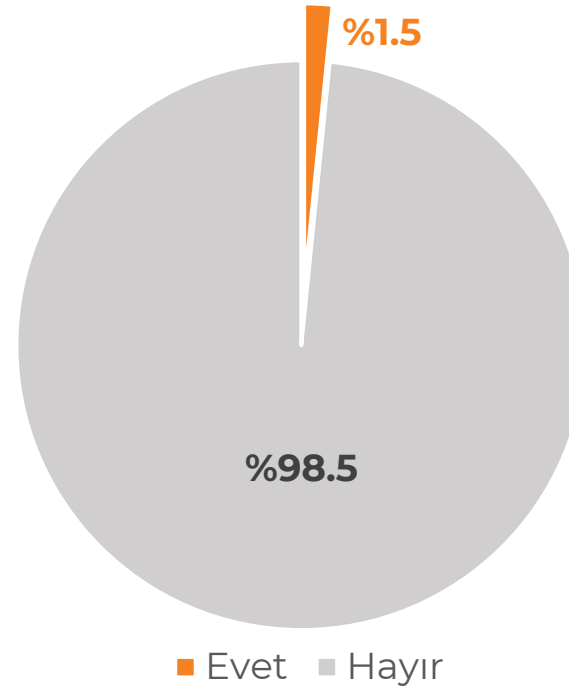
VERİ DÜZELTME VE ÖNİŞLEME

- Eğitim & Test verisinde olmayan müşterilerin işlem geçmişi silindi
- Kopya işlemler silindi
- Negatif tutara veya miktara sahip işlemler silindi



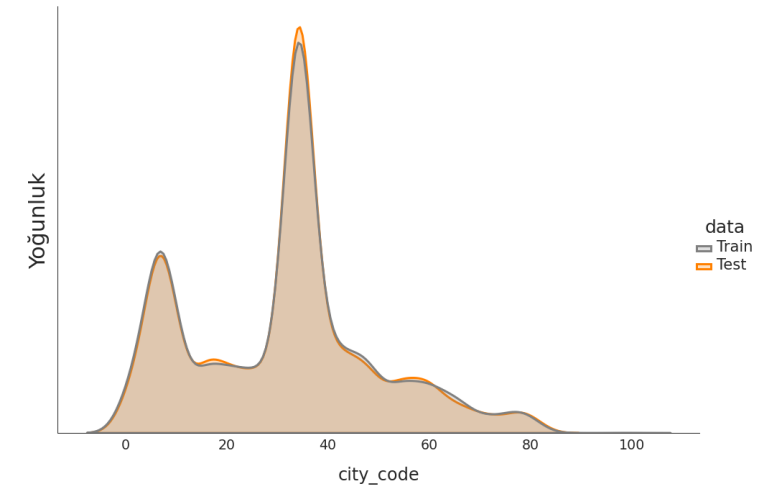
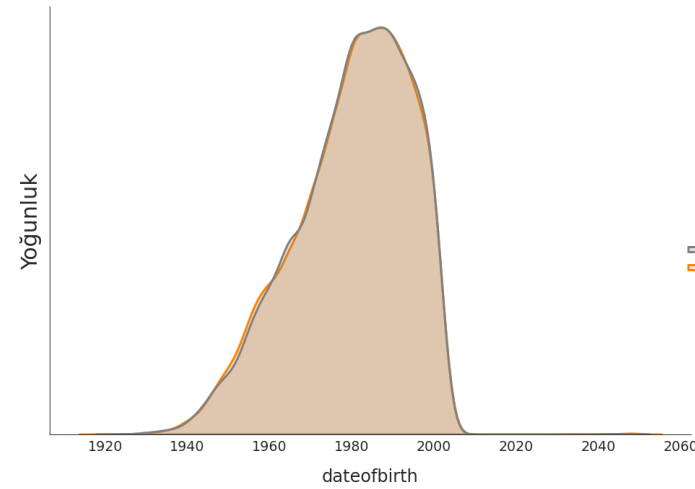
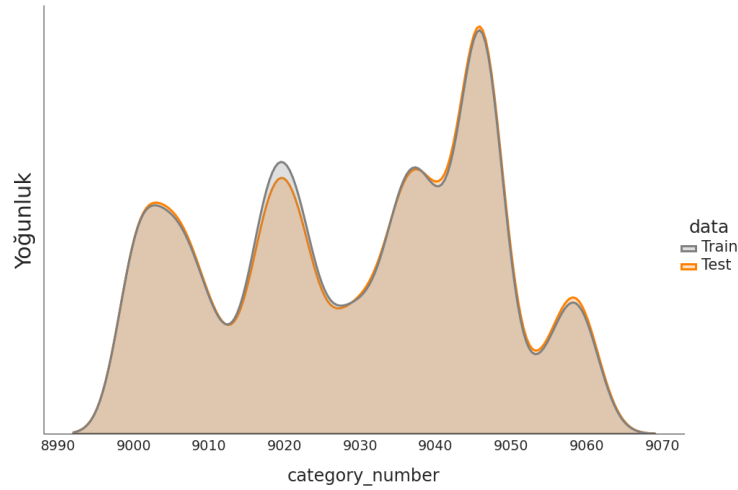
HAM VERİ ANALİZİ HEDEF DEĞERİN DAĞILIMI

Kampanyaya Dönüş Durumu



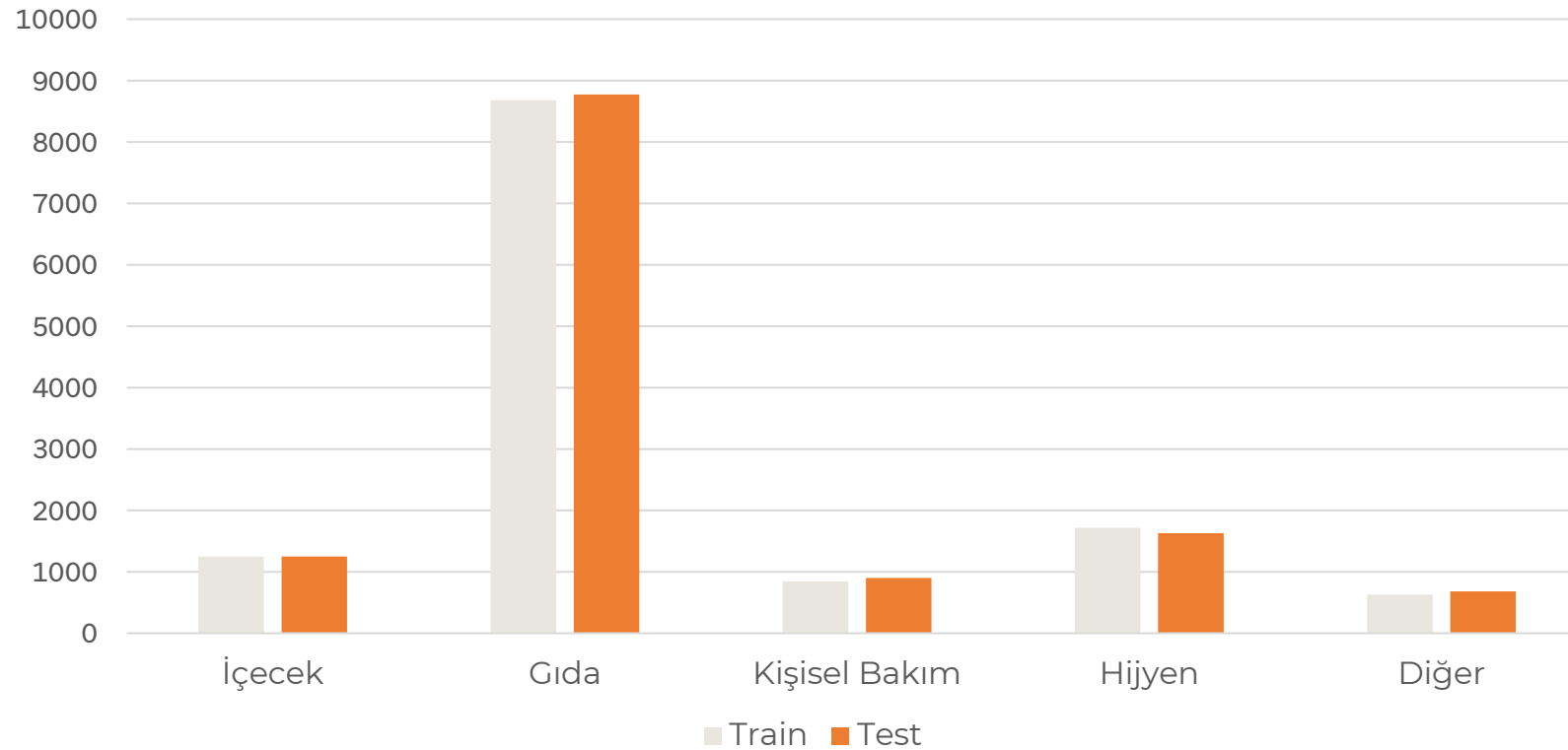
HAM VERİ ANALİZİ

ANA ÖZELLİKLERİN TRAIN-TEST DAĞILIMI



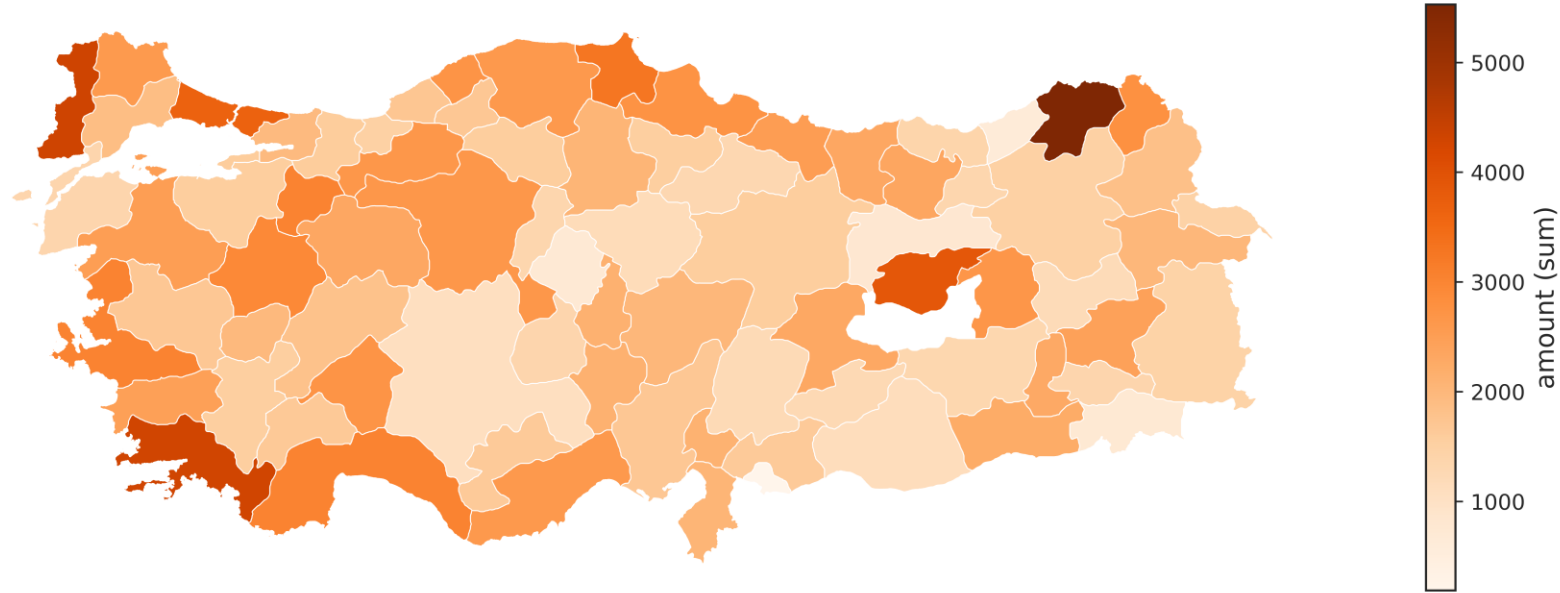
HAM VERİ ANALİZİ

İŞLEMLERDEKİ ÜRÜNLERİN GENEL KATEGORİSİ



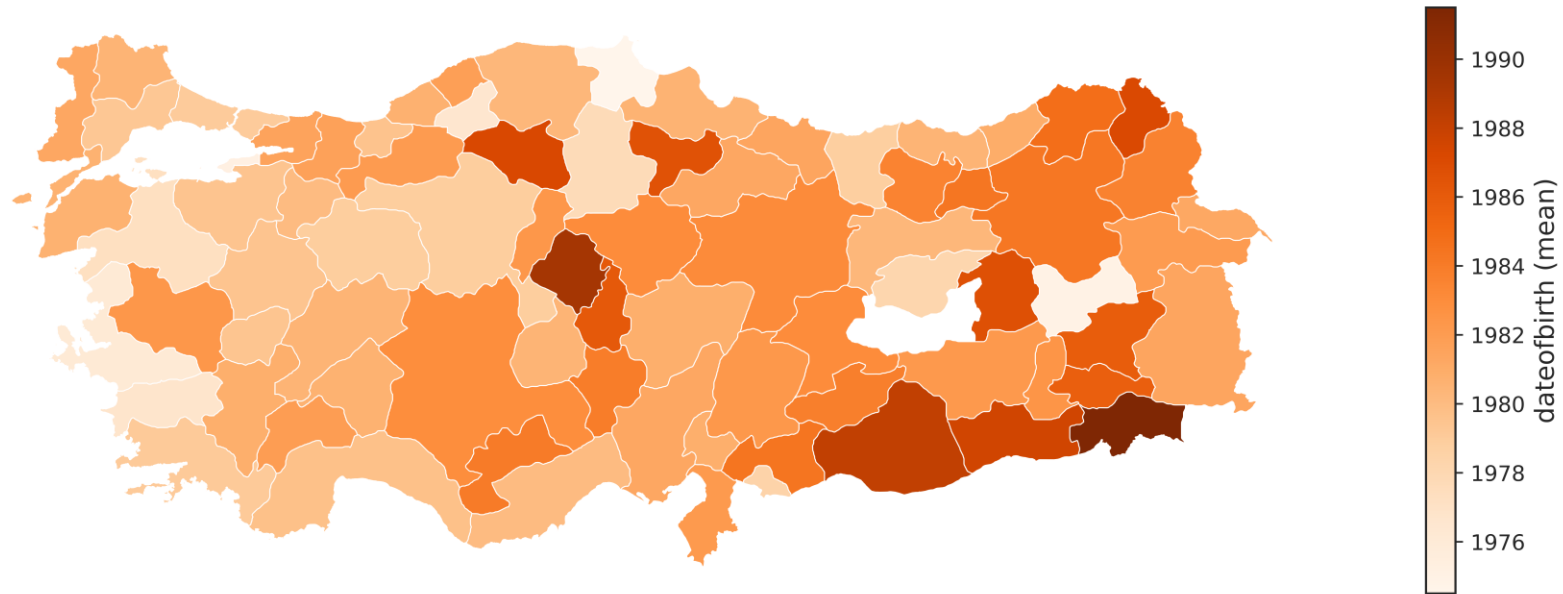
HAM VERİ ANALİZİ

TOPLAM HARCAMANIN İLLERE GÖRE DAĞILIMI



HAM VERİ ANALİZİ

MÜŞTERİ YAŞININ İLLERE GÖRE DAĞILIMI



SENTEZ VERİ

BAZ TAKVİM DEĞERLERİ

- İşlemin yapıldığı gün (One-Hot)

islem_Pzt	islem_Sal	islem_Crs	islem_Prs	islem_Cum	islem_Cmt	islem_Pzr
0	0	1	0	0	0	0

- İşlemin haftaiçi – haftasonu durumu
- İşlemin yapıldığı tarihten itibaren geçen süre

SENTEZ VERİ BAZ TUTAR DEĞERLERİ

- Ürünün birim fiyatı (*Tutar / Miktar*)
- Toplam indirim tutarı (*Tip 1 İndirim + Tip 2 İndirim + Tip 3 İndirim*)
- İndirimlerin ve toplam indirimin % cinsinden karşılığı
- Miktardan bağımsız indirimlerin varolma durumları (*>0*)

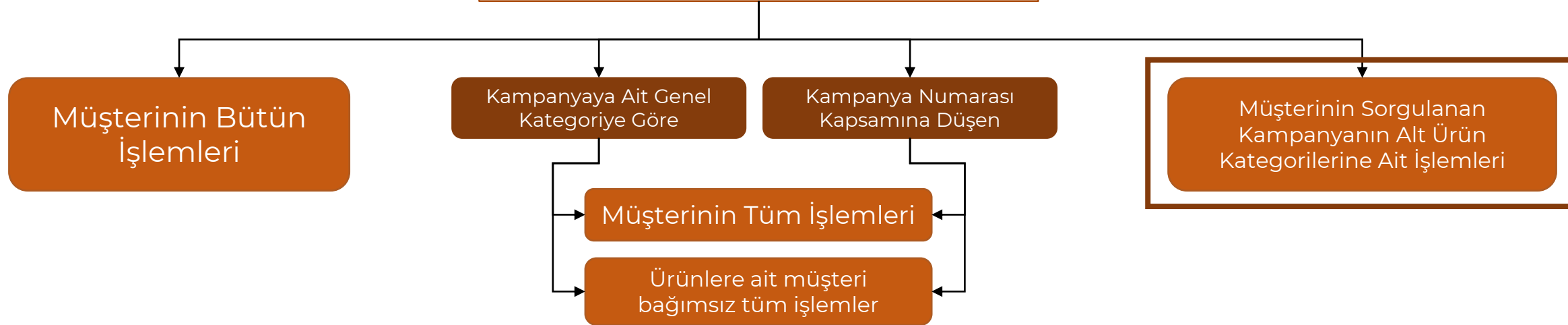
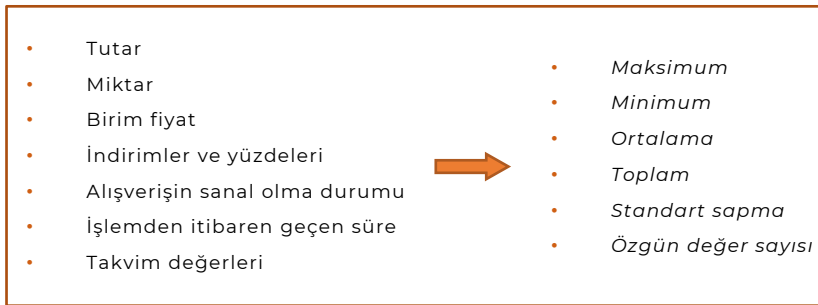
SENTEZ VERİ KULLANILAN İSTATİSTİKLER

- Tutar
- Miktar
- Birim fiyat
- İndirimler ve yüzdeleri
- Alışverişin sanal olma durumu
- İşlemden itibaren geçen süre
- Takvim değerleri



- *Maksimum*
- *Minimum*
- *Ortalama*
- *Toplam*
- *Standart sapma*
- *Özgün değer sayısı*

SENTEZ VERİ İSTATİSTİK VERİ KÜMELERİ



SENTEZ VERİ

MÜŞTERİNİN SORGULANAN KAMPANYANIN ALT ÜRÜN KATEGORİLERİNE AİT İŞLEMLERİ

Kampanya No	Seviye 1 Ürün Kategorisi	Seviye 2 Ürün Kategorisi	Seviye 3 Ürün Kategorisi	Seviye 4 Ürün Kategorisi
9000	10	20	30	40
9000	10	25	35	45
9001	20	25	35	45
9002	10	25	35	45
9002	10	30	35	45

SENTEZ VERİ

MÜŞTERİNİN SORGULANAN KAMPANYANIN ALT ÜRÜN KATEGORİLERİNE AİT İŞLEMLERİ

Kampanya No	Seviye 1 Ürün Kategorisi	Seviye 2 Ürün Kategorisi	Seviye 3 Ürün Kategorisi	Seviye 4 Ürün Kategorisi
9000	10	20	30	40
9000	10	25	35	45
9001	20	25	35	45
9002	10	25	35	45
9002	10	30	35	45

Eğitim & Test Verisi

Müşteri No	Kampanya No	Seviye 1 Ürün Kategorisi	Seviye 2 Ürün Kategorisi	Seviye 3 Ürün Kategorisi	Seviye 4 Ürün Kategorisi
1	9000	10	20	30	40
1	9000	10	20	30	50
2	9001	20	25	35	45
3	9003	22	45	55	65

Müşteriye ait birden fazla satırın oluşması

SENTEZ VERİ

MÜŞTERİNİN SORGULANAN KAMPANYANIN ALT ÜRÜN KATEGORİLERİNE AİT İŞLEMLERİ

Kampanya No	Seviye 1 Ürün Kategorisi	Seviye 2 Ürün Kategorisi	Seviye 3 Ürün Kategorisi	Seviye 4 Ürün Kategorisi
9000	10	20	30	40
9000	10	25	35	45
9001	20	25	35	45
9002	10	25	35	45
9002	10	30	35	45

Eğitim & Test Verisi

Müşteri No	Kampanya No	Seviye 1 Kesişimi Olan Ürünlere Ait İstatistikler			
1	9000	1.5852	-5.3	0.018	230

Aranan Kategori: (10)

SENTEZ VERİ

MÜŞTERİNİN SORGULANAN KAMPANYANIN ALT ÜRÜN KATEGORİLERİNE AİT İŞLEMLERİ

Kampanya No	Seviye 1 Ürün Kategorisi	Seviye 2 Ürün Kategorisi	Seviye 3 Ürün Kategorisi	Seviye 4 Ürün Kategorisi
9000	10	20	30	40
9000	10	25	35	45
9001	20	25	35	45
9002	10	25	35	45
9002	10	30	35	45

Eğitim & Test Verisi

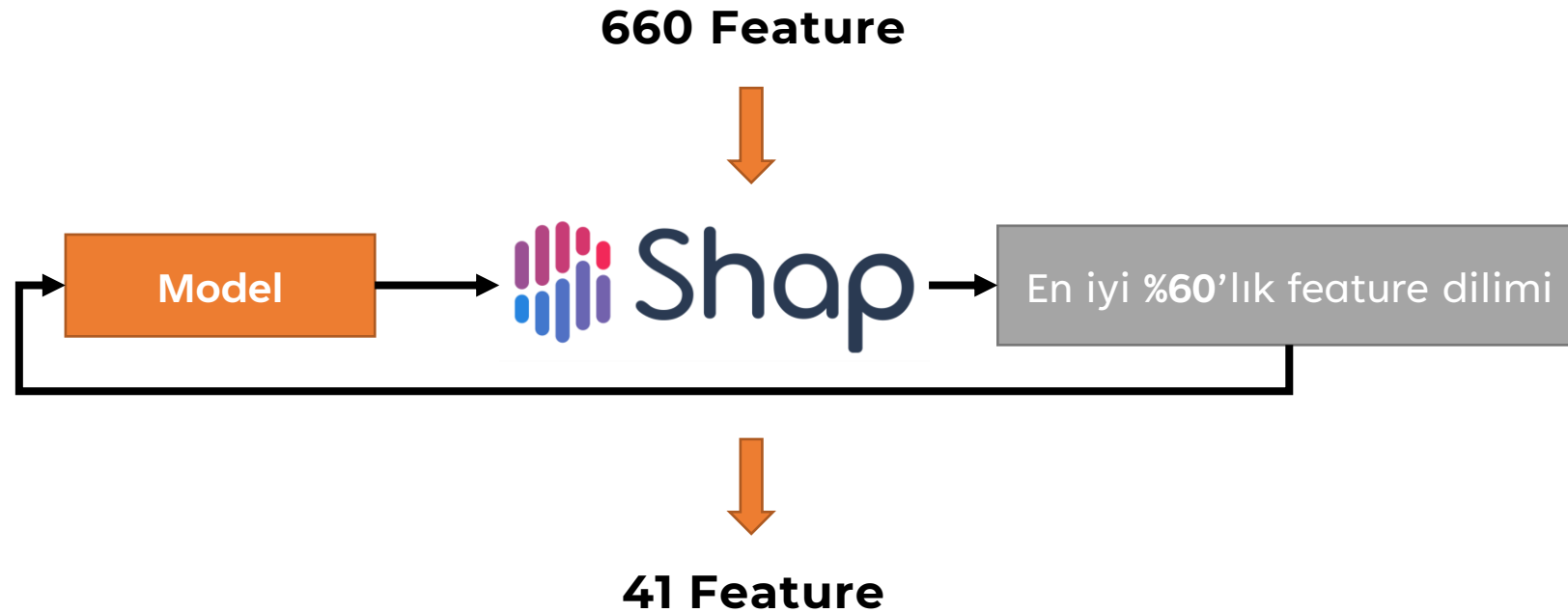
Müşteri No	Kampanya No	Seviye 1 + Seviye 2 Kesişimi Olan Ürünlere Ait İstatistikler			
1	9000	1.03	1.1	-0.222	108

Aranan Kategori: (10,20), (10,25)

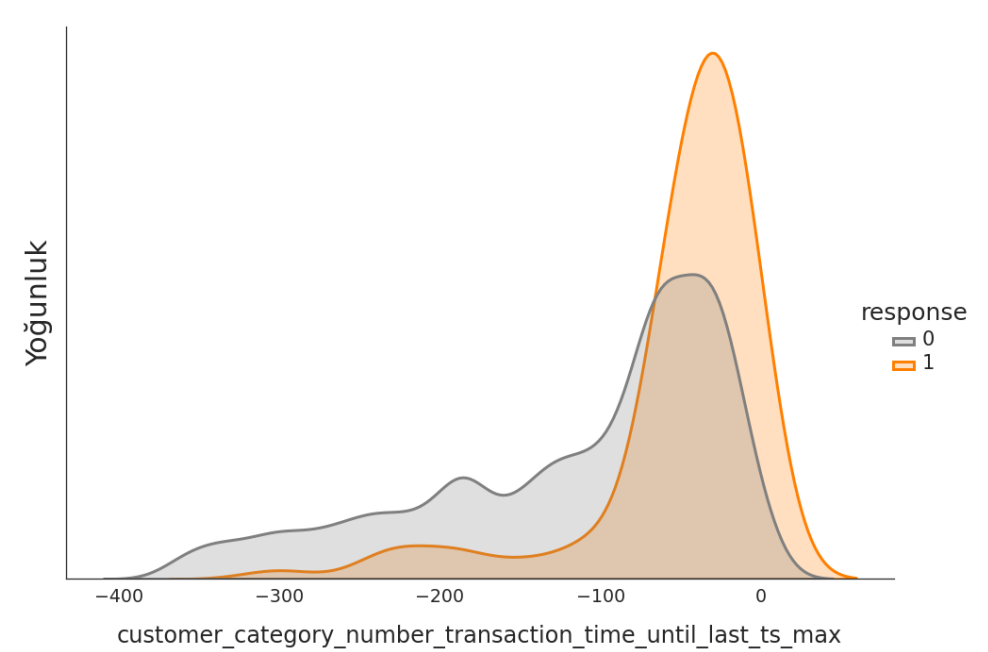
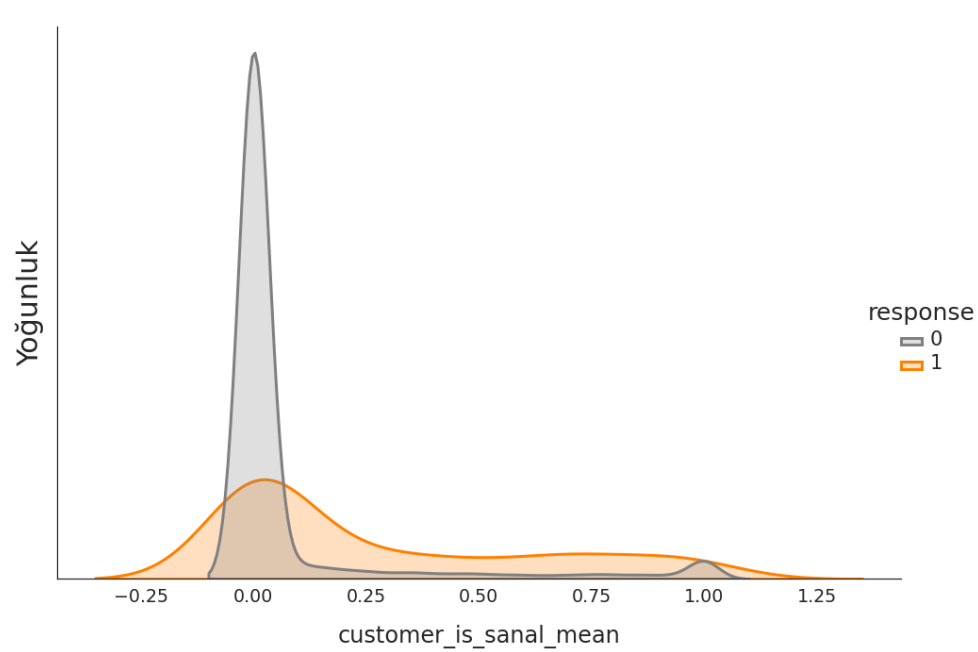
HARİCİ VERİ

- Ürünlerin dolar cinsinden dönemsel değeri **(Kullanılmadı)**
- İllerin GSYH katkısı **(Kullanılmadı)**
- İllerin nüfusu **(Kullanılmadı)**
- İllerin sahip olduğu Migros birimi sayısı **(Kullanılmadı)**

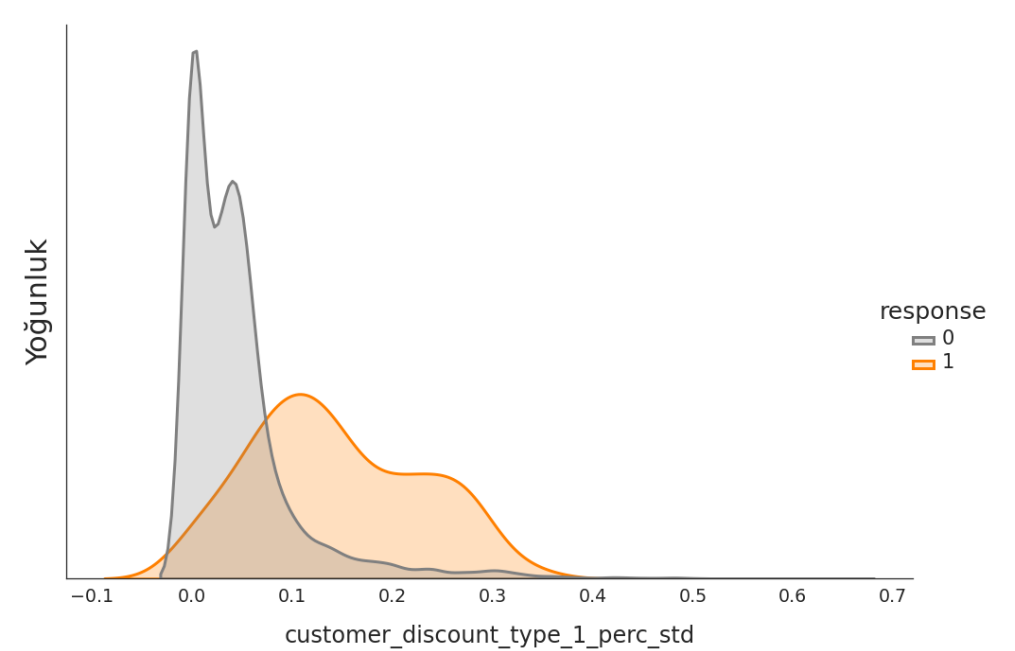
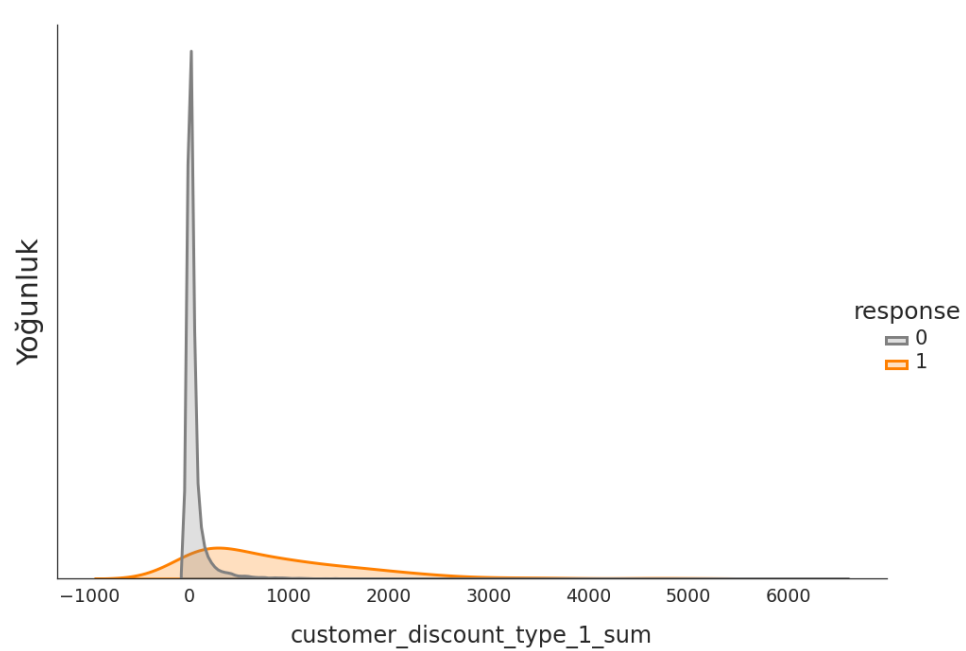
FEATURE SEÇİMİ



SEÇİLEN FEATURE SETİ BAZI ÖRNEKLER

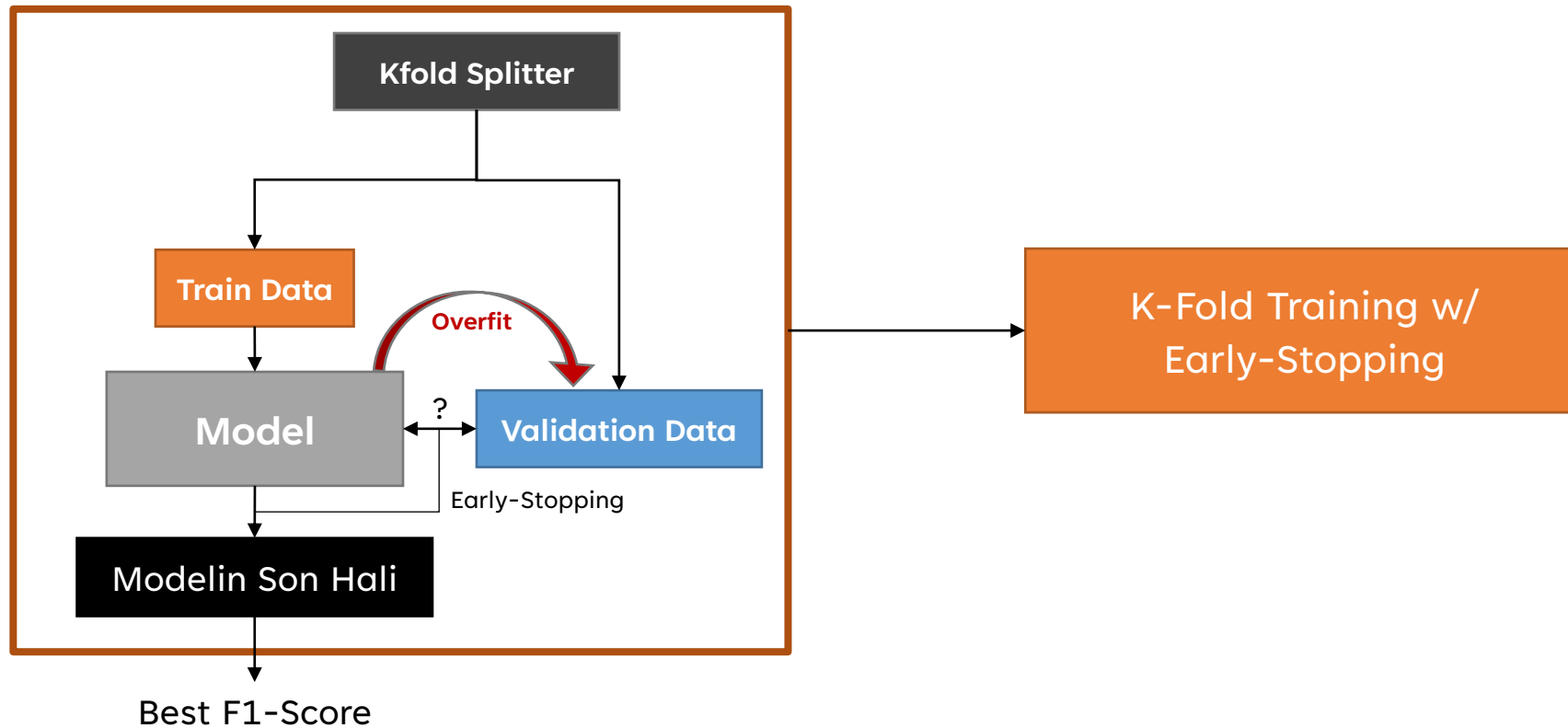


SEÇİLEN FEATURE SETİ BAZI ÖRNEKLER

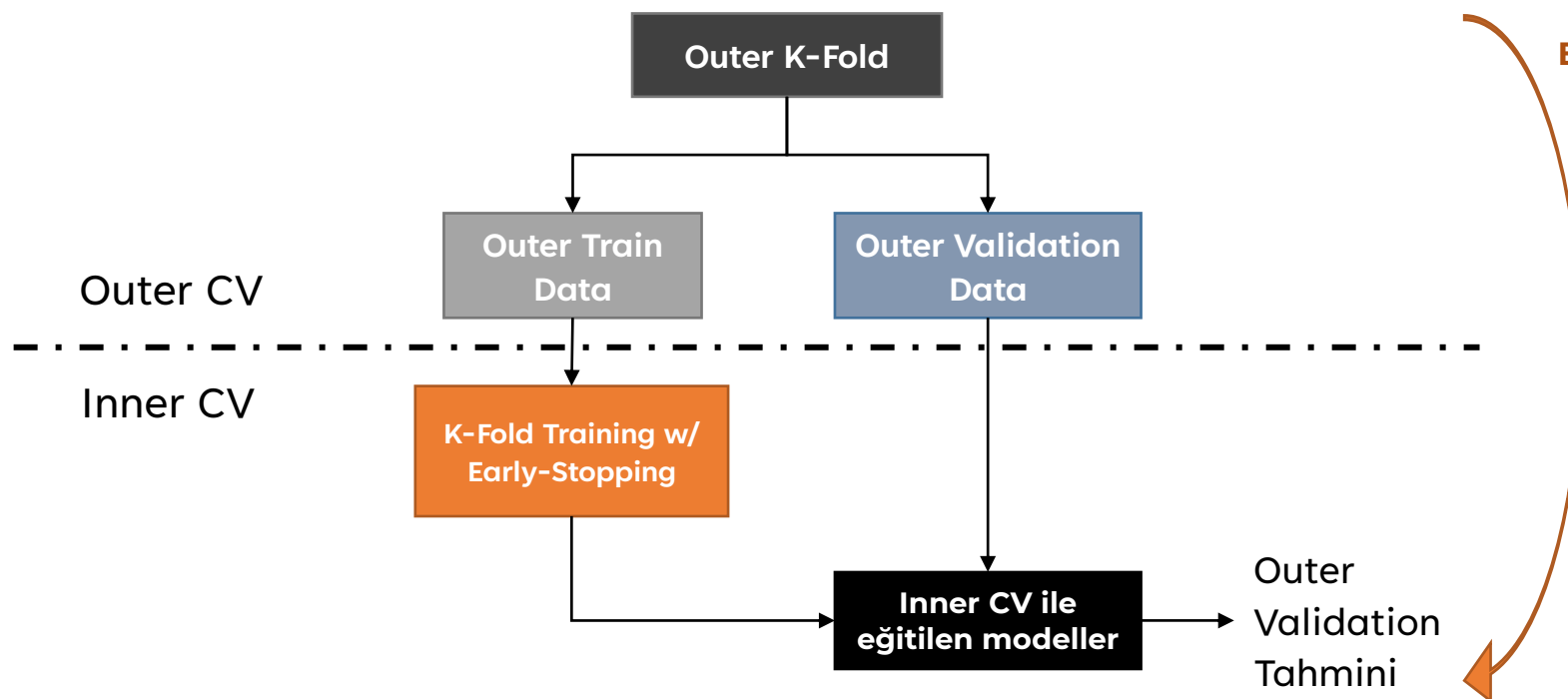


ÇÖZÜM DOĞRULAMA MİMARİSİ

KFOLD CROSS-VALIDATION



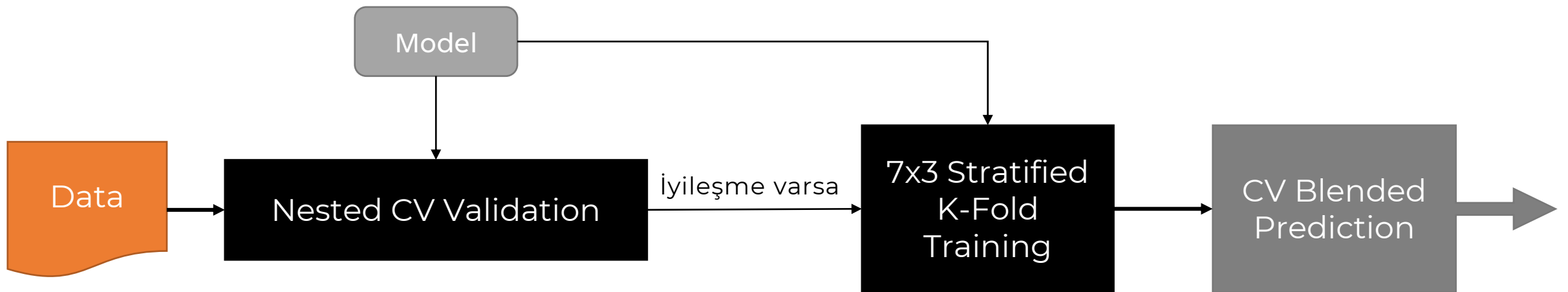
ÇÖZÜM DOĞRULAMA MİMARİSİ NESTED CROSS-VALIDATION (OUTER-INNER)



Bütün Veri

Müşteri No	Kampanya No	...	OOF Prediction
1	9001	...	1
2	9005	...	0
3	9003	...	0
4	9003	...	1

TAHMİN MİMARİSİ



MODEL SEÇİMİ

- CatBoost
- LightGBM
- XGBoost

MODEL SEÇİMİ

- **CatBoost**
- LightGBM
- XGBoost



Yandex
CatBoost

MODEL PARAMETRE SEÇİMİ



Manuel Parametreler:

- Depth
- Learning Rate
- Pos-Class-Scale



Nested CV Validation

SONUÇLAR

Nested Cross-Validation OOF Sonuçları

- **F1-Score:** 0.4779
- **Precision:** 0.4444
- **Recall:** 0.5167
- **Average Precision:** 0.3121
- **AUC:** 0.7878

Final Eğitim Sonuçları

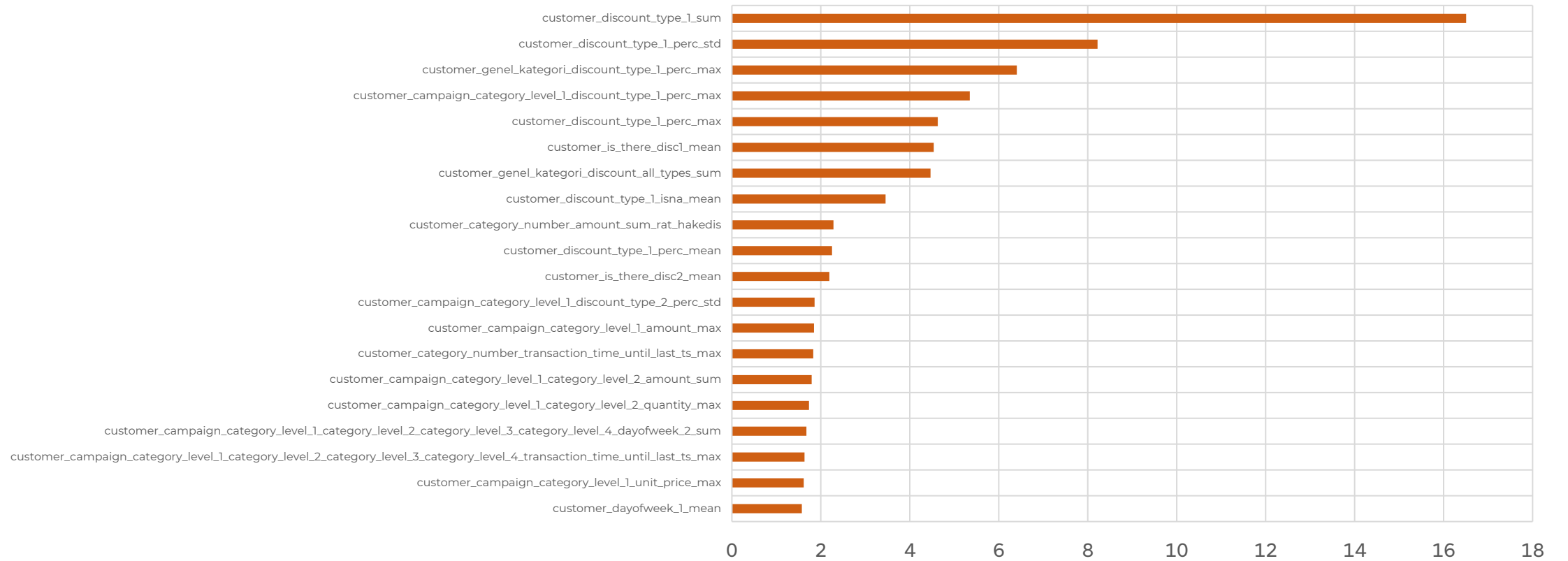
- **F1-Score:** 0.5324 ± 0.0610
- **Precision:** 0.5168 ± 0.0786
- **Recall:** 0.5551 ± 0.0616
- **Average Precision:** 0.4625 ± 0.0872
- **AUC:** 0.9440 ± 0.0212

Nested CV OOF < **LB'de Beklenen Skor** < Final Eğitim
 $0.4779 < \text{Beklenen Skor} < 0.5324$

Private Score ⓘ	Public Score ⓘ	Selected
0.47111	0.51376	<input checked="" type="checkbox"/>

Private Score ⓘ	Public Score ⓘ	Selected
0.48458	0.51376	<input type="checkbox"/>
0.48068	0.5	<input type="checkbox"/>
0.47747	0.50236	<input type="checkbox"/>

SONUÇLAR FEATURE IMPORTANCE



DENENEN ALTERNATİFLER

- *SMOTE* ile *Undersample* ve *Oversample*
- Farklı modeller (*LGBM* ve *XGBoost*)
- İstatistiklerin son 3 – 6 – 9 ay versiyonları
- Güncel zamana yaklaştıkça artan örnek ağırlıklandırma

ÇÖZÜMÜN UYGULANABİLİRLİĞİ

- Ensemble ve tekilleştirilebilir çözüm
- Düşük feature sayısı, genelleştirilmiş bir feature-set
- Canlıya alınabilir çözüm yapısı

sum_models

Purpose

Blend trees and counters of two or more trained CatBoost models into a new model. Leaf values can be individually weighted for each input model. For example, it may be useful to blend models trained on different validation datasets.

Method call format

```
sum_models(models,  
            weights=None,  
            ctr_merge_policy='IntersectingCountersAverage')
```



TEŞEKKÜRLER